THE UNIVERSITY *of* EDINBURGH

Proteogenomic Platforms Establishing Personalized and Precision Neoantigen Therapeutics in Cancer

Kamila Pawlicka

Thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy

University of Edinburgh

2023

Declaration

I declare that this thesis has been completely written and composed by myself along with the work detailed herein were entirely my own unless otherwise clearly acknowledged. This piece of work has been submitted for the degree of Doctor of Philosophy and has not been submitted for any other degree or professional qualifications.

Signed: Kamila Pawlicka

27/02/2023

Acknowledgement

This thesis would not have been possible without invaluable help of my supervisor Professor Ted Hupp who continuously supported me throughout my PhD. I am grateful for his advice and engagement during the past few years I spent in his laboratory.

To my family for being my stable core, helping me focus on pursuing my path. A special thanks to my boyfriend Mark for his never-ending support and encouragement towards completing my PhD. Thanks for being there through all the good and bad science days.

I also would like to offer my sincerest gratitude to each and every one who has contributed to this thesis. It is impossible to mention everyone involved, however I am well aware that many people supported my research journey, from the collaborators abroad to my lab colleagues.

Lay summary

Cancer is a disease of the genome. Significant technological advances have allowed researchers to molecularly characterise tumours and responding immune cells, which has resulted in breakthroughs that have translated to pharmacologically actionable markers and targets. A new paradigm in therapeutics involves stimulating the patients' own immune system to eradicate cancer. New technologies have enhanced our ability to search for new immune targets such as mutation-derived antigens (neoantigens), which has accelerated the development of novel immune therapies. In order to detect tumour cells, the immune system must recognise them as foreign; they do so by detecting molecules named neoantigens on the cell surface. Neoantigens can form when cancer cell-specific mutations occur that result in alterations of the protein from 'self'. Those tumour-specific markers can also function as immune stimulatory vaccines when administered to the body. Neoantigens can activate an immune response sufficient to clear tumour cells. Neoantigens may have diverse sources such as viral and mutated proteins. The identification of neoantigens holds great promise to further improve the efficacy of immunotherapeutic approaches, therefore the aim of this project is to develop pipelines to understand the current limitations of vaccine target discovery and to streamline future neoantigen vaccine therapy for human cancers.

.

Abstract

The evolution of new technologies such as next-generation sequencing has enhanced our ability to search for new immune targets such as mutation-derived antigens (neoantigens), which has accelerated the development of novel immune therapies. Neoantigens are tumour-specific markers that can form when cancer cell-specific mutations encode molecules that differ from 'self'. Neoantigens may have diverse sources such as viral and mutated proteins. Moreover, posttranslational modifications and altered antigen processing can contribute to the neoantigen landscape. Studies have shown that inhibition of nonsense-mediated decay (NMD) also generates new antigenic determinants in tumour cells. This RNA quality control pathway removes aberrant mRNA containing premature termination codons (PTCs) that arise through mutation or defective splicing. Neoantigens expressed on the surface of tumour cells can be detected and removed by immune cells, thus neoantigens can themselves function as immune stimulatory vaccines. Tumour vaccination aims at stimulating a systemic immune response targeted to, mostly weak, antigens expressed in the disseminated tumour lesions. The main challenge in developing effective vaccination protocols is the identification of potent and broadly expressed tumour rejection antigens and effective adjuvants to stimulate a robust and durable immune response.

The purpose of this project was to create pipelines to identify mutated proteins in human and mouse melanoma cell lines from matched genomic, transcriptomic, and proteomic data. A proteogenomic pipeline was set up including definition of cancer DNA variants, RNA variants (expressed genes under different conditions), and mass spectrometry to identify mutated protein and MHC class I expression. One key question was to what extent can we detect mutated MHC Class I binding peptide based on cancer genome and RNA sequencing. Another key question was to define to what extent a cancer gene (such as p53) or a drug lead (such as NMD inhibitor or splicing inhibitor) changes the immunopeptidome as well as the mutated immunopeptidome. These pipelines require utilising bioinformatics approach to analyse next

generation DNA and RNA sequencing and creating a mutant protein reference database specific for the isogenic cell line with or without drug treatment. The thesis outline includes; (i) Developing rules for stop codon readthrough in the presence of NMD inhibitors in order to expand mutant proteins that can be produced; (ii) Analysis of mutant tumour immunopeptidome using an isogenic wild-type and p53 null cell model to define how this tumour suppressor gene status impacts immunopeptidome and how many mutant immunopeptides can be detected when comparing two distinct methodologies for immunopeptidome identification; and finally; (iii) Defining the effects of NMD inhibition or splicing inhibition on the mutant tumour neoantigen pool in human and mouse melanoma models to form a novel paradigm for developing new drugs to stimulate tumour rejection. The developed proteogenomic platform resulted in detection of a novel potential neoantigenic peptide ASELHTSLY proving the effectiveness of the presented approach.

Abbreviations

ACT	Adoptive cell transfer
APM	Antigen presenting machinery.
ASNS	Asparagine synthetase
ATM	Ataxia-telangiectasia
AUC	Area under ROC curve
BCR-ABL	Breakpoint cluster region-proto-oncogene tyrosine-protein kinase
Cas9	CRISPR associated protein 9
CD	Cluster of differentiation
Cdk2	Cyclin-dependent kinase 2
cDNA	Complementary DNA
CFTR	Cystic fibrosis transmembrane conductance regulator
CNV	Copy number variation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSC	Cancer stem cells
CTA	Cancer/testis antigen
CTL	Cytotoxic T lymphocytes
CTLA-4	Cytotoxic T cell antigen 4
DIA	Data independent acquisition
DDA	Data dependent acquisition
DMD	Duchenne’s muscular dystrophy
DMSO	Dimethyl sulfoxide

DTT	Dithiothreitol
DUSP	Ductal specificity phosphatase 1
EBV	Epstein-Barr virus
ECL	Enhanced chemiluminescence
FACS	Fluorescence-activated cell sorting
GOF	Gain of function
HCC	Hepatocellular carcinoma
HLA	Human leukocyte antigen
ICI	Immune checkpoint inhibitor
IR	Intron retention
LOH	Loss of heterozygosity
MAE	Mild acid elution
MBP	MHC-bound peptides
MDM2	Murine double minute-2
MHC-I	Major histocompatibility complex class I
MHC-II	Major histocompatibility complex class II
mRNA	Messenger RNA
MS	Mass spectrometry
MSI	Microsatellite instable
NMD	Nonsense-mediated mRNA decay
NGS	Next generation sequencing
ORF	Open reading frame
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction

PD-1/L1	Programmed death-ligand 1
PTC	Premature termination codon
PTPs	Pioneer translation products
RF1	Release factor 1
ROC	Receiver's Operating Characteristic
ROS	Reactive oxygen species
SBP	Streptavidin-Binding Peptide
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SV40	Simian virus 40
TCGA	The Cancer Genome Atlas
TCR	T cell receptor
TEMED	Tetramethylethylenediamine
TIL	Tumour-infiltrating lymphocytes
TMB	Tumour mutational burden
TME	Tumour microenvironment
TP53	Tumour protein 53
Treg	Regulatory T cell
v/v	volume/volume
WES	Whole exome sequencing
WGS	Whole genome sequencing
WT	Wild type

Contents

List of figures

List of tables

1 Introduction

1.1 Cancer development

Despite novel developments in therapeutic strategies and diagnosis, cancer remains the second leading cause of morbidity and mortality across the world. International Agency for Research on Cancer estimated that in 2020 cancer-related death has risen to nearly 10 million annually.

Transforming normal cells into cancer cells is a multi-stage process, which progresses with the accumulation of genetic, microenvironmental and physiological changes that allow abnormal cell survival, proliferation, and often metastasis.

What makes a cell cancerous have been defined as the hallmarks of cancer. Hanahan and Weinberg created a model of shared characteristics (hallmarks) uniting all types of cancer cells based on their cellular phenotype. Initially they identified six distinct hallmarks, later expanded to eight and recently this number grew to 14 cellular parameters, that play key roles in tumourigenesis (Figure 1.1) (Hanahan, 2022; Hanahan & Weinberg, 2000). Past studies pointed to cancer being a result of the accumulation of multiple driver mutations, some within the non-coding regions (Khurana et al., 2016).

Figure 1.1 Hallmarks of cancer. Hanahan and Weinberg proposed a set of fundamental traits shared by most, if not all, human cancers. List of six cellular features set by them in 2000, was later expanded to fourteen characteristics that orchestrate the multistep process of the transformation of normal cells into malignant cells (Hanahan, 2022)

Genome instability and fast pace of cell proliferation, creates heterogenous tumours that are challenging to treat. Therefore, studying the mutational landscape of cancer holds a great promise for understanding the molecular mechanism of tumour development and progression.

1.1.1 Cancer mutational landscape

Advancements in next-generation sequencing (NGS) technology allowed for a high-throughput DNA and RNA sequencing, providing a tool for mapping and characterising mutations in an individual's tumour. Analysis revealed that alongside germline variants inherited from parents, patients with cancer carry somatic mutations in their tumour tissue. In an average human genome, there are roughly 4 million germline variants relative to the reference genome, whereas tumour genome

typically contains thousands of variants in relation to the same germline DNA (Amar et al., 2017; Koboldt et al., 2013).

Even though somatic mutations can be identified in healthy tissues, they are relatively rare in comparison with their high presence in tumour cells. Tumour mutational burden (TMB) has been defined as the number of non-synonymous mutations detected in coding regions across the whole genome. Chalmers et al., analysed 100,000 cancer cases to test for association between somatic alterations and tumour mutational burden in over 100 tumour types (Chalmers et al., 2017) and discovered that TMB increases significantly with patient's age. They also identified that cluster of somatic mutations in the promoter of PMS2 gene, found in 10% of skin cancers show high correlation with increased TMB (Chalmers et al., 2017).

Based on their impact on protein sequence, mutations can be classified as frameshift and point mutations. Frameshift variants are a result of an insertion or a deletion of a number of nucleotides that is not divisible by three. Translational machinery reads a gene's code in groups of three bases, therefore mutation disrupting this normal reading frame, causes the whole gene sequence following the mutation being read incorrectly (Chu & Wei, 2019).

Point mutations are a product of a modification of a single nucleotide and can be either synonymous or nonsynonymous. Synonymous mutation occurs when a codon is being replaced with another codon that codes for the same amino acid, thus the produced amino acid sequence is not changed. However, synonymous mutations are known to interfere gene expression through alteration in transcript stability, mRNA secondary structure, splicing processes, and translational kinetic (Sarkar et al., 2022). The modifications of splicing machinery can lead to changes in the binding pattern of splicing factors and result in exon skipping, intron retention, and various other forms of alternative splicing. It has been shown that synonymous mutations can actually influence the splicing pattern of cancer-related genes, proving that despite these mutations being silent to the protein sequence, they are not always silent to the function (Chu & Wei, 2019).

Nonsynonymous mutations on the other hand, change the protein sequence and can be divided into nonsense and missense mutations. Nonsense variants result in premature termination codons that, if translated lead to production of truncated

and often non-functional protein. Whereas missense mutations are single base pair substitutions, whose consequence is a creation of an amino acid that is different from the usual amino acid at that position.

Interestingly, Chu et al, by studying mutations across cancer-related genes, revealed that the nonsynonymous to synonymous ratio is significantly lower (~20%) in this group compared to other genes, which suggests that the nonsynonymous variants in cancer-related genes are suppressed most likely due to their deleterious effect (Chu & Wei, 2019).

Tumour suppressors and oncogenes

One of the hallmarks of cancer is evading growth suppressors by genetic or epigenetic modifications of genes responsible for keeping cell division and replication in check (Hanahan & Weinberg, 2000). These genes are named tumour suppressor genes (TSGs). Main role of TSGs is to control cellular proliferation by negatively regulating cell growth, therefore, disruption of their activity can lead to cancer promotion. Downregulation of TSGs includes mutations, deletions, inactivation of upstream or downstream effectors, DNA methylation and histone modifications (Hanahan, 2022).

In many cancers, tumour suppressor genes carry nonsense mutations. This leads to the loss of protein expression or the synthesis of a truncated protein unable to play its role in inhibition of cell proliferation and promoting apoptosis. Studies proved, that tumour suppressor genes have higher frequency of nonsense mutation than oncogenes (Chu & Wei, 2019). The most studied tumour suppressor gene is TP53, which will be described in depth within next chapters. In 1971 Alfred Knudson constructed a two-hit hypothesis, which states that to cause a phenotypic change and therefore cancer development, TSG require inactivation of both alleles, either through mutations or epigenetic silencing (Knudson, 1971).

Proto-oncogenes are normal genes promoting cellular growth and proliferation, but which also have a potential to contribute to cancer development if they lose their function through mutation, fusion with another gene or overexpression, transforming proto-oncogenes into oncogenes (Brown, 2021). Many studied proto-oncogenes encode transcription factors, regulators of cell death, growth stimulators, receptors and signal transducers. Their involvement in regulating

the behaviour of normal cells and promoting oncogenic transformation is well documented. Oncogenes typically exhibit increased production of proteins driving cell division, decrease cell differentiation, and inhibit apoptosis; taken together, these phenotypes define cancer cell (Brown, 2021; Hanahan, 2022).

It has also been shown that some oncogenes are a crucial factor in restricting the lineage versatility of cancer stem cells (Brown, 2021). In physiological state, stem cells, such as haematopoietic stem cells, replenish mature cell types to meet the demands of an organism. Some oncogenes were found to deregulate this homeostatic process by inhibiting leukaemia stem cells to give rise to a single cell lineage (Brown, 2021). However, it still remains to be explained whether the deregulation of stem cell and progenitor cell versatility and restricting the progeny of cancer stem cells (CSCs) to only one lineage is central to how some oncogenes drive the development of cancer.

Vogelstein et al. showed that patterns of mutations found in well-studied oncogenes and tumour suppressor genes are not random and in fact they are very characteristic. Oncogenic genes are often mutated at the same amino acid position, while tumour suppressors are mutated by protein-truncating alterations throughout their length (Vogelstein et al., 2013). Based on these mutation patterns rather than frequencies, Vogelstein group was able to determine which of the mutated genes can be causally linked to cancer progression, a mutation driver gene, and whether they are likely to function as oncogenes or tumour suppressor genes (Vogelstein et al., 2013)

Other genetic alterations in tumours

It has been noted that the rate of point mutation in tumours is similar to normal cells, but the rate of chromosomal changes in cancer is elevated (Vogelstein et al., 2013). Therefore, most solid tumours have been characterized by widespread changes in chromosome number, as well as deletions, inversions, translocations, and other genetic abnormalities (Chen et al., 2019). These changes affecting large part of a chromosome cause difficulty in the identification of the specific “target” gene(s) whose gain or loss creates a growth advantage to the cancer cell. In case of translocations, gene amplifications or homozygous deletions target genes are

detected more easily. Translocations can result in gene fusion and creation of an oncogene, such as Breakpoint Cluster Region-proto-oncogene tyrosine-protein kinase (BCR-ABL) in chronic myelogenous leukaemia, in a small number of cases, can also inactivate a TSG by truncating it or isolating it from its promoter (Kang et al., 2016). Homozygous deletions often target one or a few genes, usually tumour suppressor.

Oncogenes are usually transformed by amplification, causing the protein product to be abnormally active by tumour cell containing 10 to 100 copies of the gene per cell, in comparison with the two copies that are present in normal cells (Vogelstein et al., 2013). The mutational patterns of well-studied oncogenes are highly characteristic and non-random. Oncogenes are often mutated at the same amino acid positions (Vogelstein et al., 2013).

Cancer cells adapt to cellular stresses and environmental changes, such as hypoxia, by inducing metabolic changes and increasing their phenotype variability. Warburg et al. in 1950s were the first to report metabolic adaptation of cancer cells, demonstrating that cancer cells acquire metabolic alterations compared with normal cells (Warburg, 1956). Scientists discovered that in a hypoxic environment cancer cells exist at an acidic pH, because of the induction of glycolysis (Warburg effect) through Hypoxia-inducible factor 1 α (HIF1α), followed by lactate and extracellular efflux of hydrogen ions (H+), as well as the accumulation of lactate (Warburg, 1956). In addition, HIF is also known to induce the expression of genes related to, tumour vascularisation, metastasis and invasion (Wicks & Semenza, n.d.).

One of the other regulatory processes that allow cancer cells to adapt to the changing environment are RNA editing and splicing, that enable a dynamic of human transcriptome diversity. RNA editing allows alteration of the RNA sequences and introduction of mismatches, thus modifying the coding capacity of a given transcript (Frezza et al., 2023). The most prominent type of editing reaction is adenosine to inosine base-modification, , which eventually can cause an amino acid substitution in the protein product. Adenosine deaminases acting on RNA (ADARs) family of enzymes drive editing adenosine to inosine in double-stranded RNA. Structural similarities of these two nucleosides, cause inosine to be recognized as guanosine by translation machineries leading to protein recoding and further splicing change,

respectively (Nishikura, 2010; Tang et al., 2020). An alteration in RNA sequences induced by the editing machinery affects gene expression and can contribute to the acquisition of a tumorigenic and aggressive phenotype (Frezza et al., 2023). Therefore, proteins translated from edited RNA allow tumour cell to adjust and adapt to the rapidly changing microenvironment.

Cancer-associated RNA splicing is another molecular feature that characterizes almost all tumour types. Alteration in splicing machinery can arise from both mutational changes and altered expression of trans-acting factors responsible for splicing catalysis and regulation (Tang et al., 2020). Deregulation of RNA splicing can drive tumourigenesis through diverse paths, such as contribution to increasing cell proliferation, avoidance of apoptosis, chemo-resistance, evading immune surveillance, enhancing cell migration and thus potential for metastasis (Bradley & Anczuków, 2023; X. Dong & Chen, 2019). Recent studies demonstrated therapeutic benefits of correcting or antagonizing cancer-associated isoforms that play critical roles in cancer cell transformation by using small molecules that modulate or inhibit RNA splicing (Pierson et al., 2019).

Cancer genome studies

Advances in sequencing technologies over the years allowed us to broaden the knowledge about mutational load of various cancers (Broseus & Ritchie, 2020; Khurana et al., 2016; Wu et al., 2020). Without a doubt, the biggest achievement in studying cancer mutational landscape was establishment of the Cancer Genome Atlas (TCGA) project, which studied thousands of samples and led to creation of large databases containing a comprehensive catalogue of recurring mutations. Over the years, growing number of studies allowed to link certain genes to cancer development and progression, as well as identifying more complex cancer subtypes based on their somatic mutations (Amar et al., 2017).

Recent studies have also shown the tumour mutational burden can serve as a predictive marker for responsiveness to immunotherapy in multiple cancer types (Rizvi et al., 2015; Snyder et al., 2014). Therefore, analysis of mutational profile and gene expression status can provide useful predictive biomarkers for chemotherapy and immunotherapy.

1.1.2 Role of p53

p53 is a transcriptional factor encoded by the TP53 gene on human chromosome 17. Full-length isoform contains 393 amino acids with molecular weight of 53kDa. Architecture of the p53 protein consists of several crucial domains, such as N-terminal transactivation domain, core DNA binding domain and an oligomerization domain at the C-terminus (Figure 1.2). Each of the protein regions has been shown to be responsible for a different p53 function (Lane, 1992).

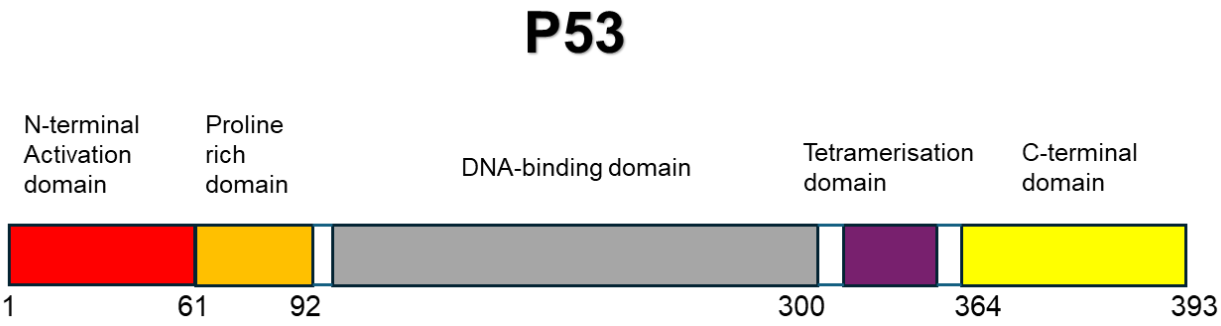


Figure 1.2 Domains of p53. Human p53 is composed of 393 amino acid residues and consists of N-terminal activation domain 1-61 (red), proline-rich domain 64-92 (orange), DNA-binding domain 100-300 (grey), tetramerisation domain 323–355 (purple), and C-terminal domain 364-393 (yellow).

The main role of p53 is regulation of the expression of genes involved in important cellular events, such as cell cycle, apoptosis, DNA repair, angiogenesis, and cellular senescence (Lane, 1992; Marei et al., 2021). In response to the aberrations in genetic material, p53 halts the cell cycle until the damage is repaired, thus preventing the propagation of DNA-defective cells and potential cancer transformation. By playing such critical role in regulation of complex intercellular networks that preserve genome integrity, p53 has been known as the “guardian of the genome”. Apart from the nucleus, p53 also localizes other cellular compartments, such as mitochondria, endoplasmic reticulum, and cytosol. Distribution of the protein in different cell

components regulates the transcriptional and non-nuclear functions of p53 (Zavileyskiy & Bunik, 2022).

Early experiments demonstrated that p53 interacts with an oncogenic simian virus 40 (SV40) to transform cells, lead to believe that p53 is a cellular oncogene (Lane & Crawford, 1979). Later studies showed that the protein acts more like a tumour suppressor based on its ability to repress oncogenic transformation in vitro and high susceptibility of tumour development in TP53 knockout mice (Attardi & Donehower, 2005; Donehower, 1996).

In the physiological state, p53 protein in maintained on the low level due to its short half-life. The turnover of p53 is regulated by its interaction with E3 ubiquitin ligase MDM2 (murine double minute-2) (S K et al., 2021, p. 2). MDM2 binds to N-terminal region and blocks the transcriptional activity of p53. This interaction targets p53 to ubiquitin-dependent degradation. However, various-DNA damaging agents lead to stabilization of p53 and increased level of the protein (H. Wei et al., 2021). It has been shown that under stress conditions, p53 dimers also accumulate at the outer membrane of mitochondria interacting with anti-apoptotic protein Bcl-2 leading to programmed cell death through mitochondrial membrane permeabilization (H. Wei et al., 2021).

In response to genotoxic stress p53 stability is regulated by various post-translational modifications (PTMs), such as phosphorylation, acetylation, methylation, glycosylation, ubiquitination, and sumoylation, of different protein regions (Marei et al., 2021). These modifications influence the protein ability to modulate cell proliferation and cell death (Marei et al., 2021). Diverse stress signals, including acute DNA damage, radiation, or oncogene activation, promote p53 phosphorylation, leading to disruption of the p53-MDM2 interaction, or directly inhibit MDM2 activity (Gencel-Augusto & Lozano, 2020). As a result, p53 protein accumulates in the nucleus, forms a stable tetramer and binds to target genes causing activation or inactivation of their transcription (Gencel-Augusto & Lozano, 2020). Among genes regulated by p53 are CDKN1A (P21), GADD45, BAX, PUMA, NOXA, and TIGAR. p53 protein binds DNA and stimulates p21 production, which then interacts with a cell division-stimulating protein (cdk2) inhibiting its kinase activity and resulting in G1-phase arrest (G. He et al., 2005).

p53 is preventing the accumulation of mutations that could drive tumour development and is counteracting the effects of oncogenic mutations. Various studies have shown additional roles of p53 protein in tumour suppression, such as regulation of tumour metabolism, microenvironment signalling, induction of ferroptosis (iron-depended programmed cell death) and cancer stem cell renewal or differentiation (Gnanapradeepan et al., 2018). Therefore, TP53 inactivation drives proliferation, invasion, and cell survival, clearing the way for cancer progression and metastasis.

Bert Vogelstein and his colleagues were first to identify somatic mutations of TP53 in the study of colorectal cancer (Baker et al., 1989), although soon it became clear that the gene is highly mutated across various cancer types (Albino et al., 1994, p. 53; Baker et al., 1989). Further systemic studies demonstrated that alterations on both alleles of p53 are the most common mutational signature of all human cancers (Hanahan, 2022). Inactivation of p53 by either mutation or deletion has been detected in more than 50% of human cancers. Some of the detected hot spot mutations in p53 gene include codons 175, 248, 273, and 282 (Baugh et al., 2018).

TP53 mutations can be classified as structural or DNA-binding mutations, which influence either p53 protein folding or its transcriptional activity. Most p53 mutations are localized within the DNA binding domain that span ~ 190 different codons (Baugh et al., 2018). Mutations within that region may either disrupt p53-DNA interactions directly or through the modification of the domain confirmation. It has been noted that the top 10 most frequent mutations account for ~ 30% of all missense mutations, which indicates that there are preferential mutation sites in TP53. This mutant preference may come from the selection of the mutated TP53 alleles that are contributing the tumour phenotype the most (Baugh et al., 2018). Additionally, carboxyl terminus of p53 protein has been shown to control p53 DNA-binding activity. Thus modifications of that region by phosphorylation, deletion or binding of antibody can convert p53 from latent to active state for DNA binding (Delphin & Baudier, 1994; Hupp et al., 1992).

According to TCGA studies, more than 64% of the mutations in p53 are missense that result in single amino acid substitutions, 33% mutations are producing

premature termination codons (PTC) that cause unstable truncated p53 protein or complete lack of its expression due to nonsense-mediated mRNA decay (NMD) (S.-Y. Lee et al., 2019). The most frequent nonsense TP53 mutation are R213X and R196X, R213X being present in about 1% of all human cancers (M. Zhang, Heldin, et al., 2018a). Lee et al showed that K120R missense mutation disables p53 protein acetylation and therefore inactivates its apoptotic function (S.-Y. Lee et al., 2019). They noticed that the mutation may create a novel splicing site, that results in an aberrant transcript variant with a frame-shifted premature termination codon which induces nonsense-mediated mRNA decay. This discovery provided a mechanism that links the missense mutation and tumourigenesis (S.-Y. Lee et al., 2019).

Discovered mutations in TP53 gene present not only loss of wild type p53 activity, but also mutant p53 protein can gain novel oncogenic properties. Studies show that overexpression of mutant p53 increases the resistance of tumour cells to drug-induced apoptosis (Blandino et al., n.d.; Datta et al., 2017). Some forms of mutant p53 protein use characteristics of wild-type p53, such as interactions with the transcriptional machinery, to induce oncogenic state (Datta et al., 2017). Molecular mechanism of the gain of function (GOF) phenotype alters with different p53 mutants and relates to the differences in structural changes caused by those mutations (Datta et al., 2017). In cell expressing tumour-associated p53 variants, altered protein was shown to promote cellular growth, enhanced drug-resistance, increased invasion, and metastasis (Pfister & Prives, 2017). Recent studies indicate that mutant p53 protein can also be selectively packed into extracellular vesicles and taken up by neighbouring cancer cells and to macrophages, thus stimulating the release of tumour-supportive cytokines and subsequently establishing a positive feedback loop for tumour progression (Bhatta et al., 2021).

Transcriptional activity of p53 relies on the formation of tetramers (Gencel-Augusto & Lozano, 2020). It has been observed that most p53 mutants show an intact oligomerization domain, therefore the mutated protein can inactivate p53 function by binding to the wild-type protein forming hetero-oligomer (Gencel-Augusto & Lozano, 2020). This phenomenon has been described as a dominant-negative effect. Boettcher et al. reported that mutant p53 inactivates WT p53 in response to DNA damage (Boettcher et al., 2019). Scientists demonstrated that mutant p53R248Q in p53R248Q/+ acute myeloid leukaemia cells inhibited the ability of the

wild-type p53 protein to activate CDKN1A expression similar to p53-null cells (Boettcher et al., 2019).

In cancers that do not carry TP53 mutation, p53 is inactivated through protein-protein interactions that lead to increase in p53 degradation or inhibition of p53 activity (Pfister & Prives, 2017). In tumours that retain wild type p53, the transcriptional activity of the protein is often impaired through alternative mechanisms, which include methylation of p53 promoter, overexpression of p53 negative regulators MDM2 and MDMX or inactivation of p14/ARF. The p14(ARF) is a small protein that regulated the activity of MDM2 by displacing p53 and preventing its proteasomal degradation (M. Zhang, Heldin, et al., 2018b).

Abnormal cellular localization of p53 protein is one of the other mechanisms related to the protein inactivation. It's been discovered that cells showing p53 cytoplasmic accumulation are less responsive to DNA damage (Cai & Liu, 2008).

Not only lack or inactive p53 causes cancer, but also overexpression of p53 can lead to various pathologies, such as premature aging, developmental changes, neurodegeneration, and can also cause adverse effects during cancer treatment (Inoue et al., 2012).

The understanding of the crucial role that p53 plays in tumourigenesis

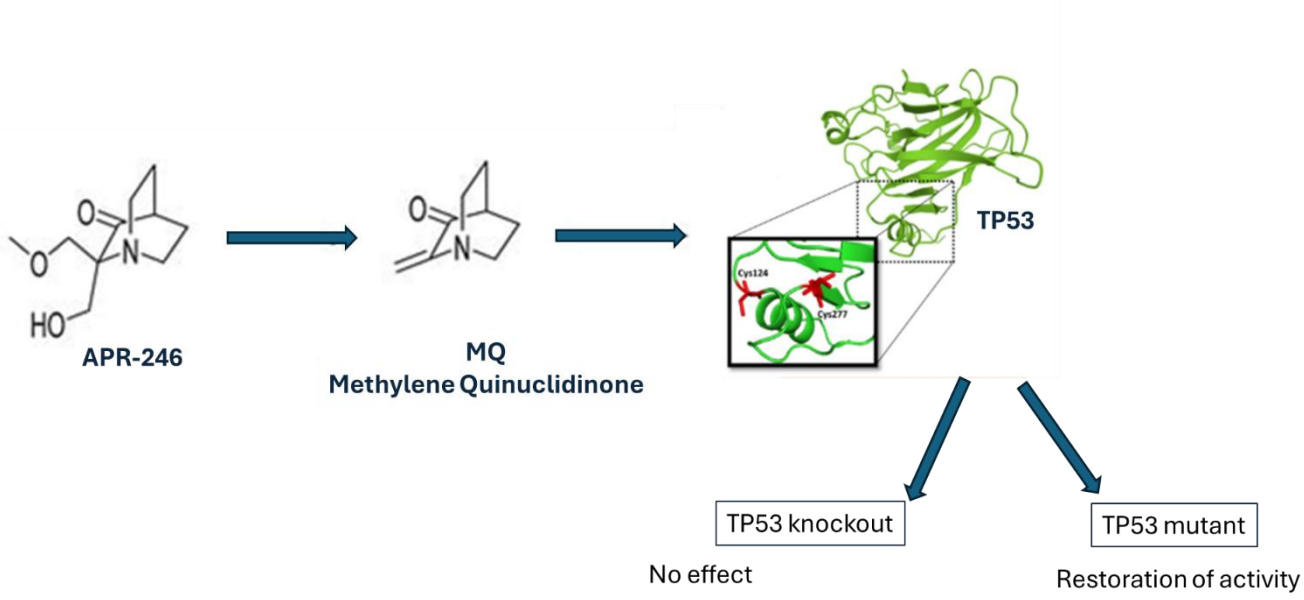


Figure 1.3 Effects of APR-246 on mutant and wild-type TP53 activity. APR-246 is converted to the reactive form methylene quinuclidinone (MQ), which binds covalently to p53 core domain. Cysteine 277 and Cysteine 124 were identified as a prime binding target for MQ and essential for MQ-mediated thermostabilization of p53 wild-type, and functional restoration of mutant p53 in cancer cells. No effect has been shown on TP53 knockout cells.

generated considerable interest in developing therapeutic approaches aiming to restore the p53 functions (M. Zhang, Heldin, et al., 2018b). Several studies have demonstrated that transfecting cancer cells with plasmids carrying wild type p53 lead to activation of apoptosis pathway and/or growth arrest. A variety of potential p53-restoring compounds are currently in the clinical trials, such as APR-246 (McCubrey et al., 2022). This prodrug once metabolised to methylene quinuclidinone (MQ), binds to the critical cysteine residues in the binding domain of mutant p53 and restores its wild-type conformation (Figure 1,3). Additionally, APR-246 generates reactive oxygen species (ROS), which alter the structure of the mutant TP53 protein (McCubrey et al., 2022; Q. Zhang et al., 2018). Therefore, understanding the precise role that p53 plays in cancer development can lead to developing more therapeutic approaches targeting this tumour suppressor gene.

1.1.3 Tumour immune environment

Tumour microenvironment (TME) plays a crucial role in tumour differentiation, epigenetics, dissemination, and immune evasion. TME is highly heterogeneous, consisting of cancer cells, tumour stromal cells including cancer-associated fibroblasts (CAFs), pericytes, endothelial cells and immune cells, such as macrophages, microglia, and lymphocytes, as well as the non-cellular components of extracellular matrix like collagen or laminin (Baghban et al., 2020).

Tumour-TME interaction regulates cancer progression in either positive or negative fashion (Baghban et al., 2020). At the early stages of malignancy, TME confers anti-tumorigenic functions, However, some cancer cells can tolerate the suppressive effect and induce reprogramming of the TME into one promoting malignancy (T. Liu et al., 2019). As the core of TME, cancer cells disturb the function of other components to use the non-cancerous cells to work for their own benefit. That dynamic interaction of tumour cells with their microenvironment promotes cancer heterogeneity, clonal evolution, deficient response to therapy or even multi-drug resistance and metastasis (Labani-Motlagh et al., 2020).

Hot and cold tumours

A plethora of studies confirmed that alongside the traditional hallmarks of cancer, immune-modulating mechanisms are also pivotal characteristics of tumour cells (Binnewies et al., 2018; Jhunjhunwala et al., 2021). Immune expression profiling, looking at variety of produced cytokines, chemokines, and cell surface proteins, can be used to assess the tumour inflammatory status and to classify the lesion as “hot” or “cold” (Binnewies et al., 2018; L. Wang et al., 2023). Hot tumours show high number of tumour-infiltrating lymphocytes (TIL) and are more sensitive to treatment with immune checkpoint blockers monotherapy or combination therapy. Whereas, cold tumours are poorly immunogenic tumours, display low number or lack of TIL and have a much lower likelihood of response to immune checkpoint blockade therapy (Binnewies et al., 2018).

Tumour-infiltrating immune cells have a significant impact on the clinical response to immunotherapy (Xie et al., 2021), and it has been identified that not only the composition of TME, but also the localization and functional orientation of immune cells determine the anti-tumour immune response (Finotello et al., 2019). Studies show, that better therapeutic prognosis display patients with high densities of specific immune cell subpopulations in the tumour core (Finotello et al., 2019).

High number of CD8+ T-cells, TH1-like chemokine, and interferon (INF) expression are characteristics of T-cell inflammation (Apcher et al., 2011). Cytotoxic T-cells are the core of the tumour immune control and response to therapies (Phan et al., 2003). Activation of CD8+ T-cell-mediated immune response requires presentation of peptides on the cell surface by major histocompatibility complex (MHC) class I molecules or as defined in human HLA (human leukocyte antigen). Processing and presentation of antigen by MHC I pathway is a complex mechanism, which requires multiple steps during which endogenous cytosolic proteins are degraded by proteasome into 8-11mer peptides, loaded onto MHC molecules, and presented on the cells surface (Apcher et al., 2011). Recognising these peptide-MHC complexes as foreign on dendritic cells in draining lymph nodes prompts CD8+ T cell activation, release of the pro-inflammatory cytokines and eradication of the affected cells. Identification of which specific peptides are being presented on MHC class I is crucial for development of diagnostics and therapeutic approaches for

cancer (Koşaloğlu-Yalçın et al., 2022). Interestingly, when Moldoveanu et al. investigated the immune microenvironment of melanoma patient samples, they observed that the patient's responsiveness to immune checkpoint blockade was correlated with proximity of antigen-experienced cytotoxic T cells to melanoma cells intratumorally, indicating that special organisation within tumour tissue also plays a part in activation of the immune response (Moldoveanu et al., n.d.).

CD4+ T cells are another type of important T cell in the TME. Population of CD4+ T cells includes T helper type 1 cells, T helper type 2 cells, T helper type 17 cells, and regulatory T-cells (Tregs) (Xie et al., 2021). Recognition of endosomal degradation products and processed extracellular proteins by CD4+ T-cells was shown to play a crucial role in cancer immune landscape (Vyasamneni et al., 2023). Activation of these special immune cells happens upon the encounter with their cognate antigen displayed on the surface of antigen-presenting cells (APCs) by major histocompatibility complex class II (MHC II) molecule (Vyasamneni et al., 2023). In tumour immunity, some CD4+ T cells are able to induce the anti-tumour function of cytotoxic T cells by promoting their infiltration of TME, maintaining their activities, and leading in antigen presentation (Vyasamneni et al., 2023). Physiologically, Tregs suppress effector T cells inhibiting autoimmunity or excessive immune response; however, immunosuppressive activity of Tregs in tumour side creates an obstacle for anti-tumour response favouring tumour progression. Therefore, inhibition of the immunosuppressive function of tumour-infiltrating regulatory T cells has become a promising subject in tumour immunotherapy research (Xie et al., 2021).

Understanding role of TME and its components in tumorigenesis has become an emerging topic in cancer research (Labani-Motlagh et al., 2020). One of the cell types found in tumour microenvironment are Cancer-associated fibroblasts (CAFs), which have been shown to provide physical support for epithelial cells, but also secrete various cancer-promoting and pro-angiogenic cytokines and chemokines, therefore having an extensive contribution to an immunosuppressive TME (T. Liu et al., 2019).

Taking in consideration different aspects of the TME is vital when designing new therapies and new treatment strategies for cancer patients. Significant

technological advances have allowed researchers to molecularly characterise tumours and responding immune cells, which has resulted in breakthroughs that have translated to pharmacologically actionable markers and targets (T. Liu et al., 2019).

Immune cell evasion

Large number of studies demonstrated that both innate and adaptive immunity are involved in immune surveillance, which is described as tumour elimination phase (Brogden et al., 2016; O'Neill et al., 2018). However, if the transformed cells evade immune control (escape phase), tumours are formed. While cancer progresses, the immunosuppression increases in magnitude, causing continued tumour growth (equilibrium phase) (XCaroline Jansen et al., 2018).

It has been shown that anti-tumour immune response is mainly led by natural killer (NK) cells, CD4+ helper T (Th) and CD8+ cytotoxic T (Tc) cells alongside proinflammatory macrophages (M1) and dendritic cells (DC) (Labani-Motlagh et al., 2020). On the other side, tumour immunity is counteracted by myeloid-derived suppressor cells (MDSCs) and Foxp3+ regulatory T cells (Tregs), which are attracted and expanded by tumour cells to control the present effector lymphocytes and to inhibit the activation of new tumour-reactive lymphocytes (Labani-Motlagh et al., 2020).

Tumour immune cell evasion is achieved by various methods such as downregulation or loss of expression of tumour antigens, release of immunosuppressive extracellular vesicles (EV) or molecules including Il-10 or transforming growth factor β (TGF-β), overexpressing programmed death ligand 1 (PD-L1), and many others (Bhatta et al., 2021; Johnson et al., 2016; O'Neill et al., 2018). Recognition by cytotoxic T cells and antigen presenting cells (APCs) can be avoided by loss of the ability of tumour antigen presentation, as well as downregulation of MHC-I and II complex (Johnson et al., 2016) Bhatta and colleagues discovered that mutant p53 proteins are encapsulated in EVs and distributed to neighbouring cells to potentially reprogram immune cells and therefore creating a positive feedback loop enhancing cancer progression (Bhatta et al., 2021).

One of the most crucial aspects of immune activation is stimulation of dendritic cells. If DCs patrolling tumour microenvironment is not properly matured, presentation of tumour antigens to T-cells can result in development of anergic/tolerant T-cells (Labani-Motlagh et al., 2020). Immature DCs in TME are characterised by low expression of costimulatory molecules such as CD80 and CD86 and high occurrence of inhibitory molecules such as PD-L1 and cytotoxic T lymphocyte associated antigen-4 (CTLA-4) (Labani-Motlagh et al., 2020). Induction of these tolerogenic dendritic cells has been shown to be achieved by various factors including cytokines released by tumour (e.g., TGF-β and IL-10), products of pathogens and regulatory T-cells secreting IL-10. Moreover, uptake of apoptotic/necrotic DC can convert immature DC into tolerogenic dendritic cell (Kushwah et al., 2010),

A previously stated, tumour cells escape immune surveillance by utilizing several mechanisms, such as production of immunosuppressive cytokines or recruitment of cells that induce immune tolerance (Rosenthal et al., 2019). Therefore, an in-depth analysis of complexity within the tumour immune composition is necessary to identify successful therapeutic strategies, that will surpass the tumour immune evasion and provide a clinical benefit.

It has been shown that anticancer immunity can be therapeutically utilized by using combinations of immune checkpoint blockers and anti-angiogenic inhibitors (W. S. Lee et al., 2020). Thus, the adequate determination of cancer immunity requires the identification of the following tumour features: neoantigens, immune complexity of tumour microenvironment (TME), patient characteristics and environmental factors. Recent technological advancement in techniques such as high-resolution sequencing, flow cytometry, imaging, mass spectrometry or immunopeptidomic studies can provide an unprecedented view of the composition, function, and location of immune cells within the tumour microenvironment (De Filippis et al., 2022; Yadav et al., 2014).

1.1.4 Immunotherapy

The immune system has been known to play an important role in the control of tumorigenic cells. The main reason why tumours are not recognised by the immune system is that, unlike pathogens, they do not express potent tumour rejection antigens. Therefore, the focus of the novel therapeutic approaches is to stimulate the patients' own immune system to eradicate cancer. Vast majority of immunotherapeutic studies focuses on CD8+ T-cells, that recognise antigenic peptides presented by MHC class I complex. Once activated, T cells undergo clonal expansion and differentiation into cytotoxic T lymphocytes (CTL), that are able to pick out and kill cancerous cells.

It has been shown that tumour microenvironment can adopt many strategies to inhibit antitumour immunity and finally achieve immune escape (M. Liu & Guo, 2018). One of these strategies focuses on the expression of inhibitory ligands that bind to the so-called 'immune checkpoints' expressed on immune cells, this leads to induction of inhibitory signals interfering with T-cell activation and function. Therefore, discovery of drugs preventing malignant cells from downregulating host cell immune responses, immune checkpoint inhibitors (ICI), has been a revolution is cancer immunotherapy (M. Liu & Guo, 2018).

The first ever ICI introduced to clinic was monoclonal antibody ipilimumab, that targets cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) (Phan et al., 2003). CTLA-4 is expressed on some T cells, such as regulatory T cells, which regulate self-tolerance (Oshi et al., 2020), as well as activated CD4+ T cells (Van Allen et al., 2015), which are responsible for generation of antitumor activation of the immune system . Expressed on tumour cells CTLA-4 allows tumours to escape the immune response. Therefore, the overall effect of CTLA-4 inhibition is an increase in the immune system's ability to recognize and eliminate cancer cells. It has been demonstrated that patients with tumours displaying high somatic mutational burden, such as melanoma, show a more favourable response when treated with PD-1 or CTLA-4 immunotherapy (Rizvi et al., 2015; Snyder et al., 2014). Despite association with the clinical response, high mutational burden alone is not sufficient to clearly point to the immune checkpoint benefits.

Another breakthrough in immunotherapy field, was a development of the anti-programmed cell death protein 1 (anti-PD-1) antibody Nivolumab, that received FDA approval in 2011 (D'Angelo et al., 2018). Expression of programmed death-ligand 1 (PD-L1) on tumour cells, leads to inhibition of the PD-1 expressing activated T-cells, weakening their function, decreasing cytokine production, and on result allowing the tumour cell to escape attack by the body's immune system. Studies demonstrate immune checkpoint inhibition therapy targeting PD-1 and PD-L1 significantly improved overall survival rate of patients with melanoma (30-40%), Hodgkin's lymphoma (80–90%), non-small cell lung cancer (15–20%) and others (Topalian et al., 2012). The activity of anti-PD-1 relies on the presence of antigen-specific T cells at the tumour site and rescuing T cells against tumour antigens spontaneously induced by the tumour. This leads to the conclusion that combining anti-PD-1 with cancer vaccine would boost already present antitumour T cells and trigger 'de novo' T cells for improved response rate.

ICI challenges and response markers

Despite being an advanced therapeutic option, only 20–30% of patients with cancer show response to ICI treatments (Snyder et al., 2014). Due to complexity of the immune system, it is difficult to predict these responders before treatment. Several factors are being discussed that might influence the clinical outcome. HLA-DR expression status has been proposed as one of the response markers. This protein is the main molecule involved in MHC-II antigen-presenting pathway and its expression has been discovered in some cancer cells, strongly correlating with the efficiency of anti-PD-1 therapy (Johnson et al., 2016). Higher response to immunotherapeutic treatment has been seen in cancers associated with oncogenic viruses (like human papillomavirus (HPV), Epstein Barr virus) . This phenomenon has been related to generation of viral antigens and elevated PD-L1 expression in cancer cells (Haddad et al., 2022).

Gong et al, focused their efforts on using bulk and single-cell transcriptomic data from cohorts of melanoma patients undergoing treatment with ICI to create a computational model that creates clusters based on the expression of an antigen-presenting machinery (APM) signature consisting of 23 genes (Gong & Karchin, 2023). Scientists were able to distinguish four APM clusters associated with distinct

immune characteristics, cancer hallmarks, and patient prognosis. The role of this model was to predict differential regulation of APM genes during ICI treatment, which drove ICI responsiveness. Interestingly, while tumours described as immunogenically hot with high baseline APM expression prior to treatment were correlated with a better response to ICI than cold tumours displaying low APM expression, researchers identified a subset of hot tumours with the highest APM expression before treatment, that failed to upregulate APM expression during treatment. Additionally, these lesions displayed infiltration of exhausted T cells. However, tumours associated with the best patient prognosis demonstrated significant APM upregulation and immune infiltration following treatment. The study provided some interesting insights into the immune escape mechanisms detected in hot tumours that compromise the efficiency of immune checkpoint blockers (Gong & Karchin, 2023).

It has been shown that disorganized tumour vasculature vessels impede CD8+ T cell migration into tumour microenvironment, disable their effector functions, and even lead to T cell death (W. S. Lee et al., 2020). Moreover, VEGF, the key factor driving angiogenesis, disturbs maturation of dendritic cells, resulting in inhibition of T cell priming (Yang & Cao, 2022).

Meanwhile, protumoural M2-like macrophages as well as helper T-cells 2 (TH2) and regulatory T cells (Tregs) secrete factors accelerating uncontrolled angiogenesis and promote vascular immaturity (W. S. Lee et al., 2020). On the other hand, CD8+ T and CD4+ TH1 cells have been shown to suppress angiogenesis and promote vascular maturation by IFN-γ secretion, however due to malformed tumour vasculature they are unable to infiltrate the TME (W. S. Lee et al., 2020). These studies opened the gate for novel therapeutic approach that would focus on simultaneous targeting of tumour vessels and immunity for cancer immunotherapy (W. S. Lee et al., 2020).

Other immunotherapeutic strategies

Aside of immune checkpoint inhibition, there are other cancer immunotherapy strategies that have found their place in clinics. Such as adoptive T-cell transfer (ACT) therapy, which involves boosting the natural ability of T cells to fight cancer. In this strategy, immune cells are sourced either from tumour as tumour-infiltrating lymphocytes (TILs) or from peripheral blood. After harvesting, T cells undergo activation

and expansion, before getting injected into patient's bloodstream (Lauss et al., 2017). ACT have demonstrated very high response rates in clinical trials with long-lasting complete tumour regression up to 20–25% of treated patients (Lauss et al., 2017; Svane & Verdegaal, 2014). However, just like with ICI, prediction of patient response is a complex subject. Lauss et al. studies patient's whole exome and transcriptomic data to find molecular characteristics correlated to the clinical benefit. Scientists demonstrated that improved progression-free and overall survival in response to ACT was associated with significantly higher mutation and predicted neoantigen load. Additionally, expression of immune activation signatures including a high MHC-I antigen processing and presentation score were also found to contribute to clinical benefit (Lauss et al., 2017).

When mentioning utilizing transfer of immune cells in therapy, we cannot forget about the revolutionary strategy of chimeric antigen receptor T cell (CAR T) treatment. CAR T approach combines the specificity of a monoclonal antibody targeting a cell surface antigen with the benefits of an ongoing T-cell-mediated response (Littman & Hexner, 2017). Collected T cells undergo retro- or lentiviral transduction with genetic constructs that carry tumour-specific extracellular antigen receptors with the intracellular workings of a T cell, that include T-cell costimulatory domain. These constructs allow for engagement of the CAR with a target cell leading to activation and proliferation of CAR T cells. After manufacturing, CAR T cells are infused back into the patient from whom the cells were collected in hopes for targeting and selectively eliminating tumour cells. Many ongoing clinical trials investigate the use of different extracellular domains for use in CAR to induce to most efficient anti-tumour response (Littman & Hexner, 2017). The approval of the first CAR T cell therapy was a landmark event in cell and gene therapy and provided a new paradigm for the treatment of many different diseases, harnessing the intricate and powerful effects of naturally occurring T cells and redirecting their response towards chosen targets.

Peptide vaccines are a promising direction to create a precision immunotherapeutic approach for cancer, autoimmune diseases, and allergies (Calvo Tardón et al., 2019). Tumour vaccination's aim is to stimulate an immune response targeted to, mostly weak, antigens expressed by tumour cells. The main challenge of

effective vaccination is to find potent and broadly expressed target, that would stimulate a robust and durable immune response. The evolution of novel technologies such as next-generation sequencing has improved our ability to search for new immune targets such as mutation-derived antigens (neoantigens) and accelerated the development of novel immune therapies.

Current strategies of cancer vaccination involve the use of different carriers, such as cationic lipids with mRNAs encoding tumour antigens, RNA-lipoplexes (Kranz et al., 2016), adenoviruses (D'Alise et al., 2019), nanoparticles, plasmid DNA and many others (Gilboa, 2016; Paston et al., 2021).

The most common approaches to cancer vaccination, including melanoma vaccines, utilize shared tumour-associated antigens (TAAs) that are self-antigens abnormally expressed or overexpressed in tumours (Vansteenkiste et al., 2016). However, due to presence of immune tolerance therapeutic vaccines based on TAAs have been largely unsuccessful (Vansteenkiste et al., 2016). Additionally, these antigens are also expressed by non-malignant cells, which increases the risk of vaccine-induced autoimmune toxicities.

This led to development of more tumour-specific strategies involving neoantigens as their target. Subject of neoantigen discovery and characteristics will be explored in detail in the next chapters. First demonstration of the antitumour efficacy of a neoantigen-based vaccine was reported in the B16F10 melanoma model by Castle et al (Castle et al., 2012). Researchers identified 563 neoantigens expressed in studied model. The ability to induce immune response was evaluated by immunizing mice with long peptides encoding the mutated epitopes.
Later, Yadav et al, combined exome sequencing and mass spectrometry approach to identify neoantigens in MC38 tumour model (Yadav et al., 2014). Vaccination of mice with designed peptide vaccine confirmed immunogenicity and tumour control yielding therapeutically active T-cell responses. (Yadav et al., 2014).

The work by D'Alise *et al.* on viral vector-based vaccines carrying multiple tumour neoantigens showed that immunization with the adenoviral construct encoding 31 neoantigens selected from the CT26 cells induced a robust CD8+ and CD4+ T-cell response and along with anti-PD-1 strategy lead to eradication of large established murine lesion (D'Alise et al., 2019).

Some success stories in the cancer vaccine field passing the FDA approval are: the use of Bacillus Calmette–Guerin (BCG) for local treatment of early-stage bladder cancer (Guallar-Garrido & Julián, 2020), and Sipuleucel-T, an autologous dendritic cell-based vaccine, for treatment of castration-resistant prostate cancer (Kantoff et al., 2010). Luckily, in recent years the number of therapeutic approaches harvesting the power of immune system have been stably increasing their presence in the clinical trials.

The success of cancer immunotherapy also relies on the understanding of tumour microenvironment and the immune evasion mechanisms in which the tumour, infiltrating immune cells, and stroma interact with each other creating a complex network (Labani-Motlagh et al., 2020). Although application of immunotherapy has been increasing over the years, we are still far from understanding how to use different anticancer strategies in the most efficient way and how to combine different therapeutic options to optimise clinical outcome.

1.2 Neoantigens

Most components of the immunopeptidomic landscape derive from ‘normal’ self-proteins processing and only small fraction originates from tumour-associated or tumour-specific proteins (Chen et al., 2019). Tumour-associated antigens (TAAs) are described as normal and unaltered proteins that are constitutively expressed or overexpressed in tumour cells but are also shown at lower levels in healthy cells. Some TAAs can be a product of genes normally expressed by cells during embryonic development, but not present in adult tissues, such as cancer/testis antigens (CTAs) (J. Liu et al., 2019).

Tumour-specific antigens or neoantigens are defined as proteins or other molecules that are found only on cancer cells and not on normal cells. Neoantigens can form when cancer cell-specific mutations occur that result in alterations of the protein from ‘self’ (Chen et al., 2019). Because neoepitopes are not presented by normal counterpart cells, or thymic epithelial cells, T cells recognising the neoepitopes are not subject to central tolerance in thymus. Hence, it is conceivable that neoantigens containing neoepitopes are in charge of eliciting strong and specific

host T cell responses against cancer (Kanaseki et al., 2019). Those tumour-specific markers can function as immune stimulatory vaccines and that property has been utilized in clinics.

1.2.1 Sources of neoantigens

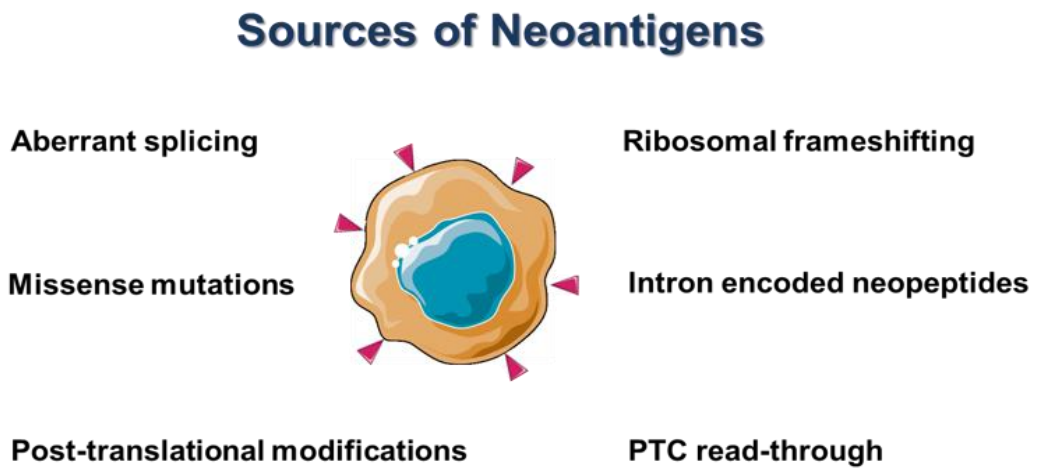


Figure 1.4 Sources of neoantigens. These immunogenic peptides have been found to be originating from various of sources, which include coding, non-coding sequences, translational errors, aberrant splicing events and many other.

The advances in deep sequencing and immunopeptidomic technologies allowed for a multi-dimensional identification of tumour-specific mutations that give rise to neoantigens. Main source of these immunogenic peptides are somatic nonsynonymous mutations in the expressed genes (Apcher et al., 2011). These changes can be identified on DNA and RNA level and the presence of neoantigen confirmed by mass spectrometry techniques.

However, exome accounts only for 2% of human genome, whereas up to 75% of the genome can undergo transcription and potentially translated into protein

(Animesh et al., 2022). Therefore, it has been demonstrated that numerous neoantigens can also originate from non-protein-coding regions (Laumont et al., 2018). So-called cryptic or noncanonical antigens can be derived as a result of posttranslational modifications, ribosomal frameshifting, RNA editing, intronic retention, codon readthrough, proteasomal splicing of peptides or altered antigen processing (Animesh et al., 2022; Bedran et al., 2022; Smart et al., 2018). Furthermore, some peptides can arise from translation of long noncoding RNAs (lncRNAs) and even pseudogenes (Figure 1.4).

Non-coding regions

It has been shown that some tumour-infiltrating lymphocytes recognise cancer-restricted non-mutated peptides bound to MHC (Laumont et al., 2018). These antigens can originate from a variety of genetic and epigenetic modification that forced the transcription and translation of genomic sequences normally not expressed in cells, such as endogenous retroelements (EREs). Laumont et al. studying murine datasets in search for tumour specific antigens, recognised that most of them derive from aberrant translation events, such as out-of-frame translation of a coding exon or the translation of non-coding regions (Laumont et al., 2018). Additionally, it was reported that any type of non-coding region can be a source of neoantigen: intergenic and intronic sequences, untranslated region/exon junctions, noncoding exons, and EREs, which showed to be a specifically rich source of tumour specific antigens. Scientists also identified novel antigen derived from structural variant, a very large intergenic deletion (Laumont et al., 2018).

Alternative splicing events

A pan-cancer TCGA study demonstrated a potential involvement of alternative splicing events in neoantigen production by detection of cancer-specific exon-exon junctions and confirming the expression of peptides originated from alternatively spliced transcripts using a proteomics database (Smart et al., 2018). Aberrant RNA splicing leads to generation of different mRNAs or transcripts containing introns.

Genome-wide analyses of alternative splicing shows that 40%-60% of human genes undergo alternative splicing, contributing to functional complexity of the human genome (Cheng et al., 2022). Aberrant splicing events in cancer cells can result in the translation of abnormal mRNAs into proteins or peptides presented by MHC class I molecules to T cells that can be detected by the immune system (R. F. Wang & Wang, 2017). Therefore, inhibition of splicing increase intron encoded peptides that leads to neoantigenic enrichment and potentially stimulating the antitumorigenic immune response. One of the drugs used for that purpose is Isoginkgetin that inhibits the function of both the major and minor spliceosomes (Tsalikis et al., 2019). Isoginkgetin has been also utilised in this study to expand landscape of potential neoantigens.

Splicing errors can lead to inclusion of an intron in the final mRNA transcripts, intron retention (IR). This phenomenon is an important mechanism for neoantigen prediction because many intron-retaining neoantigens have been detected by transcriptome sequencing and mass spectrometry (Smart et al., 2018). These aberrant mRNAs are translated and usually degraded by nonsense-mediated decay (NMD) pathway, therefore generating peptides for endogenous processing, proteolytic cleavage, and presentation on MHC-I molecules (Smart et al., 2018). Considering the fact, that introns are much longer than exons and are under less selective pressure to maintain open reading frames, there is a high chance that an IR transcript carries a PTC, and therefore is a good candidate for degradation via NMD pathway (Broseus & Ritchie, 2020). Intron retention is a common phenomenon in a wide variety of malignancies, such as pancreatic cancer, where high levels of transcripts with introns correlate with tumour progression (X. Dong & Chen, 2019). Dong et al found that in the presence of low tumour mutational burden, neoantigens originating from intron retention contribute to immune-mediated clearance of pancreatic cancer cells. In studied model, intron retention accounted for 16.5% of all alternative splicing events, which was second behind skipped exon (56%) (X. Dong & Chen, 2019). All together implied that IR-neoantigen load may affect the immune cell composition of the tumour microenvironment. These features also correspond with better prognosis and alongside expression of HLA-I could become a useful biomarker for selecting patients who may benefit from immune-checkpoint inhibitor treatments.(C. Dong et al., 2022). Cheng et al. showed computational strategies to

identify neoantigens originated from alternatively spliced transcripts, proposing a conceptual pipeline for the development of mRNA vaccine (Cheng et al., 2022).

RNA editing

Deregulation of RNA editing, as described previously, contributes to different types of human diseases, including cancers. Therefore, peptides translated from edited RNAs, edited peptides, may be presented on MHC becoming neoantigens.

Zhang et al identified using mass spectrometry five edited peptides originated from three editing sites (M. Zhang, Fritsche, et al., 2018). Comparative quantification between a set of tumour and healthy tissues allowed for assessment of potential tumour correlation. The edited peptides were elevated beyond normal levels in 7% of all tumours. RNA editing products, particularly the CCNI-ED10 peptide, were identified as immunogenic and therefore presented on HLA able to stimulate T cell immune responses (M. Zhang, Fritsche, et al., 2018). The study was a confirmation of another mechanism of neoantigen origin.

DRIPs

In the late 1980s, Thierry Boon and his team observed that cells transfected with tumour DNA fragments without a promoter sequence or carrying a frameshift or a termination codon upstream of the antigen coding sequence, could still be recognized by tumour antigen-specific CD8 + T cells (Boon et al., 1989). This discovery led to conclusion that the MHC-I peptides can be derived from small regions of the genome located in proximity to the antigen sequence and can be translated independently from the production of full-length transcripts or the corresponding protein. The concept was also supported by studies demonstrating that some tumour-infiltrating lymphocytess are able to recognise antigens derived from intronic sequences (Coulie et al., 1995), or alternative open reading frames (Rosenberg et al., 2002). In the mid-90s it was postulated that antigenic peptides for presentation on the MHC-I pathway can also originate from an unidentified source termed defective ribosomal products, or DRiPs (Yewdell et al., 1996). As DRIPs could be classified polypeptides produced as part of the pioneer round of translation,

premature termination of translation, and proteins that fail to fold properly or to assemble into their protein complexes.

Wei et al. demonstrated that ribosome modifications can selectively modify the generation of antigenic peptides derived from DRiPs, which raised the possibility of therapeutically manipulating DRiP translation to regulate immunosurveillance of pathogens, tumours, and autoantigens (J. Wei et al., 2019). Scientists reported that control of peptide generation lead by ribosome protein RPS28 links to cancer immunosurveillance, which suggests that mutations in ribosomal proteins common in cancers may be selected for immunoevasion from CD8+ T cells or NK cells (J. Wei et al., 2019).

MHC class I molecules allow CD8+ T cells to detect and eliminate cells in which the repertoire of antigens has been changed by the presence of viruses or other mechanisms altering the presentation of endogenous peptides (J. Wei et al., 2019). These antigenic peptides typically undergo proteasomal degradation, transport by TAP (transporter associated with antigen processing) into the endoplasmic reticulum (ER), trimming at their N-terminus, after processing peptides get loaded onto MHC class I molecules, and transported to the cell surface (Peters et al., 2003). However, despite playing a crucial role in immune surveillance, the origin of peptides presented by MHC class I molecules is still understudied.

The detection of peptides bound by MHC class I molecules is a highly sensitive process. Activation of specific T cells requires recognition of only a few peptide–MHC class I complexes and therefore can detect rare peptide products originated from noncanonical translation during the mRNA maturation process. Apcher et al., reported that pioneer translation products (PTPs) one of the substrates for MHC-I are produced by translation that is different to canonical event giving rise to full-length proteins (Apcher et al., 2011). Studies show that this translation event does not require the cap-binding translation factor eIF4E to occur and that mRNAs which stopped producing antigens for the MHC class I pathway still are able to produce full-length proteins (Apcher et al., 2011).

Studies indicated that synthesis of antigenic peptide substrates occurs on non-spliced transcripts and relies on whether the epitope is in frame with the start codon of the main open reading frame (ORF), which supported the hypothesis that

scanning of pre-spliced mRNAs by a noncanonical translation event lead to production of PTPs that form an important source of epitopes for the MHC class I pathway (Apcher et al., 2011). Scientists observed that inhibition of splicing machinery by drug isoginkgetin increased antigen presentation (Apcher et al., 2013). The base of this phenomenon has also been explored in the presented work.

The identification of immunogenic peptides, whatever their source, holds a great promise to development of efficient and targeted immunotherapeutic approaches.

1.2.2 Neoantigen discovery pipelines

The success of immune checkpoint antibodies indicate that the immune system can serve as a therapeutic tool. Studies demonstrate that tumours with high mutational burden and high numbers of predicted neoantigens have a more favourable response to immune checkpoint inhibitors (Animesh et al., 2022). Neoantigen cancer therapeutics are expected to be of benefit in cancers with a defined mutagen load or defined chromosome instability that generate novel fusion/mutant antigens. That is why, in order to improve the efficacy of the immunotherapeutic approach, it is crucial to better understand what are the characteristics of cancer cells that allow the immune system to recognise them as foreign.

Variant discovery

First step of neoantigen discovery pipeline requires identification of somatic mutations found in cancer. Most approaches include studies of DNA and/or RNA to detect variants on the genomic and transcriptomic level (Rezaei et al., n.d.; Sahin et al., 2017). Computational methodologies as applied to the cancer research field are emerging as approaches to define the expressed, mutated genome.

With achievements in sequencing technologies there has been also a plethora of software developed for comparisons of somatic changes to reference sequences. In the field of cancer studies, the most commonly used variant detectors are Mutect,

Varscan and CLC Bio (Benjamin et al., 2019, 2019; Faktor et al., 2018). As part of Genome Analysis Toolkit (GATK) Mutect 2 is coding-based software that utilizes several probabilistic models for genotyping and filtering, which could be used with and without a matched normal sample (Benjamin et al., 2019). However, without a doubt more user-friendly software is CLC Bio, which has been developed by Qiagen. CLC Bio is an integrated DNA and RNA variant identification software platform that allows for a comprehensive and accurate data analysis for genetic aberrations, transcriptomic analysis and many more (Faktor et al., 2018). A big advantage of the CLC bio application is that it is a software that does not require computational coding, which as a result opens the door for life-scientists to identify tissue or cell-specific variants relevant to the studied biological system.

Considerable efforts are being devoted to discovering neoantigenic peptides efficiently stimulating immune response that can be utilised in therapeutic cancer vaccines (Ashi et al., 2022). In silico prediction of neoantigens originating from mutated genes involves, identification and annotation of mutations found through whole-genome sequencing (WGS) or whole-exome sequencing (WES) data by comparison with normal tissue or publicly available reference data. WES or RNAseq are also use for genotyping of the patient's HLA genes (Finotello et al., 2019) . Knowing patient's HLA expression helps in prediction of peptides able to bind to these molecules. WES provides the deepest mutation coverage, however limiting the assay only to protein-coding regions may also restrict the number of predicted peptides originating from non-coding regions (Finotello et al., 2019). This provides an argument for identification of mutational changes found in both DNA and RNA.

Neoantigen predictions

Binding of peptides to HLA molecules can be predicted by various tools, that use machine-learning approach trained on data sets from large in vitro peptide–HLA binding studies (Jurtz et al., 2017a; X. Zhang et al., 2023). The most widely used methods are NetMHC and NetMHCpan, both based on artificial neural networks (Jurtz et al., 2017a). Binding affinity here is predicted as the half-maximal inhibitory concentration (IC50) represented in nanomolar units. To account for allele-specific bias the rank of predicted affinity is compared with a set of random natural peptides.

Peptides are classified as strong binders when their binding affinity is lower than 500nM or if they rank 0.5%. Recent advancements allowed for development of an improved tool NetMHC 4.0 in which the binding prediction algorithm is trained on binding affinity and eluted ligand mass spectrometry data, leveraging the information from both data type (Jurtz et al., 2017a).

With the identification of functional neoantigen-specific T cell receptors (TCR) a huge number of data has been gathered thus, to unify the knowledge NeoTCR database has been developed (Zhou et al., 2023). This immunoinformatic database collected publicly available analysis of TCR sequences with neoantigen specificities as well as characteristics of targeted peptides, from experimentally supported studies across variety of cancer subtypes. NeoTCR uses raw sequencing data to identify neoantigen specific TCRs, their clonotypes and excludes bystander viral-associated T cell receptors (Zhou et al., 2023).

Presentation of antigenic peptides requires transport from cytoplasm to endoplasmic reticulum (ER) by transporter associated with antigen processing (TAP) (X. Zhang et al., 2023). Bioinformatic approach platform DeepTAP has been applied for prediction of neoantigen presentation (X. Zhang et al., 2023). This software utilizes a sequence-based multi-layered recurrent neural network for estimation of TAP-neoantigen binding (X. Zhang et al., 2023). These studies show the increased interest in application of more advanced computational and machine learning methods neoantigen discovery.

In most studies neoantigen detection is based on predicted affinity of HLA molecules (Jensen et al., 2018; Jurtz et al., 2017b; O'Donnell et al., 2018). And although it could give a forecast of the potential neoantigen load, it does not reflect the in vitro state. Interestingly, only ~1–5% of the MHC class I binders predicted using different computational tools have been experimentally validated (Finotello et al., 2019). It is due to the fact that available computational methods may predict MHC binding, but they cannot predict other steps involved in MHC peptide processing. To overcome this limitation, the emerging field of immunopeptidomics comes to place. Direct detection of neoantigens by mass spectrometry is becoming a feasible method (Bassani-Sternberg et al., 2016).

One of the obstacles facing neoantigen prediction algorithms from the translational point of view is ensuring that potential vaccine epitopes are presented to T cells by MHC, what requires accurate predictions of proteasomal cleavage (Ziegler et al., 2022). Experimental procedures identifying proteasomal cleavage in vitro are inefficient and low throughput, however thanks to recent advances in high-throughput MHC ligandomics, it is possible to predict cleavage events from the termini of MHC-presented epitopes (Ziegler et al., 2022).

Identification of truly tumour specific antigens, that are shared among patients, expressed in all malignant cells and immunogenic is crucial to rapidly advance the development of cancer vaccines.

1.2.3 Proteogenomics

Mass spectrometry (MS) methods have many applications, including exploration of proteomics field from the first demonstration of peptide sequencing in 1970s (Nau & Biemann, 1976). Traditionally, mass spectrometry analysis output consists of a list of detected peptides and proteins they derive from. The abundance of each analyte is calculated by substation of the background signal from the peak area of the identified ion chromatogram(s). Determination of the peak area varies with the method used. It can be obtained from the unfragmented MS1 spectra or from tandem mass spectra (MS/MS or MS2) collected using methods like data independent acquisition (DIA) or data dependent acquisition (DDA) (Figure 1.5) (Reynisson et al., 2020). The peak area of detected ion chromatogram derives from the detector ion current, which is a measure of the number of charged molecules identified normalized by the time of signal sampling, therefore the measured signal is proportional to ions/second (Reynisson et al., 2020).Technology shift allowed for establishing a new field, proteogenomics, which focuses on combining mass spectrometry acquisition with information gathered by next generation sequencing (Kanaseki et al., 2019). Standard shotgun proteomic methods base peptide sequencing on matching tandem MS spectra from analysed sample to a reference protein sequence database (for example, UniProt), limiting peptide identification to only those encoded by the canonical reading frame of classic exons. Whereas in

proteogenomic studies, peptide sequencing is achieved by searching customized databases that include the six-frame translation of genomic or transcriptomic sequences, allowing for identification of peptides originated from all reading frames of any genomic region (Laumont et al., 2016).

Proteogenomic field enabled detection of peptides from regions that were classified as non-coding, such as non-coding RNAs, and are therefore not included in annotated protein databases (Xiang et al., 2021). Moreover, non-coding regions including long noncoding RNAs, short open reading frames (ORFs) and pseudogenes, were found producing proteins after all. Moreover, numerous peptides were shown to originate from non-canonical reading frames with non-AUG start

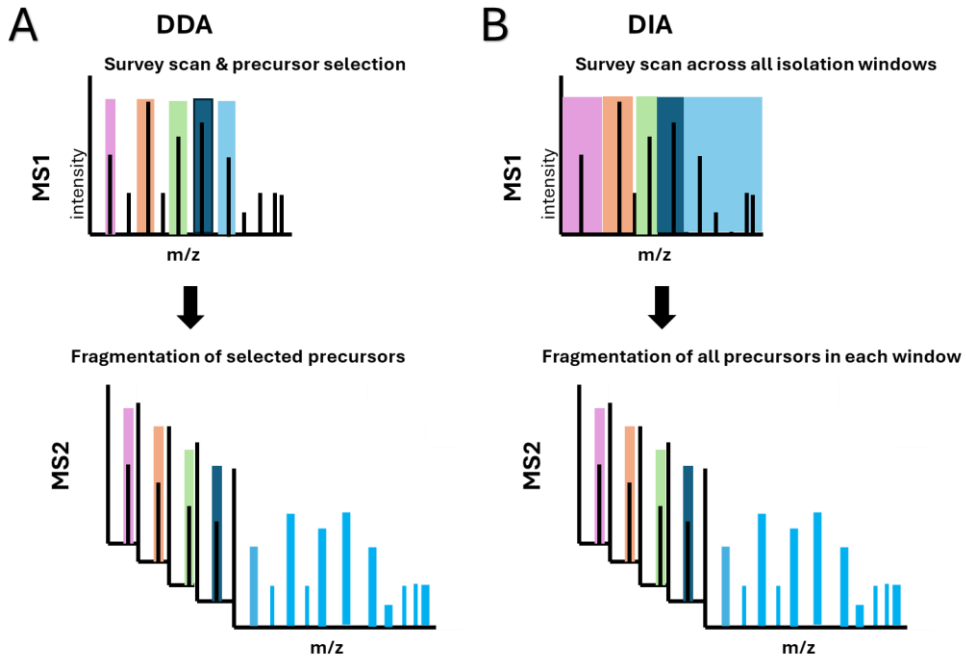


Figure 1.5 Schematic comparison of the DDA and DIA mass spectrometry methods. A) In Data Dependent Acquisition approach, the most abundant precursor ions are selected based on the survey scan (MS1) and these selected ions are later fragmented in MS2, B) In Data Independent Acquisition method the survey scan provides a snapshot of the precursor ions (MS1). Pre-defined wide isolation windows cover the whole MS1 m/z range and all precursor ions within each isolation window are fragmented in MS2.

codons (Ingolia et al., 2014).

There are several platforms developed recently for integration of DNA, RNA and protein data. 'Proteoformer' produces a complete protein synthesis database that can be used to identify peptides with mass spectrometry through the use of ribosome profiling (Verbruggen et al., 2019). The limitation of this approach is that

living cells are required to isolate bioactive ribosomes and this method might not be utilized for analysis of frozen tissues from the clinic. Other tool focuses on automation of spliced variants detection in cell systems which is especially powerful in cancer genomes where there might be DNA fusions, cancer-specific splicing, and trans-splicing (Krasnov et al., 2015). Spliced variant detection algorithms are always improving, especially those that capture the pathological, heterogeneous splicing specific to cancer cells. Development of novel R-packages iterates innovations in identifying expressed mutated genes (Wen et al., 2014). Translation toolkits have been generated that aim to produce a theoretical total polypeptide space of a genome using RNAseq that captures six-frame genome translation (Zickmann & Renard, 2015). These examples highlight the types of several bespoke algorithms that generate cutting-edge information on the proteogenomic landscape. As the diversity in software and computational tools tend not to be benchmarked against each other, end-user compatibility, especially crossing different disciplines, can be limited or non-accessible.

Emerging proteogenomic analyses led to conclusion that the proteome is more complex than previously thought. It has been also hypothesised that proteogenomics allows to discover the contribution of proteins translated from non-canonical mRNAs to the repertoire of major histocompatibility complex (MHC) class I-associated peptides (Laumont et al., 2016). These endogenous peptides bound to MHC are collectively referred to as the immunopeptidome and will be discussed in the next chapter.

1.2.4 Immunopeptidomics

Immunopeptidomic studies focus on identification and quantification of sample-specific repertoires of MHC-presented peptides following purification and measurements with liquid chromatography coupled with mass spectrometry. Immunopeptidome sequencing was pioneered by Hunt et al. in the 1990s (Hunt et al., 1992).

Elution HLA-bound peptides from tumour cells and subjecting these to mass spectrometry analysis, allows for the identification of peptide sequence by matching

the resulting spectra against reference databases (Fujiwara et al., 2022). Isolation of MHC-bound peptides (MBPs) has been performed mainly using two established methods named immunopurification/immunoprecipitation (IP) and mild acid elution (MAE). IP is an antibody-dependent method, that utilizes HLA-specific antibodies that recognize the pan-MHC or specific HLA alleles. Fujiwara et al, utilized immunoprecipitation to identify HLA class I (HLA-I) and HLA class II (HLA-II)-restricted peptides from pancreatic ductal adenocarcinoma patient biopsies using affinity purification columns and MS peptidome analysis (Fujiwara et al., 2022). Scientists detected T cell epitopes shared across multiple patients with expression of different HLA types and those containing sequences of both anti-HLA-I and HLA-II peptides. Epitopes identified in the study bound non-matched HLA molecules and activated T cell response in peripheral T cells from both HLA-type matched and non-matched patients. Additionally, peptides that contained HLA-I and HLA-II epitopes were able to induce polyfunctional cytokine responses in peripheral T cells (Fujiwara et al., 2022). Immunoprecipitation enables identification of a higher number of MHC peptides; however, it usually requires the use of more than 100 million antigen-presenting cells (APCs), which can be a limiting factor with the amount of the sample available from patients.

In late 80’s a simple and reproducible method to eliminate the antigenicity of MHC I molecules was introduced, called mild acid elution (Shunji Sugawara et al., 1987). The essence of this approach is incubation of the viable cells in a citric acid buffer (pH 3), which leads to effective elution of MHC I molecules. Interestingly, it has been noticed that neither MHC class II antigens nor other non-MHC antigens were removed from the cell surface when using this technique (Shunji Sugawara et al., 1987).

Most immunopeptidomic analyses are discovery-oriented, and therefore data-dependent (DDA) MS acquisition is most common (Kote et al., 2020). These approaches allow for generation of high-quality references of peptide tandem mass spectrometry (MS/MS); however, the downfall is their low sensitivity and reproducibility between samples. Use of tandem MS (MS/MS) tools, like Peaks software (B. Ma et al., 2003), means that the method depends on a user’s protein database to match each acquired MS/MS spectrum to a peptide sequence. In DIA approach, all ion precursors are detected and fragmented in an unbiased way within

isolation windows. Therefore, peptide identification rate, reproducibility and quantitation across multiple samples can be boosted by DIA methods (Kote et al., 2020). However, in clinical setting, the amount of HLA peptides eluted from biological samples such as small tissue biopsies is not sufficient to acquire both DDA data essential for generation of comprehensive spectral libraries and DIA MS measurements, the application of DIA in the immunopeptidomics translational research domain has remained limited (Pak et al., 2021). Pak et al. proposed a novel approach, where they used multi-HLA library for DIA method in combination with MS/MS prediction (Pak et al., 2021). They showed that, analysis of DIA immunopeptidomics data against a complex multi-HLA spectral library increases peptide identification by two-fold when compared with sample-specific library, whereas compared with DDA, a three-fold increase was seen. What lead to conclusion that a comprehensive multi-HLA library for DIA approach in combination with MS/MS prediction is highly advantageous for clinical immunopeptidomics, especially when low amounts of biological samples are available (Pak et al., 2021).

As previously mentioned, reference proteome does not account for tumour specific antigens and therefore custom databases derived from RNAseq should contain even unannotated proteins expressed in the measured tumour sample (Kote et al., 2020). Creating a hybrid approach that integrates both spectral library and database search was reported to improve quantitative precision and increase the number of quantified peptides (J et al., 2019). Matching spectra against extensive noncanonical reference database may lead to generation of false positives, however combination of sensitive mass spectrometry methods, analytical validation and advance bioinformatic tools are expected to reveal the full landscape of presented antigens and clinically relevant targets (Chong et al., 2021).

Aside of direct detection of immunopeptides by mass spectrometry, a number of predictive tools exists to identify genomic and transcriptomic mutations that would be presented to the immune system, such as PVacTools (Hundal et al., 2020). This software can be used to establish variants producing peptides likely to be presented in the immunopeptidome, as well as to confirm those that have affinity for MHC (Kote et al., 2020). Additionally, application of machine-learning tools to predict MS/MS spectra, such as the publicly available Prosit software with DIA-based approach

could potentially allow for the discovery of entirely novel sequences (Gessulat et al., 2019).

Many steps of the MHC class I antigen processing pathway can be predicted using computational methods, including prediction of the proteasomal cleavage (Nielsen et al., 2005), transport into the ER by TAP (Peters et al., 2003), peptide-MHC binding, and predicting the stability of the peptide-MHC complex (Rasmussen et al., 2016). However, the most discriminative tools in predicting immunogenic epitopes are those calculating potential peptide-MHC binding (Koşaloğlu-Yalçın et al., 2022).

There are still many challenges facing precise identification of tumour specific peptides bind to HLA and are able to mediate T cell-based tumour rejection. To date, mass spectrometry is the only analytical method that enables the direct identification of the HLA-bound peptide repertoire in vivo. Immunopeptidomic discoveries use to be only limited to using the standard, available protein sequence database, that include mostly proteome-derived sequences (Chong et al., 2021). However, thanks to advances in proteogenomic field recent studies improved their method by including protein sequences derived from the translation of transcripts identified from RNAseq, or ribosome profiling in their MS searches (Chong et al., 2021). This led to detection of novel, yet unannotated, noncanonical antigens from biological samples.

In 2018, two murine cancer cell lines were used to validation of the existence and relevance of tumour-specific noncanonical peptides (Laumont et al., 2016). To assure tumour specificity, RNAseq reads common between transcriptome of the cancer cell lines and in matched murine thymic epithelial cells were removed, which led to identification of 21 tumour-specific murine noncanonical peptides (Laumont et al., 2018).

Recent years brought revolution in immunopeptidomics field, that enabled discovery of noncanonical antigens—antigens derived from sequences outside protein-coding regions or generated by noncanonical antigen-processing mechanisms. However, there are still some limitations of using MS techniques in immunopeptidomics, including the need for a large amount of starting material or dependence on protein sequence databases for data analysis, which narrows down peptide identification to the annotated human proteome (Finotello et al., 2019).

These challenges can be resolved in the upcoming years with the development of novel sequencing approaches and improvements in sensitivity of mass spectrometry machines. The next decade will see the immense advantages coming from using mass spectrometry techniques supported by computational tools, for discovery of tumour-specific antigens.

1.2.5 Immunogenicity

Identifying peptides able to induce immune response is currently an unsolved problem in cancer field. Majority of neoantigens predicted in silico from mutational data fail to induce an immune response in vivo (Finotello et al., 2019). This creates the need to test experimentally neoantigens predicted computationally or derived from mass spectrometry data. Unfortunately, there is still a big gap in our knowledge when it comes to neoantigen immune responses and what drives their immunogenicity.

Characteristics that have been found to correlate with immunogenicity are the stability of the peptide–HLA complex bond and the functional avidity of the CD8+ T cell receptor interacting with it (Yadav et al., 2014). Stability of the binding between peptide and HLA complex can be predicted in silico using tools such as publicly available NetMHC, described in previous chapter. Collaborative consortium such as the Tumour Neoantigen Selection Alliance (TESLA) is collecting large number of data sets for the validation of these in silico predicted antigens helping to identify the best predictors of mutated peptides binding in vivo (Finotello et al., 2019). However, there is no currently available high-throughput that would predict the recognition of peptide–HLA by TCRs. Resources such as The Immune Epitope Database (IEDB) gather experimental data on antibody and T cell epitopes studied in humans, non-human primates, and other animal species providing valuable information for training computational predictors (Finotello et al., 2019).

Another feature that has been linked to neoantigen immunogenicity is its “novelty” (Blank et al., 2016). It was determined that mutations generating novel HLA-binding sites are more likely be immunogenic because they allow for the presentation of antigens never seen before by the immune system, therefore

inducing strong immune responses. Łuksza et al., created a prediction model that measures neoantigen fitness by comparing HLA-binding affinity of mutated and wild-type peptides and by identifying sequence similarity of neoantigens to known immunogenic microbial epitopes (Łuksza et al., 2017). The model was used to predict the survival of patients treated with immune checkpoint inhibitors. Nonetheless, the predictive value of this method requires validation on other cancer types.

Experimental evaluation of immune response to neoantigens has been performed using variety of methods (Chudley et al., 2014; Malaina et al., 2023). However, to date the most commonly used approach is measuring T-cell reactivity by IFN-γ ELISpot assay (Roudko et al., 2020). Interaction of neoantigen-MHC I complex with T cell receptors induces Tc cell priming, which results in either TNF-α, IFN-γ, or double TNF-α and IFN-γ cytokine responses, IL-2 secretion, and T cell proliferation. Level of these cytokines can be analysed through ELISA-type assays or binding to tetrameric peptide-MHC complexes. This last approach relies on the folding properties of MHC I complex with peptide or UV-cleavable substate which is later exchanged for the studied peptide (Roudko et al., 2020).

Knowing that an individual amino acid within a peptide can affect its immunogenicity by altering peptide–MHC or peptide–T-cell receptor interactions, in cancer studies it is crucial to create platforms that would allow for identification of neoantigen able to induce anti-tumour immune response (Kalaora et al., 2016). This would increase the success rate of neoantigenic vaccinations as well as reduce the time, cost of its development.

1.2.6 Neoantigen-based vaccines

Accumulating genetic alterations in cancers result in the production of tumour-specific antigens (TSAs) or neoantigens, which can be presented by major histocompatibility (MHC) molecules of tumour cells. Due to these characteristics, neoantigens have been long envisioned as the promising targets for anti-tumour immune response. It has been hypothesized that immunisation with neoantigens can

both lead to expansion of the pre-existing neoantigen-specific T cells and activation of a broader spectrum of new T cell specificities in cancer patients (Ott et al., 2017).

As opposed to prophylactic vaccines, cancer vaccines are administered to patients with developed tumours. The efficiency of this therapeutic approach often relies on the difference in the expression of the targeted neoantigen vs tumour and normal cells (Ott et al., 2017).

Characterisation and evaluation of neoantigens has only became more achievable with the availability of deep sequencing platform and machine learning approaches to predict mutated peptides with high affinity to HLA molecules.

In recent study of Hilf et al., a neoantigen vaccination was shown to induce a therapeutic response in glioblastoma with low mutational load and no T cell infiltration (Hilf et al., 2019). Inoculation of 15 patients with two highly personalized vaccines activated T cell response and led to improvement in patients' median total survival time to 29.0 months (Hilf et al., 2019).

In breast cancer, Morisaki et al. studied whole exome and RNA sequencing of 31 breast cancer tissues and performed prediction of neoantigen determinants from non-synonymous single nucleotide variants (nsSNVs) among mutations found in exome (Morisaki et al., 2021). Neoantigen determination was followed by prediction of HLA binding affinity with mRNA expression measurements. From the neoantigen profiles 10 mutant peptides were synthesized, which displayed high affinity for HLAA02:06 and transcript expression with a read count of ≥2. Dendritic cells were pulsed with identified neoantigens and cocultured with autologous peripheral lymphocytes to induce activation of cytotoxic T lymphocytes (CTLs). Successful CTLs induction was later evaluated by using cytotoxicity and IFN-γ ELISpot assays, confirming that neoantigen analysis may in developing strategies to elicit T cell responses in breast cancer (Morisaki et al., 2021).

Study by Ott et al. highlighted an interdisciplinary approach for vaccinology that requires the translation of the genomic changes to mutant MHC class I mutated antigens from which to develop precision cancer vaccines (Ott et al., 2017). The researchers showed the pipeline for melanoma vaccine preparation that involved

analysis of the whole-exome sequencing (WES), RNAseq, somatic mutation calling, identification of target epitopes for peptide design, synthesis of long peptides, pooling and final vaccine production. Vaccine targeted up to 20 predicted neoantigens and induced polyfunctional CD4+ and CD8+ T cells response. Among 6 vaccinated patients, 4 showed no recurrence 25 months after vaccination and 2 with progressive form of disease were additionally treated with anti-PD-1 therapy. This combination led to a complete tumour regression, with the detected expansion of neoantigen-specific T cells (Ott et al., 2017). With our understanding of neoantigen origin, there are many pipelines, including presented here, that aim to discover the most promising target for immunotherapeutic elimination.

However, it's been shown that vaccines alone cannot achieve complete elimination of tumour cells (Rosenthal et al., 2019). One of the reasons behind it is that vaccines can only kill a small fraction of cancer cells that express neoantigens targeted by the vaccine, which restricts clinical effect. Moreover, these immunogenic peptides are very tumour-specific, showing little to no sharing between patients. Additionally, it has been reported that the selection pressures from tumour microenvironment can influence neoantigen presentation, promote immune escape and affect clinical outcomes in NSCLC (Rosenthal et al., 2019).

These highlighted weaknesses show that there is still a lot to achieve when it comes to identification, evaluation, and therapeutic effects of neoantigen vaccinations. Thus, improving the neoantigen discovery pipelines was one of the aims of my PhD.

1.3 Nonsense-mediated mRNA decay pathway (NMD)

Nonsense-mediated decay (NMD) is a crucial host RNA control pathway that removes aberrant mRNA containing premature termination codons (PTCs) that arise through mutation or defective splicing (Celik et al., 2017). NMD was found to affect one-third of mutated transcripts (Popp & Maquat, 2018). mRNAs with PTC in coding regions originate from endogenous genes with nonsense or frameshift mutations, pseudogenes (Karousis et al., 2016), or as a product of alternative splicing that

causes intron retention or inclusion of PTC-containing exons (Celik et al., 2017; Pawlicka et al., 2020). To avoid producing C-terminally truncated proteins that can have deleterious effects for the organism, those transcripts harbouring PTC are recognized and subsequently degraded.

NMD signalling pathway was initially discovered in Saccharomyces cerevisiae and Caenorhabditis elegans (Leeds et al., 1991; Peltz et al., 1993) and later was proven to be conserved from yeast to humans (Baudu et al., 2021; Behm-Ansmant et al., 2007). NMD has been involved in various stages of the embryonal development, DNA damage response, regulation of cell cycle and functioning of immune system (Huth et al., 2022; Nakano et al., 2022; Nasif et al., 2018; Weischenfeldt et al., 2008). Due to its role in maintaining the genome stability, aberrations of this strictly regulated control pathway lead to pathologies including neurological disorders, immune disfunctions and cancers.

1.3.1 Mechanism of NMD

NMD pathway was first studies in yeast, where it was discovered that introduction of UAG nonsense codon in ura3 gene lead to reduction of its mRNA without interfering with the transcription rate (Losson & Lacroute, 1979) . It was also observed that the transcript level decrease depended on the position of the nonsense mutation within the gene. PTC introduced close to the 5’ end of mRNA lead to decreased production of the transcript, while PTC close to the end of ura3, caused levels close to that of wild-type (Losson & Lacroute, 1979). Studies in human model only confirmed these observations (Kinniburgh et al., 1982).

NMD is cytoplasmic and translation-dependent process. During pre-mRNA splicing, the exon junction complex (EJC) is deposited upstream of the exon-exon junction (Schlautmann & Gehring, 2020). EJC is a multi-subunit protein complex, spanning ∼20–24 nucleotides. In uninterrupted translation process, bound to mRNA, EJCs are transported into the cytoplasm, where the force of the ribosome, as it translates the transcript, is sufficient to remove it from the transcript. During translation of an unaltered mRNA, the stop codon in the last exon ensures that no

EJCs remain associated to transcript upon translation termination. Translation termination is also regulated by the position of the ribosome at the end of the transcript where interactions to proteins bound to the mRNA poly(A) tail and release factors are crucial (Pawlicka et al., 2020).

If a PTC is present at least 50 nucleotides (nt) upstream of the last exon junction, an exon junction complex (EJC) will remain bound after the pioneering round of translation has been terminated. This results in assembly of the NMD-related proteins and recruitment of other cofactors (Pawlicka et al., 2020).

Target characteristics

NMD substrates can be divided in several groups, the most commonly studied are mRNAs with a destabilizing PTCs in their coding region generated from endogenous genes carrying nonsense or frameshift mutations, pseudogenes or from alternative splicing events that lead to intron retention or inclusion of a PTC-containing exon (Lykke-Andersen & Jensen, 2015). A second group includes mRNA-like transcripts with limited or no apparent coding capacity, such as long noncoding RNAs, small RNAs originated from intragenic regions or mRNAs of inactivated transposable elements (F. He et al., 2003; Lykke-Andersen & Jensen, 2015). Characteristic for the third category of NMD targets is that they appear to be “normal,” such as transcripts with upstream open reading frames, or with unnaturally long 3′ UTRs, or normal-looking wild-type mRNAs with no apparent atypical features (Celik et al., 2017; F. He et al., 2003)

Mutations inducing NMD share some characteristics, such as PTC situated more than 50–54 bp upstream of the last exon–exon junction, targeted gene comprising of at least two exons and PTC present more than 200 bp downstream of the start codon. These features can explain up to 80% of the decreased expression of the genes carrying nonsense mutation (Zhao et al., 2020).

It has been also shown that transcripts not containing an EJC can be targeted by NMD (Bongiorno et al., 2021). This phenomenon has been proposed to explain how NMD is able to physiologically regulate transcriptome without the need for EJC induction. Studies show that uORF increases transcript sensitivity to NMD without

the need for exon-exon junction in the 3'UTR (Bongiorno et al., 2021; Schilff et al., 2021). As well as retained introns and the presence of exons in 3'UTR are all mechanisms increasing NMD activity (Bongiorno et al., 2021).

Activation of NMD pathway

In humans, proteins involved in NMD pathway include the hUPFs—human up-frameshift (UPF) proteins (UPF1, UPF2, UPF3a and UPF3b), the suppressors with morphological effects on genitalia proteins (SMG1, SMG5, SMG6, SMG7, SMG8 and SMG9), and the exon junction complex (EIF4A3, MAGOH, RBM8A and Barentsz (BTZ)) as reviewed (Pawlicka et al., 2020).

To date, there have been recognised two NMD activation patterns that coexist in the cell (Huth et al., 2022). In the first one, as a consequence of intron splicing, EJC is deposited 20–24 nucleotides upstream of exon-exon junctions. If transcripts have a PTC upstream of the last exon, one or multiple EJCs remain bound to the mRNA at the moment when translation terminates. This leads to recognition by a number of proteins including the core NMD effectors UPF1, UPF2, and UPF3/3X. UPF1 associate with proteins involved in translation termination, while UPF2 and UPF3/3X interact with the EJC (Jia et al., 2017).

In the presence of a premature termination codon, translation pauses upstream of an EJC and the eukaryotic release factors (eRF) physically attach to the transcript and recruit UPF1 that functions as the RNA helicase (Jia et al., 2017). The eRFs recognize the PTC, and when the mRNA stop codon enters the ribosomal A site, the termination of the protein synthesis occurs. The single eukaryotic class-I RF eRF1 recognizes all three stop codons: UAG, UGA, UAA as reviewed (Pawlicka et al., 2020).

During the pioneer round of translation UPF1 interaction with CBP80 protein on 5' cap enables its recruitment, alongside the SMG1/SMG8/SMG9 proteins and the release factors eRF1 and eRF3 (Kalathiya et al., 2019). Translation termination factor eukaryotic release factor (eRF) 1 has a similar structure to an aminoacyl-tRNA, forms a complex with, and is activated by the GTPase activity of eRF3, which

allows interaction with UPF1 and later UPF2 (Huntzinger et al., 2008). Activation of the NMD pathway accelerate remodelling of the surveillance complex (SURF), which consists of the UPF1, eRF1, eRF3 proteins and a phosphatidylinositol-3-kinase-related kinase, SMG1. UPF3b binds to the EJC and anchors UPF2 protein. The SURF complex connects with the UPF2, UPF3b and an EJC downstream of the PTC, forming the decay-inducing complex (DECID) (Yamashita et al., 2009). UPF3X which participates in recruitment of eRF3, positions it at the A site of the ribosome. SMG1-mediated phosphorylation of UPF1 is the key step NMD-activation and enables stabilisation of the UPF1-mRNA binding of UPF1 determining the transcript for decay (Cowen et al., 2019; Kalathiya et al., 2019). Together with the UPF proteins the SURF complex induces the phosphorylation of UPF1 by SMG1, which results in losing its affinity for SMG1/SMG8/SMG9 and the eRFs and causing their departure and recycling of the ribosome. Phosphorylated UPF1 protein, stimulated by its interaction with UPF2, deploys its RNA helicase activity in order to remove the proteins located downstream of the PTC and preparing this end for degradation by ribonucleases (Kalathiya et al., 2019). In addition, the SMG6 protein, due to its endoribonuclease activity, induces cleavage in the proximity of the PTC, generating free, accessible 5’ and 3’ ends for exoribonucleases. Phosphorylation of UPF1 increases its ability to bind to a protein complex consisting of SMG5, SMG7, SMG6, and protein phosphatase 2A, which then acts as a negative feedback loop dephosphorylating the UPF1 protein (Kalathiya et al., 2019). This UPF1 phosphorylation-dephosphorylation cycle induces activation of NMD pathway and recruitment of enzymes involved in degrading the 5 ‘and 3’ ends. Allowing for the adjustments of the NMD activity, the UPF3a protein inhibits NMD, and this activity is regulated by the UPF3b protein (Huang et al., 2012).

The main role of SMG1 protein is phosphorylation of UPF1, however it’s been also shown to phosphorylate other cellular proteins, which indicates its function in different cellular mechanisms, such as response to stress induced by DNA-damaging agents (Brumbaugh et al., 2004). SMG1 induced p53 stabilization and phosphorylation at serine 15 in radiated cells (Brumbaugh et al., 2004).

In the second and less studied model of NMD pathway activation, PABPC1, a protein located on polyA tail, is involved (Huth et al., 2022). The probability of NMD induction increases with the distance between PABPC1 and PTC. In this process,

UPF1 protein accumulating at 3'UTR compete with PABPC1 for recruitment of translation termination complex. When competition favours PABPC1, termination of translation does not induce NMD but instead initiates new translation cycles. If UPF1 recruits the complex, NMD of this transcript is induced and initiation of new translation cycles is inhibited (Huth et al., 2022).

Transcript degradation

There are many mechanisms that lead to degradation of NMD-targeted RNAs. In primary models, such as yeast, it has been shown that PTC-containing transcripts are degraded mainly by deadenylation-independent process, which involves decapping by enzymes Dcp1p/Dcp2p and 5′–3′ exonuclease digestion by Xrn1p (Muhlrad & Parker, 1994). In human cells, these mechanisms are more complex. mRNA degradation can be led by endonucleolytic cleavage (Huntzinger et al., 2008), exosome mediated 3′–5′ decay (Lejeune, 2022) or deadenylation-dependent decapping (Muhlrad & Parker, 1994). Degradation of transcript follows, SMG6 endonuclease binding to phosphorylated residues in the N-terminal part of UPF1 and catalysis of a cleavage in the proximity of the PTC (Huntzinger et al., 2008). Next, the SMG5/ SMG7 heterodimer attaches to phosphorylated domain in the C-terminal part of UPF1 and activates the CCR4–NOT deadenylase complex, which the catalyzes polyA tail shortening that induces decapping by DCP1/ DCP2 (Lejeune, 2022; Okada-Katsuhata et al., 2012). Decapping complex can also be recruited to UPF1 either directly or in a PNRC2-dependent manner, activating the deadenylation-independent decapping (Lykke-Andersen & Jensen, 2015). For the removal of decay intermediates in all of the described pathways exonuclease activity of XRN1 is responsible.

Transcriptome-wide analysis of NMD substrates and their 5′–3′ decay intermediates revealed that SMG6-catalyzed endonucleolysis widely activates the degradation of human nonsense RNAs, whereas decapping is used on a smaller scale (Lykke-Andersen & Jensen, 2015; Pawlicka et al., 2020). Transcript decay depends on interactions between core NMD factors and a terminating ribosome.

1.3.2 PTC readthrough

Behind physiological cell function lies genetic information that has been accurately expressed in RNAs or proteins. Translation from messenger RNA (mRNA) onto protein is a crucial step in gene expression pathway (Huth et al., 2022). This process follows very specific rules, to ensure the correct protein sequence and length. Translation starts when an initiation codon (most commonly AUG) enters the A-site of the ribosome, with a complimentary base pairing of aminoacyl-tRNA anticodon, loaded with methionine, and terminates when site A of the ribosome encounters one of the stop codons UAA, UAG, or UGA (Labunskyy et al., 2014). In humans, the only tRNA to recognize these termination codons is tRNA[Ser]Sec, which carries the amino acid selenocysteine (Labunskyy et al., 2014). Decoding of each codon varies, depending on the tRNA availability.

However, as well as physiologically, as artificially induced, stop codons can be read through with high efficiency. This phenomenon can be achieved by using suppressor tRNAs, which carry mutations in the anticodon to match the stop codons, readthrough inducers, such as gentamycin and other techniques (Baradaran-Heravi et al., 2016; Eggertsson & Söll, 1988; H. Zhang et al., 2020). The use of suppressor tRNAs for induction of readthrough have been widely used to site-specifically insert noncanonical amino acids into proteins of interest (Eggertsson & Söll, 1988; Lueck et al., 2019). Stop codon readthrough leads to an extension of amino acid sequence to the C terminus and has reported to be critical for expression of functional protein variants.

It was reported that in Escherichia coli excess of carbon in culture media significantly increased the readthrough levels of UAA, UAG, and UGA codons, showing that external factor can modulate transcript expression levels (H. Zhang et al., 2020). It was later shown that the elevation depended on a drop of pH during bacterial growth in excess carbon (H. Zhang et al., 2020).

Physiological readthrough

Study of mRNA half-lives indicated that a small percentage of transcripts carrying PTCs escape surveillance and do not degrade (Schilff et al., 2021) Readthrough occurs as a natural process at the termination codon, as a result of an aminoacyl-tRNA base pairing outcompeting the release factor eRF1, during the proofreading stage. In effect an insertion of an amino acid occurs instead of translational halt. This phenomenon has been detected in eukaryotic cells with a frequency of less than 0.1% (Schilff et al., 2021).

It has been reported that some nucleotides in the proximity of the termination codon support endogenous readthrough (Schilff et al., 2021). The rate of programmed readthrough has been shown to be dependent on the stop codon type (UGA being the most prone to being misread), as well as certain nucleotides upstream and downstream of the stop codon (Schilff et al., 2021).

Physiological readthrough plays role in expression of some protein variants, such as lactate dehydrogenase B (LDHB) (Schueren et al., 2014). Experimental studies show that LDH contains a ‘hidden’ peroxisome-targeting region and more translational readthrough leads to an increase in lactate dehydrogenase was found in the peroxisomes. (Schueren et al., 2014). Lactate dehydrogenase found in peroxisomes are conserved in mammalian cells and most likely regulates redox equivalent regeneration in peroxisomes. This phenomenon demonstrates how the cell regulates the use of the genetic information encoded in the genome and in mRNA.

Induction of translational readthrough

Readthrough-inducing agents influence ribosome function and shift the balance between termination and (mis)reading of the termination for a sense codon (Crick, 1966). Efficiency of the readthrough depends on competition between stop codon recognition by a class I release factor and decoding of the stop codon by a near-cognate tRNA, paired using two of the three bases, knows as wobble hypothesis (Crick, 1966). Drugs that have been proven to have the highest impact on the stop codon readthrough are aminoglycosides (Baradaran-Heravi et al., 2016). Aminoglycosides use their PTC readthrough activity by binding at the decoding

center of the eukaryotic ribosome. This binding alters the ability of translation termination factors to accurately recognize a PTC. Consequently, aminoglycosides increase the frequency of pairing of near-cognate aminoacyl-tRNAs to the PTC and enables formation of full-length protein (Baradaran-Heravi et al., 2016).

Despite the fact that, that aminoglycosides have been in clinical use for many years, their application is limited due to shown oto- and nephrotoxicity (Bitner-Glindzicz & Rahman, 2007; Mingeot-Leclercq & Tulkens, 1999). To overcome this obstacle, number of aminoglycoside-derivatives are being developed and examined, such as ataluren which has been approved for treatment of certain Duchenne muscular dystrophy cases in several countries (Baradaran-Heravi et al., 2016; Namgoong & Bertoni, 2016).

One of the features of genetic code is codon degeneracy, which means that 20 amino acids require up to 5 tRNAs with distinct anticodons to recognise codons for each amino acid (Beznosková et al., 2021). This characteristic has been utilized to modify the stop codon readthrough. Studies have shown that levels of tRNA recognising tryptophan and tyrosine-tRNA impacted stop codon readthrough on UGA or UAG and UAA stop codons in human cells (Beznosková et al., 2021). Overexpression of Trp-tRNA boosted readthrough of reporters carrying termination sequences of selected cellular and viral genes known to undergo this type of regulation. It also lead to restoration of a functional p53 protein production from transcripts carrying PTC (Beznosková et al., 2021). Stimulation of translational readthrough by using readthrough-inducing drugs may serve as a possible therapeutic strategy for the treatment of genetic PTC diseases.

Wobble hypothesis

Translation termination requires class I release factors (RFs) to recognize stop codons. Classical eRF1s recognize the three termination codons UGA, UAG and UAA, however they don’t detect UGG encoding tryptophan. This shows that these eRF1 have the ability of distinguishing between the bases A and G in both, 2nd and 3rd position of a termination codon, as described by wobble hypothesis (Crick, 1966). Additionally, in order to exclude UGG, the recognition of these two positions of stop codon cannot be performed independently from each other. The

fact that release factors can detect all three termination codons, but omit the tryptophan codon has been attributed to the flexibility of the helix containing these positions, that is able to take a partly relaxed or tight conformation depending on the recognized stop codon (Muramatsu et al., 2001).

In this work, we have developed a cancer cell model to determine the rules that define the type of amino acid placed at a stop codon during termination readthrough mediated by an aminoglycoside antibiotic.

Genetic disorders caused by PTC

It has been demonstrated that around 30% of inherited disorders results from premature termination codons (Mort et al., 2008). Nonsense mutation can cause severe genetic diseases, such Duchenne muscular dystrophies, ataxia-telangiectasia (ATM), cystic fibrosis (CFTR), beta-thalassemia or Wilm's tumour (WT1) (Malik et al., 2010). Most of these disorders are due to a failure of the NMD machinery to efficiently degrade the PTC-containing transcripts. Leading to production of truncated, non-functional proteins.

Number of compounds have been studied for their readthrough-inducing functions in the context of PTC-caused genetic disorders, such as aminoglycosides or ataluren. Readthrough drugs force the ribosome through PTC, protecting peptide from NMD and induces expression of almost full proteins (Malik et al., 2010).

Duchenne muscular dystrophy (DMD) is the most common type of muscular dystrophy caused by dystrophin deficiency (Namgoong & Bertoni, 2016). It is estimated that 10–15% of patients with DMD carry a nonsense mutation which introduces a PTC into the dystrophin transcript, resulting in translation of a truncated, non-functional protein. It has been demonstrated that ataluren binds to the ribosome and disrupts the recognition of termination codons thus inducing readthrough of the nonsense mutation to produce a full-length functional protein (Namgoong & Bertoni, 2016). In 2017 ataluren was approved by the European Medicines Agency for treatment of DMD patients, however further phase III studies didn’t reach the statistical significance (McDonald et al., 2017).

Another PTC-caused disorder is cystic fibrosis (CF) that results in different mutations within the CFTR gene resulting in a non-functional Na+ ion channel (De Boeck et al., 2014). PTC-containing transcript activate NMD pathway, which degrades most of mRNAs leaving little to no CFTR translated protein. Studies have shown that aminoglycoside gentamicin activated CFTR protein expression from CFTR gene carrying nonsense mutations in bronchial cells taken from cystic fibrosis patients (De Boeck et al., 2014). Ion channels formed with the protein that has been read-through are still altered and have limited function, however they show beneficial outcome when compared to complete absence of protein (De Boeck et al., 2014). However, due to high oto- and nephrotoxicity of gentamycin, novel derivatives of aminoglycosides with readthrough potential have been studies (Bidou et al., 2017; Bitner-Glindzicz & Rahman, 2007; Mingeot-Leclercq & Tulkens, 1999).

Stop-codon readthrough agents have been proposed as the main therapeutic option for treatment of NMD-implicated diseases (Malik et al., 2010). In this approach small readthrough-inducing drugs promote the translation machinery to ignore the nonsense codon and instead recode it as an amino acid. The produced full-length functional protein has the potential of ameliorating the disease as reviewed (Pawlicka et al., 2020).

1.3.3 Dual role of NMD pathway in cancer

It is becoming increasingly evident that the many proteins involved in NMD pathway are inhibited in cancer cells (Popp & Maquat, 2018). However, studies also indicate that the role of NMD in cancer development is complex. Some cases show that cancer cells have exploited NMD by selecting for mutations causing destruction of key tumour-suppressor mRNAs (Kalathiya et al., 2019). In other studies, tumours adjust NMD activity to adapt to their microenvironment (Park et al., 2018). NMD also serves as a tool for tumours to adjust themselves to the microenvironment (Kalathiya et al., 2019; Popp & Maquat, 2018).

Cancer-protective function of NMD in Cancer

NMD inhibition in cancer has been reported to stabilise transcripts important for cancerogenesis, such as MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) in hepatocellular carcinoma (Malakar et al., 2017). A synergistic regulation of splicing and NMD pathway has been reported for the alternatively spliced isoform β of tumour suppressor gene p53 (Gudikote et al., 2021).

Studies demonstrated that high number of UPF1 mutations found in pancreatic adenosquamous carcinomas, result in deregulation of the NMD pathway in these tumour asparagine synthetase (ASNS) (Hu et al., 2022). ASNS regulates protein translation under stress conditions controlling the biosynthesis of selected proteins, such as the cell cycle regulator cyclin B1, which has been connected to poor cancer patient outcome. Studies highlighted also the influence of deregulated NMD pathway on increased level of ASNS, suggesting that knockdown of this protein may disrupt the adaptive responses of pancreatic cancer cells to metabolic stress, which also may encourage development of new treatments and improve patients benefits (Hu et al., 2022). This has been shown in blood cancer models, where knockdown experiments of ASNS led to an improved response to (Hu et al., 2022)

Interestingly, reports from osteosarcoma studies show that neoantigen expression is reduced by NMD ultimately causing an accumulation of genomic alterations not paralleled by immune infiltration (Wu et al., 2020). NMD activity has also been a negative predictive factor of immune response against microsatellite-instable colorectal cancer, where elevated levels of UPF1 were seen (El-Bchiri et al., 2008). Increased levels of NMD factors increases efficiency of NMD pathway, thus allowing cancer cells to continue tumorigenesis.

O'Leary et al, identified that in gastric cancer cells loss of UPF2 expression leads to induction of resistance against inhibitors of ATR (ATRi), a key regulator of the DNA damage response (O'Leary et al., 2022). Additionally, global transcriptomic and proteomic profiling experiments revealed that UPF2-mutant cells display alteration in cell-cycle progression and DNA damage responses, that have the tumour-promoting effect (O'Leary et al., 2022).

Role of NMD in anti-cancer response

Various studies indicated that overexpression of UPF1 leads to reduction of hepatocellular carcinoma (HCC) growth through multiple cell signalling pathways (Chang et al., 2016; S. Lee et al., 2022). Transcriptomic analysis revealed that the level of dual specificity phosphatase 1 (DUSP1) was elevated by UPF1 variants through posttranscriptional regulation. DUSP1 belongs to tumour suppressor genes and has been shown to inhibit cell cycle progression in HCC cells. Increased expression of DUSP1 has been observed more frequently among patients with higher overall survival. Furthermore, overexpression of DUSP1 correlates with increase in antitumour genes, such as phos-p53, p21 and p27. Therefore, the UPF1 overexpression-mediated elevation of DUSP1 activates tumour suppressor signalling leading to inhibition of cancer cell growth. Additionally, overexpression of UPF1 variants was shown to reduce cell motility and invasiveness of studied cells (S. Lee et al., 2022)

Impaired NMD was also found to promote the accumulation of NFκB-inducing kinase that is a potent activator of the non-canonical NF-kB pathway and stimulator of several pro-inflammatory cytokines such as tumour necrosis factor (TNF). Lu et al reported that UPF1 mutations downregulate NMD, resulting in NIK-dependent NF-κB induction, which increased the immune cell infiltration in inflammatory myofibroblastic tumours (Lu et al., 2016). This regulatory link with the NMD pathway connected with diagnosis and therapeutic approaches of these tumours (Lu et al., 2016).

By degradation of mRNAs carrying PTC, NMD pathway shields heterozygous germline carriers of these cancer mutants from developing cancer (Pawlicka et al., 2020). This regulative circuit has been well studied for dominant negative heterozygous mutants of the BRCA1 (Fan et al., 2001). Germline mutations in this tumour suppressor gene induce familial cases of ovarian and breast cancers (Perrin-Vidoz et al., 2002). Expression of truncated version of BRCA1 protein was correlated with chemoresistance, reduced susceptibility to apoptosis, decrease in vivo tumour growth suppression and many other. This observed phenomenon suggested that that truncated BRCA1 proteins inhibit the activity of wild-type BRCA1. It was also hypothesised that some inherited BRCA1 mutations are able to pro-actively induce oncogenesis by deactivating the remaining wild-type BRCA1 allele (Fan et al., 2001).

These studies indicated, that NMD may take either a tumorigenesis-amplifying or an antioncogenic role depending on the cancer type and its expression profile (Lejeune, 2022).

1.3.4 NMD inhibition in cancer therapy

Studies have shown that inhibition of NMD also generates new antigenic determinants in tumour cells (Park et al., 2018). Cancer-targeted NMD inhibition has been shown to be a promising approach to increase the antigenicity of disseminated tumours by induction of neoantigen expression on the surfaces of cancer cells allowing for the recognition by immune system and rejection (Pastor et al., 2010; Popp & Maquat, 2018).

It has been demonstrated that in mouse models inhibition of UPF2 or SMG1 led to immune-dependent reduction of tumour progression (Pastor et al., 2010). On the other hand, silencing of UPF1 in vitro increased the production of potentially immunogenic neoantigens (Oka et al., 2021). In osteosarcoma expression of immunogenic peptides was shown to be downregulated by NMD resulting in an accumulation of genomic alterations unrecognised by the immune system (Wu et al., 2020). In the light of immunotherapy advancements, these observations open the gate combining NMD inhibitors with immune checkpoint blockade, which would make the most of the potential antigenic enrichment of tumour targets. Despite the fact that not all mutations induce immune response, frameshift mutations have been shown to be the most immunogenic.

In cancers carrying nonsense mutations in the TP53 gene, combination of readthrough induction and NMD inhibition led to restoration of full-length p53 protein (Martin et al., 2014). As a result, NMD inhibition promotes induction of apoptosis, reduction of cell viability, and enhanced tumour radiosensitivity in p53-dependent manner (Martin et al., 2014). Gudikote et al demonstrated, that in non-small-cell lung cancer cells NMD inhibition led to induction of radiosensitivity in cells bearing either WT p53 or mutant p53 (Gudikote et al., 2021). It was also reported that radiation (IR)

inhibits SMG1 and induces the expression of p53β, which in turn triggers IR-induced cellular senescence in MCF-7 cells (Gudikote et al., 2021).

Antigenic peptides originated from frameshift-bearing transcripts when presented by HLA class I molecules are able to induce tumour-specific T cell-mediated immune responses (Becker et al., 2021). However, in many cases expression of these potential neoantigens is limited by NMD. 5-azacytidine is a cytidine analogue that incorporates into DNA, inhibiting methyltransferase activity and activating expression of genes, like tumour suppressor ones, which were previously suppressed by methylation (Becker et al., 2021). It has been reported that inhibition of NMD pathway by 5-azacytidine stabilizes frameshifted sequences and elevates presentation of InDel neoepitopes on the HLA-I in MSI colorectal cancer (Becker et al., 2021). Immunization with the identified neoantigen resulted in strong CD8+ T cell responses in an HLA-A*02:01 transgenic mouse model, which highlights the potential of NMD inhibition in cancer therapy (Becker et al., 2021).

NMD inhibition represents an interesting strategy for tumour eradication by re-expression of a plethora of genes suppressed by NMD activity. Several NMD-inhibiting agents have already been introduced to clinic for treatment of genetic diseases, therefore these compounds could be repurposed to use as a novel cancer therapeutic approach (Bongiorno et al., 2021).

1.4 Aim

Despite that neoantigens are proving to represent a mechanism whereby immune cells can detect and eradicate cancer cells, there is very little known about how many neoantigens a cancer cell can produce, how to manipulate their abundance using targeted therapies, or the optimized way to detect them. Developing such improved strategies will be important as cancer vaccines begin to be heavily used in clinical trials. Main goal of this precision medicine PhD was to implement and develop pipelines for discovery and studies of neoantigens. More specifically:

1. Modelling potential neoantigen target identification by using DNA and RNA variant detection pipelines to define the mutant expression landscape in cancer cells. Most current neoantigen-based vaccines use DNAseq-RNAseq along with MHCnet to define mutant immunopeptides. This study extends this approach by utilising mass spectrometry datasets. The experimental approach focused on defining the depth of SNV, indel, and frameshift-stop codons in studied p53 wild-type and knockout cell model to characterise the potential sources of neoantigenic peptides and assess the influence of tumour suppressor gene deletion on neoantigen production.

2. The presence of frameshift mutations was shown to theoretically lead to a large number of ‘foreign’ neoantigens. As drugs exist for readthrough of such stop codons that can occur during frameshift, they were utilised to try to develop ‘rules’ to determine the types of amino acids incorporated at stop codon readthrough to improve the mutant refseq database for future neoantigen search.

3. Using optimized pipeline including DNAseq-RNAseq databases, isogenic wild-type and p53 null melanoma cell model will be developed to define the normal and mutant expressed immunopeptidomic landscape in a p53-dependent manner. This approach also includes optimising two different isolation methods (Immunoprecipitation and acidic washes) to enrich the immunopeptidome for identification using mass spectrometry.

4. Using the same cell model, drug leads impacting other potential neoantigen sources such as splicing or NMD were used to determine whether such treatments impact on the normal and mutant immunopeptidome detected by mass spectrometry.

Altogether, combined DNAseq, RNAseq, immunopeptidomics pipeline will be utilised to identify strengths and weaknesses of current neoantigen discovery technologies under different genetic and physiological states.

.

2 Material and methods

2.1 Microbiological techniques

Bacterial cultures were grown in Luria-Bertani (LB) broth (1%(w/v) Tryptone, 0.5% (w/v) Yeast extract, 1% (w/v) NaCl, 25g LB in 1l dH_20; autoclaved 121°C for 20 min) in incubator-shaker at 150rpm in 37°C.

2.1.1 Glycerol stocks

For the long-term storage of bacteria, glycerol stocks have been made by mixing 800µl of 24h bacterial culture to a cryotube (Nunc) with 200µl of sterile glycerol. The cells were rapidly frozen in liquid nitrogen and stored at -80°C.

2.1.2 Bacterial transformation

Escherichia coli DH5α competent cells (ThermoFisher Scientific, 18265017) from a glycerol stock were thawed on ice. Per transformation 50µl of thawed cells were used. After addition of DNA, competent cells were placed on ice for 30-40min. Next, the tube was subjected to 42°C for 60 sec in a water bath, followed by 2min incubation on ice. 200µl of SOC medium (2% (w/v) Tryptone, 0.5% (w/v) Yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl2, 20mM Glucose) was added into the tube and incubated for 1h in 37°C. 20-100 µl of transformed cells were plated on either Ampicillin or Kanamycin LB agar plates (40 g LB agar 1l dH_20, 1% (w/v)

Tryptone, 0.5% (w/v) Yeast extract, 1% (w/v) NaCl, 1.5% (w/v) Agar), as appropriate, and incubated overnight at 37°C for bacterial growth.

2.1.3 Plasmids

Plasmids used throughout the study were: pcDNA3.1 wild type p53 SBP tagged, pcDNA3.1 p53 R213X SBP tagged (Thermo Fisher), pCMV-hygro-p53-217-TGA-SBP, pCMV-hygro-p53-218-TGA-SBP, pCMV-hygro-p53-219-TGA-SBP, pCMV-hygro-p53-219-TAG-SBP, pCMV-hygro-p53-219-TAA-SBP (Twist Bioscience).

2.2 Molecular biology techniques

2.2.1 Plasmid DNA isolation and purification

From overnight bacterial culture (subsection 2.1.2) a single colony was picked and inoculated onto a tube containing 5ml LB medium supplemented with antibiotic. The culture was grown overnight at 37°C with shaking. Bacterial cells were centrifuged at 8000rpm, and plasmid DNA was isolated using QIAprep Spin Miniprep Kit (Qiagen, 27104). DNA concentration was measured by NanoDrop 2000c Spectrophotometer (ThermoScientific).

2.2.2 RNA isolation

Cells grown in 10cm Petri dishes at 90% confluency were harvested and frozen until further processing. Total RNA extraction was performed using the RNeasyMini kit (Qiagen). To remove DNA contaminants RNA samples were pre-treated with RNase-free DNase I (Qiagen). The concentration and purity of the total RNA were analysed using a NanoDrop 2000c Spectophotometer (ThermoScientific).

2.3 Cell culture

A375, HDQ-P1 and B16 cell lines were grown in DMEM media (Gibco) containing, and H1299 and ESS1 required RPMI (Gibco). Both media were supplemented with 10% (v/v) foetal bovine serum (FBS) (Labtech) and 1% (v/v) 0.5U/ml penicillin/ 500 ng/ml streptomycin (P/S) (Invitrogen). Cell culture was maintained in the incubator at 37°C with 5% CO_2. Once confluent, cells were sub-cultured using Trypsin–EDTA 0.5% (Gibco, Life Technologies). When necessary, cells were frozen in supplemented culture media containing 10% DMSO and stored at -80°C or liquid nitrogen.

2.3.1 Cell transfections

Prior to transfection, cells were seeded at 0.2 x 10^6 density onto 6-well plate (Corning). Plasmid DNA was transfected using Attractene (Qiagen, 301007) according to manufacturer's protocol on cells at 80% confluency. The reaction mix was added dropwise onto the plate and cells were left for 24 or 48 hours, depending on experimental requirements.

2.3.2 Collecting cells and lysis

Cells were harvested on ice. The medium was aspirated, cells rinsed twice with 1 ml ice-cold PBS, then detached from the culture dish by scraping of the surface and collected by centrifugation at 3000 rpm at 4ºC for 15 min. The cell pellet was lysed in four times their volume of lysis buffer (50mM HEPES pH 7.5, 1% v/v Triton-X, 0.5% NP-40, 150mM NaCl, 2mM DTT, 0.1mM EDTA, 1mM Benzamidine, 1x Protease Inhibitor Mix) incubated for 40min on a wheel at 4ºC followed by 15min centrifugation at 13,000 rpm at 4ºC. Supernatants were transferred to a new Eppendorf tube. Lysates were flash-frozen on dry-ice and kept at -80ºC for future analysis. Protein concentration was measured by using Protein Assay Dye Reagent (Bio-Rad).

2.4 Western blotting

2.4.1 SDS-PAGE

Protein samples were prepared by mixing 1:1 cell lysate with loading buffer (100 mM Tris-Cl (pH 6.8), 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, 200 mM DTT (dithiothreitol), 4% (w/v) SDS (sodium dodecyl sulphate; electrophoresis grade) and heating for 3min at 95˚C. 20µg of protein were loaded on gel. Polyacrylamide gels 12% (1.5mm thickness) were prepared in a Mini-PROTEAN 3 Cell apparatus (Biorad) according to the recipe (Table 2.1). Pre-stained protein standard (PageRuler Plus, Thermo Fisher) was loaded as marker and proteins were separated at 80-140V in SDS running buffer (Table 2.2

Table 2.1 Reagents to prepare gels for SDS-PAGE

Resolving gel	Stacking gel
15% Acrylamide mix (Protogel, 30% (v/v)) 1.5 M Tris-HCl pH 8.8 0.1% (w/v) SDS 0.1% (w/v) Ammonium persulphate (APS) 0.1% TEMED (v/v) Top up to final volume with dH_2O	5% Acrylamide mix (Protogel, 30% (v/v)) 0.13 M Tris-HCl pH 6.8 0.1% (w/v) SDS 0.1% (w/v) Ammonium persulphate (APS) 0.1% TEMED (v/v) Top up to final volume with dH_2O

2.4.2 Protein detection

Resolved proteins were transferred from gels onto nitrocellulose membranes (Amersham Protean, GE Healthcare) using a wet electroblotting system (Biorad) in

transfer buffer (*Table 2.2*) at 100V for 1h. Membranes were blocked with 5% dried skimmed milk (Marvel) in PBST at room temperature for 1h. Appropriate primary antibodies were diluted in blocking solution and incubated with membrane overnight in 4°C. Next, blots were washed in PBST three times for 5 minutes. Appropriate horseradish peroxidase-conjugated secondary antibodies in dilution 1:1000 in milk were added onto membranes and incubated for 1h at the room temperature. The blots were washed three more times with PBST. Enhanced chemiluminescence (ECL) reagents 1:1 (*Table 2.3*) were added onto a blot for 1min. Signal detection was performed by exposition of the blot to X-ray film (SLS) or Amersham hyperfilm ECL (GE Healthcare) and developing with a film processor (SRX-101A, KonicaMinolta). Protein expressions were normalized to that of house-keeping gene GAPDH.

Table 2.2 SDS-PAGE Buffers recipes

Running Buffer	**Transfer Buffer**
192mM glycine 25mM Tris-HCl pH 6.8 0.1% (w/v) SDS	192mM glycine 25mM Tris-HCl pH 6.8 20% (v/v) methanol
top up to final volume with dH_2O	

Table 2.3 ECL reagents

ECL Solution 1	**ECL Solution 2**
100mM Tris-HCl pH 8.5 2.5mM luminol stock 0.4mM p-coumaric acid top up to final volume with dH_2O	100mM Tris-HCl pH 8.5 0.02% (v/v) H_2O_2 top up to final volume with dH_2O

2.4.3 Antibodies

Proteins were detected using the primary antibodies described in Table 2.4. Detection of primary antibody was performed using the following secondary antibodies: rabbit anti-mouse 1:1,000 (Dako,PO260) and swine anti-rabbit 1:1,000 (Dako, PO217).

Table 2.4 List of antibodies used

Antibody	Company	Dilution
Mouse monoclonal anti-GAPDH	Abcam (a5441)	1:2000
Anti-P53 (DO-1)	Supplied by collaborators *	1:500
Monoclonal anti-HLA-DR	Thermo Fisher (MA5-11966)	1:100
Anti-SBP-tag	Merck (MAB10764)	1:1000
Rabbit polyclonal anti-HLA-B	Thermo Scientific (PA5-29911)	1:1000

*(Midgley et al., 1992)

2.5 Mild acid elution and immunoprecipitation experimental setup

The method for acidic elution was essentially as reported previously (Sturm et al., 2021). To control the pH of the elution method, collaborators in Gdansk sent to Edinburgh their Citrate buffer (pH 3.2) due to possible differences in pH meters between labs. We thus prepared citrate buffer to the same pH as that buffer sent to our lab in Edinburgh (pH3.35). A375 cell lines p53 WT and p53 KO were plated on 10cm Petri dishes and grown until 80% confluency. Cells were treated with 10µM NMDI14 (Sigma-Aldrich) for 48h. Three Biological replicates of 30 x 10^6 A375 cells (three 10 cm plate) from all listed conditions were used. The dishes were placed on ice, the medium was removed, and cell monolayers were gently rinsed three times

with 2 ml ice-cold PBS. After aspiration, 2ml of ice-cold citrate-phosphate buffer pH 3.3 were pipetted onto the cells and incubated for 3 minutes with agitation. Liquid was collected to Falcon tubes and spun at 1000 rpm/5min. Supernatant was moved to new tubes, flash-frozen and placed in -80°C to further analysis.

After acid elution, cells were washed three times with PBS, fresh media was added, and plates were placed back into the incubator for 24h. The next day acid elution was repeated, and cells scraped for proteomic analysis. Eluted peptides were flash frozen and sent for mass spectrometry analysis.

Preparation of samples for immunoprecipitation analysis involved growing A375 p53 wild type and knockout cells on 10cm Petri dishes until they reached 10^9 counts. Next, the cells were washed three times with PBS and collected by scraping onto Falcon tubes. Pellet was flash-frozen, and samples sent to collaborators group led by Tim Fugmann for immunoprecipitation analysis.

2.6 Pulldown

2.6.1 Protein G beads DO1 crosslinking

For crosslinking of protein G beads to chosen antibodies protocol by (DeCaprio & Kohl, 2019) was followed. Protein G agarose (Sigma) beads were washed three times with PBS and once with Buffer A (0.2% Triton, 1M Tris-HCl pH 8.0, 5M NaCl, 0.5M EDTA, ddH_20). Next beads were incubated with DO-1 antibody (1µg/20µl beads) overnight in cold room with rotation. Conjugated beads were washed once with buffer A, followed by two washes with 0.1M borate buffer at pH 9.0 consisting of 0.2M boric acid and 0.05M sodium tetraborate. 40mM dimethyl pimelimidate (DMP) was added into solution containing borate buffer and beads and kept at room temperature for 1h with rotation. After aspiration of the supernatant beads were washed twice with 1M Tris-HCl pH 8.0, then twice with PBS. DO1-crosslinked beads were kept at 4°C until further use.

2.6.2 Pulldown

DO1-conjugated beads or streptavidin agarose beads (Sigma) were placed in low-binding microtube and washed twice with PBS. Cell lysates (prepared according to Collecting cells and lysis) were incubated with beads overnight in 4°C with rotation. After series of washes, beads were flash-frozen on dry-ice. Mouse DO1 antibody (dilution 1:500) was used for p53 detection. GAPDH was detected with mouse anti-GAPDH antibody (dilution 1:1000).

2.7 Mass spectrometry techniques

2.7.1 Sample preparation after pulldown

Due to restrictions in travels during pandemic, I was unable to travel as intended for my training in hands-on mass spectrometry. Nevertheless, prior to the pandemic travel restriction I attended the Dubrovnik MS Summer School which involved a weeklong training in fundamentals of mass spectrometry. During the travel restrictions, I prepared MS samples, and their analysis was performed as described below by our collaborators group led by Borivoj Vojtesek at the Masaryk Memorial Cancer Institute, Brno, Czech Republic.

The elution of proteins from beads was performed by adding 1x NuPAGE LDS buffer and boiling samples for 5 min. Nextly, samples were separated on polyacrylamide gel. SDS-PAGE run lasted 10 min at 120 V and the gels were stained by Coomassie Brilliant Blue G-250. Gel bands were excised out of the gel, washed with deionized water, cut into small pieces and decoloured with a freshly prepared 200 mM solution of ammonium hydrogen carbonate (NH4HCO3, pH 7.8) in 40% aqueous acetonitrile (ACN) (v/v) for 20 min at 30°C and equilibrated in 50 mM NH4HCO 3 (pH 7.8) in 5% aqueous ACN (v/v) for 30 min at 30°C. The supernatant was aspirated, and the gels were dehydrated with ACN. The supernatant was removed, and the samples were reduced by the addition of 10 mM DTT for 1 h at 60 °C, followed by alkylation with 55 mM iodoacetamide in the dark for 45 min at room

temperature. The supernatant was aspirated, and the gel pieces were washed three times with equilibration buffer and dehydrated with acetonitrile. Trypsin digestion was carried out at 37 °C overnight using Promega sequencing-grade trypsin. Digested peptides were extracted using ACN, vacuum dried and desalted using C-18 micro spin columns (Harvard Apparatus) according to the manufacturer´s guidelines. Before mass spectrometry analysis, the evaporated peptide samples were dissolved in 2% ACN (v/v) with 0.05% aqueous trifluoroacetic acid (TFA) (v/v).

2.7.2 Mass spectrometry measurement on termination codon read through experiments

LC-MS/MS analysis was carried out using an Orbitrap Fusion™ mass spectrometer (Thermo Fisher Scientific) with a New Objective digital PicoView 565 nanospray source (Scientific Instrument Services) coupled to a Dionex™ UltiMate™ 3000 RSLC Nanoliquid chromatograph. The peptides were loaded into an Acclaim PepMapTM 100 nano trap column (nanoViper™ C18, 0.3 × 5 mm, 5 µm particle size, 100 Å pore size; Thermo Fisher Scientific) with loading buffer (2% ACN with 0.05% aqueous TFA (v/v)) for 5 min desalting at a flow rate 5 µl/min. Next, the peptides were eluted onto an Acclaim PepMap™ RSLC C18 (nanoViper™ 75 µm × 25 cm, 2 µm particle size, 100 Å pore size, Thermo Fisher Scientific) kept at 50 °C and separated by linear gradient elution over 64-min from 2-25% B and 6-min gradient from 25-60% B, followed by a 10-min wash step with 98% B, and 34-min of equilibration with 2% B. Mobile phase A was composed of LC-MS grade water and 0.1% formic acid (FA) while B was ACN 80% with 0.1% aqueous FA (v/v). The flow rate was 300 nl/min. The Orbitrap mass analyser was operated in positive ion mode, with the static positive ion spray voltage set to 2.45 kV, and ion transfer tube temperature to 275 °C. The master scan was acquired at resolving power settings of 120,000 (FWHM @ m/z 200), precursor mass range 350-1400 m/z. The MS/MS spectra of multiply charged ions were collected in data-dependent mode (top20 method of the most intensive precursors). Dynamic exclusion was set to 20 s. The

peptides were fragmented using collision-induced dissociation (CID) with the normalized collision energy setting at 35 %. The peptide fragments generated via CID were detected in an ion trap (rapid scan rate). The profile data was recorded in the master scan and the centroid data in MS/MS scans.

2.7.3 Data analysis of p53 readthrough peptides

Data analysis was performed with the software Proteome Discoverer™ version 2.5 (Thermo Fisher Scientific). The database search was performed with Sequest HT search engine against the cRap protein database (ftp://ftp.thegpm.org/fasta/cRAP) containing all p53 possible read-through fasta sequences. The search engine settings employed 10 ppm precursor mass tolerance and 0.6 Da fragment mass tolerance, enzyme Trypsin (full), considering up to two missed cleavages and the following dynamic modifications: methionine oxidation (+15.995 Da); protein N-terminal acetylation (+42.011 Da), and static modification: cysteine carbamidomethylation (+57.021 Da). The results of the search were further submitted to generate the final report with the strict false discovery rate (FDR) threshold of 1% using a fixed value PSM validator. The IMP-ptmRS node was used for automated and confident localization of modification sites with calculating individual probability values for each putatively modified site based on the given MS/MS data.

The original RAW files collected by mass spectrometry were processed and retrieved using PEAKS Studio 10.6 (Bioinformatics Solutions Inc. Waterloo, Canada) software. The search was done against a manually prepared database containing all p53 read-through and frame-shifted fasta sequences. The cRap protein database (ftp://ftp.thegpm.org/fasta/cRAP) with commonly appeared laboratory contaminant proteins was used too. The error tolerance for the precursor mass was set 10 ppm using monoisotopic mass and 0.6 Da for the fragment ion. Trypsin was used for a specific digestion of proteins with 3 maximum missed cleavages per peptide and modifications: methionine oxidation (+15.995 Da); protein N-terminal acetylation (+42.011 Da), and cysteine carbamidomethylation (+57.021 Da). The Spider search was used for finding more protein/peptide mutations. The results from data

evaluation were accompanied by the following parameters: (a) scan (indicates when MS/MS spectrum was measured); (b) -10lgP PEAKS peptide score (calculated for every peptide precursor spectrum match (PSM), the score is derived from the p-value that indicates the statistical significance of the peptide PSM, the acceptance threshold was set to correspond to 1% FDR; (c) mass of the precursor peak and its m/z; (d) error of the precursor peak in ppm; (e) retention time (RT), and (f) area under the chromatographic peak. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (PMID: 22073976)**.**

2.7.4 Immunopeptidomic MS analysis

I was also planning to travel to our collaborators lab in Gdansk to process immunopeptidomics. I was awarded NAWA PROM student exchange grant for travel and training at Institute of Cancer Vaccine Science; instead, I prepared acidic washes in Edinburgh and the material was sent to Gdansk for processing by Jakub Faktor as per 2.5. Purification of immunopetides from contaminants was performed on Oasis HLB cartridges as follows: briefly, cartridges were equilibrated according to manufacturer instructions. Peptides were washed and desalted on Oasis column followed by elution with methanol gradient. Further cleaning of peptides was performed on 3kDa mass cutoff filters and dried. Immunopeptides were resuspended in 2.5% acetonitrile (ACN) 0.08% trifluoroacetic acid. Approximately one fourth of the sample was injected onto Ultimate 3000 RSLCnano System (Thermo Scientific, MA, USA) liquid chromatograph. PepMap™ 100 C18 stationary phase was used to separate peptides using a gradient with increasing percentage of Acetonitrile, 0.1% formic acid (FA) in water with 0.1% formic acid. Eluting peptides were detected using Exploris 480 Mass Spectrometer.

The data dependent method has been constructed according to Chong et al (Chong et al., 2018). The data dependent method full scan was performed within m/z 350 Th to m/z 1650 Th mass range. The top 20 the most intense precursor ions were fragmented followed by their exclusion was set for 20 sec. Minimum allowed precursor ion intensity was 3.0e3, normalized collision energy type with fixed

collision energy mode was selected. The collision energy was set to 30%. Orbitrap resolution was set to 60000. Data type was centroid.

Data independent acquisition method was prepared inspired by (Gillet et al., 2012). Optimizations of DIA were made on precursor range, the initial method cycled within from 350 Th up to 750 Th and was only used for A375 dataset. Latest method covered also higher masses up to m/z 1100 Th. Setting the AGC target to 100% resulted in a method retrieving approximately 10 datapoints per peak which is in line with FDA requirements.

Spectral library generation, database search

Spectral library was generated by searching the DDA data in MSFragger and COMET search engines on-specific peptide digestion was set. Precursor mass tolerance was set to from -8 to 8 ppm and fragment mass tolerance was set to 10ppm. N-term acetylation and oxidation of methionine were set as variable modifications. Homo sapiens Uniprot database was used as search reference database. Spectral library was built from retrieved .pepXML files.

Peptide quantification

Peptide quantification was performed in Skyline software. Briefly, extracted product ion chromatograms were generated in Skyline software for all peptide hits listed in spectral library with 0.055 m/z mass tolerance. The confidence of extracted peakgroups was calculated as Qvalue in mPROPHET algorithm. Peakgroups with high Qvalue (>0.01) were suspended from experiment. Area under the ROC (Receiver's Operating Characteristic) curve (AUC) was determined for peakgroups, and summed intensities were determined, and log transformed. Statistical analysis was performed in MSstats module. Differential immunopeptide quantitation across specified conditions was performed pairwise via mixed-effect models. P values of fold changes were adjusted using Benjamini-Hochberg (Benjamini & Hochberg, 1995). Graphical output was generated in R packages plyr 1.8.6., ggplot2 3.3.5., Heatplus 3.0.0., ProBatch 1.8.0

2.8 Bioinformatic analysis

2.8.1 RNA-seq analysis

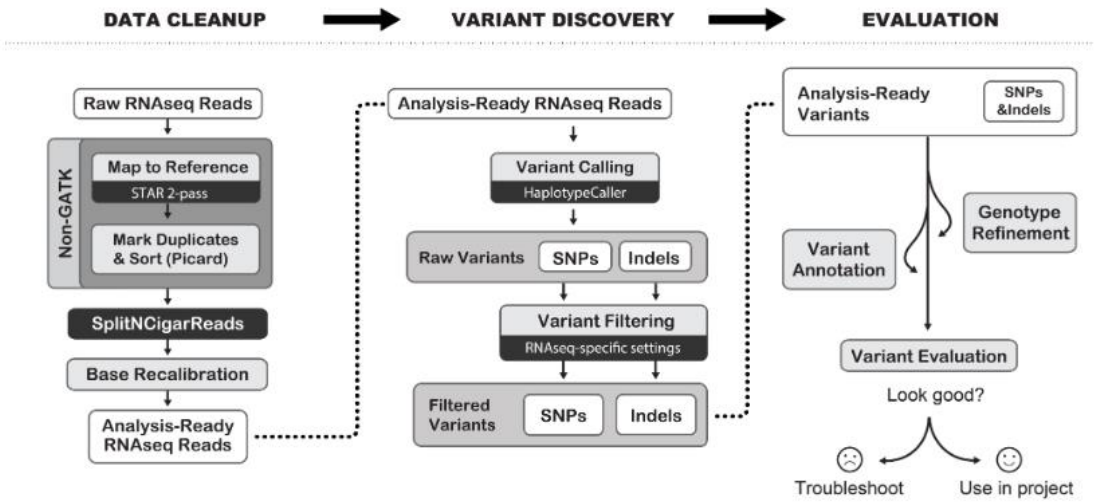
RNA was processed by Otogenetics for paired-end RNA sequencing analysis, using an Illumina HiSeq 2500 and designated 20 million reads. Paired fastq files from RNAseq reads were imported and processed using CLC Genomic Workbench 12.0 to measure total RNA expression levels. Sequencing reads were mapped to the human reference genome GRCh38 was taken for comparison with the following mapping settings: mismatch cost: 2, insertion cost: 3, deletion cost: 3. The results were compared using RNA gene expression for each condition, taking TPM (Transcript Per Million) values as reporting abundances and log2 (TPM condition 1 versus log2 (TPM condition 2) as comparison values.

2.8.2 Variant detection

The process of obtaining the list of sequence variants is complex and consists of several steps. Detection of mutational load was performed with the use of two platforms: CLC Bio and Mutect2 (Figure 2.1).

Approach supported by GATK, requires access to a big volume of computational power, therefore the analysis was performed on Prometheus server, to which access was granted through collaboration with Bioinformatic Group from the International Centre for Cancer Vaccine Science in Gdansk, Poland.

First step of RNA variant calling involved processing of RNA fastq files by FASTQC for quality control. Once the control is passed, the files are aligned to a reference genome using STAR version 2.6.1d, specifically using the 2-pass system where first, the novel detected junctions will be inserted into the genome indices, and then, all reads will be re-mapped using annotated and novel junctions. This process results in creation of a bam file, which later undergoes sorting, conversion and marking of the duplicates. Next, a base recalibration is performed, variants are called using HaplotypeCaller and then filtered using tumour/normal pairs.



.

Figure 2.1 GATK RNA variant discovery pipeline (source http://www.broadinstitute.org/gatk/guide/best-practices). Software firstly trims the adapter sequences and bases with low quality, to remove sources of potential error. Sequencing reads are then mapped to human reference hg38. Next, the software removes duplicate mapped reads to remove paired reads that have the same start and end coordinates. Sequences are filtered through dbSNP databases to remove known germline variants. Last, the variants are annotated to match mutated genes with their function.

In the CLC Bio approach, the software firstly trims the adapter sequences and bases with low quality, to remove sources of potential error. Sequencing reads were mapped to human reference hg38. Then the software removes duplicate mapped reads to remove paired reads that have the same start and end coordinates and are, thus, probably PCR duplicates. Sequences were also filtered through dbSNP databases to remove known germline variants.

The output files containing the recognised mutations, have been extracted and analysed using Variant Effect Predictor that determines the effect of the variants (SNPs, insertions, deletions, CNVs) on genes, transcripts, and protein sequence, as well as regulatory regions. Detected variants have been categorised according to the mutation types and frequency. For analysis of changes occurred in signalling pathways Ingenuity Pathway Analysis (IPA) software was used.

Discovered RNA variants were in silico translated and cut to obtained 8-12mer peptides, which could potentially be presented by MHC machinery.

Additionally, comparison of detected RNA variants with previously identified DNA variants by (Faktor et al., 2018) has been performed. CLC Bio platform allows for simultaneous analysis of DNA and RNA samples in order to detect changes occurring at genomic or transcriptomic level. Comparison tool was used with these set parameters for the low frequency variant detection: minimal sequence coverage of 5, minimum count of 1 for DNA variants and 2 for RNA, minimum frequency of variant at 5%. The most important tracks generated by this pipeline are information about variants found in both DNA and RNA and detailed information about variants found in DNA or RNA track.

2.8.3 NetMHC affinity analysis

Immune epitope prediction was performed on translated peptide sequences harbouring detected mutations using the NetMHCpan v4.0. The software was inspired by the principles introduced by Brusic et al. (Brusic et al., 1998), who augmented the information about peptide binding, that were used to train the prediction model with information about the amino acids defining the MHC binding cleft. Since, it has been shown that A375 cell line expresses HLA-A0301 and HLA-B0702, the samples were analysed in consideration of those HLA two isotypes. Peptides with a predicted affinity rank > 0.5 were classified as strong binders and will be used later for the construction of neoantigen vaccine.

3 Developing rules for stop codon readthrough in the presence of NMD inhibitors

3.1 Introduction

Translation from messenger RNA (mRNA) starts when an initiation codon enters the A-site of the ribosome and terminates when site A of the ribosome encounters one of the stop codons UAA, UAG, or UGA (Labunskyy et al., 2014). It has been shown that decoding of each codon varies, depending on the tRNA availability. After the translational machinery reaches a termination site, competition occurs between the termination complex and near-cognate tRNAs (Palma & Lejeune, 2021).

However, in some cases translation machinery may not recognise terminating codons and continue to extend the protein in C-terminal domain, the phenomenon is called stop codon readthrough (Schueren et al., 2014). Efficiency of this process depends on a competition between stop codon recognition by a class I release factor and decoding of the stop codon by a near-cognate tRNA. This decoding leads to the readthrough, by incorporating an amino acid in place of the stop and to the synthesis of an extended protein that ends at the next stop codon present in the same reading frame (Figure 3.1) (Baradaran-Heravi et al., 2016). Previous studies reveal that stop codons differ in readthrough efficiency, with UGA resulting in the highest, and UAA the lowest level (Bonetti et al., 1995; Loughran et al., 2014).

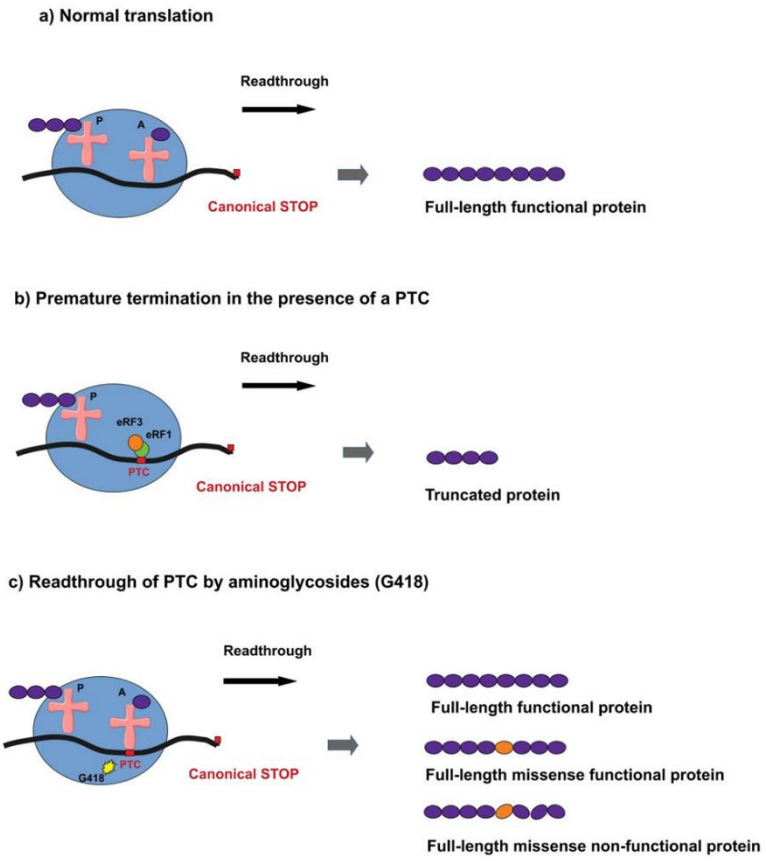


Figure 3.1 Protein translation in the presence and absence of PTC. a) Translation from messenger RNA starts when an initiation codon enters the A-site of the ribosome and terminates when site A of the ribosome encounters one of the stop codons UAA, UAG, or UGA. b) In the presence of PTC, translational machinery stalls leading to production of truncated non-functional protein.c) Aminoglycosides bind at the decoding centre of the eukaryotic ribosome altering the ability of translation termination factors to accurately recognize a PTC. That leads to the increase of pairing of near-cognate aminoacyl-tRNAs with the PTC and enables formation of full-length protein. Depending on the amino acid inserted at the PTC position, formed protein can be functional or non-functional.

Premature stop codons (PTCs) are a result of nonsense mutations, that occur when a codon for an amino acid is changed to a STOP codon, which in turn stops the translation of the protein earlier than expected. Transcripts carrying PTC form defective truncated proteins, that are non-functional and have been shown to lead to many pathologies (Bidou et al., 2004, 2017; De Boeck et al., 2014). Plethora of cancers are linked to a PTC found in tumour suppressor genes, resulting truncated protein unable to either inhibit cell proliferation or promote apoptosis. Substitution nonsense mutations account for 8% of TP53 mutations in cancer. The most common

nonsense mutation occurring in TP53 gene is R213X, affecting 1.3% of patients (Baradaran-Heravi et al., 2016).

Readthrough of PTCs can serve as a promising therapeutic approach in restoring the physiological role of affected proteins as well as, through incorporation of amino acid different from the wild-type sequence instead of STOP codon, generating potentially immunogenic isoforms (neoantigens). Small molecules exist that interfere with PTC recognition and lead to production of full-length, often functional protein (Clancy et al., 2001). Agents that have been proven to have the highest impact on the STOP codon readthrough are aminoglycosides. Their efficiency to interfere with PTC recognition has been evaluated for different genes (Bidou et al., 2004; Borgatti et al., 2020). Aminoglycosides bind at the decoding centre of the eukaryotic ribosome, which alters the ability of translation termination factors to accurately recognize a PTC. That leads to an increase of pairing of near-cognate aminoacyl-tRNAs with the PTC and enables formation of full-length protein (Baradaran-Heravi et al., 2016).

The PTC suppression is controlled not only by the readthrough drugs, but also by the PTC mRNA abundance, which depends on the efficiency of nonsense-mediated decay pathway (Palma & Lejeune, 2021). Therefore, NMD inhibition is considered an important target to mediate gene expression, and current strategies aiming to treat the negative effects of nonsense mutations search for chemical compounds targeting either premature translation termination, NMD, or both (Andreutti-Zaugg et al., 1997). It has been shown, that when used in cells carrying PTC mutated p53, NMD inhibition combined with the use of a PTC readthrough drug led to restoration of full-length p53 protein, upregulation of p53 downstream transcripts, and cell death (Martin et al., 2014).

The mechanism of amino acid incorporation during the readthrough process is still not fully understood. As cancers can induce novel frameshift-stop codon mutations, it is important to develop deeper rules that predict the type of mutated peptide sequences that can be induced by nonsense-suppression. Such insights might accelerate the development of drugs that can induce production of novel neoantigens for the induction of tumour rejection.

3.2 Aims

One of the main aims of this Precision Medicine PhD was to develop a pipeline of methods to identify mutated neoantigens that could form the basis for personalised cancer vaccines. In this case two cancer models were used: human melanoma cell line A375 and mouse melanoma B16 (not described in this thesis). In addition, NMD inhibitor was used as a “drug lead” due to the fact that genetic suppression of NMD was shown to result in tumour rejection in mouse models (Becker et al., 2021). Created pipeline also involved the application of DNAseq and RNAseq to make a cancer specific reference database that identifies not only mutated proteins but premature stop codons (in frame or out of frame) which readthrough could generate novel amino acids (neoantigens or tumour specific antigens). This first chapter therefore aimed to determine rules for readthrough of a model tumour suppressor gene TP53 carrying a nonsense mutation and trying to define the codon choice a cell makes in reading through a PTC after the treatment with G418 and NMDI14. The purpose of identifying such rules is to produce a ‘more improved’ reference database of mutated proteins including those incorporated during stop codon readthrough when we search for novel peptides in the immunopeptidome, as described later in Chapter 4 and 5.

3.3 Results

3.3.1 G418 and NMDI14 treatment leads to restoration of p53 expression in cell lines

Some cancers such as oesophageal adenocarcinoma have a high frequency of nonsense mutations (~30%; from pan-cancer atlas libraries; cBIOPORTAL); but in particular approximately 10% of all human cancers are reported to contain p53 PTCs (M. Zhang, Heldin, et al., 2018b). As such, the p53 gene is a good model to begin to

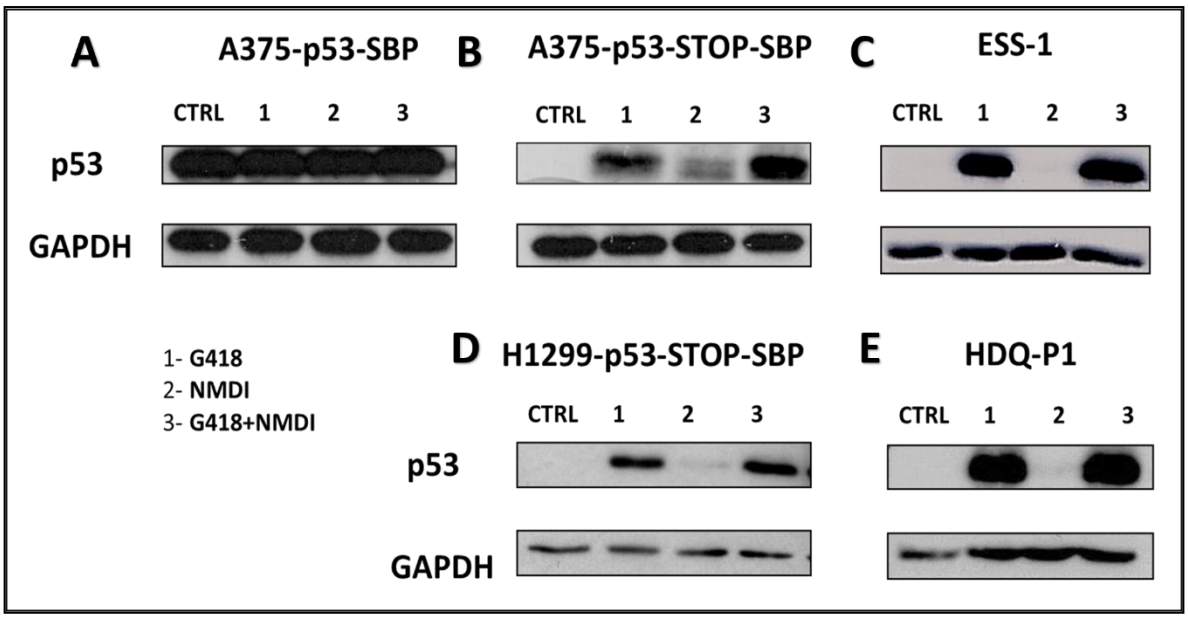


Figure 3.2 Western blot analysis of A375-p53-SBP (A), A375-p53-R213X-SBP (B), NH1299-p53-R213X-SBP (C), HDQ-P1 (D) and ESS1 (E) cell lines treated for 48h with G418 (1), NMDI14 (2) and combination of those compounds (3).The effects of p53 protein production after treatment with either an NMD inhibitor or G418 (Gentamycin) was assessed by western blot analysis as indicated. Lysates were immunoblotted with antibodies to either p53 protein or house-keeping gene product GAPDH.

address what types of amino acids can be incorporated at a PTC upon treatment with readthrough inducers that stimulate amino acid misincorporations at these codons. The ultimate therapeutic goal is to control the type of amino acid inserted

during readthrough to reactivate the wild-type tumour suppressor functions of p53. Since the most common nonsense mutation occurring in the TP53 gene is at R213X (M. Zhang, Heldin, et al., 2018a), the type of the amino acid incorporated at that site (such as the wild-type Arg itself) upon drug treatment could give a therapeutic solution by the reactivation of wild-type p53 function. To assess, whether NMD inhibition and/or PTC readthrough using the aminoglycoside G418 leads to the restoration of p53 protein expression, ESS1 and HDQ-P1 cell lines carrying R213X have been firstly used (Baradaran-Heravi et al., 2016) (Figure 3.2 C, E). Cells were treated with NMD inhibitor NMDI14 and/or G418 drug for 48h. P53 protein expression of treated cells was analysed by western blot and compared with untreated controls The NMD inhibitor (NMDi14) is not as effective in the production of full-length p53 (Figure 3.2 C, E). Interestingly, combining G418 with NMDi14 treatment did not lead to detectable synergy in the production of full-length p53 protein (Figure 3.2).

However, to model the PTC readthrough and test the sequence of the translated protein, cell lines not exhibiting p53 expression, A375 p53-/- and NH1299, were transiently transfected with plasmid encoding wild-type or mutated (R213X) p53 tagged with the Streptavidin-binding protein (SBP). The SBP-tag at the C-terminus was incorporated to potentially enable isolation of the full-length p53 protein from the cell lysates using streptavidin beads after the successful readthrough for subsequent trypsinization and peptide identification by mass spectrometry. Western blot analysis revealed that treatment of cells expressing the R213X gene from the plasmid with the NMD inhibitor (NMDI14) induced weak readthrough. On the other hand, G418 induced more pronounced restoration of the full-length p53 protein after 48h. Combination of two agents NMDI14 and G418 did not synergize to induce more full-length p53 to appreciable levels, than G418 alone. Additionally, treatment with NMDI14 and G418 did not affect the expression of the p53 gene without the premature termination codon (Figure 3.2 A). Interestingly, NMD inhibition in cell lines stably expressing R213X led to a much weaker p53 restoration, than in the case of plasmid-expressed protein (Figure 3.2 B, D).

Together, these data indicate that the H1299-p53 null cell line can be used as a model upon transfection with a plasmid containing C-terminally tagged SBP-tagged p53 with a stop codon at the 213 position and treatment with G418 should be used

for readthrough induction. The weakness of this approach is that G418 and NMDi may not induce the same set of readthrough rules, nevertheless, this was a proof of concept to determine whether we can detect readthrough amino acids sequences on a human target protein. The vast majority of cancer vaccines do not use mass spectrometry derived peptide data at all; relying on RNAseq libraries (Sahin et al., 2017). Therefore, to improve on future cancer vaccines it would be optimal to know the types of neoantigens generated by readthrough induction.

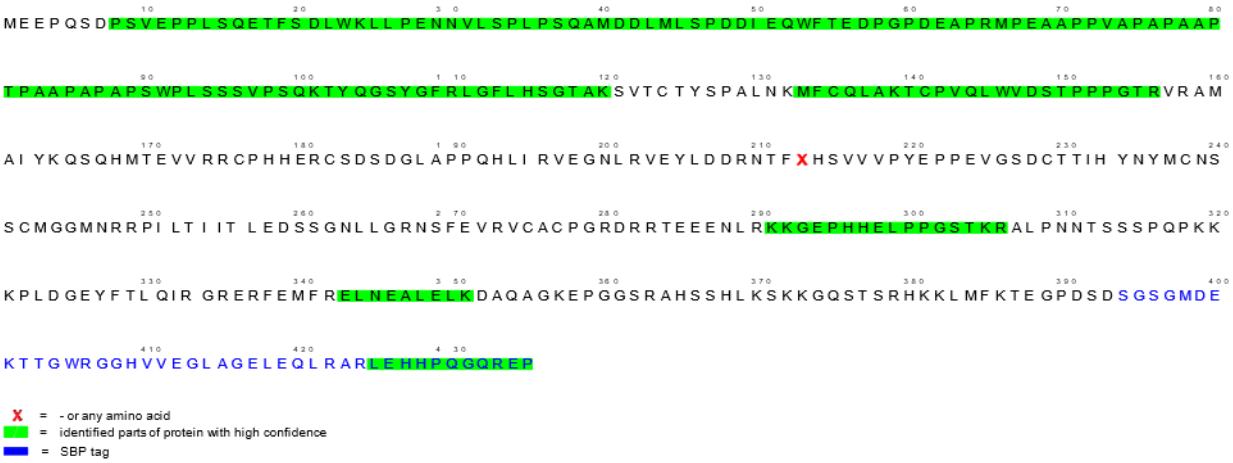


Figure 3.3 Tryptic peptide coverage of the p53 protein expressed from the R213X expression plasmid by TripleTOF analysis. H1299 cells were transfected with plasmid expressing the UGA termination codon at codon 213 and treated with Gentamycin. Samples were processed as indicated in the methods to produce tryptic peptides from the gel slice. The figure shows an example of one mass spectrometric run that detected readthrough of the premature termination codon as defined by detection of the SBP tag sequence at the very C-terminus (highlighted in green; 425-LEHHPQGQREP-435) but where there was no tryptic peptide coverage over the 213 stop codon. The remaining amino acids highlighted in green, mostly in the N-terminus prior to codon 213 highlight the sum of peptide sequence identified. Data analysis was performed with the software Proteome Discoverer™ version 2.5 (Thermo Fisher Scientific)

After establishing the efficiency of the readthrough, cell lysates were subjected to streptavidin-coated beads in order to pull down SBP-tagged p53 protein. The captured proteins were then trypsinized on the beads and analysed using mass spectrometry by our collaborators at Masaryk Memorial Cancer Institute in Brno, Czech Republic (Figure 3.3). The yield of the SBP-p53 protein pulldown was relatively low (data not shown), when using this method, which led to many rounds of

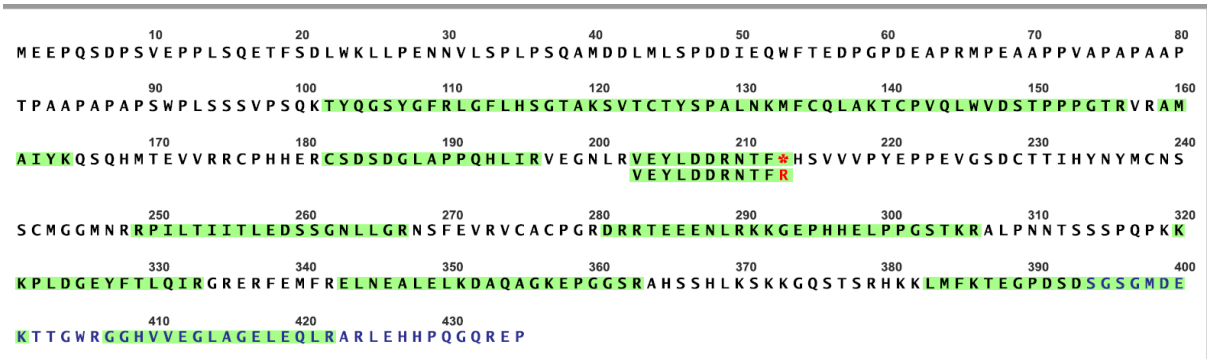
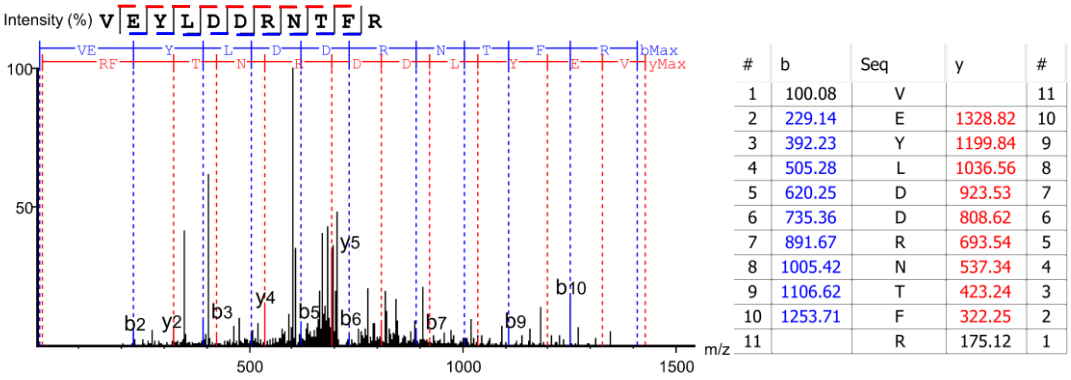


Figure 3.4 Tryptic peptide coverage of p53 protein expressed from the R213X expression plasmid eluted from DO-1-conjugated beads and analysed by Orbitrap Fusion.
Example of another independent representative mass spectrometric run with tryptic peptides from DO-1 bead pull down, that detected readthrough of the 213 premature termination codon resulting in the detection of a R incorporation at the 213 codon and creating the tryptic peptide 203-VEYLDDRNTFR-213. Data analysis was performed with the software Proteome DiscovererTM version 2.5 (Thermo Fisher Scientific).

unsuccessful optimization before we decided to modify the approach and to use anti-p53 antibody (DO-1)-cross-linked protein G beads. Using the DO-1 monoclonal antibody, allowed to recover more p53 protein. Using the DO-1 monoclonal antibody (which binds to the N-terminus of p53, before the PTC), we recovered more p53 protein (data not shown); however, on-bead trypsinization was poor (data not shown). Thus, the material captured in the DO-1 antibody bead precipitation was eluted with SDS sample buffer and we used SDS-gel electrophoresis to recover protein followed by in-gel trypsinization (Methods 2.7.1). This method yielded better peptide recovery as evidenced by the number of tryptic peptides identified after the addition of G418. TripleTOF mass spectrometric analysis showed that after treating cells with G418, immunoprecipitation with DO-1, processing by SDS-gel electrophoresis, and trypsinization, the peptide sequence surrounding premature stop codon at the 213 position was often not detected reproducibly (Figure 3.3). Interestingly, the spectra search revealed the presence of peptides closer to the C-terminus, which demonstrates that that the readthrough event occurred *(*Figure 3.3). Due to this low coverage and upon recommendations from our collaborators in Masaryk Institute in Brno, TripleTOF method was not used in the further sample analysis. Instead, the samples were re-analysed using Orbitrap Fusion mass

spectrometry machine, which showed higher p53 tryptic peptide coverage than the previous method (Figure 3.4).



#	b	Seq	y	#
1	100.08	V		11
2	229.14	E	1328.82	10
3	392.23	Y	1199.84	9
4	505.28	L	1036.56	8
5	620.25	D	923.53	7
6	735.36	D	808.62	6
7	891.67	R	693.54	5
8	1005.42	N	537.34	4
9	1106.62	T	423.24	3
10	1253.71	F	322.25	2
11		R	175.12	1

Figure 3.5 PEAKS analysis of 203-VEYLDDRNTF*-212 peptide detected after p53 STOP codon readthrough. MS/MS spectrum for VEYLDDRNTFR peptide and table with ion matches (red are y- ions and blue are b- ions). Further are introduced parametrs from the measurement and evaluation: scan: 16 697; -10lgP PEAKS score: 47.67; mass of the pre-cursor peak: 1426.63, m/z 714.35 (z = 2); error in ppm: 1.1; RT 27.10 min.

Unfortunately, introduced STOP codon at position 213 had a very low detection rate, compared to other peptides. In the proximity of position 213, only low frequency tryptic peptide 203-VEYLDDRNTF*-212 has been detected. Identified Arg (R) incorporation at the 213-stop codon produces a tryptic cleavage site at Arg-213 (Figure 3.4). However, we could not detect other possible amino acids incorporated at codon 213, presumably because the resulting tryptic peptide was too long with the next available trypsin cleave site at codon 248 and possible yield recovery of this larger tryptic peptide is poor. Interestingly, incorporation of R restores the wild-type sequence to p53 suggesting that the UGA stop codon, at least at codon 213 can be targeted to restore wt-p53 function. Using a different software (PEAKS) the spectra resulting from this is more visible (Figure 3.5). However, we could not detect p53 activity in cells treated with Gentamycin (data not shown), suggesting that other amino acids might be incorporated at the 213 codon that create a mutant p53; and as mutant p53 can inactivate wt-p53 through mixed oligomers, Gentamycin cannot reactivate full-length p53 function in these conditions. There are studies reporting reactivation of p53 function using a gene with a R213X mutation (Floquet et al.,

2011; M. Zhang, Heldin, et al., 2018a) but the type of amino acids incorporated at the premature termination codon position were undefined.

Additionally, the assessment of the effect of surrounding sequence on amino acids inserted at the PTC as a result of the readthrough was impossible, due to such low sequence coverage.

3.3.2 Mass spectrometry analysis reveals amino acids inserted at the PTC position

Apart from the variable R incorporation at codon 213 in the presence of Gentamicin (Figure 3.4), the poor tryptic peptide sequence coverage around the 213-position led to exploration of other PTC positions that could be developed to allow the detection of the amino acid inserted during the readthrough event. Since the tryptic peptide 122-SVTCTYSPALNK-132 was recovered reproducibly (Figure 3.4), we decided to incorporate termination codons within this tryptic peptide since it showed higher stability in the mass spectrometer. p53 expression plasmids were developed that incorporate a string of single synthetic UGA stop codons at $S^{127}P^{128}A^{129}$ within the tryptic p53 peptide 122-SVTCTYSPALNK-132.

By using these plasmids, we first tested the influence of the stop codon position on the readthrough. To increase the yield of the p53 pulldown, after transfection, cells were selected for 4 weeks with hygromycin to maintain in culture only those carrying hygromycin resistance gene alongside tagged p53 with nonsense mutation. The treatment of cells stably expressing the p53-UGA^{129} (Ala stop) mutation, immunoprecipitation and trypsinization of p53, resulted in the incorporation of R, C, A and W in the tryptic peptide at codon-UGA^{129} relative to the original p53 sequence 121-SVTCTYSPALNK-132. Readthrough of the PTC at codon 128 to yield the tryptic peptides substituted with R, C and W, compared to full-length p53 sequence 121-SVTCTYSPALNK-132. And finally, codon 128 readthrough lead to PTC substitution with R, C and W; relative to the wild-type p53 sequence 121-SVTCTYSPALNK-132.

Data analysis was initially performed with the use of Proteome Discoverer™ software version 2.5. Two independent biological duplicates, measured in triplicate, gave

variable yield of the novel peptides *(*Figure 3.6*).* However, since analysis of mass spectra is a very intricate field. Another software was utilized to confirm or disprove the sequence of peptides created during the readthrough process *(*Figure 3.7, Figure 3.8, Figure 3.9).The original RAW files collected by mass spectrometry were processed and retrieved using PEAKS Studio 10.6 (Bioinformatics Solutions Inc. Waterloo, Canada) software. The search was done against a manually prepared database containing all p53 readthrough and frameshifted fasta sequences. The cRap protein database (ftp://ftp.thegpm.org/fasta/cRAP) with commonly appeared laboratory contaminant proteins was also applied*.*

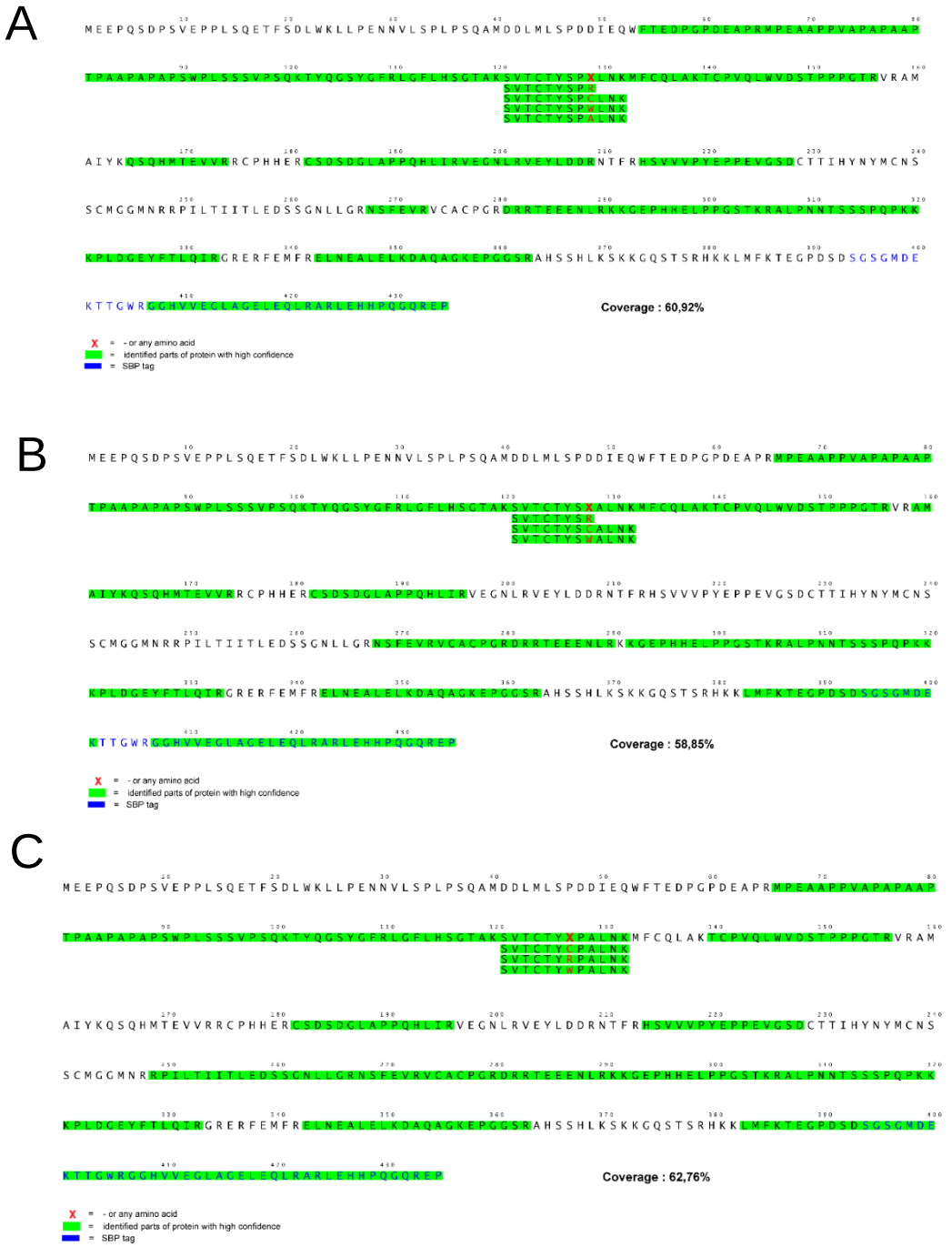
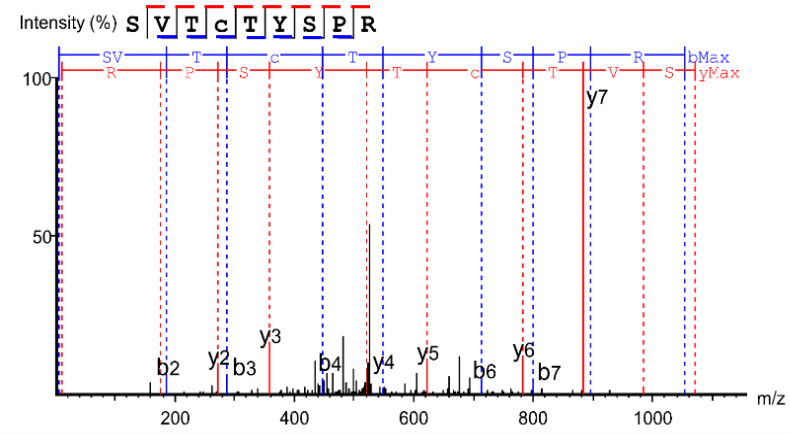


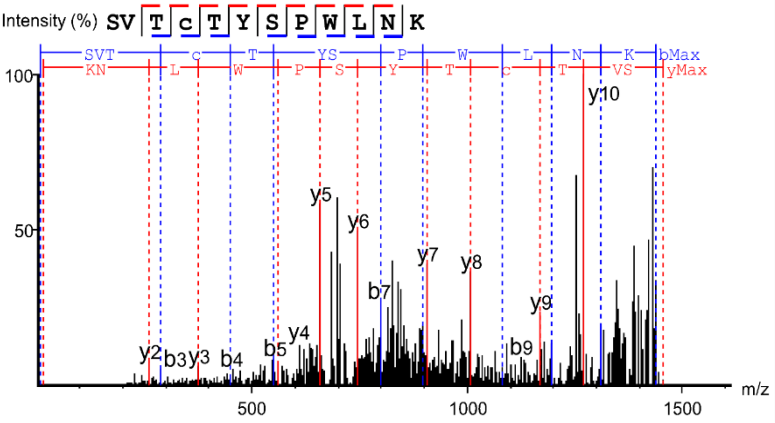
Figure 3.6 Tryptic peptide coverage of the p53 protein expressed from UGA premature termination codons at codons 127, 128, and 129. H1299 cells were stably transfected with plasmid expressing the UGA termination codon at codons (A) 129; (B) 128; and (C) 127 and treated with G418. Samples were processed as indicated in the methods to produce tryptic peptides from the gel slice. In the construction of an artificial termination codon model, we chose to focus on the tryptic peptide 121-SVTCTYSPALNK-132 because it was most often reliably detected in many independent mass spectrometric runs. (A). Example of one mass spectrometric run that detected readthrough of the premature termination codon at codon 129 to yield the tryptic peptides substituted with R, C, W and A, relative to the wild-type p53 sequence 121-SVTCTYSPALNK-132. The C-terminal SBP tag sequence was also recovered in this run (in green, as defined by detection of the SBP tag sequence at the very C-terminus (highlighted in green; 425-LEHHPQGQREP-435). (B). Example of one mass spectrometric run that detected readthrough of the premature termination codon at codon 128 to yield the tryptic peptides substituted with R, C and W; relative to the wild-type p53 sequence 121-SVTCTYSPALNK-132. (C) Example of one mass spectrometric run that detected readthrough of the premature termination codon at codon 127 to yield the tryptic peptides substituted with R, C and W; relative to the wild-type p53 sequence 121-SVTCTYSPALNK-132.

A



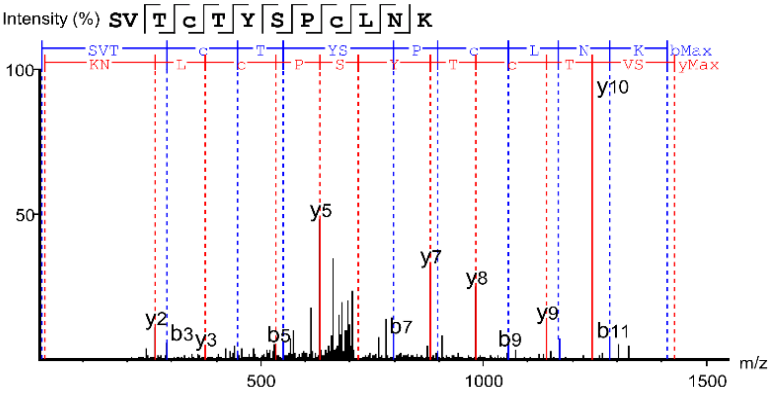
#	b	Seq	y	#
1	88.04	S		9
2	187.12	V	983.80	8
3	288.26	T	884.49	7
4	448.30	C(+57.02)	783.40	6
5	549.41	T	623.38	5
6	712.43	Y	522.33	4
7	799.52	S	359.27	3
8	896.53	P	272.20	2
9		R	175.19	1

B



#	b	Seq	y	#
1	88.04	S		12
2	187.11	V	1310.95	11
3	288.16	T	1211.70	10
4	448.19	C(+57.02)	1110.54	9
5	549.27	T	950.57	8
6	712.49	Y	849.60	7
7	799.37	S	686.44	6
8	896.65	P	599.46	5
9	1024.47	Q	502.30	4
10	1137.56	L	374.25	3
11	1251.65	N	261.15	2
12		K	147.11	1

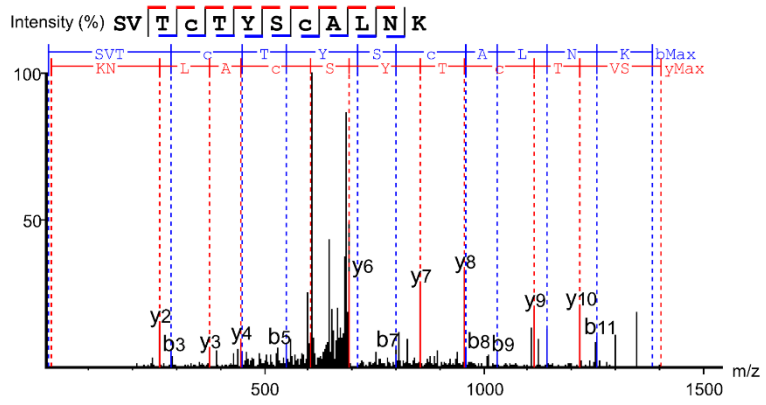
C



#	b	Seq	y	#
1	88.04	S		12
2	187.11	V	1342.61	11
3	288.20	T	1243.65	10
4	448.27	C(+57.02)	1142.58	9
5	549.37	T	982.58	8
6	712.30	Y	881.50	7
7	799.45	S	718.58	6
8	896.59	P	631.42	5
9	1056.57	C(+57.02)	534.33	4
10	1169.61	L	374.24	3
11	1283.71	N	261.19	2
12		K	147.11	1

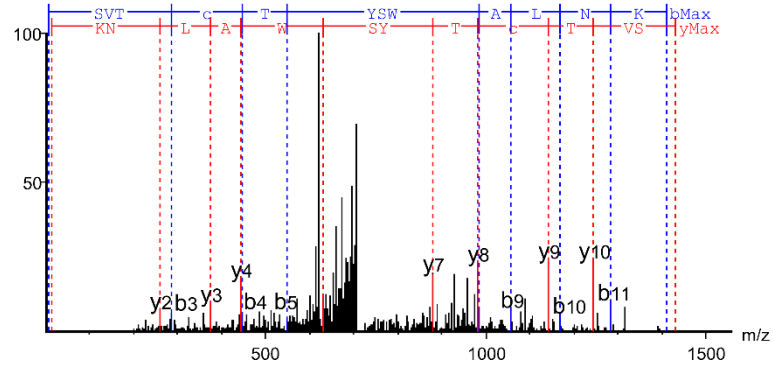
Figure 3.7 PEAKS analysis of p53-129-STOP codon readthrough. Peptides identified by Orbitrap Fusion mass spectrometry measurement from readthrough of plasmid encoding p53 with 129-STOP codon and analysed with PEAKS software MS/MS spectrum for SVTCTYSPR (A) and SVTCTYSPWLNK (B) and SVTCTYSPCLNK (C) peptide and table with ion matches (red are y- ions and blue are b- ions).

A



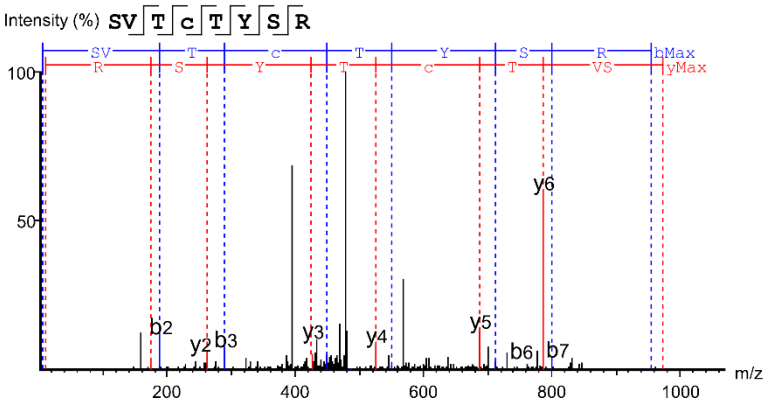
#	b	Seq	y	#
1	88.04	S		12
2	187.11	V	1316.60	11
3	288.24	T	1217.64	10
4	448.42	C(+57.02)	1116.61	9
5	549.34	T	956.63	8
6	712.29	Y	855.53	7
7	799.61	S	692.35	6
8	959.41	C(+57.02)	605.55	5
9	1030.44	A	445.30	4
10	1143.61	L	374.28	3
11	1257.62	N	261.16	2
12		K	147.11	1

B



#	b	Seq	y	#
1	88.04	S		12
2	187.11	V	1342.65	11
3	288.17	T	1243.80	10
4	448.21	C(+57.02)	1142.73	9
5	549.33	T	982.65	8
6	712.30	Y	881.59	7
7	799.33	S	718.39	6
8	985.50	W	631.50	5
9	1056.54	A	445.40	4
10	1169.74	L	374.30	3
11	1283.58	N	261.20	2
12		K	147.11	1

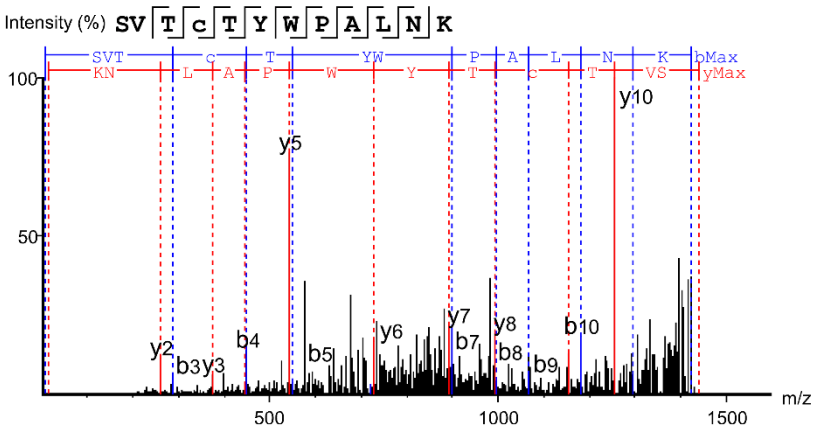
C



#	b	Seq	y	#
1	88.04	S		8
2	187.14	V	886.41	7
3	288.20	T	787.38	6
4	448.31	C(+57.02)	686.34	5
5	549.29	T	526.32	4
6	712.46	Y	425.26	3
7	799.41	S	262.23	2
8		R	175.13	1

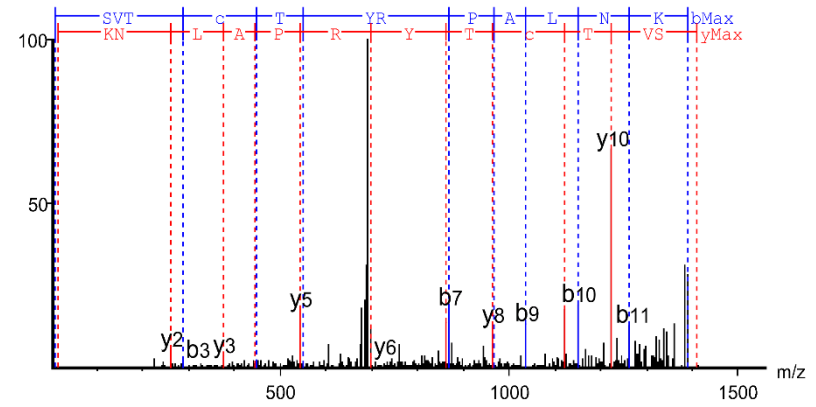
Figure 3.8 PEAKS analysis of p53-128-STOP codon readthrough. Peptides identified by Orbitrap Fusion mass spectrometry measurement from readthrough of plasmid encoding p53 with 128-STOP codon and analysed with PEAKS software. MS/MS spectrum for SVTCTYSWALNK (A), SVTCTYSWALNK (B) and SVTCTYSR (C) peptide and table with ion matches (red are y- ions and blue are b- ions).

A



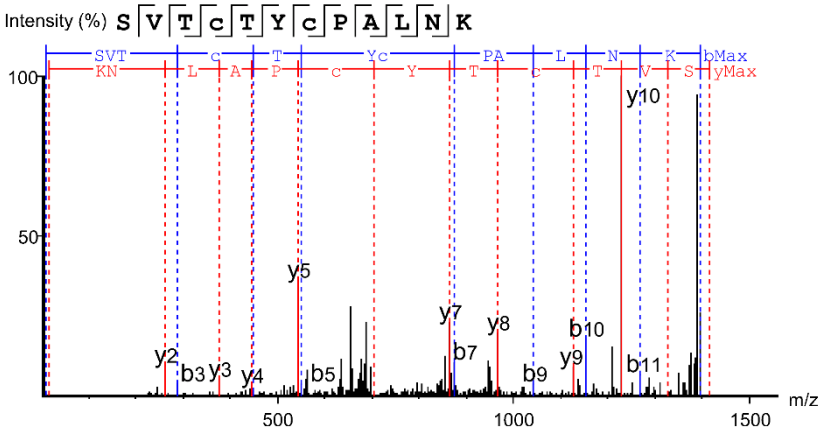
#	b	Seq	y	#
1	88.04	S		12
2	187.11	V	1352.67	11
3	288.18	T	1253.53	10
4	448.21	C(+57.02)	1152.60	9
5	549.10	T	992.54	8
6	712.30	Y	891.63	7
7	898.47	W	728.43	6
8	994.91	P	542.39	5
9	1066.56	A	445.49	4
10	1179.63	L	374.20	3
11	1293.67	N	261.17	2
12		K	147.11	1

B



#	b	Seq	y	#
1	88.04	S		12
2	187.11	V	1322.69	11
3	288.24	T	1223.61	10
4	448.27	C(+57.02)	1122.65	9
5	549.46	T	962.62	8
6	712.30	Y	861.56	7
7	868.47	R	698.47	6
8	966.01	P	542.38	5
9	1036.58	A	445.28	4
10	1149.73	L	374.32	3
11	1263.74	N	261.16	2
12		K	147.11	1

C



#	b	Seq	y	#
1	88.04	S		12
2	187.11	V	1326.67	11
3	288.19	T	1227.59	10
4	448.11	C(+57.02)	1126.45	9
5	549.29	T	966.50	8
6	712.30	Y	865.36	7
7	872.33	C(+57.02)	702.08	6
8	969.38	P	542.40	5
9	1040.44	A	445.44	4
10	1153.56	L	374.22	3
11	1267.58	N	261.12	2
12		K	147.11	1

Figure 3.9 PEAKS analysis of p53-127-STOP codon readthrough. Peptides identified by Orbitrap Fusion mass spectrometry measurement from readthrough of plasmid encoding p53 with 127-STOP codon and analysed with PEAKS software. MS/MS spectrum for SVTCTYWPALNK (A), SVTCTYRPALNK (B) and SVTCTYCPALNK (C) peptide and table with ion matches (red are y- ions and blue are b- ions). Further are introduced parameters from the measurement.

Semi-quantitative analysis in triplicates (Supplementary Figure 7.2) showed that R and W were the most frequently introduced at the codon-UGA129, which fitted the wobble model, where for the tRNA anticodon-mRNA pairing to occur even two out of three positions are sufficient.

We evaluated the p53-UGA127 (Ser stop) and p53-UGA128 (Pro stop) (Figure 3.7, Figure 3.8, Figure 3.9). In one independent experiment, we found Cysteine alone was found the most abundant amino acid inserted at the p53-UGA128 (Pro stop) *(*Figure 3.11 *A*). However, in another experiment, R incorporation dominated at the p53-UGA128 codon with levels of C insertions equivalent to W (Figure 3.10 C, Figure 3.9). We think the variability between independent experiments done on different dates stems from the multiple steps in the pipeline; (i) variability in the covalent conjugation of the DO-1 antibody to the beads can vary the amount of p53 recovered in the immunoprecipitation, and (ii) variability in trypsinization from the gel slice can also contribute to difference in p53 peptides recovered.

In the case of the p53-UGA127 codon, in both independent experiments, R was the most dominant amino acid inserted at approximately 50% with equivalent levels of W and C detected (Figure 3.9). Thus, it appears that all three codons at positions UGA127, UGA128, and UGA129, generally tolerated an R as the dominant amino acid incorporated as a result of a readthrough. A prior report using yeast and luciferase as the protein model with a UGA codon identified W as the predominant amino acid inserted in PTC over R (Roy et al., 2015). In addition, a recent report using bacteria cells as a model identified the W as a dominant amino acid replacement at the UGA codon using (mCherry-TGA-YFP reporters) or the RpsG nonsense mutation gene as encoding model proteins (M. Zhang, Heldin, et al., 2018a). There could be several reasons for why the R>W dominates in our model, whereas W>R dominates in the other models, The dominance of R over W at the UGA codon might be in the more stable G-U base pair between the position 1 (U) of the stop codon (UGA) and the position 3 (G) in the anticodon of Arginine, which is 3'-GCU-5'. It might also reflect differences in the cells used, as our model cell is the H1299 human cancer cell and the prior models used yeast cells as the model (Roy et al., 2015) or bacteria (M. Zhang, Heldin, et al., 2018a).

3.3.3 Influence of stop codon sequence on the choice of amino acids inserted at the PTC position

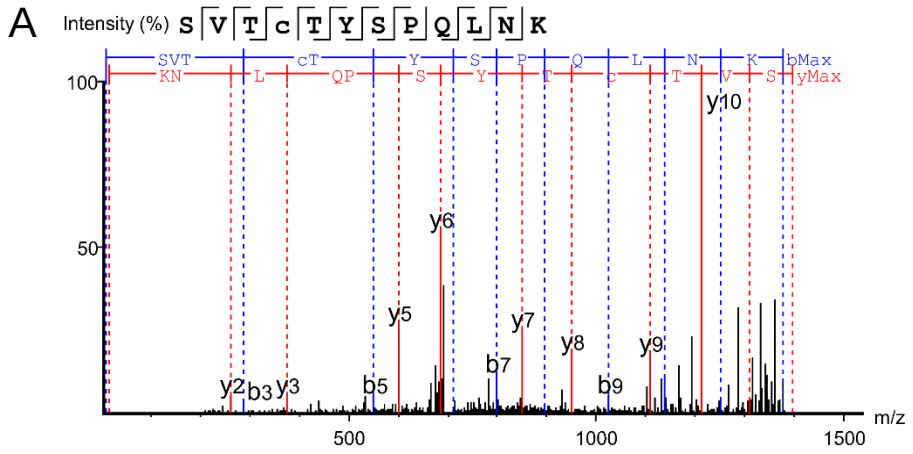
Various studies showed that different stop codons impact the efficiency of readthrough. To study this phenomenon, we designed plasmids p53-STOP^{129} that encoded three different termination codons UAA, UAG, UGA to test wether the preference of amino acid incorporation shifts with the stop codon change.

H1299 cell line was transfected with plasmids carrying p53 with nonsense mutation and hygromycin resistance gene for selection. Populations expressing studied constructs were then treated for 48h with G418 readthrough drug. P53 protein was isolated via beads-DO1 pulldown as described previously (Methods 2.6.2) and prepared for mass spectrometry measurement.

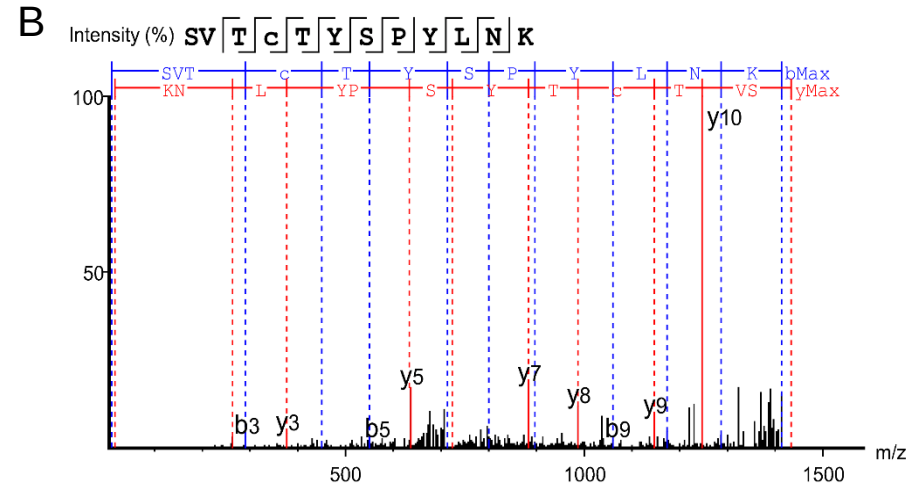
Analysis revealed that UAA and UAG stop codons demonstrated a preferance towards Q insertion and less frequently Y (Supplementary Figure 7.2 A, B, Figure 3.11). Both of these amino acids show two base pairing of these anticodons to UAA and UAG. In another independent experiment, R dominated at the p53-UGA^{128} codon with C levels equivalent to W (Supplementary Figure 7.2 *F).* At the UAA codon, in technical triplicates, Q was detected 71% frequency with the Y at 28% *(*Supplementary Figure 7.2 A*)* whilst at the UAG codon, in technical triplicates, Q was detected 77% frequency with the Y at 23% (Supplementary Figure 7.2 *A, B*). A prior report using yeast as the model also generally detected Q as dominating over Y at the UAG stop, with Q and Y incorporated at relatively similar levels at the UAA stop (Roy et al., 2015). The domination of Q in our model might be explained by the more stable G-U base pair between the position 1 (U) of the stop codon (UAA) and the position 3 (G) in the anticodon of Glutamine, which is 3'-GUU-5'. For example, presumably E is not detected as a readthrough amino acid replacement because the anti-codon for E is 3'-CUU-5' and the C-U bonding it not as stable as a 'wobble' replacement.

These data confirms that position of premature termination codon, as well as its sequence influences the choice of amino acid inserted during readthrough (Figure 3.12). These results also open possibilities of developing promising treatments of cancers linked with nonsense mutations in tumour suppressor genes. Therefore,

understanding the mechanism of this process is a vital point to optimise restoration of proteins carrying those mutation.



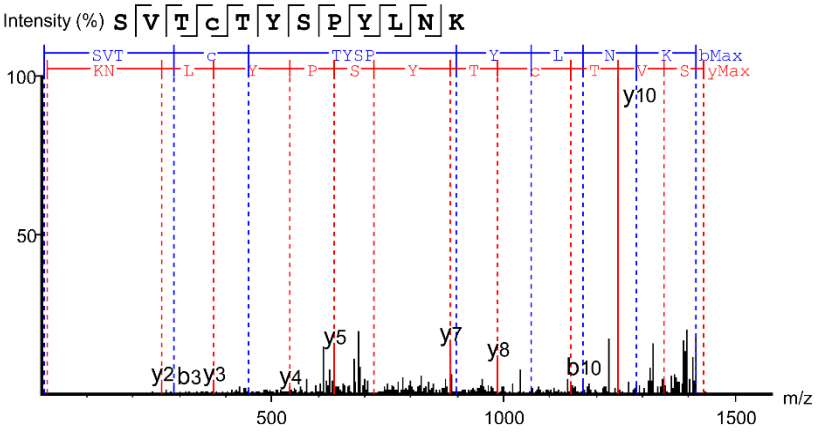
#	b	Seq	y	#
1	88.04	S		12
2	187.11	V	1310.95	11
3	288.16	T	1211.70	10
4	448.19	C(+57.02)	1110.54	9
5	549.27	T	950.57	8
6	712.49	Y	849.60	7
7	799.37	S	686.44	6
8	896.65	P	599.46	5
9	1024.47	Q	502.30	4
10	1137.56	L	374.25	3
11	1251.65	N	261.15	2
12		K	147.11	1



#	b	Seq	y	#
1	88.04	S		12
2	187.11	V	1345.65	11
3	288.13	T	1246.73	10
4	448.21	C(+57.02)	1145.67	9
5	549.30	T	985.59	8
6	712.88	Y	884.53	7
7	799.71	S	721.52	6
8	896.21	P	634.45	5
9	1059.56	Y	537.30	4
10	1172.56	L	374.33	3
11	1286.65	N	261.18	2
12		K	147.11	1

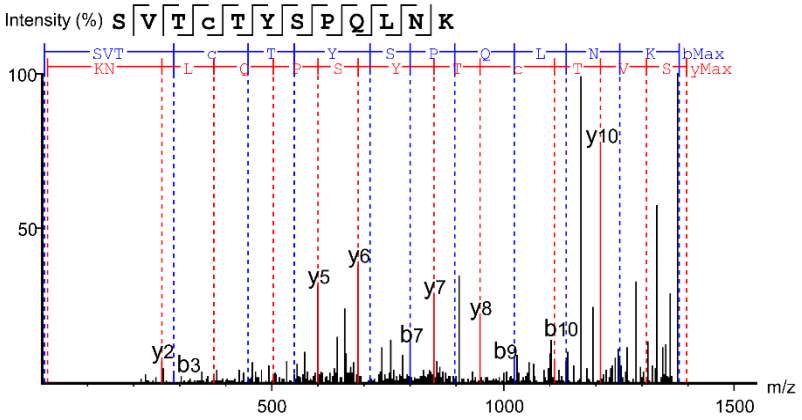
Figure 3.10 PEAKS analysis of peptides detected after p53-UGA127 codon readthrough A) MS/MS spectrum for SVTCTYSPQLNK peptide and table with ion matches (red are y- ions and blue are b- ions). B) MS/MS spectrum for SVTCTYSPYLNKpeptide and table with ion matches (red are y- ions and blue are b- ions)

A



#	b	Seq	y	#
1	88.04	S		12
2	187.11	V	1345.46	11
3	288.22	T	1246.80	10
4	448.23	C(+57.02)	1145.72	9
5	549.23	T	985.61	8
6	712.30	Y	884.52	7
7	799.33	S	721.41	6
8	896.68	P	634.43	5
9	1059.52	Y	537.32	4
10	1172.56	L	374.29	3
11	1286.62	N	261.15	2
12		K	147.11	1

B



#	b	Seq	y	#
1	88.04	S		12
2	187.11	V	1310.48	11
3	288.18	T	1211.70	10
4	448.20	C(+57.02)	1110.59	9
5	548.83	T	950.57	8
6	712.50	Y	849.59	7
7	799.28	S	686.41	6
8	896.60	P	599.42	5
9	1024.89	Q	502.08	4
10	1137.63	L	374.35	3
11	1251.63	N	261.10	2
12		K	147.11	1

Figure 3.11 PEAKS analysis of peptides detected after p53-UAA127 codon readthrough. A) MS/MS spectrum for SVTCTYSPYLNK peptide and table with ion matches (red are y- ions and blue are b- ions) B) MS/MS spectrum for SVTCTYSPQLNK peptide and table with ion matches (red are y- ions and blue are b- ions)

There are still many questions to answer regarding the mechanism of the PTC readthrough, such as whether the flanking amino acids or local folding motifs impact on the choice of tRNA-aa type, or can the choice be manipulated by regulating levels of tRNA synthetases/ligases.

3.4 Discussion

Readthrough of premature termination codons has long been recognized as a potential therapeutic strategy for diseases where a nonsense mutation leads to expression of a truncated protein. In these cases, induction of stop codon readthrough would allow for the production of a full-length protein that may retain wild-type function and ease the disease symptoms (Jagannathan & Bradley, 2016). Aminoglycosides are the most widely studied readthrough-inducing drugs that hold therapeutic potential in a wide variety of contexts, such as epilepsy (Huang et al., 2012), Duchenne muscular dystrophy (Malik et al., 2010), and cancer (Bordeira-Carriço et al., 2012). First studies showing that some aminoglycosides could partially restore the synthesis of an entire protein from a gene carrying nonsense mutation were published in 1985 (Burke & Mogg, 1985). With the expansion of knowledge about protein translation, novel molecules promoting PTC readthrough were identified, such as proposed by Baradaran-Heravi et al. the use of the minor component of gentamycin, gentamicin B1 (B1), as a potent readthrough inducer (Baradaran-Heravi et al., 2016).

Here, we aimed to develop a PTC model using a specific gene (p53) in a human cell model that exploits the ability of mass spectrometry to identify the amino acids inserted when a PTC occupies the ribosomal A site. Methodologies we developed using this cell model can be used to begin to manipulate or fine-tune the quantitative incorporation of specific amino acids of choice during the readthrough process and control the type of functional protein produced after premature termination codon read-through and/or the C-terminal termination codon read-through.

It has been shown that, similarly to natural ribosomal readthrough, the drug-induced readthrough is likely to be influenced by stop codon sequence context (Wangen & Green, 2020). Efficiency of stop codon readthrough has been shown to depend on various factors. Genome-wide studies in both yeast and human cells have identified the UGA stop codon as more permissive to readthrough than the other two stop codons (Mangkalaphiban et al., 2021; Wangen & Green, 2020). Studies proved that sequence surrounding the stop codon, both upstream and downstream, can significantly influence the rate of readthrough (Wangen & Green,

2020), highlighting that sequences directly downstream of the stop codon have the strongest effect on readthrough rate.

However, when we discuss the therapeutic effect, not only rate, but also quality of readthrough is of high importance. By quality we can understand the insertion of amino acids at the stop position and recovery of the wild-type function. Here we report on the development of a model cell line using nonsense mutations in p53 as a tool to begin to define the type of amino acids incorporated during readthrough. In our model we observed the incorporation of R, W, or C as the dominant amino acids at a UGA termination codon, as well as Q or Y at the UAA and UAG termination codons, similarly to those reported in other models (Roy et al., 2015; M. Zhang, Heldin, et al., 2018a). Although there were some quantitative differences (R over W at UGA and Q over Y at the other two) that suggest position 1 of the codon and position 3 of the anticodon play a more important role in readthrough in the presence of Gentamycin (Figure 3.12).

Identification of peptide sequence can be a challenging process. Starting with the method of sample preparation for the mass spectrometry analysis. Majority of proteomics studies use a bottom-up strategy that involves digesting proteins to peptides, usually with trypsin, to overcome the vast physiochemical diversity of proteins in the cell. Samples prepared like that undergo Liquid Chromatography with tandem mass spectrometry (LC-MS/MS), which involves physical separation of target compounds (or analytes) followed by their mass-based detection. However, not all tryptic peptides can by analysed through LC-MS/MS, and due to this, proteins found in complex samples are usually annotated through partial sequences. The coverage of detected protein can estimate between 10-100% and depends on the protein type and experimental setup (Alfaro et al., 2021). Advancements in computation approach to MS data had a massive impact on LC-MS/MS improvement (Alfaro et al., 2021). These methods are used to assign peptide sequence to the measure ion current.

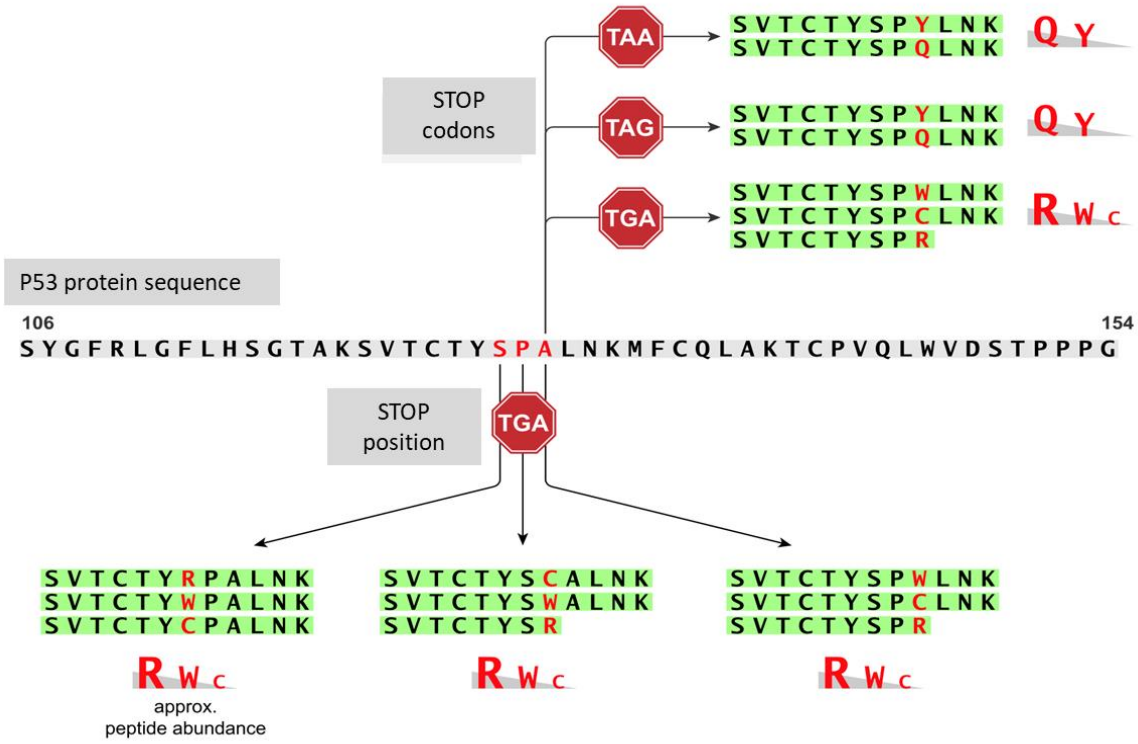


Figure 3.12 Summary of amino acids incorporated at premature termination codon in p53 depending on the position and type of STOP codon. Experiments focused on amino acid incorporation after induced readthrough of PTCs at position 217, 218, and 219 of p53, as well differences between amino acid insertions during readthrough of TAA, TGA, TAG termination codons.

Our model shows variability in the p53 detection, which might be due to several steps including variability in the amount of DO-1 antibody bound to the beads as a batch-to-batch variation or in the amount of p53 produced during the read though process. Thus, because of this variability, technical triplicates were required in any given experiment to allow for semi-quantitation.

Ma et al, in their study created a hypothetical peptide library containing all combinations of LEHHHHHHXXX peptide, wherein X was any amino acid incorporated at the three residue positions directly downstream of the UAG codon (N. J. Ma et al., 2018). This reference data, served as a platform to identify novel peptides. They used tagged bacteria proteins to study mechanism of readthrough, which gave them a huge amount of starting material for trypsinization and resulted in high signal and coverage of studies proteins (N. J. Ma et al., 2018).

This study points out the limitation of the method, which is that even when we use pulldown to isolate p53, sample abundance is still limiting to cover fully the protein landscape. And identification of the noncanonical peptides may be even more difficult from real cells because the abundance too low to detect.

One of the other difficulties in identifying the amino acid incorporated at any nonsense codon is that the readthrough product needs to be stable using mass spectrometry. One cannot guarantee that particular region of any protein can produce a stable tryptic peptide. Indeed, in our p53 map, we observed less than 50% coverage of the entire protein over several runs, thus reducing the regions we could use to study readthrough. As such we focused on a region in the N-terminus of p53 that produces a relatively reproducible tryptic peptide that allowed us to analyse the effects of codon change in the context of a UGA nonsense codon (127, 128. Or 129) and the effects of codon changes at one codon (129).

Recent publication by Ma et al, shed a new light at translational readthrough (N. J. Ma et al., 2018). The study concluded that when organisms are recoded to obtain non-assigned codons, compensatory mechanisms emerge, such as frameshifts and stop codon readthrough. Study showed that the lack of release factor 1 (RF1) expression in recoded Escherichia coli resulted in ribosome stalling, mis-translation (e.g., frameshift), or misincorporation of natural amino acids at encoded UAG termination codon. It has been proposed that some codons remained unassigned or decoded by lack of specificity or ribosome frameshifts (N. J. Ma et al., 2018). Utilizing organisms with simplified genetic codes could allow for clarification of the exchangeability (or lack thereof) of different amino acids and enable studies of mechanisms by which the translation apparatus recovers from problems in translation. This also highlights novel amino acid information that can be created using the drugs like G418 for which this study provides a proof of concept. Future developments of this pipeline could lead to creation of a Ref seqs of read through at the two stop codons, where we could ask if we see any novel termination or readthrough, or misincorporation events.

Our understanding of mechanism underlining the readthrough and rules of the amino acid incorporation in place of PTC are still far from complete. However, for the therapeutic aspect of readthrough studies it is important to know, which

incorporation would lead to overcoming a nonsense codon and restoring the function of the normal protein. In the case of p53, it is sensitive to inactivation at most of its codons thus if we cannot rely on these amino acids to reactivate function-W, R, C, Y, or Q-then we would need strategies that can make amino acid choice at any one codon tuneable. We hope this model will help develop such tuneable strategies to develop the PTC readthrough language.

In later chapters, as produced RNA reference sequence using the genomics pipeline had many nonsense mutations, we did make a theoretical reference sequence which included W, R, C, Y, or Q incorporation in PTCs. However, this did not result in us detecting any such neopeptides (see next chapters). Indeed, from hundreds of mutated proteins, we only detected a very small number of mutated peptides which once again highlights one limitation of mass spectrometry. That is mRNA refseq based on cancer genome can indeed give rise to theoretical mutated neopeptides that could be future vaccine candidates; however, the direct detection of such mutated peptides using mass spectrometry remains limiting. This will be discussed in the next chapters.

4 Analysis of mutant tumour immunopeptidome using an isogenic wild-type and p53 null cell model

4.1 Introduction

The immunopeptidome is a highly variable group of peptides bound to MHC complexes and presented on the cell surface to T-cell receptors, which, after their recognition, activate T-cells (Chong et al., 2018). Mass spectrometry examination of immunopeptidome facilitates pinpointing to the peptides that more likely induce the antitumorigenic immune response. Experiments show that MHC-peptide complexes can be dissociated from a cell surface and the bound epitopes eluted (Chong et al., 2021; Sturm et al., 2021). Isolation of MHC-bound peptides (MBPs) is typically performed with two established methods: immunoprecipitation (IP) and mild acid elution (MAE). These methods have been used to reveal the neoantigentic load of cancer cells in search for specific peptides inducing anti-tumour immune response (Kote et al., 2020).

Novel sequencing technologies has allowed to define patient-specific cancer features thus highlighting the unique genetic signature of any given tumour (Rizvi et al., 2015). This presents an opportunity to develop precision, personalized therapeutics based on expressed, mutant proteins from an individual tumour. Such mutant proteins could inform target pathway choice for the development of novel biologics that aim at the mutated cancer landscape in a patient-specific manner. Understanding the expression of mutant proteins could also form a platform for the development of mutant neoantigen based anti-cancer vaccines, that could be based on synthetic proteins (Schumacher et al., 2014), dendritic cells primed with neoantigen (Carreno et al., 2015), nucleic acids such as RNA (Deng et al., 2022), or synthetic viral vectors (Colloca et al., 2012). However, the study of mutant proteomes is only in its infancy. The relative difficulty in this task is that building

mutant proteomes requires integration of robust and user-friendly methods linking the fields of informatics, mass spectrometry, and cancer biology.

Computational methodologies as applied to the cancer research field are emerging as approaches to define the expressed, mutated genome. There are several challenges with optimizing platforms that integrate DNA sequencing, RNA sequencing, and mass spectrometric datasets (Kanaseki et al., 2019). One overall challenge is integrating these molecular data into one pipeline; and depends on the variant detection platform used for DNA sequencing; the algorithms for defining mutations with RNA sequencing datasets in exons, non-coding RNAs, and introns (Chong et al., 2020). For example, the expression of intron encoded mutant peptides is almost completely unexplored at a systems biology level (Apcher et al., 2015), as are cancer-specific RNA editing events and tumour-specific spliced mRNAs that create cancer-associated polypeptide epitopes (Bradley & Anczuków, 2023). In addition, different mass spectrometers and sample preparation methods, coupled to tumour heterogeneity, result in an incomplete understanding in the source and extent of expressed mutated proteins using cancer-specific DNA and/or RNA sequencing reference databases.

Tumour suppressor TP53 is the most frequently mutated gene across variety of human cancer types. Interestingly, studies show that melanomas frequently retain p53 wild-type allele, however considering the malignant character and genomic instability of melanomas, p53 fails to play its role, suggesting that p53 activity could be suppressed by a mechanism different from p53 mutation (Albino et al., 1994; L. Ding et al., 2014). Various studies have revealed additional roles that p53 play in tumour suppression, including regulation of tumour metabolism, microenvironment signalling, induction of programmed cell death dependent on iron (ferroptosis) and cancer stem cell renewal or differentiation (Gnanapradeepan et al., 2018; Midgley et al., 1992, p. 53). Therefore, targeting this vastly altered protein can hold a promise for potential cancer vaccine markers.

4.2 Aims

Purpose of this chapter was to create a mutant reference dataset to define mutated proteins in a key cancer cell model, a melanoma cell model named A375. This serves as an expected model that will be used in personalized vaccine target discovery; how to take a tissue or cell line from a patient's tumour and define the mutant immunopeptide space. In addition, there is an emerging question in the cancer research field; how does inactivation of a major cancer-associated gene impact on the mutant proteome landscape?

In this chapter identification of immunopeptidome landscape was performed using two different methods, as well as defining the types of immunopeptides that are p53-dependent.

Detection of the immunopeptidome may support the inclusion of tumour-associated neoantigens for the vaccine preparation, as well as expand the knowledge about the pathways leading to the peptide presentation on the cell surface. However, a key issue is the depth of the immunopeptidome per volume of cell input: Currently RNAseq – ELISPOT are the two tools used to predict cancer vaccine mutant peptide candidates. We would need to be able to go 'deeply' into the mutant immunopeptidome to add value. Thus, the secondary purpose of this chapter is to define the weaknesses of mass spectrometry and how one can improve on its usefulness.

4.3 Results

4.3.1 Transcriptional changes detected between p53 null and wild type cell line model and creation of reference database.

4.3.1.1 Differential expression analysis

RNA sequencing has been an emerging tool of transcriptomic analysis, that enables identification of mutational changes occurring on RNA level, as well as measuring gene expression (Blaydes & Hupp, n.d.). For our model, we used human melanoma A375 cell line that has classically served as a model of p53 activity in response to DNA damage (Blaydes et al., 2001; Blaydes & Hupp, n.d.). RNAseq was performed on total RNA isolated from A375 cell line. Transcriptomic analysis was carried out using CLC Biomedical Genomics Workbench 12 software. The program

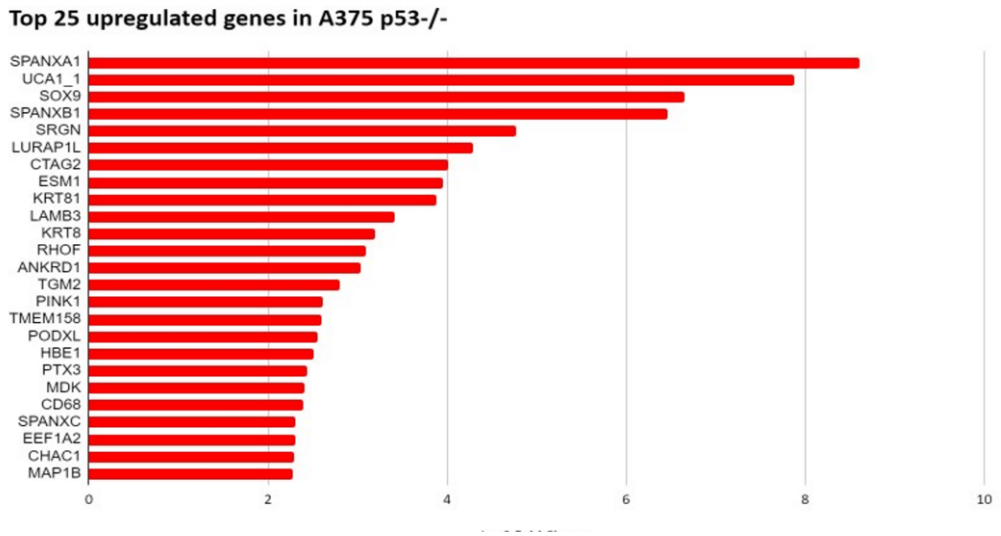


Figure 4.1 Summary of quantitative transcript expression of top 25 upregulated genes in A375 p53 null cell line compared to p53 wild-type model. Values presented are calculated as log2 fold change of transcript expression of A375 p53 KO cell line vs WT.

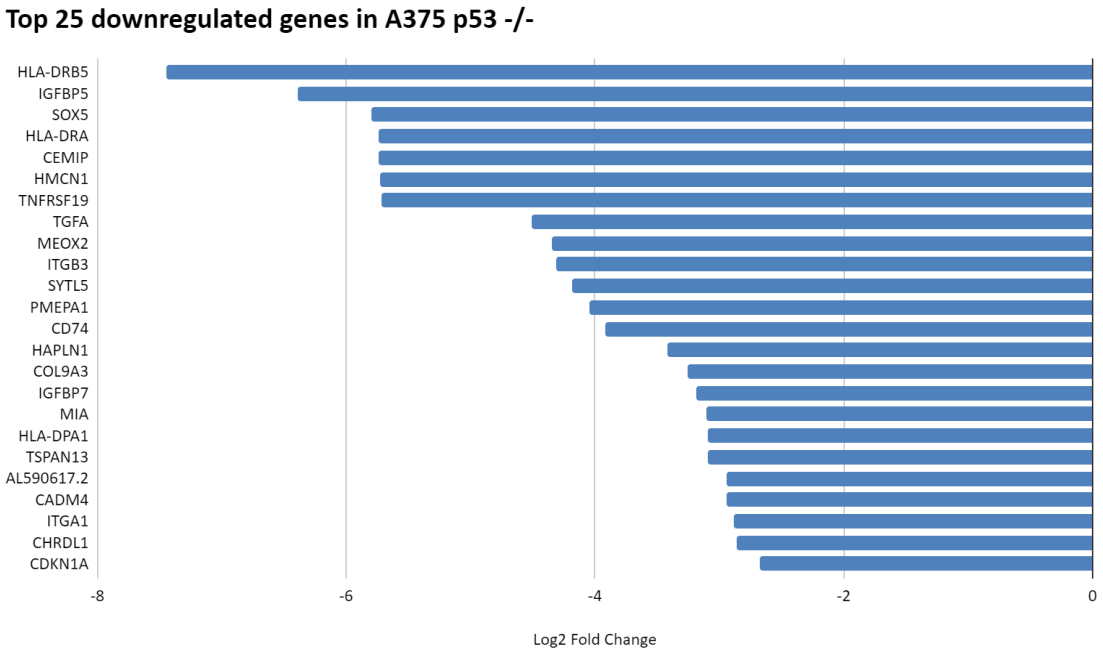


Figure 4.2 Top 25 downregulated genes in A375 p53 KO cell line. Values presented are calculated as log2 fold change of transcript expression of A375 p53 KO cell line vs WT.

enables a comprehensive and accurate data analysis for genetic aberrations. Imported fastq files were aligned to the human reference genome hg38 and expression values for each gene and each transcript were calculated compared with the reference.

Differential expression of A375 p53 KO cell line genes was carried out by using wild-type transcript levels as the baseline model to identify transcriptomic changes occurring after loss of p53 gene. Analysis revealed that among top 25 upregulated genes in A375 p53 knockout cells, were transcripts encoding Cancer/Testis Antigens, such as SPANX family or CTAG2, which was shown to be expressed in a wide array of cancers including melanoma, breast cancer, bladder cancer and prostate cancer(J. Liu et al., 2019), as well as Urothelial Cancer Associated 1 (UCA1) that produces a long non-coding RNA found to be upregulated in bladder cancer and plays a regulatory role in cell proliferation (Z. Ding et al., 2021; Lebrun et

al., 2018). In addition, an alternative open reading frame product of CTAG2 was termed CAMEL and classified as a tumour antigen that is recognized by melanoma-specific cytotoxic T-lymphocytes (Aarnoudse et al., 1999). Evaluation of those genes in more depth as potential therapeutic targets, could be further explored especially in correlation with p53 depletion (Figure 4.1).

Whereas lack of functional p53 gene lead to downregulation of crucial genes involved in MHC class II processing, such as HLA-DRB5, HLA-DRA, HLA-DPA1 (Figure 4.2). As well as decrease in transcripts coding Cell migration-inducing hyaluronidase 1 (CEMIP), which has been shown to be implicated in the process of metastasis in a variety of malignant tumours, including breast cancer (X. Dong et al., 2021).

Ingenuity Pathway Analysis (IPA) software served as a tool for identification of changes in specific pathways occurring after p53 deletion. IPA is a web-based functional analysis tool for interpretation of omics data. Software uses accessible datasets to identify the most relevant signalling and metabolic pathways, molecular networks, and biological functions for the list of supplied genes. Comparison of expression of genes in A375 p53 null cell line with wild-type indicated that the highest number of deregulated genes where part of the oestrogen-mediated S-phase entry as well as upregulation of PD-1-PD-L1 pathway. As predicted, top downregulated pathway was involved with p53 signalling. Interestingly, other supressed pathways related to regulation of immune system, such as dendritic cell maturation, Th1 pathway or leucocyte extravasation signalling (Figure 4.3).

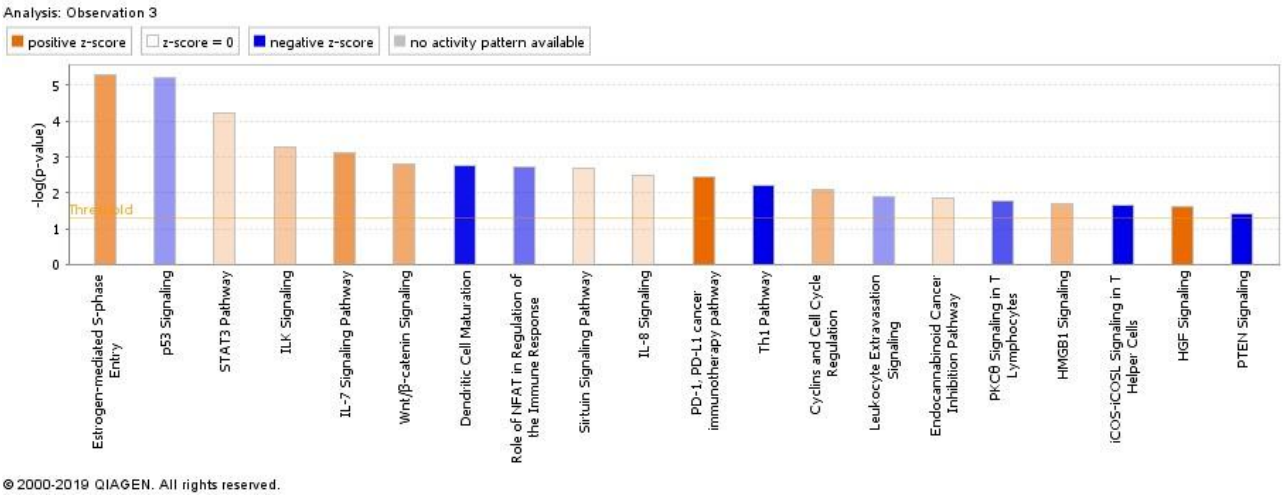


Figure 4.3 Ingenuity pathway analysis of A375 p53 null cell line versus wild type. Canonical pathways significantly enriched in A375 p53 null cell line when compared with wild type. Significantly enriched canonical pathways were identified with a right-tailed Fisher's Exact Test that calculates a P-value determining the probability that each canonical pathway associated with the dataset was due to chance alone. The P-values were corrected for multiple testing using the Benjamini-Hochberg method for correcting the false discover rate (FDR). The z-score indicates predicted activation state of the canonical pathway. Blue colour or lighter shades of blue indicate a negative z-score and down-regulation of the pathway, and orange colour or lighter shades of orange indicate a positive z-score and up-regulation of the pathway. Ratio shows the number of significantly expressed genes compared with the total number of genes associated with the indicated canonical pathway.

4.3.1.2 RNA variant analysis

The next step of RNAseq analysis included identifying mutational landscape of the transcriptomic data, variant calling. This process was carried out with the use

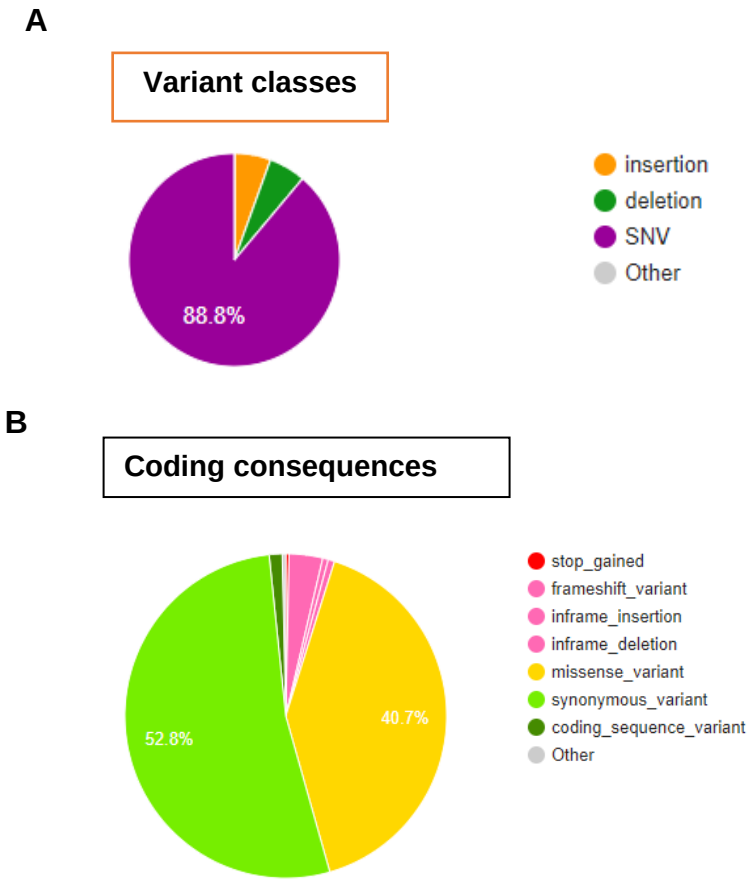


Figure 4.4 Annotation of identified RNA variants in A375 p53 null cell line by Variant Effect Predictor software. A) Changes detected on transcriptomic level were classified into four categories: insertion, deletion, single nucleotide variant (SNV) and other. B) Pie chart distribution of specific mutation types found in coding regions.

of two bioinformatic tools: CLC Bio software and Genomic Analysis Toolkit (GATK), Mutect2. CLC Bio, among other application, is used as a variant calling tool and does not require computational coding, which makes it a user-friendly software accessible to all life-scientists (Figure 2.1).

The process of obtaining the list of sequence variants is complex and consists of several steps. The software firstly trims the adapter sequences and bases with low quality, to remove sources of potential error, and maps reads to human reference. Then removes duplicate mapped reads to remove paired reads that have the same start and end coordinates and are, thus, probably PCR duplicates.

The output file containing the recognised modifications, was extracted and analysed using Variant Effect Predictor that determines the effect of the variants (SNPs, insertions, deletions, CNVs) on genes, transcripts, and protein sequence, as well as regulatory regions. Detected variants were categorised according to the mutation type and frequency of occurrence.

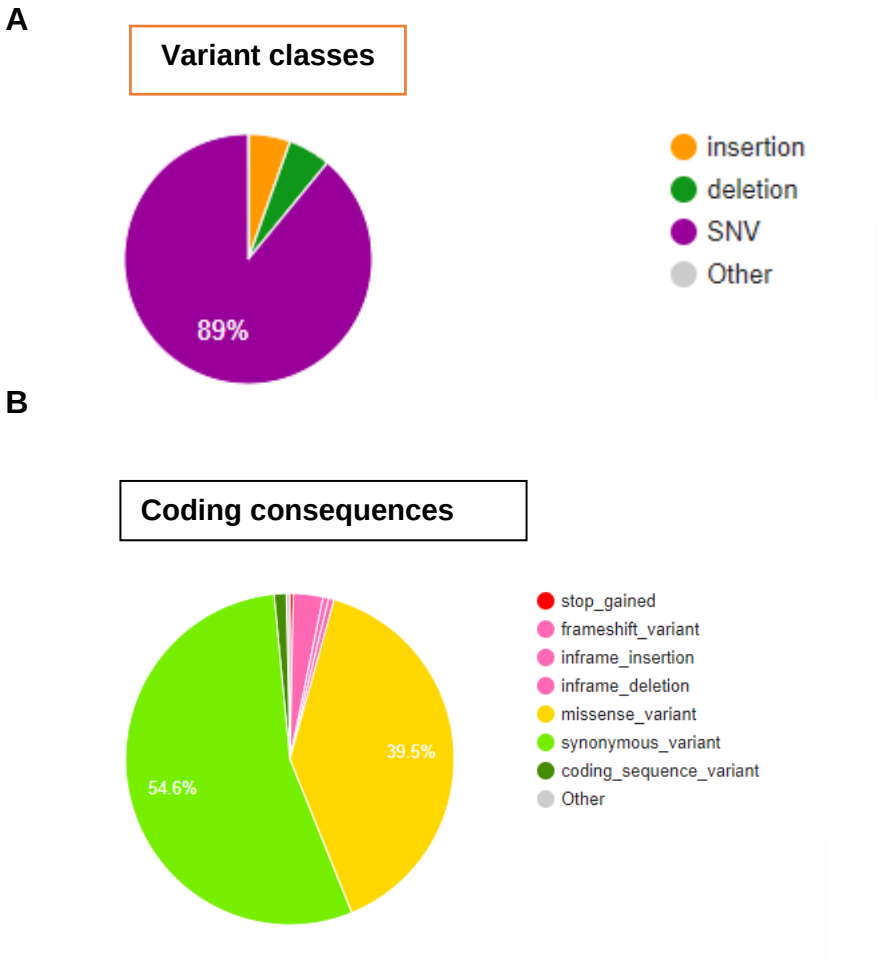


Figure 4.5 Annotation of identified RNA variants in A375 p53 wild-type cell line by Variant Effect Predictor software. A) Changes detected on transcriptomic level were classified into four categories: insertion, deletion, single nucleotide variant (SNV) and other. B) Pie chart distribution of specific mutation types found in coding regions

Of the filtered RNA variants detected by CLC Bio in p53 null cell line a total of 100,896 expressed non-synonymous variants were identified on by CLC Bio, which accounted for approximately 12% of high frequency variants (frequency > 40 reads). This number of non-synonymous mutations is within the expected range of a tumour like melanoma which has a relatively high number of single nucleotide variants (SNPs). VEP analysis revealed that almost 53% of variants found in coding regions

were synonymous, thus not causing changes on the protein level. However, missense variants accounted for over 40%. Less frequent mutations in coding regions included frameshift variants (3.4%), stop gained (0.3%), in frame insertion (0.6%) or in frame deletion (0.6%) (Figure 4.4). These annotations show that detected variants show a lot of potential for production of mutated peptides.

In isogenic p53 wild type cell line a total of 101847 non-synonymous RNA variants were identified, out of which 12.4% were of high frequency. As predicted, the highest number of variants classified as SNV accounting for 89%. Distribution of variants in the coding region showed that there are slightly more synonymous variants than in the p53 null model (54.6%). Missense variants ranked second with frequency of 39.5%. Less frequent mutations in coding regions were frameshift variants (2.9%), stop gained (0.3%), in frame insertion (0.5%) or in frame deletion (0.5%) (Figure 4.5). Variant annotation allows for classification of changes occurred either on DNA or RNA level, which later can serve as an overview of the mutational load and estimation of potential neoantigen sources.

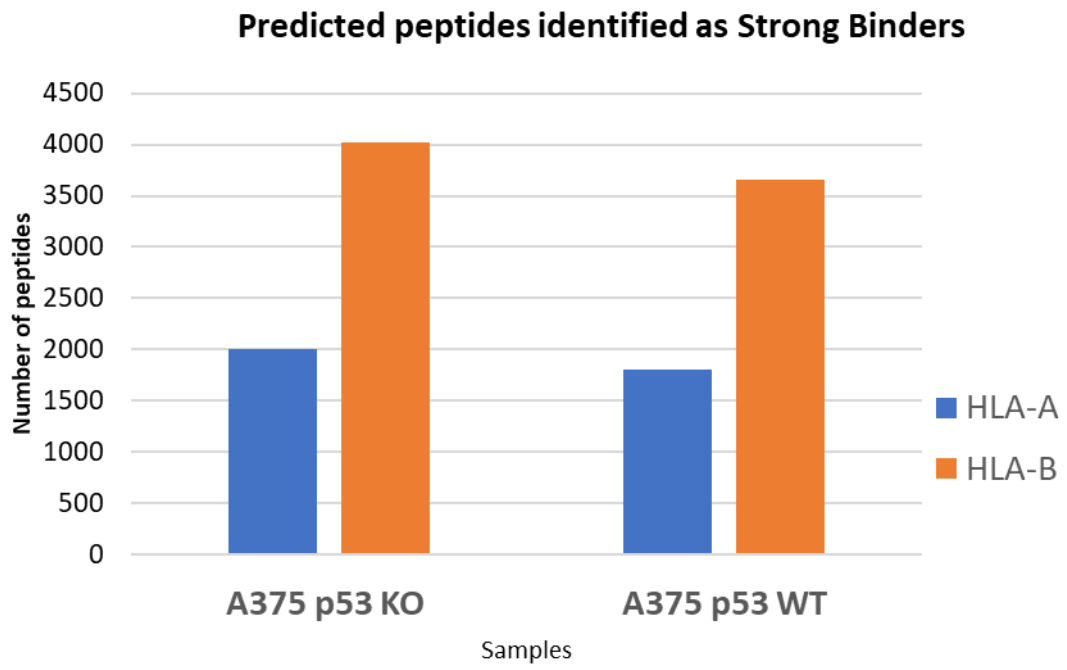


Figure 4.6 Number of peptides identified by NetMHCPan software. Identified nonsynonymous mutations were in-silico translated and cut into peptides to mimic the proteasomal cleavage. Strong binding 9-mer peptides were analysed by affinity to two allotypes HLA-A0301 and HLA-B0702 expressed by A375 cell line. Strong binders were at %rank>0.5 of set of random natural peptides.

Sequences carrying identified mutations underwent *in-silico* translation to peptides. Afterwards, predicted peptide sequences were cut in different combinations to mimic proteasome processing during antigen presentation. Nine-mer sequences were analysed by NetMHCPan software predicting MHC binding affinity. For HLA-A0301 number of peptides classified as Strong Binders (%rank>0.5) oscillated around 2000 peptides, whereas for HLA-B0702 ranked around 4000. This gives a promising result, that confirmed by in vitro studies, can be used as a platform for inducing a production of immunogenic peptides in tumours with low mutational load (Figure 4.6).

4.3.2 Identification and characterisation of MHC-bound peptides using immunoprecipitation and mild acid elution

Analysis of peptides presented as antigens on the cell surface is crucial for expanding our knowledge about neoantigen presentation in cancers. A375 p53 wild-type and null cell lines were subjected to mild acid elution and immunoprecipitation with PAN-MHC antibodies followed by mass spectrometry detection. The purpose of this section was to (1) identify mutated proteins and potential neoantigenic properties that could be used as a vaccine target, (2) finding out the main source of presented peptides, and (3) determining how a regulatory gene deletion (p53) alters the im-munopeptidome.

For the assessment using immunoprecipitation (IP) methodology, I cultured and collected $1x10^9$ cells in triplicate, that were next lysed and processed by team of our collaborator Tim Fugmann. MS analysis was performed using data independent acquisition (DIA) mass spectrometry and a total of 10,129 peptides from 4532 pro-teins were identified from 44,980 spectral counts. Peptides presented by MHC I most commonly rank between 8 and 12 amino acids length, which corresponded with the previous findings (Cole et al., 2007). In both studied samples identified peptides with the highest frequency were 9-mers, followed by 10-mers, 8-mers and 11-mers (Fig-ure 4.7).

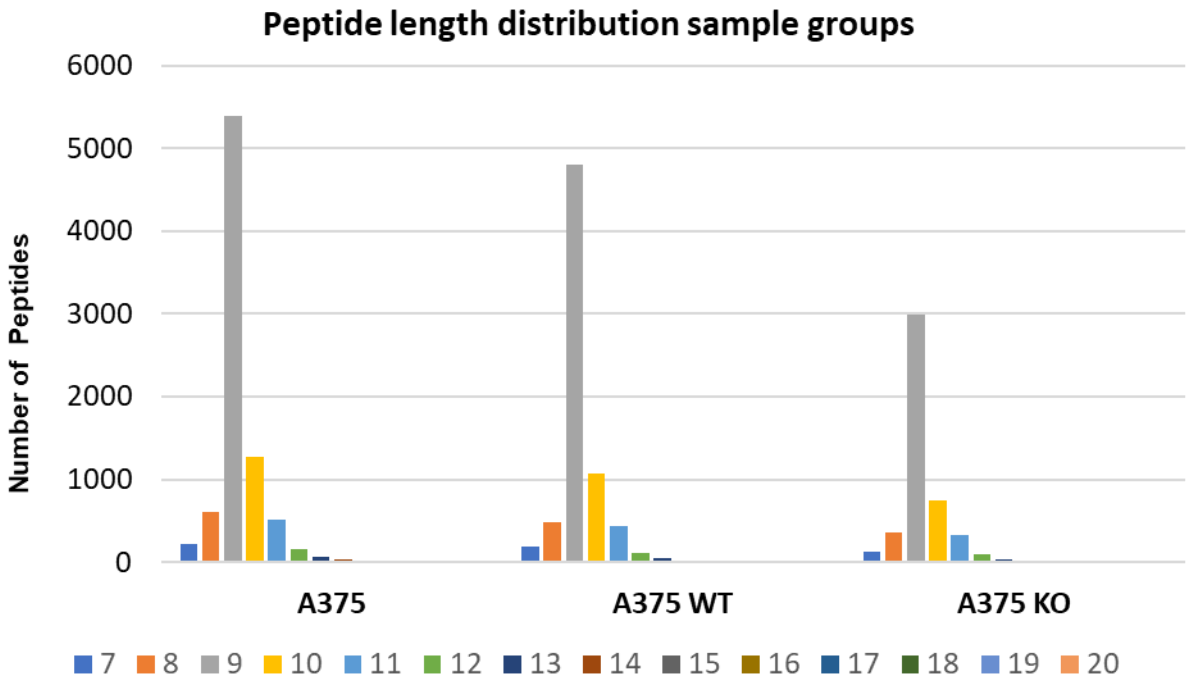


Figure 4.7 Peptide length distribution identified in A375 p53 wild-type and null cells using immunoprecipitation. HLA complexes were isolated from A375 cell lines by pan-HLA immunoprecipitation (IP) followed by extensive wash to remove the unbound mixture, and then acidic elution to dissociate the binding antigen peptides from HLA molecules. The purified peptides were subjected to analysis followed by database/library search of the MS raw files and identified peptides were separated by their length and compared between A375 p53 wild-type and knockout samples.

In parallel, the same batch of A35 p53 wild-type and knockout was grown until each of three replicates reached 10^7 cells and processed by MAE. Samples were subjected to mass spectrometry analysis and total of 2,815 peptides from 1,338 proteins were identified from 27,910 spectral counts. Peptide abundance was calculated, compared between A375 p53 null cell line and p53 wild-type and displayed on the Volcano plot (Figure 4.8).

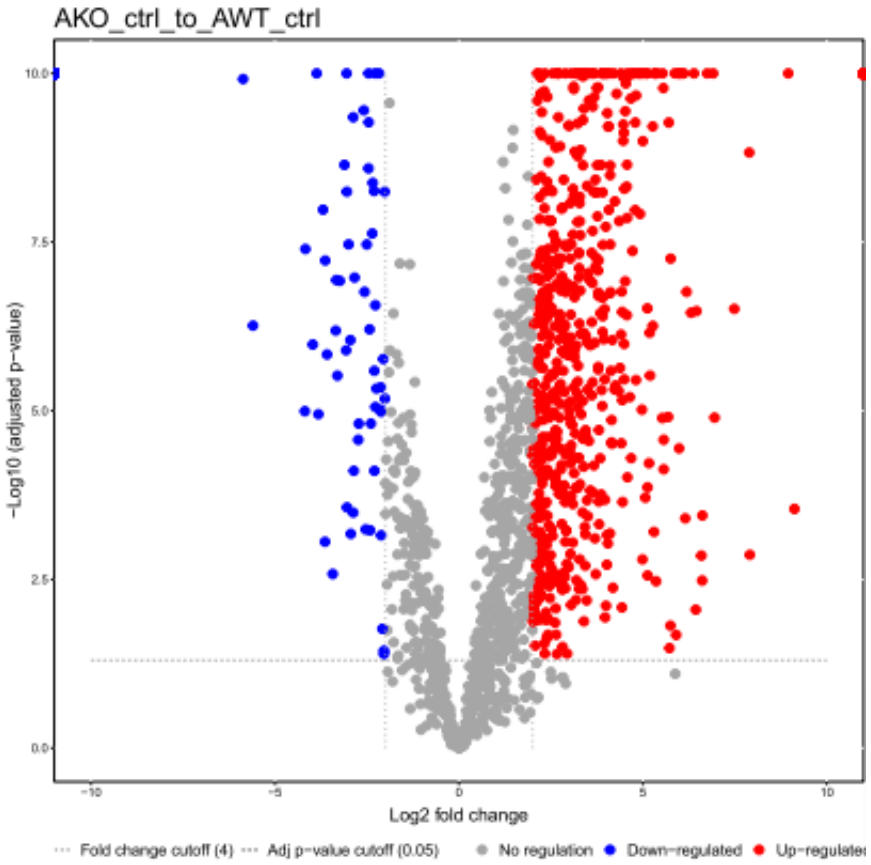


Figure 4.8 Volcano plot comparing expression of peptides detected through MAE in A375 p53 null (AKO_ctrl) to p53 wild type (AWT_ctrl). The volcano plot was constructed using fold-change (LogFC) values and adjusted P-values. The vertical and horizontal grey lines correspond to 4.0-LogFC up and down and the horizontal grey line represents an adjusted P-value of 0.05. In A375 p53 knockout sample there were more upregulated peptides (red) when compared to the p53 wild-type samples than downregulated (blue).

Clearly the advantage of the IP method increases in the number of peptides detected due to elevated sample starting material. However, the MAE method allows more manipulation of cell culture conditions as well as 100-fold smaller cell numbers. In addition, below we examine whether the MAE method indeed captures a subset of peptides present in the IP method and therefore whether we can routinely use the MAE method in future as a more flexible approach to study dynamics on smaller cell numbers. We also interpret the MAE derived peptides which are not part of the IP dataset to determine their characteristics and whether they might comprise MHC Class I peptides or they might just represent secreted peptides.

The peptide sequence overlap between the IP method and MAE method was compared using a web application BioVenn software, where 1095 peptides were identified through both methods (Figure 4.10); this suggests that almost half of the MAE derived peptides are likely MHC Class I binders as they overlap with the IP dataset. Firstly, we evaluated peptide length as a function of the IP method and the MAE method (Figure 4.9, blue). From the 8-mers to the 11-mers the IP method detects elevated peptides relative to the MAE method. However, at peptide lengths greater than 11-mers and up to 20-mers, the MAE method showed enriched peptide recovery (Figure 4.9, orange). This might reflect the weaker affinity of the longer peptides to the MHC Class I peptide binding pocket and the time frames involved in the sample purification. In the case of MAE, the method requires 3 minutes of acid exposure which can recover essentially all peptides, whilst the IP method involves washing steps that likely wash away the larger peptides with weaker affinity. Nevertheless, these possibly weaker binding peptides are interesting to us as this allows us to understand the biology of what peptides can be presented.

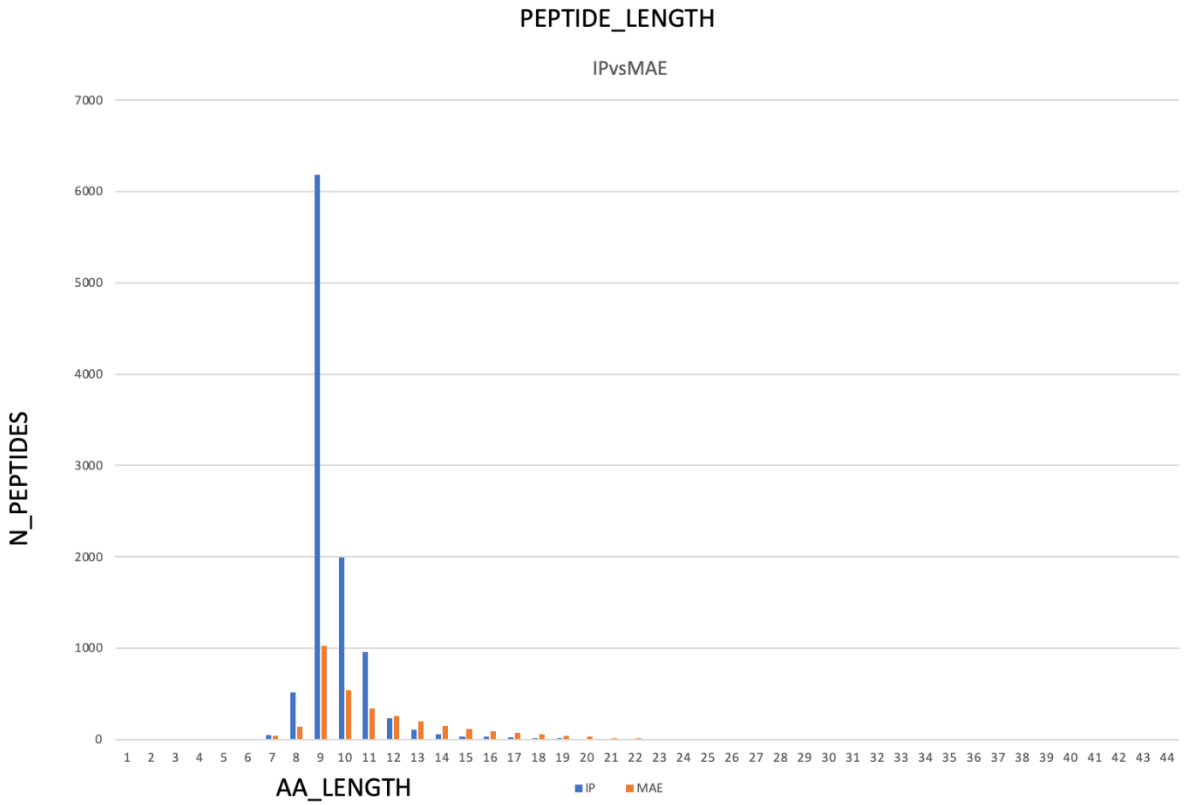


Figure 4.9 Comparison of peptide length distribution between samples analysed through immunoprecipitation vs mild acid elution method (IPvsMAE) Library of peptides identified in both methods was plotted according to their length (AA_LENGTH) on the x-axis and count (N_PEPTIDES) on y-axis. In blue, peptides detected using immunoprecipitation (IP) method and in orange, peptides identified using mild-acid elution method (MAE). As shown above, higher number of longer peptides was detected using MAE, whereas more peptides overall were identified using IP method.

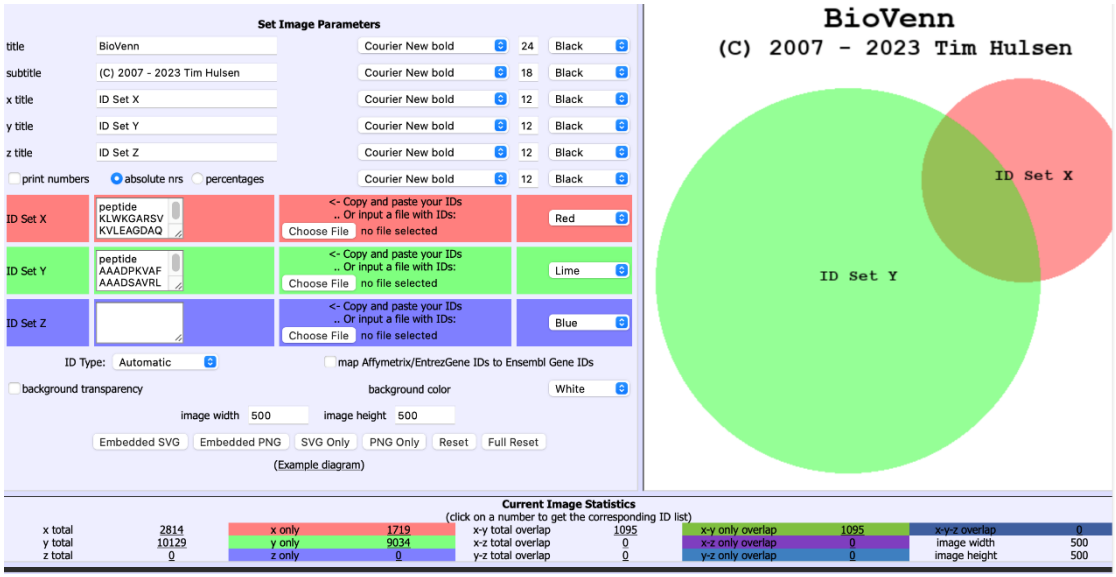


Figure 4.10 Quantitative comparison of peptides identified by IP and MAE method using Biovenn software. List of peptides identified using two methods was uploaded onto Biovenn software to assess similarites and differences between detected peptides.

Another interesting feature was that a subset of peptides was enriched at the C-terminus; 122 immunopeptides were mapped to the termination codon at the C-terminus of the proteins using IP datasets and using MAE, 316 peptides were mapped to termination codons. It is interesting that there was more termination codon dependent immunopeptides using the MAE dataset. In addition to this, the stop codon immunopeptides from the MAE had enrichment of acidic residues E and D at the cleavage site (Figure 4.11 A), compared to the IP method. This was specific to stop codons because when examining all other immunopeptides their results were more normalized (Figure 4.11 B). Such data would be important to predict vaccine neoantigen targets, as one could argue that if we chose a c-terminal stop codon

immunopeptide as a target, if the cleavage amino acid was for example a poorly cleaved amino acid such as W, C, or H, then it would be a less likely neoantigen.

This focus of immunopeptides at the stop codons also pinpoints potential novel class of neopeptides that could be targeted and exploited using readthrough drugs that would create novel peptide vaccines that target the readthrough event. Thus, future improvements on vaccine neoantigen predictors would be to develop rules on when the stop codon immunopeptide space can be exploited; I gave here just one example where we can focus on high and low frequency amino acid abundance at the cleavage site at the C-terminal stop codon.

Presented data demonstrate several features in the pipeline to consider in developing future neoantigen vaccine targets; (i) informatic pipelines that can be used to identify mutant RNAs, how a specific gene deletion (in this case p53) can impact on the frequency of specific peptides; how the method of elution can bias, and how the cleavage site can begin to be incorporated into predictions. Improvement on such methods should also allow the exploitation of these immunopeptide targets for identifying genes that regulate immunopeptide processing.

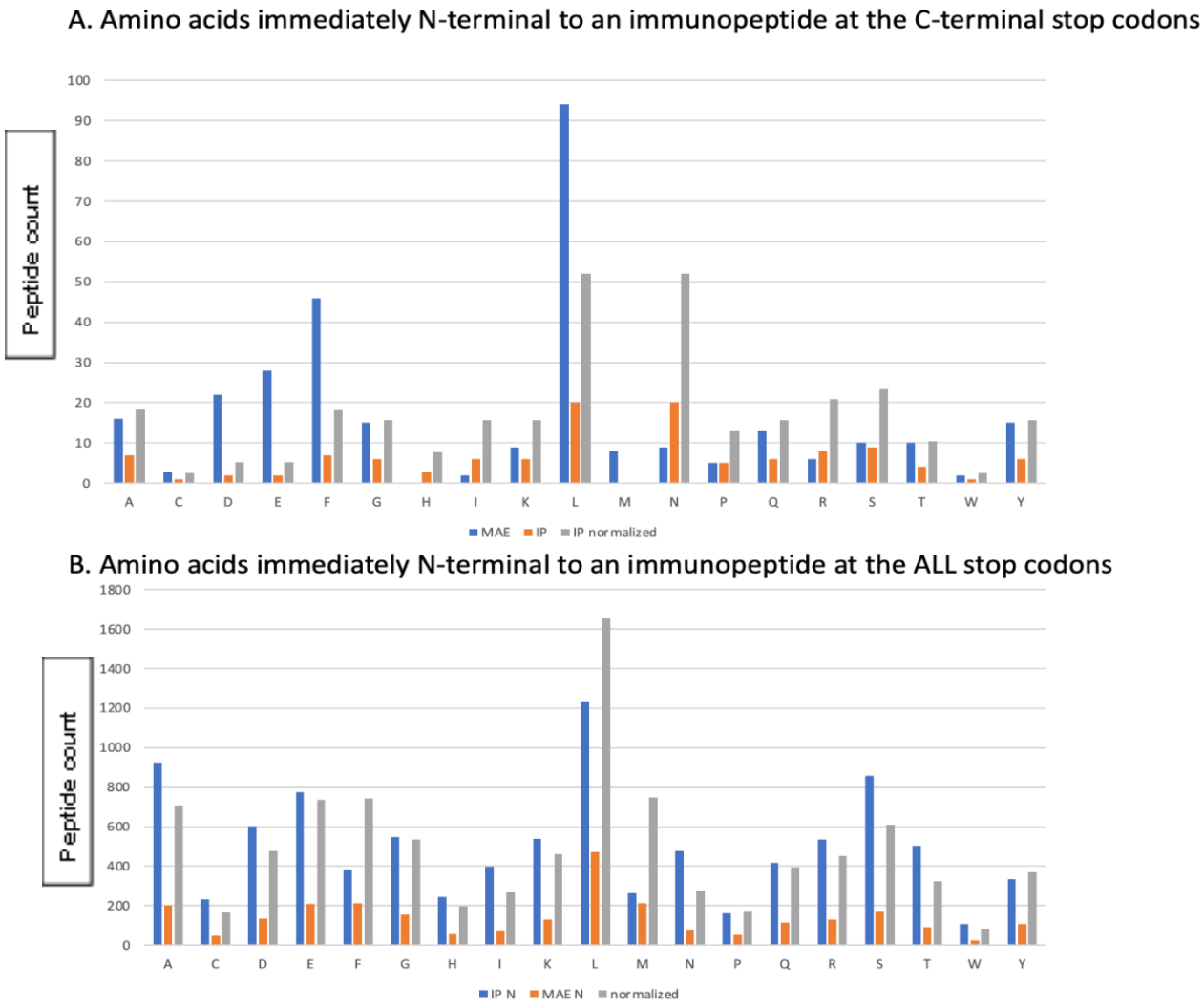


Figure 4.11 Frequency of amino acid N-terminal to cleavage site at the STOP codon containing immunopeptides. A) A subset of immunopeptides were mapped to the termination codon at the C-terminus of the proteins using IP (orange) and using MAE (blue) B) Immunopeptides from all stop co-

4.4 Discussion

Detection of the immunopeptidome may support the inclusion of tumour-associated neoantigens for the vaccine preparation, as well as expand the knowledge about the pathways leading to the peptide presentation on the cell surface. This chapter developed a pipeline that defined the number of

immunopeptides one can detect from a fixed cell number using different methods and some novel features that could be considered including protease processing motifs out with the immunopeptide itself. In addition, we examined how a major tumour suppressor can impact on the immunopeptidome.

Prior studies have shown how p53 itself can have knock on effects on cancer regulatory gene expression. Kichina et al., performing a study on a panel of seven human melanoma cell lines found, that cells carrying wild-type p53 were also characterized by depletion or loss of p14/ARF, a small MDM2-binding protein that controls the activity of MDM2 preventing p53 degradation (Kichina et al., 2003). Scientists also found correlation between the presence of wild-type p53 and complete loss or significant downregulation of another tumour suppressor gene p16 (Kichina et al., 2003). Which indicates that p53 status has a strong influence on tumour fate. Moreover, it has been demonstrated that loss or mutation of p53 in cancers can affect the recruitment and activity of myeloid and T cells, allowing immune evasion and promoting cancer progression (Blagih et al., 2020). It supports the use of our p53 wild-type and knockdown model to indicate changes in immunopeptidomic landscape of melanoma cell line.

Computational methodologies as applied to the cancer research field are emerging as approaches to define the expressed, mutated genome. There are several challenges with optimizing platforms that integrate DNA sequencing, RNA sequencing, and mass spectrometric datasets (H. Zhang et al., 2016). One overall challenge is integrating these molecular data into one pipeline; and depends on the variant detection platform used for DNA sequencing; the algorithms for defining mutations with RNA sequencing datasets in exons, non-coding RNAs, and introns (Laumont et al., 2016). For example, the expression of intron encoded mutant peptides is almost completely unexplored at a systems biology level (Apcher et al., 2015), as are cancer-specific RNA edits and tumour-specific spliced mRNAs that create cancer-associated polypeptide epitopes (M. Zhang, Fritsche, et al., 2018). In addition, different mass spectrometers, sample preparation and pre-fractionation methods, coupled to tumor heterogeneity, result in an incomplete understanding in the source and extent of expressed mutated proteins using cancer-specific DNA and/or RNA sequencing reference databases. However, our study and its follow-up

in the next chapter, showed that the integration of the proteogenomic platform with immunopeptidomic approach allows for detection of potential neoantigenic peptides.

There are several other types of platforms recently developed for integrating DNA, RNA and protein data integration. ‘Proteoformer’ produces a complete protein synthesis database that can be used to identify peptides with mass spectrometry through the use of ribosome profiling (Verbruggen et al., 2019). The limitation of this approach is that living cells are required to isolate bioactive ribosomes and this method might not be conducive to frozen tissues from the clinic. Methods have been established for automating spliced variants in cell systems which is especially powerful in cancer genomes where there might be DNA fusions, cancer-specific splicing, and trans-splicing (Krasnov et al., 2015). Spliced variant detection algorithms are always improving, especially those that capture the pathological, heterogeneous splicing specific to cancer cells (Afsari et al., 2018). Development of novel R-package automating the construction of variation-associated databases from public single nucleotide variant repositories or sample-specific next-generation sequencing (NGS) data iterates innovations in identifying expressed mutated genes (Wen et al., 2014). Translation toolkits have been generated that aim to produce a theoretical total polypeptide space of a genome using RNAseq that captures six-frame genome translation (Zickmann & Renard, 2015). These examples highlight the types of several bespoke algorithms that generate cutting-edge information on the proteogenomic landscape. As the diversity in software and computational tools tend not to be benchmarked against each other, end-user compatibility, especially crossing different disciplines, can be limited or non-accessible.

Here, we analysed transcriptomic data and performed in-silico translation of detected variants to predict potential neoantigenic peptides originating from characterised mutational landscape. However, one drawback of using predictions based only on peptide-MHC binding data (such as Net-MHC) is that they omit the antigen processing and presentation pathway features. Wherever immunopeptidomic experiments are not available, using ligand elution data for training has been proposed. These analysed ligands are naturally shown to be presented on cell surface by MHC molecules, which indicates that they went through the antigen processing and presentation pathway. This type of data contains information on sequence motifs influencing the processing that is not available when solely peptide-

MHC binding is considered. Additionally, the eluted ligands do not have any bias of pre-selection when using prediction methods (Koşaloğlu-Yalçın et al., 2022).

The direct identification of peptides expressed on tumour cell surface would give direct evidence for short-listing neoantigens that could promote T-cell-mediated tumor rejection. Main assays used for identification of MHC-bound peptides are immunoprecipitation and mild acid elution. Both approaches have their advantages, however as demonstrated here, mild acid elution showed higher efficiency of peptide detection with the consideration of the use of smaller sample size than in case of IP. The main downside of utilizing MS detection in clinical setup is limited patient sample amounts, which affects the sensitivity and reproducibility of the method. Based on the cell number used in the MAE method, this opens many more ‘smaller’ resected cancer specimens to analysis using mass spectrometry. On the other hand, where tumour sizes are very small such as melanoma or glioblastoma, then the ability to expand primary cell in culture would increase neoantigen detection using mass spectrometry. This leaves a huge space for advancements in development of novel sample preparation protocols to increase the yield of isolated MHC peptides from various cancer types and reduce the number of preparation steps might minimize the loss of low abundant peptides. These advancements followed by improved instrument sensitivity, which will reduce the tissue quantity requirements, will lead to tremendous advantages in neoantigen characterization by mass spectrometry.

5 The effects of NMD inhibition on the mutant tumour neoantigen pool in human and mouse melanoma model

5.1 Introduction

Next-generation sequencing (NGS) has become an important method to answer various biological research questions at different omics levels like genomics, epigenomics, transcriptomics, metabolomics, etc. In cancer studies, such in depth characterisation enhanced our ability to search for new immune targets such as mutation-derived antigens (neoantigens), which has accelerated the development of novel immune therapies.

Neoantigens are tumour-specific markers that can form when cancer cell-specific mutations encode molecules that differ from ‘self’ (Ott et al., 2017). Neoantigens expressed on the surface of tumour cells can be detected and removed by immune cells, thus neoantigens can themselves function as immune stimulatory vaccines.

Tumour vaccination aims at stimulating a systemic immune response targeted to, mostly weak, antigens expressed in the disseminated tumour lesions. The main challenge in developing effective vaccination protocols is the identification of potent and broadly expressed tumour rejection antigens and effective adjuvants to stimulate a robust and durable immune response. Further, studies have reported that also the likelihood of successful immunotherapeutic targeting of cancer by many different methods was reliant on immune response to neoantigens (Łuksza et al., 2017).

The most prominent source of potential neoantigenic peptides has shown to be missense mutations in exonic regions, however there are other mechanisms that exist including ribosomal frameshifting, proteasomal splicing of peptides, intron encoded neopeptides, altered antigen processing and post-translational modifications such as di-methylation of arginine (Apcher et al., 2011; Laumont et al., 2018; Smart et al., 2018).

Nonsense-mediated decay (NMD) is a host RNA control pathway that removes aberrant mRNA containing premature termination codons (PTCs) that arise through mutation or defective splicing (Park et al., 2018). If a PTC is present at least 50 nucleotides (nt) upstream of the last exon junction, an exon junction complex (EJC) will remain bound after the pioneering round of translation has been terminated. The interaction between termination factors and the downstream EJC subsequently triggers NMD (Popp & Maquat, 2018). Role of NMD in cancer development is complex (Popp & Maquat, 2018). In some cases, cancer cells have exploited NMD by selecting for mutations causing destruction of key tumour-suppressor mRNAs. In other cases, tumours adjust NMD activity to adapt to their microenvironment. NMD also serves as a tool for tumours to adapt to the microenvironment (Popp & Maquat, 2018).

Studies have shown that inhibition of NMD also generates new antigenic determinants in tumour cells (Becker et al., 2021). Cancer-targeted NMD inhibition forms the basis of a simple, broadly useful, and clinically feasible approach to enhance the antigenicity of disseminated tumours leading to their immune recognition and rejection (Popp et al., 2018). Aberrant RNA splicing can also expose neoantigens that can lead to generating different mRNAs or mRNAs containing introns. Aberrant splicing events in cancer cells can result in the translation of abnormal mRNAs into proteins or peptides presented by MHC class I molecules to T cells that can be detected by the immune system (R. F. Wang & Wang, 2017). Therefore, inhibition of splicing might increase intron encoded peptides that leads to neoantigenic enrichment and potentially stimulating the antitumorigenic immune response. One of the drugs used for that purpose is Isoginkgetin that inhibits the function of both the major and minor spliceosomes (Tsalikis et al., 2019). To determine whether drugs implicated in neoantigen expression can de-repress potential vaccine targets, we defined the genes which expression changes in response to NMD or splicing inhibitors. These studies serve as proof-of-concept that NMD and splicing inhibitors lead to the increase in neoantigen pool and could be used as part of an immune therapy strategy. The identification of peptides inducing anti-cancer response holds great promise to further improve the efficacy of immunotherapeutic approaches.

5.1.1 Aims

The main aim of this study was to create pipelines to identify mutated immunopeptides in a human melanoma cell line model from matched genomic, transcriptomic, and proteomic data in response to a specific drug lead. The drug leads chosen (splicing and NMD inhibitors) would be expected to produce altered amino acid sequences and potentially novel neoantigens. A proteogenomic pipeline was set up including definition of cancer DNA variants, RNA variants (expressed genes under different conditions), and mass spectrometry to identify mutated protein and MHC class I expression. This included working on the bioinformatics software to analyse next generation DNA and RNA sequencing and creating a mutant protein reference database specific for the cell line. This pipeline would also form the concept that specific drugs that target NMD can be used to stimulate the tumour cell to make its own vaccine, i.e., mutated immunopeptide.

5.2 Results

5.2.1 Transcriptomic analysis of differentially expressed genes after NMD and splicing inhibition

In a large majority of melanomas, the wild-type p53 gene is retained suggesting a functional role of this tumour suppressor gene in regulation of tumour-immune interactions. To determine whether drugs implicated in neoantigen expression can de-repress potential neopeptide vaccine targets, we first defined the genes encoded by mutated genes whose expression changes in response to NMD or splicing inhibitors. The cell model used was the melanoma cell line A375 p53 wild-type, and its isogenic p53-null counterpart (A375 p53-/-).

NMD inhibitor and splicing inhibitor titrations were optimized to identify concentration to be used. The aim was to increase the expression of mutated mRNAs and ask to what extent the mutant peptidome is translated from such normally degraded mRNAs. Genetic material isolated from human melanoma cell line A375 treated with splicing or NMD inhibitor was sent for Next-Generation

Sequencing to Otogenetics. Obtained RNAseq files were subjected to computational analysis using Biomedical Genomics Workbench, software created by CLC Bio (Qiagen). The program enables a comprehensive and accurate data analysis for genetic aberrations, therefore identifying mutational landscape.

Transcriptomic analysis revealed RNA quantity changes after stimulation. Differential expression analysis allowed for calculation of variability between different samples. The analysis demonstrated, that after treatment with NMDI, among the top upregulated genes in A375 p53 -/- cell line were genes involved in heat-shock response, such as HSPA1A/B, DNAJB1, HSPA6/8, genes coding for histone proteins, proto-oncogene FOS and Cancer/Testis Antigen Family 12, Member 1b XAGE1B (Figure 5.1*).*

Interestingly, a lot of overexpressed genes have been shown to take part in oncogenesis and influence cancer patients survival, what gives hope for identification of immunogenic peptides specific to cancer cells (Li et al., 2019; Sato & Singer, 2021; Zhao et al., 2020).

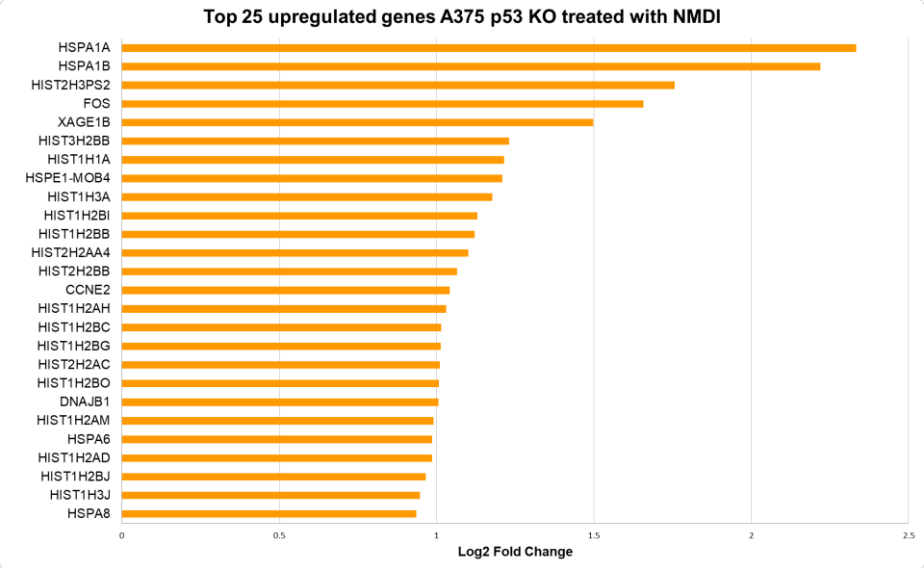


Figure 5.1 Top 25 upregulated genes in A375 p53 knockout (KO) cell line after NMDI14 treatment. Values presented are calculated as log2 fold change of transcript expression of A375 p53 KO cell line treated with NMD inhibitor vs untreated cells.

Treatment of A375 p53 wild-type cell line with the same drug, lead to upregulation of genes engaged in mediation of inflammatory response and regulation

of the immune system, such as IL1A/B, CXCL8, ICAM, IL24, CD68 and control of the NF-kB pathway such as BIRC3, NFKB1A, NFKB2, LURAP1L (Figure 5.2).

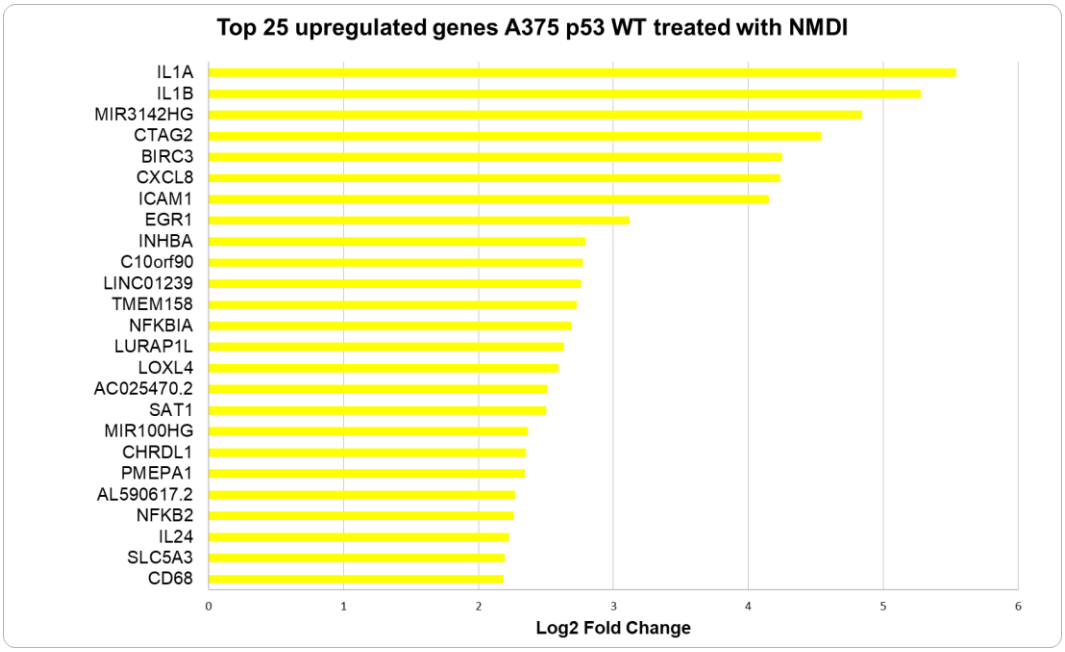


Figure 5.2 Top 25 upregulated genes in A375 p53 wild-type cell line after NMDi14 treatment. Values presented are calculated as log2 fold change of transcript expression of A375 p53 wild-type treated cell line vs wild-type (WT) untreated samples.

Treatment with the splicing inhibitor showed an intriguing pattern. In A375 p53-/- cell line, among top overexpressed genes was BEX2, that was previously found to be a tumour suppressor silenced in human glioma, whereas in breast cancer cells, this gene product modulates apoptosis in response to oestrogen and a chemotherapeutic agent tamoxifen enhancing its anti-proliferative effect (Naderi et al., 2007). Other upregulated gene was found to be SOX11. In breast cancer elevated SOX11 expression is significantly associated with smaller tumour size, earlier tumour grade, and inversely correlated with lymph node metastasis (D.-T. Liu et al., 2014) (Figure 5.3).

On the other hand, treating melanoma cells expressing p53 increased transcripts level of Cancer/Testis Antigens, such as SPANX family or CTAG2, which was shown to be expressed in a wide array of cancers including melanoma, breast cancer, bladder cancer and prostate cancer, and Urothelial Cancer Associated 1 (UCA1) that produces a long non-coding RNA found to be upregulated in bladder

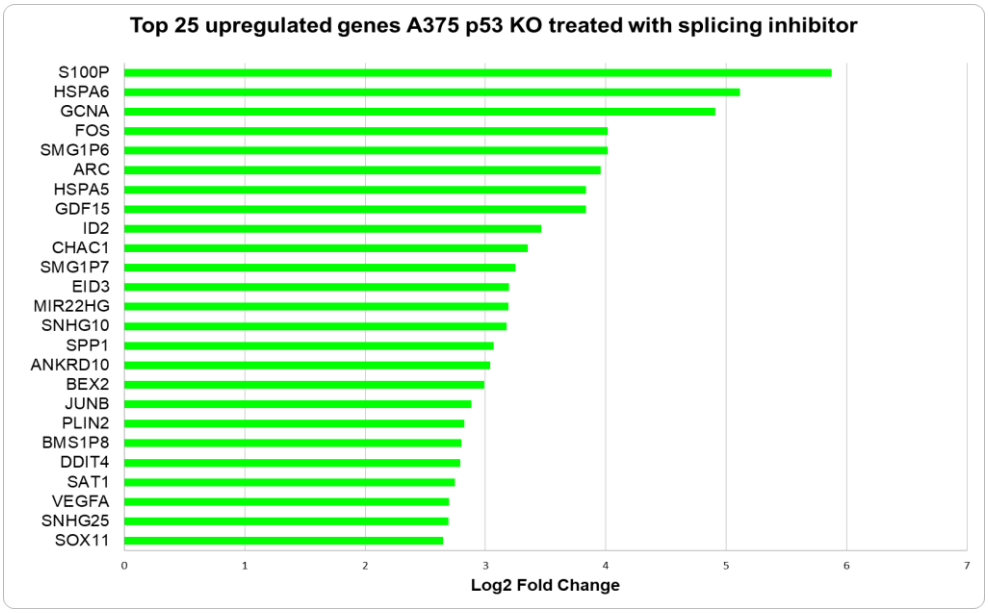


Figure 5.3 Top 25 Upregulated genes in a375 p53 knockout (KO) cell line treated with splicing inhibitor isoginkgetin. Values presented are calculated as log2 fold change of transcript expression of A375 p53 knockout cell line treated with splicing inhibitor vs untreated control.

cancer and plays a regulatory role in cell proliferation (Figure 5.4). In addition, an alternative open reading frame product of CTAG2 was termed CAMEL and classified as a tumour antigen that is recognized by melanoma-specific cytotoxic T-lymphocytes (Aarnoudse et al., 1999). It would be very interesting to explore those genes in more depth as potential therapeutic targets.

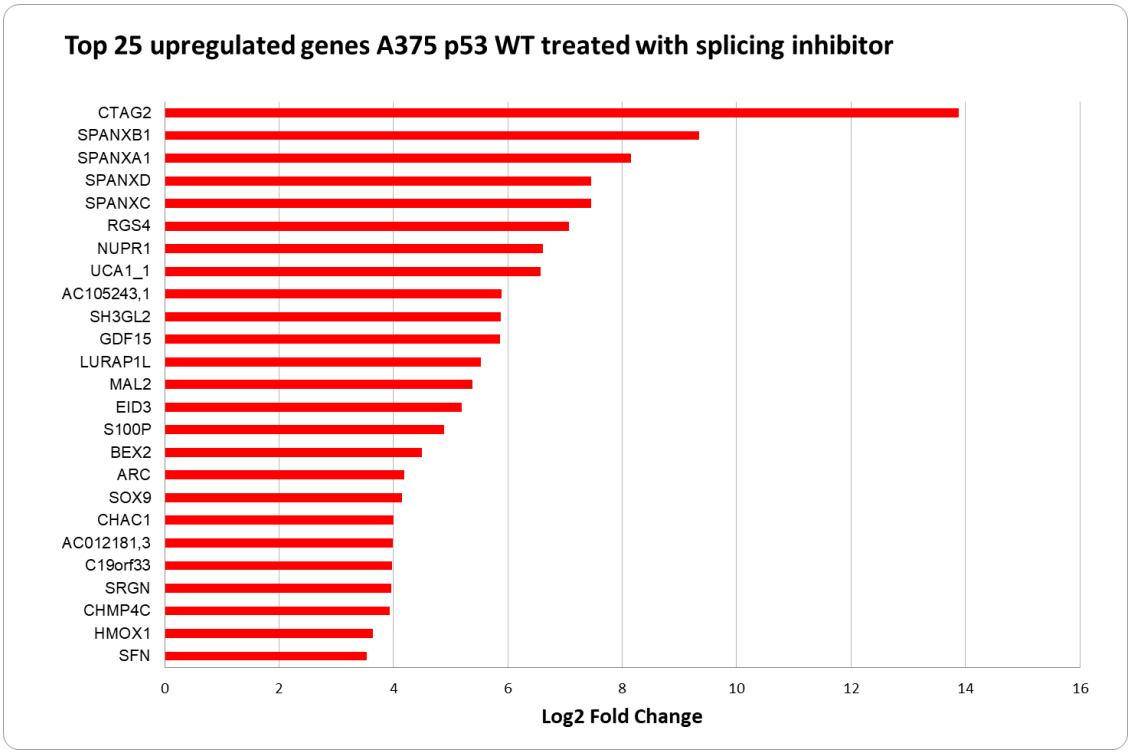


Figure 5.4 Top 25 upregulated genes in A375 p53 wild-type cell line treated with splicing inhibitor isoginkgetin. Values presented are calculated as log2 fold change of transcript expression of A375 p53 wild-type cell line treated with splicing inhibitor vs untreated control.

Closer analysis revealed that as a result of splicing inhibition a group called small nucleolar RNA host genes (SNHG) was highly upregulated in both studied cell lines. The role of genes from this family have not yet been fully understood. However, it was shown that elevated level of SNHG1 in cancers promotes cell proliferation and negatively regulates tumour suppressor genes such as p53 (Zhu et al., 2017). Expression of this transcript may serve as a marker of tumour progression (Zhu et al., 2017). Studies reveal that SNHG15 play an important role in promoting colon cancer and mediating drug resistance, suggesting its potential as prognostic marker and target for RNA-based therapies (Saeinasab et al., 2019)

One of the genes that showed particularly high expression change was SNHG10, reaching 9-fold for A375 p53 -/- cell line and 3-fold for A375 p53 WT in comparison to the control. Recent study has identified SNHG10 expression to be positively correlated with its homolog SCARNA13 expression in 64 hepatocellular carcinoma cases. Additionally, high expression of SNHG10 or SCARNA13 was associated with poor overall survival (Lan et al., 2019).

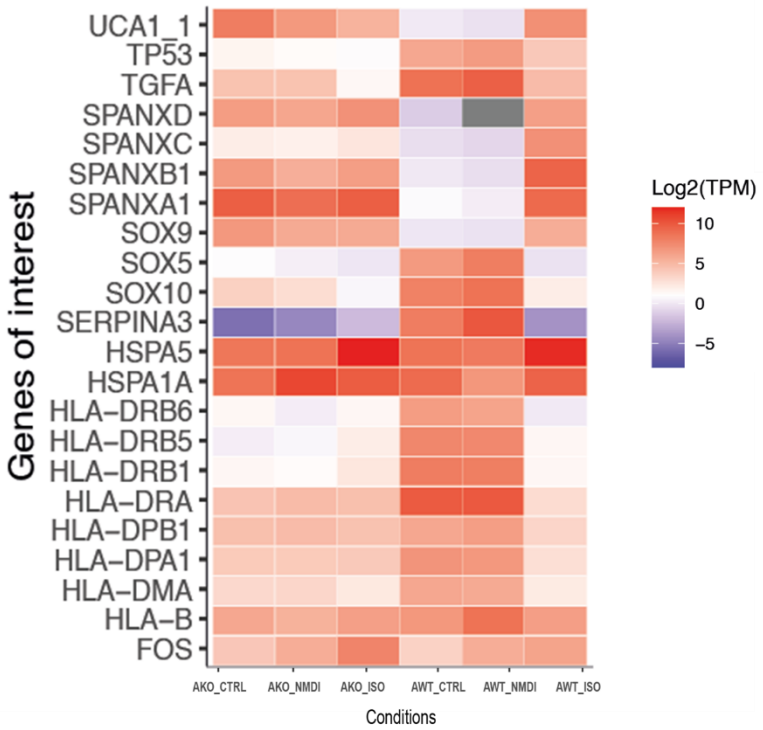


Figure 5.5 Expression analysis of A375 p53 wild-type (AWT) and knockout (AKO) cell lines treated with NMD and splicing inhibitor (ISO). The colours of the heat map represent the log2(TPM) gene expression, becoming red when it is highly expressed and purple to blue for non-expressed and under-expressed values. The values in white represent a low level of expression becoming non-significant. Finally, the values in grey are for the genes which TPM was equal to 0.

Expression of chosen genes was compared between samples and visualised using a heatmap (Figure 5.5). Comparison was based on log2 of Transcript Per Million (TPM) value. Visualizing data using this method allowed for a clear overview of the differences and similarities between cell lines and treatment schemes. Transcriptomic analysis revealed that the expression profile of A375 p53WT after splicing inhibition resembles a control sample of A375 p53 -/-. Interestingly in p53-null cells, expression of genes associated with MHC class II cell surface receptors is downregulated as well as SOX5 and SOX10. Whereas, transcript level of Cancer/Testis Antigens family SPANX genes, UCA1 and SOX9 is elevated in the knockout. After isoginkgetin treatment the HLA-DR expression goes down in A375 p53 WT cell line and HLA-B transcript levels show an increase. Additionally, analysis of pathways associated with expression changes was performed and displayed in Supplementary section (Figure 7.3).

Transcriptomic analysis carried for the number of investigated samples is a costly procedure to be ran in technical triplicates. Therefore, for sequencing, three independent biological samples from each treatment were combined and sent for sequencing, thus not allowing to assign statistical measures.

Western blot analysis was performed to assess whether transcript measurements correlate to protein production of the chosen genes. Immunoblotting showed that level of HLA-B protein is lower in the p53-null cell line (Figure 5.6). However, an increase in this protein production was demonstrated after NMDI14 treatment. HLA-B is a crucial component of HLA I presentation pathway and elevation has been associated with more favourable response to treatment with as

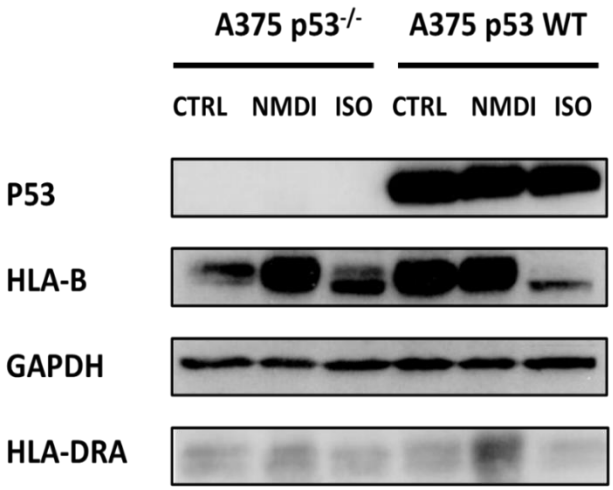


Figure 5.6 Western blot analysis of A375 p53 -/- and A375 p53 WT cell line treated with NMD inhibitor and splicing inhibitor. Analysis of protein expression in treated cell line was performed by western blot. Measurements included proteins involved in antigen presentation pathway HLA-B and HLA-DR, as well as p53 and a product of a housekeeping gene GAPDH.

well as HLA-DRA. Whereas, in cells treated with a splicing inhibitor, HLA-B protein expression goes down in both cell lines. In the near future, further analysis will be conducted. These results however imply that the transcript level does not always translates to the protein level. The next ongoing step is integrating mass

spectrometry data into this equation, giving a broader overview of the proteomic landscape.

These findings suggest that NMD and splicing inhibitor treatment can potentially be used as a tool to increase the neoantigen pool in p53-dependent manner, therefore increasing the number of potential cancer vaccine candidates.

5.2.2 Variant detection and characterization

Mutations in RNA of A375 cell lines were identified using CLC Bio and GATK pipeline. Two approaches were tested to compare sensitivity and accuracy of variant calling. Only one platform is more likely to omit some variants, or on the contrary annotate false-positive mutants, due to the low filtering rate (Figure 2.1).

In the CLC Bio approach, RNA sequence is firstly trimmed from the adapters. Bases with low quality are removed to assure no room for potential error arising from low quality reads. Next, prepared sequenced were mapped to human reference hg38. Software gets rid of duplicate mapped sequences to remove paired reads that have the same start and end coordinates and are, thus, probably PCR duplicates. dbSNP databases were used to filter out known germline variants.

In A375 p53 knockout samples where NMD pathway was inhibited by NMDI14, CLC Bio identified total of 790,897 RNA variants out of which 12.59% were nonsynonymous, thus able to produce novel peptides. On the other hand, RNA of NMDi-treated p53 wild type cell line displayed 1,049,790 mutation and 10.41% nonsynonymous. Samples treated with proteasomal splicing inhibitor displayed 89425 for p53 knockout cell line and for wild-type 58,971 nonsynonymous mutations.

The use of GATK pipeline led to identification of 298,684 nonsynonymous variants for p53 knockout and 50,1274 for wild-type samples treated with NMD inhibitor. Cells undergoing splicing inhibition showed 516,736 and 497,807 nonsynonymous mutations for, accordingly A375 wild-type and knockout. The

analysis showed that CLC-Bio called significantly more variants than GATK pipeline (Figure 5.7), which indicates that either GATK filters out the low frequency mutations or CLC Bio detects some false positive events.

It is important now to compare the coverage of those platforms, discover the high confidence variants identified by both and asses the unique variants called

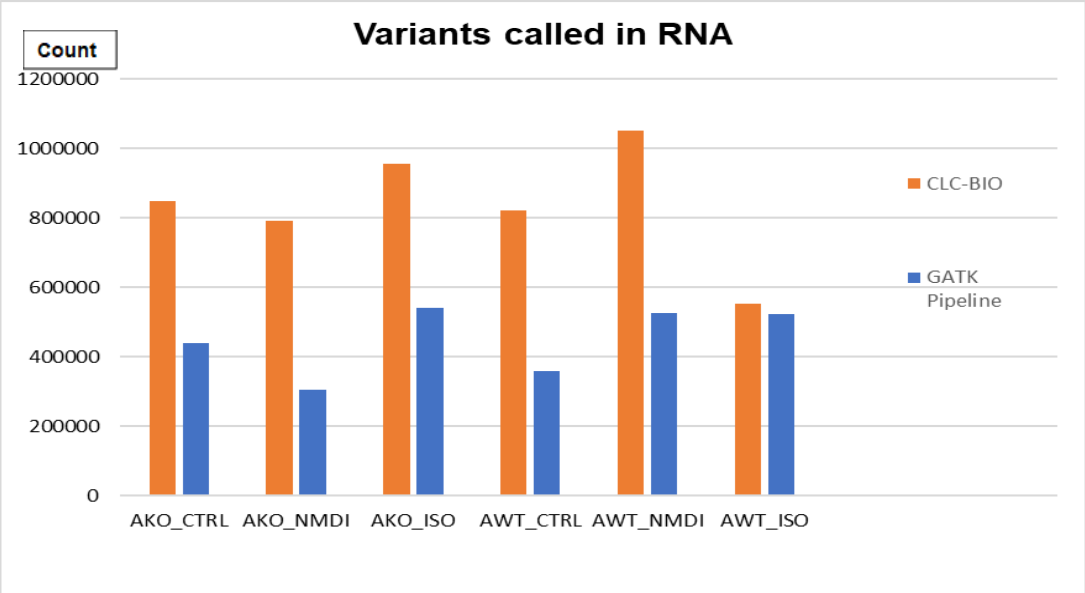


Figure 5.7 Identification of mutational changes in RNA using CLC-Bio and GATK pipeline. Number of variants identified using CLC-BIO (orange) and GATK Pipeline (Blue) in A375 p53 KO (AKO) and A375 p53 wild-type (AWT) samples treated with NMDI or splicing inhibitor (ISO).

through only one approach.

Mutations were then annotated by Variant Effect Predictor. Prediction of the effect of the identified changes in the sequences showed that most variants (around 60% in all samples) came from intronic regions. In coding regions, majority of identified variants came from the Single Nucleotide Variations (SNVs) that occur when a single nucleotide (e.g., A, T, C, or G) is altered in the DNA or RNA sequence (Figure 5.8). The biological impact of SNVs in coding regions depends on their type (synonymous versus missense), and in noncoding regions depends on their impact on RNA processing or gene regulation. Synonymous SNVs do not affect the protein sequence, while nonsynonymous SNVs change the amino acid sequence of protein. SNVs that are not in protein-coding regions may still affect gene splicing, transcription factor binding, messenger RNA degradation, or the sequence of noncoding RNA. On the other hand, missense variants are point mutations in a

sequence of DNA in which change in a single nucleotide, results in a codon that codes for a different amino acid. In these cases, more than 50% of SNV mutations in coding regions were synonymous, whereas missense mutations accounted for around 40% of the identified variants.

Detected variants were categorised according to the mutation type and frequency of occurrence. With these parameters defined, an RNAseq database was produced that integrates into the previously analysed DNAseq variant database. Variants found in DNA were compared with those identified on RNA level to point at the source of the potential mutated peptides. With such a combined RNA and DNA derived reference library, mutant peptides will be identified using mass spectrometry.

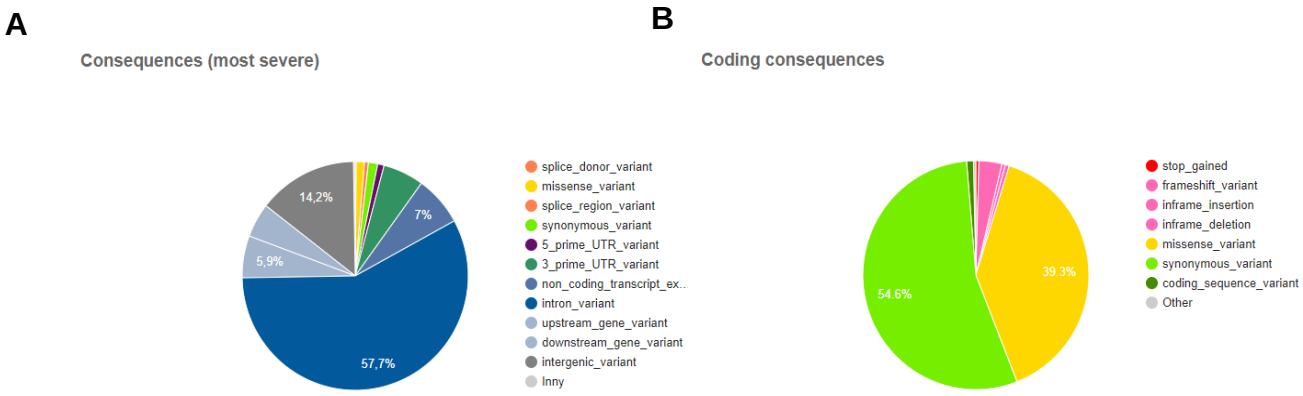


Figure 5,8 Variant effect prediction of A375 p53 WT cell line treated with splicing inhibitor. Here presented exemplary VEP output charts for identified variants in treated cell lines A) Shows most severe consequences of detected mutations B) Annotates variants found in coding regions.

Sequences carrying identified mutations underwent *in-silico* translation to peptides. Afterwards, predicted peptide sequences were cut into 9-mers to mimic proteasome processing during antigen presentation. Nine-mer sequences were analysed by NetMHCPan software predicting MHC binding affinity. For HLA-A0301 number of peptides classified as Strong Binders (%rank>0.5) oscillated around 2000 peptides, whereas for HLA-B0702 ranked around 4000 (Figure 5.9). Identified peptides were assessed and compared between treatment schemes, what showed

that the largest number of peptides with high MHC binding affinity was found to arise after splicing has been inhibited by isoginkgetin. This gives a promising result, that confirmed by in vitro studies, can be used as a platform for inducing a production of immunogenic peptides in tumours with low mutational load.

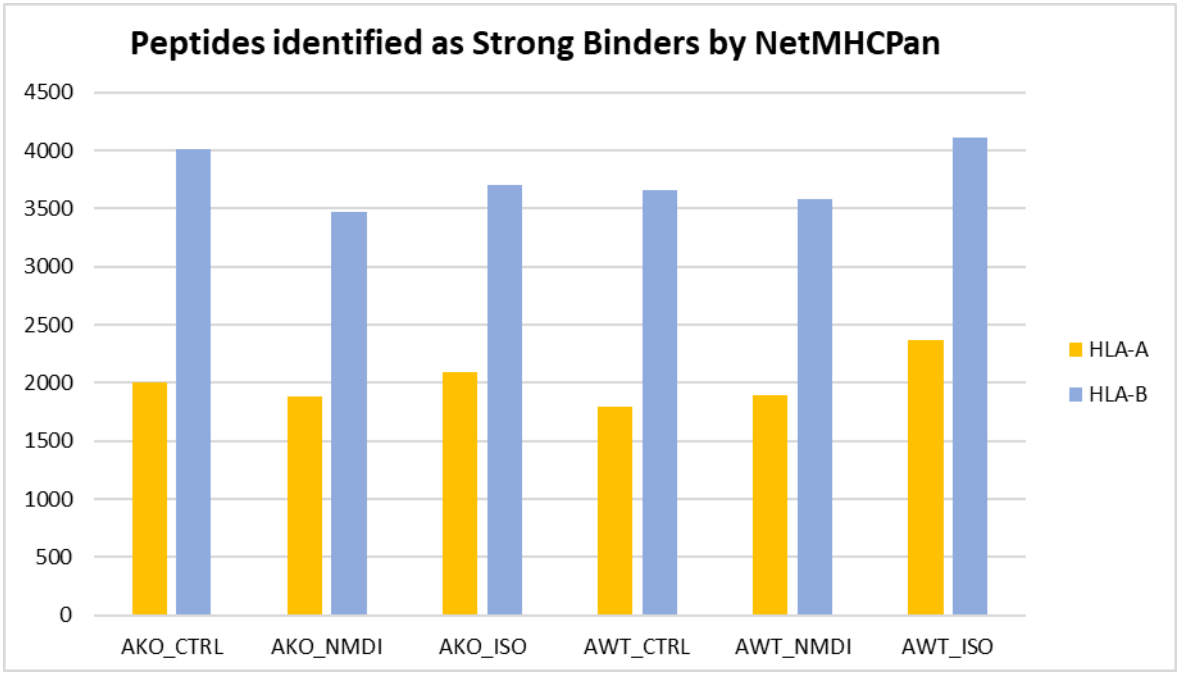


Figure 5.9 Peptides identified as Strong Binders by NetMHCPan. Peptides with the %rank of 0.5 were classified as Strong Binders. Predictions were made for HLA-A0301 and HLA-B0702

5.2.3 The immunopeptidome landscape in cells treated with NMD and splicing inhibitors.

Variants discovered on RNA level, followed by expression analysis allowed us to detect differences in the immunopeptidome in the basal state or in drug treated cells. The RNAseq reference files were derived from shotgun RNAseq using RNA from untreated or treated cells and were processed using the CLCbio Genomics Workbench. In addition, where there were stop codons due to frameshift mutations in

the genome, a second RNA reference sequence was produced that incorporated amino acids that can be detected at a stop codon readthrough in response to NMD inhibitor treatment. These RNA reference sequences were used to search for mutated peptides in the immunopeptidome.

A375 p53 wild-type and knockout cell lines treated with NMD or splicing inhibitors were subjected to mild acid elution and immunoprecipitation followed by mass spectrometry detection. The analysis showed number of peptides eluted from MHC complexes. Comparison of samples with controls allowed for identification of peptides that were up- or downregulated as a result of treatment. Interestingly, NMD inhibition had a significantly stronger effect on elevating immunopeptide levels in both melanoma cell models (Figure 5.10 A, B). We also see slightly more potent effect of NMDI in p53 knockout cells compared to wild-type cells. NMD upregulated immunopeptides more potently almost regardless of status of p53. After 24h with isoginkgetin an upregulation of the immunopeptidome was seen only in knockout cells and had almost no effect on these changes in p53 wild-type model. After 24h with isoginkgetin an upregulation of the immunopeptidome was seen only in knockout cells and had almost no effect on these changes in p53 wild-type model (Figure 5.10 C,D).

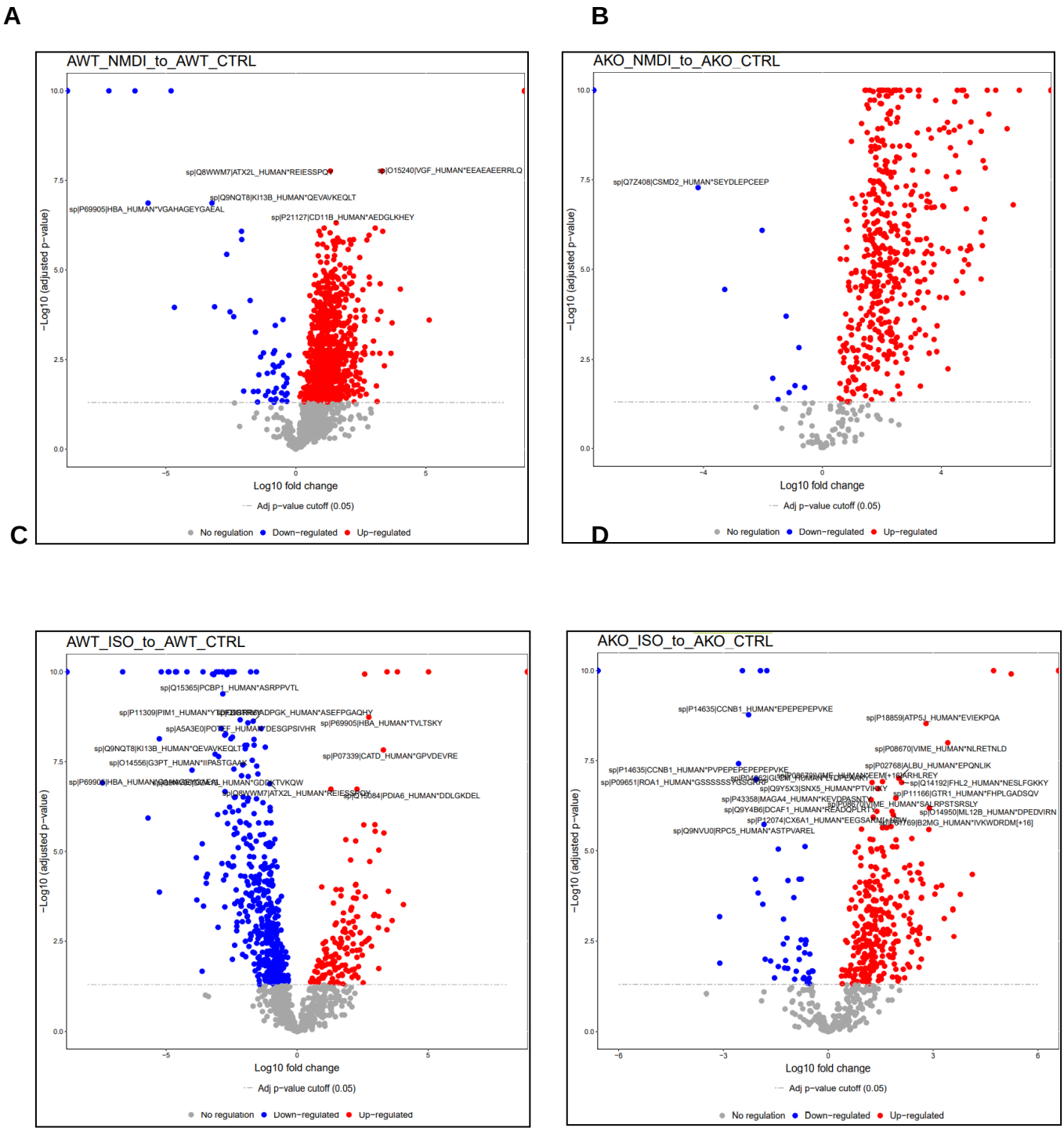


Figure 5.10 Volcano plot of peptides identified after recovery NMDi or splicing inhibitor treatment. Samples were treated with inhibitors for 24h and eluted MHC peptides via MAE. Peptides identified in A375 p53 wild-type (A) and knockout samples (B). Peptides after treatment with splicing inhibitors still showed changes in peptide abundance in p53 wild-type (C) and null (D) cell line.

After acid elution, cells were washed with PBS and placed back in culture for 24 hours to test the recovery of presented peptide landscape. After MAE and mass spectrometry evaluation, it was demonstrated that inhibition of NMD shown the same trend of upregulated peptides as the initial samples. On the other hand, splicing inhibitor treatment demonstrated lower detection rate and high number of downregulated peptides in p53 wild-type cell line. However, we see slightly more potent effect of NMDI in p53 knockout cells at 24h compared to wild-type cells (Figure 5.11). Recovery experiments serves as an indicator of MHC- peptide turnover in growing cells. NMD upregulated immunopeptides more potently compared to isoginkgetin regardless of the count of recovery and almost regardless of status of p53.

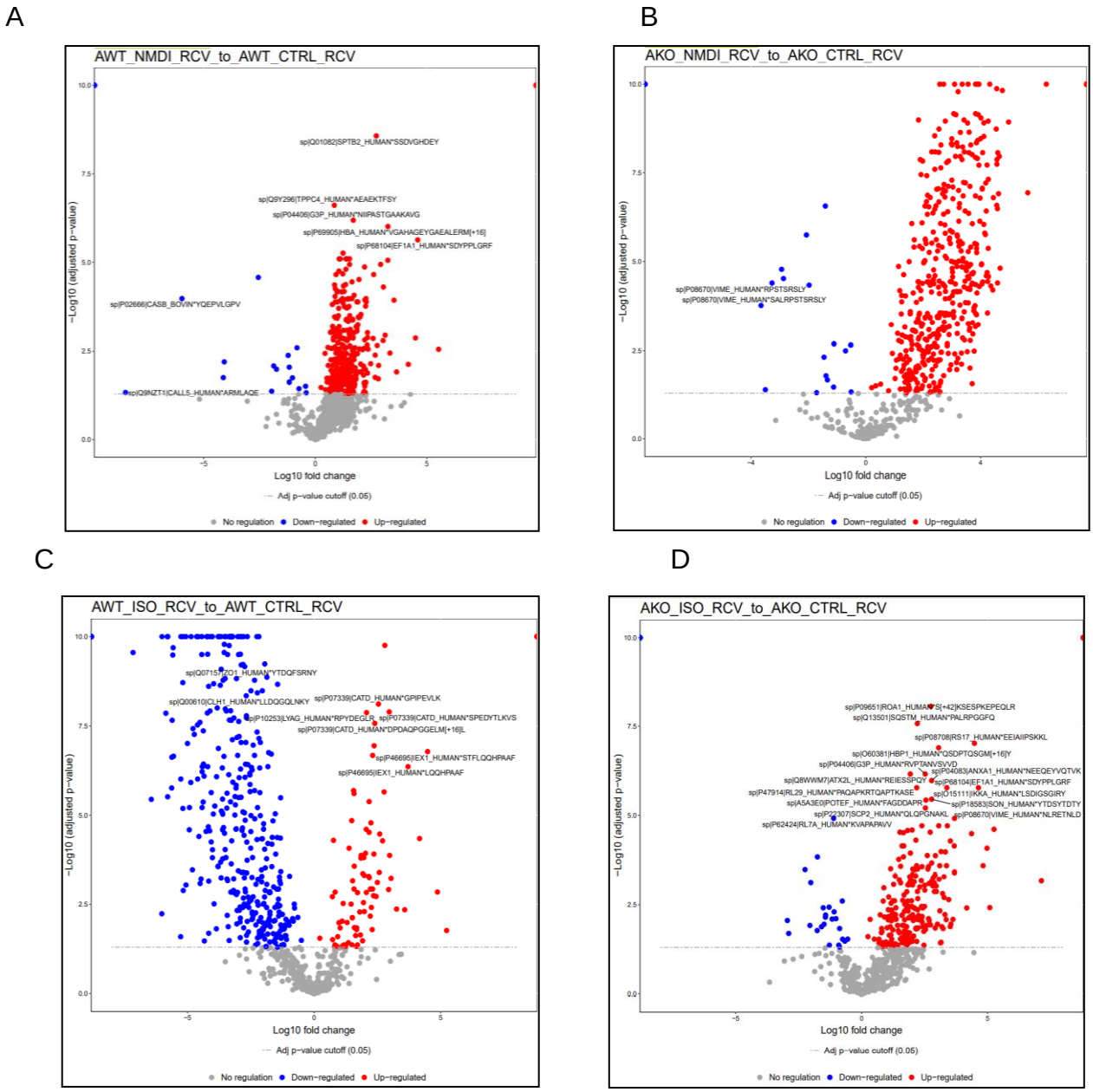


Figure 5.11 Volcano plot of peptides identified after recovery (RCV). Samples were treated with inhibitors for 24h and eluted MHC peptides via MAE. Peptides identified after 24h recovery from NMD inhibition in A375 p53 wild-type (A) and knockout samples (B). Recovery after treatment with splicing inhibitors still showed changes in peptide abundance in p53 wild-type (C) and null (D) cell line.

Additionally, p53 knockout cells recover from the NMDI effect (global upregulation of immunopeptidome effect) more quickly compared to WT, in another word, the NMDI effect is persistence strongly in WT even at 48 h. Therefore, it is the

evidence that NMDI has different action of mechanism in each cell line (WT, KO) (Figure 5.11).

Next, immunopeptides were compared with the corresponding mutated protein sequence from mutated protein database that was produced by the aggregate file

Table 5.1 Unique peptides discovered in immunopeptidomic data.

Peptide sequence	Protein of origin	Mutation
AATPAIRTV	PABPC1	V505I
AILPRALQ	KIF9	G114A
APHGPGSLTTLVPW	ATRAID	D60G
ASELHTSLY	MDN1	H3423Y
EIEILPLPV	OS9	G42E
EPEVQDPP	BOD1L2	G161V
GRADLIRLL	ANKRD27	P761R
HLQELYLA	TIMELESS	R831Q
HQIDLIMQL	CDK10	V205M
HSTPTGTIPPP	MUC6	V2046I
HTAPTMTVTT	MUC6	P2063T
IFQLFLK	MDN1	P5276L
KEKREYL	WDFY4	D2107Y
KEMNNAFSHL	UGT3A2	A321S
KLDAGEQRL	AMZ2	N30D
LDGHRAGSPN	LAMTOR5	S17N
LELMKLK	AK9	V727L
LMISDFR	EYS	S1884R
LPDLLKKVL	ATP11A	V1091L
LTEEVAQEY	TRIM5	R136Q
LTNADETWYL	CD1D	P255T
MVAIRHRC	HTR3D	R6H
NLDKNTMGY	DNAJC11	V243M
NLDKNTMGY	DNAJC11	V243M
NSEEHSARY	PXK	K481R
PHVSEKLIQLFQNK	NELL1	R82Q
QELDGVFQKL	ZWINT	R187G
QIIQGIPQL	PNPT1	Q260P
RLRGHLEAV	TRIM65	V222G
RSVDVTNITF	COPE	T117I
SELVKNFRRES	STON2	W738L
SKGLRG	PARP10	L395R
SLFFRKVAF	MTCH2	P290A
TFSEVIRNK	GRIA4	A854V
TLPHAEIFL	CTH	Q291P
TWLLHIFIPS	PDCD1LG2	F229S
VIFDRFLGV	PHB	R43L
VTEPGTAQY	AKAP13	M452T
VYLDRYR	SLC15A3	G99D
YLEDRPMFL	LOXL2	M570L

that combined my mutant RNArefseq from DMSO treated cells, NMD inhibitor treated cells, isoginkgetin treated cells, all in both wild-typeand p53-knockout states (created by Jakub Faktor). This led to identification of 40 unique peptides (Table 5.1).

Out of which, **ASELHTSLY** peptide originating from mutation in MDN1 displayed high abundance, consistently found on a relatively high level in 2 different set of experiments. MDN1 is a nuclear chaperone required for maturation and nuclear export of pre-60S ribosome subunits. The originating mutation of ASELHTSLY peptide is a point mutation replacing H/histidine at position 3423 by Y /tyrosine. In the analysis, potential of the aforementioned peptide to be a neoantigen has been confirmed by several characteristics. It has been identified (DDA data) in both independent experiments 24h and recovery after 48h, which also points to another neoantigenic feature of having a high binding stability. It has been quantified in (DIA data) in both datasets at least in two conditions, especially after NMDI and splicing inhibition. Both DIA quantitation showed the results that agree in NMDI vs CTRL comparison and there is chance that peptide emerge after NMDI or at least it is upregulated compared to control (Figure 5.12). Additionally, the identified peptide is a ninemer, which is a known sequence length to be preferential for MHC presentation. Next, to fully confirm neoantigen features of detected peptides is to experimentally test its immunogenicity by using methods, such as Enzyme-Linked Immunosorbent Spot assay measuring IFN-γ secretion from T-cells incubated with tested peptide (Chudley et al., 2014). Secretion of IFN-γ from T-cells indicates their activation and therefore would show whether ASELHTSLY induced immune response.

Identification of ASELHTSLY peptide has been a very exciting development of the pipeline. Confirmation of its presence with different methods and in different samples, leads to conclusion that it can be a potential neoantigen that could be used for production of cancer vaccine and in vivo tumour-rejection studies. This development indicates the importance of using proteogenomic pipelines with mass spectrometry-aided immunopeptidomic studies for identification of neoantigens.

In parallel analysis of b16 mouse melanoma cell line was carried out. Using the same pipelines and prediction algorithms to facilitate identified peptides for vaccine formulation and in vivo studies of tumour rejection and immune response. However, results of this analysis are beyond the scope of this PhD as the data are still in preparation.

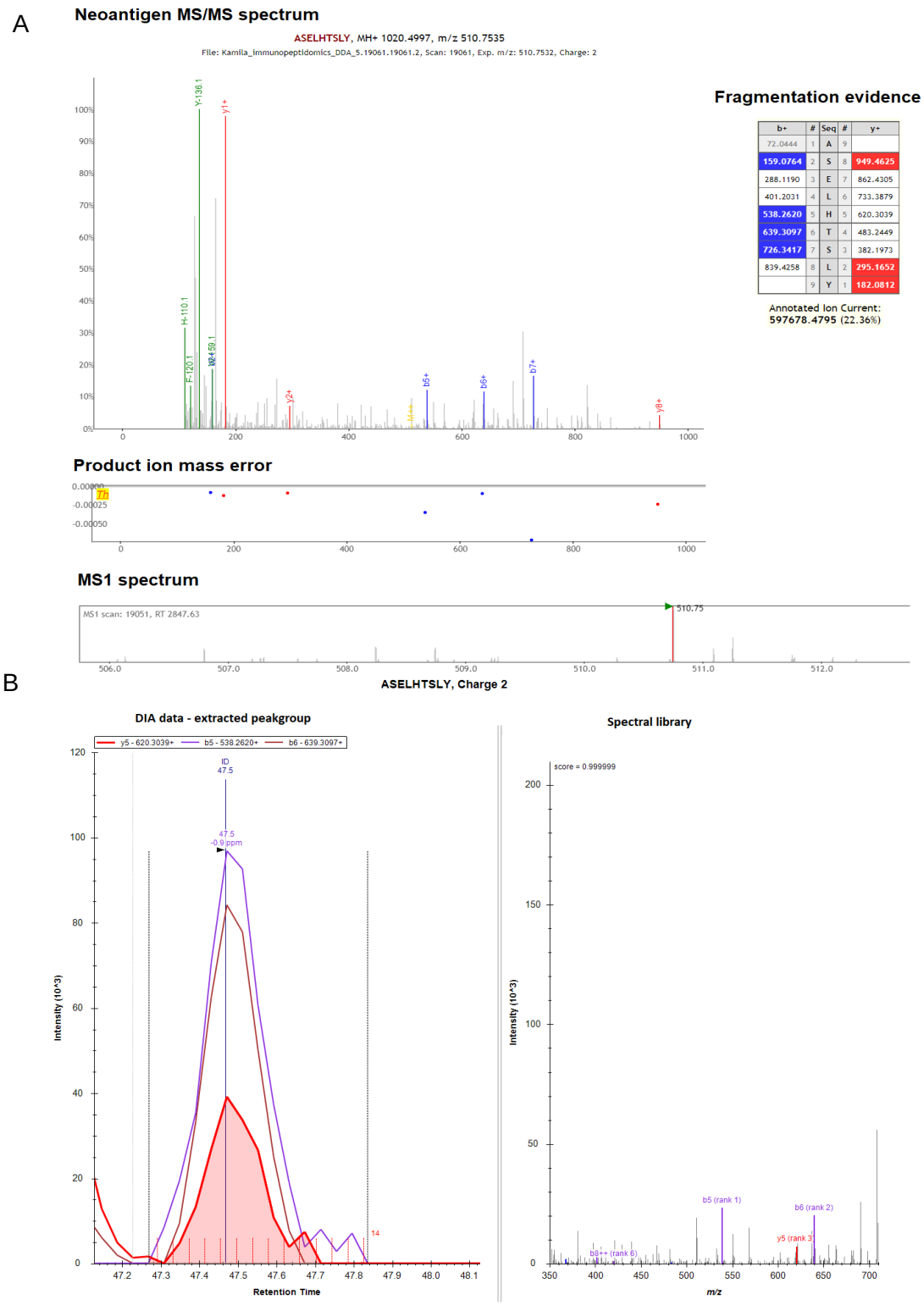


Figure 5.12 DDA and DIA evidence of neoantigen presence. A) Data dependent acquisition mass spectrometry analysis identified **ASELHTSLY** peptide in multiples samples as indicated in MS1 spectrum B) Data independent mass spectrometry acquisition as shown by extracted peak groups and produced spectral library, confirms identification of **ASELHTSLY** in samples.

5.3 Discussion

Personalized immunotherapies became an attractive alternative to the conventional treatment options available to fight cancer. However, prediction of cancer-specific antigens has been a challenging task. As indicated by the presented study,,development of personalized therapeutics should be aided by information recovered from NGS and immunopeptidomic tools. Identified mutated peptides could be exploited as a patient-specific target pathway, such as mutations in EGFR in lung cancers (An et al., 2019). Mutations in the TP53 tumour suppressor gene are very frequent in cancer, and therefore the attempts to restore the functionality of p53 in tumours as a therapeutic strategy have been studied for decades (Ishizaki et al., 2010). Aside of its tumour suppressor function, p53 has been shown to be involved in variety of mechanisms, including immune regulation (Levine, 2020). Therefore, creating a model that shows transcriptomic and immunopeptidomic changes upon manipulations of NMD pathway and splicing inhibition were focus of this study.

Neoantigens able to induce anti-tumour immune response have been vastly explored in recent years. With the identification of these immunogenic peptides always comes a question of their origin. The most prominent source of potential neoantigenic peptides are missense mutations in exonic regions, however there are other mechanisms that give rise to neoantigens including ribosomal frameshifting or proteasomal splicing of peptides (Roudko et al., 2020). Modification of these pathways in theory would be an excellent idea to increase the neoantigenic landscape of cancer cell and therefore leading to its elimination by immune system.

To detect these antigenic peptides and interdisciplinary approach is required. Utilizing information from patient's genomic and transcriptomic data has been widely used for this purpose (Sahin et al., 2017). With the need for identification of mutational load in a large number of samples came the development of novel computational platforms aimed at variant detection, annotation and prediction of potential MHC-binding peptides inducing immune response (Benjamin et al., 2019; Koboldt et al., 2013; Krasnov et al., 2015).

However, the false discovery rates of current bioinformatics methods are poorly understood. Assarsson et al. showed that the proportion of peptides presented among all possible peptides may be in the order of 15% (Assarsson et al., 2007). Because of that, there is a need for choosing the potential ‘candidates’ carefully. That is why, adding another level of the analysis, coming from the mass spectrometry examination of immunopeptidome will facilitate pinpointing to the peptides that more likely induce the antitumorigenic immune response. MHC complexes can be washed off a cell surface and the bound epitopes eluted, both methods MAE and immunoprecipitation have been useful in my PhD studies. However, as indicated in the previous chapters, both approaches have their disadvantages, including a large volume of a sample needed and the sensitivity of a mass spectrometry machine. Nonetheless, computational tools and development of better sample processing protocols could help solve these issues.

Created in this study pipeline resulted in identification of mutational load of A375 p53 wild-type and knockout cell line. Induction of tumour-specific determinants have been explored in several studies (Pastor et al., 2010; Pierson et al., 2019). Here, aberrant splicing and NMD inhibition were used to determine whether we can see an increase the expression of mutated neoantigens. This approach has formed the basis of a simple, broadly useful, and clinically feasible approach to enhance the antigenicity of disseminated tumours leading to their immune recognition and rejection (Popp & Maquat, 2018).

Detection of the immunopeptidome may support the inclusion of tumour-associated neoantigens for the vaccine preparation, as well as expand the knowledge about the pathways leading to the peptide presentation on the cell surface. The presented neoantigen discovery pipeline forms the concept that specific drugs that target NMD can be used to stimulate the tumour cell to make its own vaccine, i.e. mutated immunopeptide.

6 Final remarks

The identification of neoantigens, whatever their source, holds a great promise to further improve the efficacy of immunotherapeutic approaches. In order to achieve that, there is a need to create computational pipelines that would involve analysis of the data from next generation DNA and RNA sequencing and comparison with the proteomic data. Below is the outline of the conclusions of this thesis and weaknesses in a technological pipeline that could be improved in future.

1. Using DNA and RNA variant detection pipelines to define the mutant landscape. This aim was an essential part of the 'personalized' reference database with one advantage being including low DNA variant reads with high RNA variant reads because the depth of RNAseq appears to be greater than the DNAseq depth used based on our current models.

2. Because frameshifts were a key mutation load identified, I set up a project in parallel to the main aim (define immunopeptides) to define 'rules' for what can be incorporated where I found that specific amino acids can be incorporated at a stop codon based on rules but also novel frameshift readthrough can be identified even at the very C-terminus of a protein (Figure 4.11). However, despite the fact that we added these rule-based amino acids to our frameshift RNArefseq database (Chapter 5) we detected NO readthrough immunopeptides (data not shown). Thus, this rules-based readthrough refseq was successful in practise, but a weakness of including this in a future vaccine target discovery pipeline is that detecting these low abundance peptides requires "enrichment' (in this case the pulldown of p53). Therefore, including a stop codon read through database in future will require additional mass spectrometry methodologies to capture lower abundance immunopeptides.

3. Using this optimized pipeline including DNAseq-RNAseq databases, demonstrated utilizing the A375 isogenic panel of cells, that the acidic wash step was as effective as the more 'popular' IP methodology. This

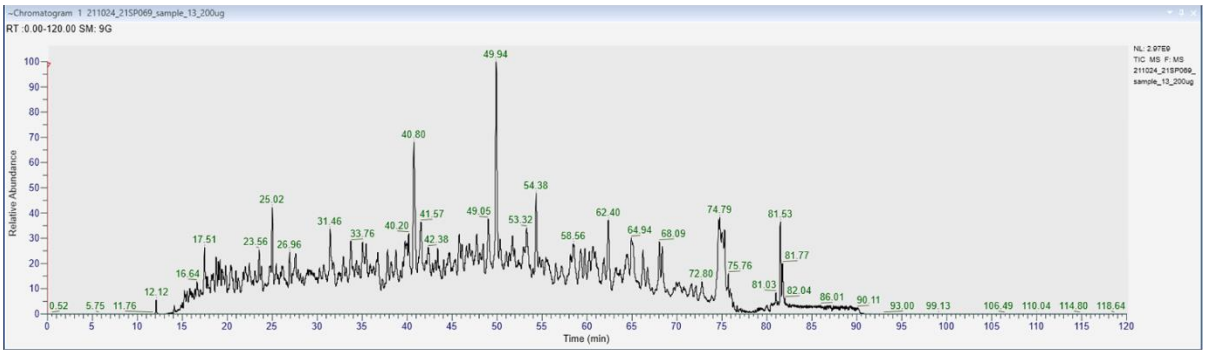
also has the advantage of using 100x fold less cells. Thus, I would propose that a future personalized vaccine target discovery pipeline can actually use only 1 x 10^7 cells from patients, with the 3-minute acid wash, to recover immunopeptides. Of course, if such cells can be amplified from patients even for a few divisions it would increase cell yield and also allow for more mass spectrometry injection or a second dimension prior to mass spectrometry to go deeper into the immunopeptidome detection.

4. Utilizing this same cell model, was finally used with drug leads that impact on splicing or NMD to determine whether such treatments impact on the normal and mutant immunopeptidome using mass spectrometry. These data are at their infancy, but we could detect drug-dependent changes in the immunopeptidome. In addition, the mutated immunopeptides that we detected (Chapter 5) required the full RNAseq database (data not shown). This implies that the RNA read depth still remains limiting so that buy fusing the six different cell states-- (wt and null p53), DMSO-only, NMDi treatment, isoginkgetin treatment—the RNArefseq was comprehensive enough to allow mutant immunopeptide detection. As a result, future stydies should involve ‘ultra-deep’ RNAseq from an individual patient’s tumour to create a more accurate refseq when we search for mutant immunopeptides using mass spectrometry.

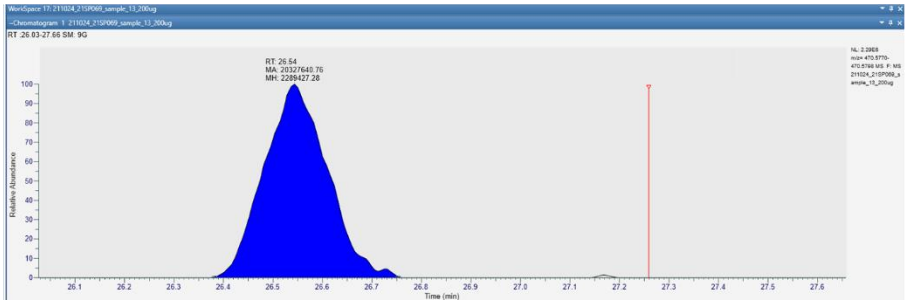
.

7 Supplementary figures

A



B



C

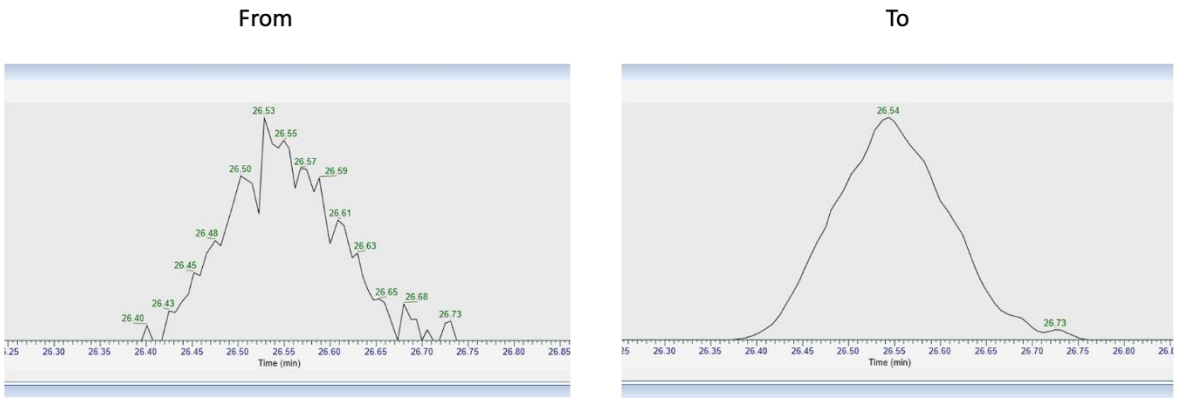


Figure 7.1 Methods for semi-quantitative analysis of peptides identified at the p53 STOP codon positions. A) An example of MS1 total ion chromatogram showing good spectra separation without contaminants with the peak $2.9*10^9$ B) Determination of peak area from extracted ion chromatograms (EIC) of p53 peptide ions in MS1 with the use of the Thermo FreeStyle Software C) Spectra smoothing with Gaussian 9 point was applied to all EIC.

A

Sample:			TAA YSP*L								
			S1			S2			S3		
Sequence	Charge	m/z of precursor ion	RT	Area MI	Ratio	RT	Area MI	Ratio	RT	Area MI	Ratio
[K].SVTCTYSP**Y**LNK.[M]	2	716.8425	39.1424	2387035	**37%**	38.6771	1628360	**31%**	38.8245	1301315	**17%**
[K].SVTCTYSP**Q**LNK.[M]	2	699.3401	33.0285	4037734	**63%**	32.7167	3680403	**69%**	32.7742	6182903	**83%**

B

Sample:			TAG YSP*L								
			S4			S5			S6		
Sequence	Charge	m/z of precursor ion	RT	Area MI	Ratio	RT	Area MI	Ratio	RT	Area MI	Ratio
[K].SVTCTYSP**Y**LNK.[M]	2	716.8425	38.7917	6194752	**26%**	38.8221	4648152	**21%**	38.8328	8152441	**22%**
[K].SVTCTYSP**Q**LNK.[M]	2	699.3401	32.7050	17524788	**74%**	32.7222	17473129	**79%**	32.7309	29728200	**78%**

C

Sample:			TGA YSP*L								
			S7			S8			S9		
Sequence	Charge	m/z of precursor ion	RT	Area MI	Ratio	RT	Area MI	Ratio	RT	Area MI	Ratio
[K].SVTCTYSP**W**LNK.[M]	2	728.3505	47.4298	1678544	**30%**	47.45**	848058	**23%**	47.4801	1512763	**19%**
[K].SVTCTYSP**R**.[L]	2	535.7504	22.87*	3894371	**70%**	22.88**	2820608	**77%**	22.9839	6596849	**81%**

* not identified, based on MS1 only
** not identified, based on MS1 only, lower quality

D

Sample:			TGA YS*AL								
			S7			S8			S9		
Sequence	Charge	m/z of precursor ion	RT	Area MI	Ratio	RT	Area MI	Ratio	RT	Area MI	Ratio
[K].SVTCTYS**C**ALNK.[M]	2	702.3183	31.5416	3660841	-	ND	-	-	ND	-	-

E

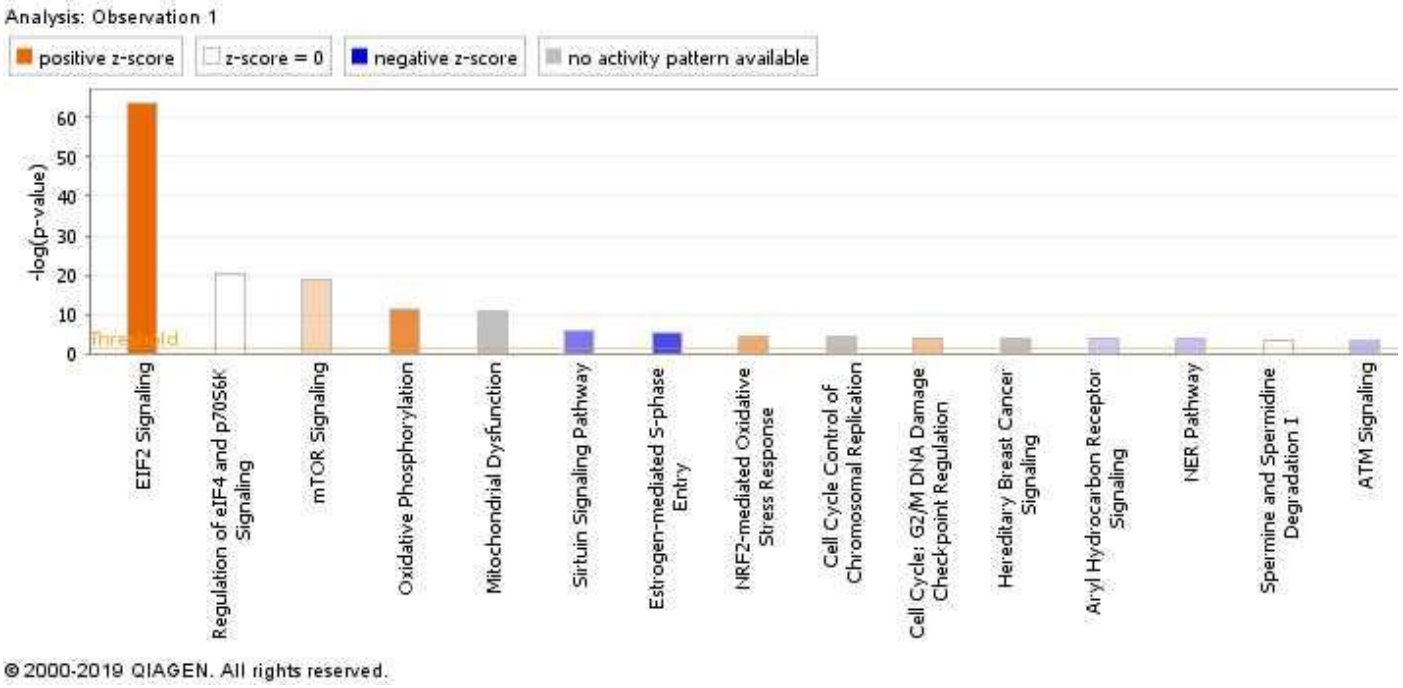
Sample:			TGA Y*SPAL								
			S13			S14			S15		
Sequence	Charge	m/z of precursor ion	RT	Area MI	Ratio	RT	Area MI	Ratio	RT	Area MI	Ratio
[K].SVTCTY**R**PALNK.[M]	2	705.3639	26.4834	5473438	14%	26.44**	166079	16%	ND	-	-
[K].SVTCTY**W**PALNK.[M]	2	720.3530	51.9013	22109575	58%	52.02**	876743	84%	ND	-	-
[K].SVTCTY**C**PALNK.[M]	2	707.3287	35.1758	10793744	28%	ND	-	-	ND	-	-

[K].SVTCTY**R**PALNK.[M]	3	470.5784	26.5	20327641		26.4	1671036	
				R	**44%**		R	68%
				W	**38%**		W	32%
				C	**18%**			

** not identified, only based on MS1, lower quality

Figure 7.2 Summary of semi-quantitative analysis of amino acid incorporation in two independent experiments. Semi-quantitative approach was based on the calculation of the ratios of spectral peak area. A) Semi-quantitation of peptides identified in p53 protein as a result of TAA-219 readthrough, in triplicates, highlighting more Q substitutions (63%, 69%, and 83%) over Y substitutions (37%, 31%, and 17%), B) TAG codon 129, in triplicates, highlighting more Q substitutions (74%, 79%, and 78%) over Y substitutions (26%, 21%, and 22%), C) TGA codon 129, in triplicates, highlighting more R substitutions (70%, 77%, and 81%) over W substitutions (30%, 23%, and 19%), D) TGA codon 128, in triplicates, highlighting mass of C substitutions, E) TGA codon 127, in triplicates, highlighting distribution (in %) or R, W, and C (charge +2 state), as well R residue in the charge +3 state.

A



B

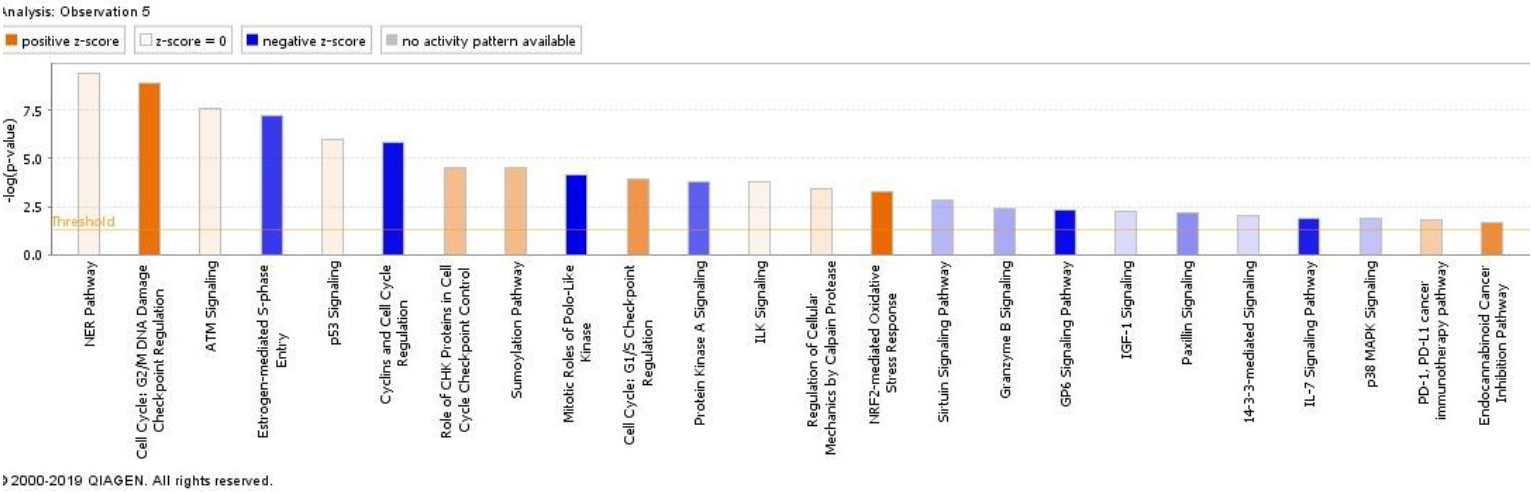


Figure 7.3 Ingenuity pathway analysis of A375 p53 null and wild-type cell line treated with splicing inhibitor. Canonical pathways significantly enriched in A375 p53 null cell line (A) and wild type (B) after 24h treatment with isoginkgetin. Significantly enriched canonical pathways were identified with a right-tailed Fisher's Exact Test that calculates a P-value determining the probability that each canonical pathway associated with the dataset was due to chance alone. The P-values were corrected for multiple testing using the Benjamini-Hochberg method for correcting the false discover rate (FDR). The z-score indicates predicted activation state of the canonical pathway. Blue colour or lighter shades of blue indicate a negative z-score and down-regulation of the pathway, and orange colour or lighter shades of orange indicate a positive z-score and up-regulation of the pathway. Ratio shows the number of significantly expressed genes compared with the total number of genes associated with the indicated canonical pathway.

8 References

Aarnoudse, C. A., van den Doel, P. B., Heemskerk, B., & Schrier, P. I. (1999). Interleukin-2-induced, melanoma-specific T cells recognize CAMEL, an unexpected translation product of LAGE-1. *International Journal of Cancer*, *82*(3), 442–448. https://doi.org/10.1002/(sici)1097-0215(19990730)82:3<442::aid-ijc19>3.0.co;2-z

Afsari, B., Guo, T., Considine, M., Florea, L., Kagohara, L. T., Stein-O'Brien, G. L., Kelley, D., Flam, E., Zambo, K. D., Ha, P. K., Geman, D., Ochs, M. F., Califano, J. A., Gaykalova, D. A., Favorov, A. V., & Fertig, E. J. (2018). Splice Expression Variation Analysis (SEVA) for inter-tumor heterogeneity of gene isoform usage in cancer. *Bioinformatics (Oxford, England)*, *34*(11), 1859–1867. https://doi.org/10.1093/bioinformatics/bty004

Albino, A. P., Vidal, M. J., McNutt, N. S., Shea, C. R., Prieto, V. G., Nanus, D. M., Palmer, J. M., & Hayward, N. K. (1994). Mutation and expression of the p53 gene in human malignant melanoma. *Melanoma Research*, *4*(1), 35–45. https://doi.org/10.1097/00008390-199402000-00006

Alfaro, J. A., Bohländer, P., Dai, M., Filius, M., Howard, C. J., van Kooten, X. F., Ohayon, S., Pomorski, A., Schmid, S., Aksimentiev, A., Anslyn, E. V., Bedran, G., Cao, C., Chinappi, M., Coyaud, E., Dekker, C., Dittmar, G., Drachman, N., Eelkema, R., … Joo, C. (2021). The emerging landscape of single-molecule protein sequencing technologies. *Nature Methods*, *18*(6), 604–617. https://doi.org/10.1038/s41592-021-01143-1

Amar, D., Izraeli, S., & Shamir, R. (2017). Utilizing somatic mutation data from numerous studies for cancer research: Proof of concept and applications. *Oncogene*, *36*(24), 3375–3383. https://doi.org/10.1038/onc.2016.489

An, B., Pan, T., Hu, J., Pang, Y., Huang, L., Chan, A. S. C., Li, X., & Yan, J. (2019). The discovery of a potent and selective third-generation EGFR kinase inhibitor as a therapy for EGFR L858R/T790M double mutant non-small cell lung cancer. *European Journal of Medicinal Chemistry*, *183*, 111709. https://doi.org/10.1016/j.ejmech.2019.111709

Andreutti-Zaugg, C., Scott, R. J., & Iggo, R. (1997). Inhibition of nonsense-mediated messenger RNA decay in clinical samples facilitates detection of human MSH2 mutations with an in vivo fusion protein assay and conventional techniques. *Cancer Research*, *57*(15), 3288–3293.

Animesh, S., Ren, X., An, O., Chen, K., Lee, S. C., Yang, H., & Fullwood, M. J. (2022). *Exploring the Neoantigen burden in Breast Carcinoma Patients* [Preprint]. Cancer Biology. https://doi.org/10.1101/2022.03.03.482669

Apcher, S., Daskalogianni, C., & Fåhraeus, R. (2015). Pioneer translation products as an alternative source for MHC-I antigenic peptides. *Molecular Immunology*, *68*(2, Part A), 68–71. https://doi.org/10.1016/j.molimm.2015.04.019

Apcher, S., Daskalogianni, C., Lejeune, F., Manoury, B., Imhoos, G., Heslop, L., & Fåhraeus, R. (2011). Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. *Proceedings of the National Academy of Sciences*, *108*(28), 11572–11577. https://doi.org/10.1073/pnas.1104104108

Apcher, S., Millot, G., Daskalogianni, C., Scherl, A., Manoury, B., & Fåhraeus, R. (2013). Translation of pre-spliced RNAs in the nuclear compartment generates peptides for the MHC class I pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(44), 17951–17956. PubMed. https://doi.org/10.1073/pnas.1309956110

Ashi, M. O., Mami-Chouaib, F., & Corgnac, S. (2022). Mutant and non-mutant neoantigen-based cancer vaccines: Recent advances and future promises. *Exploration of Targeted Anti-Tumor Therapy*, *3*(6), 746–762. https://doi.org/10.37349/etat.2022.00111

Attardi, L. D., & Donehower, L. A. (2005). Probing p53 biological functions through the use of genetically engineered mouse models. *Mutation Research*, *576*(1–2), 4–21. https://doi.org/10.1016/j.mrfmmm.2004.08.022

Baghban, R., Roshangar, L., Jahanban-Esfahlan, R., Seidi, K., Ebrahimi-Kalan, A., Jaymand, M., Kolahian, S., Javaheri, T., & Zare, P. (2020). Tumor microenvironment complexity and therapeutic implications at a glance. *Cell Communication and Signaling*, *18*(1), 59. https://doi.org/10.1186/s12964-020-0530-4

Baker, S. J., Fearon, E. R., Nigro, J. M., Hamilton, S. R., Preisinger, A. C., Jessup, J. M., vanTuinen, P., Ledbetter, D. H., Barker, D. F., Nakamura, Y., White, R., & Vogelstein, B. (1989). Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science (New York, N.Y.)*, *244*(4901), 217–221. https://doi.org/10.1126/science.2649981

Baradaran-Heravi, A., Balgi, A. D., Zimmerman, C., Choi, K., Shidmoossavee, F. S., Tan, J. S., Bergeaud, C., Krause, A., Flibotte, S., Shimizu, Y., Anderson, H. J., Mouly, V., Jan, E., Pfeifer, T., Jaquith, J. B., & Roberge, M. (2016). Novel small molecules potentiate premature termination codon readthrough by aminoglycosides. *Nucleic Acids Research*, *44*(14), 6583–6598. https://doi.org/10.1093/nar/gkw638

Bassani-Sternberg, M., Bräunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., Weber, J., Slotta-Huspenina, J., Specht, K., Martignoni, M. E., Werner, A., Hein, R., H. Busch, D., Peschel, C., Rad, R., Cox, J., Mann, M., & Krackhardt, A. M. (2016). Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nature Communications*, *7*(1), 1. https://doi.org/10.1038/ncomms13404

Baudu, T., Parratte, C., Perez, V., Ancion, M., Millevoi, S., Hervouet, E., Peigney, A., Peixoto, P., Overs, A., Herfs, M., Fraichard, A., Guittaut, M., & Baguet, A. (2021). The NMD Pathway Regulates GABARAPL1 mRNA during the EMT. *Biomedicines*, *9*(10), 1302. https://doi.org/10.3390/biomedicines9101302

Baugh, E. H., Ke, H., Levine, A. J., Bonneau, R. A., & Chan, C. S. (2018). Why are there hotspot mutations in the TP53 gene in human cancers? *Cell Death & Differentiation*, *25*(1), 154–160. https://doi.org/10.1038/cdd.2017.180

Becker, J. P., Helm, D., Rettel, M., Stein, F., Hernandez-Sanchez, A., Urban, K., Gebert, J., Kloor, M., Neu-Yilik, G., von Knebel Doeberitz, M., Hentze, M. W., & Kulozik, A. E. (2021). NMD inhibition by 5-azacytidine augments presentation of immunogenic frameshift-derived neoepitopes. *IScience*, *24*(4), 102389. https://doi.org/10.1016/j.isci.2021.102389

Bedran, G., Wang, T., Pankanin, D., Weke, K., Laird, A., Battail, C., Zanzotto, F. M., Pesquita, C., Axelson, H., Rajan, A., Harrison, D. J., Palkowski, A., Pawlik, M., Parys, M., O'Neill, R., Brennan, P. M., Symeonides, S., Goodlett, D. R., Litchfield, K., … Alfaro, J. A. (2022). *A comprehensive library of canonical and non-canonical MHC class I antigens for cancer vaccine development* (p. 2022.01.13.475872). bioRxiv. https://doi.org/10.1101/2022.01.13.475872

Behm-Ansmant, I., Kashima, I., Rehwinkel, J., Saulière, J., Wittkopp, N., & Izaurralde, E. (2007). mRNA quality control: An ancient machinery recognizes and degrades mRNAs with nonsense codons. *FEBS Letters*, *581*(15), 2845–2853. https://doi.org/10.1016/j.febslet.2007.05.027

Benjamin, D., Sato, T., Cibulskis, K., Getz, G., Stewart, C., & Lichtenstein, L. (2019). *Calling Somatic SNVs and Indels with Mutect2* [Preprint]. Bioinformatics. https://doi.org/10.1101/861054

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Beznosková, P., Bidou, L., Namy, O., & Valášek, L. S. (2021). Increased expression of tryptophan and tyrosine tRNAs elevates stop codon readthrough of reporter systems in human cell lines. *Nucleic Acids Research*, *49*(9), 5202–5215. https://doi.org/10.1093/nar/gkab315

Bhatta, B., Luz, I., Krueger, C., Teo, F. X., Lane, D. P., Sabapathy, K., & Cooks, T. (2021). Cancer Cells Shuttle Extracellular Vesicles Containing Oncogenic Mutant p53 Proteins to the Tumor Microenvironment. *Cancers*, *13*(12), 2985. https://doi.org/10.3390/cancers13122985

Bidou, L., Bugaud, O., Belakhov, V., Baasov, T., & Namy, O. (2017). Characterization of new-generation aminoglycoside promoting premature termination codon readthrough in cancer cells. *RNA Biology*, *14*(3), 378–388. https://doi.org/10.1080/15476286.2017.1285480

Bidou, L., Hatin, I., Perez, N., Allamand, V., Panthier, J.-J., & Rousset, J.-P. (2004). Premature stop codons involved in muscular dystrophies show a broad spectrum of readthrough efficiencies in response to gentamicin treatment. *Gene Therapy*, *11*(7), 619–627. https://doi.org/10.1038/sj.gt.3302211

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., Coussens, L. M., Gabrilovich, D. I., Ostrand-Rosenberg, S., Hedrick, C. C., Vonderheide, R. H., Pittet, M. J., Jain, R. K., Zou, W., Howcroft, T. K., Woodhouse, E. C., Weinberg, R. A., & Krummel, M. F. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature Medicine*, *24*(5), 541–550. https://doi.org/10.1038/s41591-018-0014-x

Bitner-Glindzicz, M., & Rahman, S. (2007). Ototoxicity caused by aminoglycosides. *BMJ : British Medical Journal*, *335*(7624), 784–785. https://doi.org/10.1136/bmj.39301.680266.AE

Blagih, J., Buck, M. D., & Vousden, K. H. (2020). P53, cancer and the immune response. *Journal of Cell Science*, *133*(5), jcs237453. https://doi.org/10.1242/jcs.237453

Blandino, G., Levine, A. J., & Oren, M. (n.d.). *Mutant p53 gain of function: Di€erential e€ects of di€erent p53 mutants on resistance of cultured cells to chemotherapy*. 9.

Blank, C. U., Haanen, J. B., Ribas, A., & Schumacher, T. N. (2016). CANCER IMMUNOLOGY. The 'cancer immunogram'. *Science (New York, N.Y.)*, *352*(6286), 658–660. https://doi.org/10.1126/science.aaf2834

Blaydes, J. P., & Hupp, T. R. (n.d.). *ONA damage triggers DRS-resistant phosphorylation of human p53 at the CK2 site*. 8.

Blaydes, J. P., Luciani, M. G., Pospisilova, S., Ball, H. M., Vojtesek, B., & Hupp, T. R. (2001). Stoichiometric phosphorylation of human p53 at Ser315 stimulates p53-dependent

transcription. *The Journal of Biological Chemistry*, *276*(7), 4699–4708. https://doi.org/10.1074/jbc.M003485200

Boettcher, S., Miller, P. G., Sharma, R., McConkey, M., Leventhal, M., Krivtsov, A. V., Giacomelli, A. O., Wong, W., Kim, J., Chao, S., Kurppa, K. J., Yang, X., Milenkowic, K., Piccioni, F., Root, D. E., Rücker, F. G., Flamand, Y., Neuberg, D., Lindsley, R. C., … Ebert, B. L. (2019). A dominant-negative effect drives selection of TP53 missense mutations in myeloid malignancies. *Science (New York, N.Y.)*, *365*(6453), 599–604. https://doi.org/10.1126/science.aax3649

Bonetti, B., Fu, L., Moon, J., & Bedwell, D. M. (1995). The Efficiency of Translation Termination is Determined by a Synergistic Interplay Between Upstream and Downstream Sequences inSaccharomyces cerevisiae. *Journal of Molecular Biology*, *251*(3), 334–345. https://doi.org/10.1006/jmbi.1995.0438

Bongiorno, R., Colombo, M. P., & Lecis, D. (2021). Deciphering the nonsense-mediated mRNA decay pathway to identify cancer cell vulnerabilities for effective cancer therapy. *Journal of Experimental & Clinical Cancer Research*, *40*(1), 376. https://doi.org/10.1186/s13046-021-02192-2

Boon, T., Van Pel, A., De Plaen, E., Chomez, P., Lurquin, C., Szikora, J. P., Sibille, C., Mariamé, B., Van den Eynde, B., & Lethé, B. (1989). Genes coding for T-cell-defined tum transplantation antigens: Point mutations, antigenic peptides, and subgenic expression. *Cold Spring Harbor Symposia on Quantitative Biology*, *54 Pt 1*, 587–596. https://doi.org/10.1101/sqb.1989.054.01.070

Bordeira-Carriço, R., Pêgo, A. P., Santos, M., & Oliveira, C. (2012). Cancer syndromes and therapy by stop-codon readthrough. *Trends in Molecular Medicine*, *18*(11), 667–678. https://doi.org/10.1016/j.molmed.2012.09.004

Borgatti, M., Altamura, E., Salvatori, F., D'Aversa, E., & Altamura, N. (2020). Screening Readthrough Compounds to Suppress Nonsense Mutations: Possible Application to β-Thalassemia. *Journal of Clinical Medicine*, *9*(2), 289. https://doi.org/10.3390/jcm9020289

Bradley, R. K., & Anczuków, O. (2023). RNA splicing dysregulation and the hallmarks of cancer. *Nature Reviews Cancer*, 1–21. https://doi.org/10.1038/s41568-022-00541-7

Brogden, K. A., Vali, S., & Abbasi, T. (2016). PD-L1 is a diverse molecule regulating both tumor-intrinsic signaling and adaptive immunosuppression. *Translational Cancer Research*, *5*(7), S1396-S1399–S1399. https://doi.org/10.21037/11305

Broseus, L., & Ritchie, W. (2020). Challenges in detecting and quantifying intron retention from next generation sequencing data. *Computational and Structural Biotechnology Journal*, *18*, 501–508. https://doi.org/10.1016/j.csbj.2020.02.010

Brown, G. (2021). Oncogenes, Proto-Oncogenes, and Lineage Restriction of Cancer Stem Cells. *International Journal of Molecular Sciences*, *22*(18), 9667. https://doi.org/10.3390/ijms22189667

Brumbaugh, K. M., Otterness, D. M., Geisen, C., Oliveira, V., Brognard, J., Li, X., Lejeune, F., Tibbetts, R. S., Maquat, L. E., & Abraham, R. T. (2004). The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. *Molecular Cell*, *14*(5), 585–598. https://doi.org/10.1016/j.molcel.2004.05.005

Brusic, V., Rudy, G., Honeyman, G., Hammer, J., & Harrison, L. (1998). Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics (Oxford, England)*, *14*(2), 121–130. https://doi.org/10.1093/bioinformatics/14.2.121

Burke, J. F., & Mogg, A. E. (1985). Suppression of a nonsense mutation in mammalian cells in vivo by the aminoglycoside antibiotics G-418 and paromomycin. *Nucleic Acids Research*, *13*(17), 6265–6272.

Cai, X., & Liu, X. (2008). Inhibition of Thr-55 phosphorylation restores p53 nuclear localization and sensitizes cancer cells to DNA damage. *Proceedings of the National Academy of Sciences*, *105*(44), 16958–16963. https://doi.org/10.1073/pnas.0804608105

Calvo Tardón, M., Allard, M., Dutoit, V., Dietrich, P. Y., & Walker, P. R. (2019). Peptides as cancer vaccines. *Current Opinion in Pharmacology*, *47*, 20–26. https://doi.org/10.1016/j.coph.2019.01.007

Carreno, B. M., Magrini, V., Becker-Hapak, M., Kaabinejadian, S., Hundal, J., Petti, A. A., Ly, A., Lie, W.-R., Hildebrand, W. H., Mardis, E. R., & Linette, G. P. (2015). Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science (New York, N.Y.)*, *348*(6236), 803–808. https://doi.org/10.1126/science.aaa3828

Castle, J. C., Kreiter, S., Diekmann, J., Löwer, M., van de Roemer, N., de Graaf, J., Selmi, A., Diken, M., Boegel, S., Paret, C., Koslowski, M., Kuhn, A. N., Britten, C. M., Huber, C., Türeci, Ö., & Sahin, U. (2012). Exploiting the Mutanome for Tumor Vaccination. *Cancer Research*, *72*(5), 1081–1091. https://doi.org/10.1158/0008-5472.CAN-11-3722

Celik, A., Baker, R., He, F., & Jacobson, A. (2017). High-resolution profiling of NMD targets in yeast reveals translational fidelity as a basis for substrate selection. *RNA*, *23*(5), 735–748. https://doi.org/10.1261/rna.060541.116

Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., Schrock, A., Campbell, B., Shlien, A., Chmielecki, J., Huang, F., He, Y., Sun, J., Tabori, U., Kennedy, M., Lieber, D. S., Roels, S., White, J., Otto, G. A., … Frampton, G. M. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Medicine*, *9*(1), 34. https://doi.org/10.1186/s13073-017-0424-2

Chang, L., Li, C., Guo, T., Wang, H., Ma, W., Yuan, Y., Liu, Q., Ye, Q., & Liu, Z. (2016). The human RNA surveillance factor UPF1 regulates tumorigenesis by targeting Smad7 in hepatocellular carcinoma. *Journal of Experimental & Clinical Cancer Research*, *35*(1), 8. https://doi.org/10.1186/s13046-016-0286-2

Chen, F., Wei, J., & Liu, B. (2019). *Neoantigen identification strategies enable personalized immunotherapy in refractory solid tumors Find the latest version :* https://doi.org/10.1172/JCI99538

Cheng, R., Xu, Z., Luo, M., Wang, P., Cao, H., Jin, X., Zhou, W., Xiao, L., & Jiang, Q. (2022). Identification of alternative splicing-derived cancer neoantigens for mRNA vaccine development. *Briefings in Bioinformatics*, *23*(2), bbab553. https://doi.org/10.1093/bib/bbab553

Chong, C., Coukos, G., & Bassani-Sternberg, M. (2021). Identification of tumor antigens with immunopeptidomics. *Nature Biotechnology*. https://doi.org/10.1038/s41587-021-01038-8

Chong, C., Marino, F., Pak, H., Racle, J., Daniel, R. T., Müller, M., Gfeller, D., Coukos, G., & Bassani-Sternberg, M. (2018). High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferonγ-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome. *Molecular & Cellular Proteomics: MCP*, *17*(3), 533–548. https://doi.org/10.1074/mcp.TIR117.000383

Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B. J., Michaux, J., Bilic, I., Hirsekorn, A., Calviello, L., Simó-Riudalbas, L., Planet, E., Lubiński, J., Bryśkiewicz, M., Wiznerowicz, M., … Bassani-Sternberg, M. (2020). Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nature Communications*, *11*(1), 1293. https://doi.org/10.1038/s41467-020-14968-9

Chu, D., & Wei, L. (2019). Nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation. *BMC Cancer*, *19*(1), 359. https://doi.org/10.1186/s12885-019-5572-x

Chudley, L., McCann, K. J., Coleman, A., Cazaly, A. M., Bidmon, N., Britten, C. M., van der Burg, S. H., Gouttefangeas, C., Jandus, C., Laske, K., Maurer, D., Romero, P., Schröder, H., Stynenbosch, L. F. M., Walter, S., Welters, M. J. P., & Ottensmeier, C. H. (2014). Harmonisation of short-term in vitro culture for the expansion of antigen-specific CD8+ T cells with detection by ELISPOT and HLA-multimer staining. *Cancer Immunology, Immunotherapy*, *63*(11), 1199–1211. https://doi.org/10.1007/s00262-014-1593-0

Clancy, J. P., Bebök, Z., Ruiz, F., King, C., Jones, J., Walker, L., Greer, H., Hong, J., Wing, L., Macaluso, M., Lyrene, R., Sorscher, E. J., & Bedwell, D. M. (2001). Evidence that systemic gentamicin suppresses premature stop mutations in patients with cystic fibrosis. *American Journal of Respiratory and Critical Care Medicine*, *163*(7), 1683–1692. https://doi.org/10.1164/ajrccm.163.7.2004001

Cole, D. K., Pumphrey, N. J., Boulter, J. M., Sami, M., Bell, J. I., Gostick, E., Price, D. A., Gao, G. F., Sewell, A. K., & Jakobsen, B. K. (2007). Human TCR-Binding Affinity is Governed by MHC Class Restriction1. *The Journal of Immunology*, *178*(9), 5727–5734. https://doi.org/10.4049/jimmunol.178.9.5727

Colloca, S., Barnes, E., Folgori, A., Ammendola, V., Capone, S., Cirillo, A., Siani, L., Naddeo, M., Grazioli, F., Esposito, M. L., Ambrosio, M., Sparacino, A., Bartiromo, M., Meola, A., Smith, K., Kurioka, A., O'Hara, G. A., Ewer, K. J., Anagnostou, N., … Nicosia, A. (2012). Vaccine vectors derived from a large collection of simian adenoviruses induce potent cellular immunity across multiple species. *Science Translational Medicine*, *4*(115), 115ra2. https://doi.org/10.1126/scitranslmed.3002925

Coulie, P. G., Lehmann, F., Lethé, B., Herman, J., Lurquin, C., Andrawiss, M., & Boon, T. (1995). A mutated intron sequence codes for an antigenic peptide recognized by cytolytic T lymphocytes on a human melanoma. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(17), 7976–7980. https://doi.org/10.1073/pnas.92.17.7976

Cowen, L. E., Luo, H., & Tang, Y. (2019). Characterization of SMG7 14-3-3-like domain reveals phosphoserine binding-independent regulation of p53 and UPF1. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-49229-3

Crick, F. H. (1966). Codon--anticodon pairing: The wobble hypothesis. *Journal of Molecular Biology*, *19*(2), 548–555. https://doi.org/10.1016/s0022-2836(66)80022-0

D'Alise, A. M., Leoni, G., Cotugno, G., Troise, F., Langone, F., Fichera, I., De Lucia, M., Avalle, L., Vitale, R., Leuzzi, A., Bignone, V., Di Matteo, E., Tucci, F. G., Poli, V., Lahm, A., Catanese, M. T., Folgori, A., Colloca, S., Nicosia, A., & Scarselli, E. (2019). Adenoviral vaccine targeting multiple neoantigens as strategy to eradicate large tumors combined with checkpoint blockade. *Nature Communications*, *10*(1), 1. https://doi.org/10.1038/s41467-019-10594-2

D'Angelo, S. P., Mahoney, M. R., Van Tine, B. A., Atkins, J., Milhem, M. M., Jahagirdar, B. N., Antonescu, C. R., Horvath, E., Tap, W. D., & Schwartz, G. K. (2018). Nivolumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): Two open-label, non-comparative, randomised, phase 2 trials. *The Lancet Oncology*, *19*(3), 416–426.

Datta, A., Ghatak, D., Das, S., Banerjee, T., Paul, A., Butti, R., Gorain, M., Ghuwalewala, S., Roychowdhury, A., Alam, S. K., Das, P., Chatterjee, R., Dasgupta, M., Panda, C. K., Kundu, G. C., & Roychoudhury, S. (2017). P53 gain-of-function mutations increase Cdc7-dependent replication initiation. *EMBO Reports*, *18*(11), 2030–2050. https://doi.org/10.15252/embr.201643347

De Boeck, K., Zolin, A., Cuppens, H., Olesen, H. V., & Viviani, L. (2014). The relative frequency of CFTR mutation classes in European patients with cystic fibrosis. *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society*, *13*(4), 403–409. https://doi.org/10.1016/j.jcf.2013.12.003

De Filippis, R., Wölflein, G., Um, I. H., Caie, P. D., Warren, S., White, A., Suen, E., To, E., Arandjelović, O., & Harrison, D. J. (2022). Use of High-Plex Data Reveals Novel Insights into the Tumour Microenvironment of Clear Cell Renal Cell Carcinoma. *Cancers*, *14*(21). https://doi.org/10.3390/cancers14215387

DeCaprio, J., & Kohl, T. O. (2019). Cross-Linking Antibodies to Beads Using Dimethyl Pimelimidate (DMP). *Cold Spring Harbor Protocols*, *2019*(2). https://doi.org/10.1101/pdb.prot098624

Delphin, C., & Baudier, J. (1994). The protein kinase C activator, phorbol ester, cooperates with the wild-type p53 species of Ras-transformed embryo fibroblasts growth arrest. *Journal of Biological Chemistry*, *269*(47), 29579–29587. https://doi.org/10.1016/S0021-9258(18)43919-1

Deng, Z., Tian, Y., Song, J., An, G., & Yang, P. (2022). mRNA Vaccines: The Dawn of a New Era of Cancer Immunotherapy. *Frontiers in Immunology*, *13*. https://www.frontiersin.org/articles/10.3389/fimmu.2022.887125

Ding, L., Kim, M., Kanchi, K. L., Dees, N. D., Lu, C., Griffith, M., Fenstermacher, D., Sung, H., Miller, C. A., Goetz, B., Wendl, M. C., Griffith, O., Cornelius, L. A., Linette, G. P., McMichael, J. F., Sondak, V. K., Fields, R. C., Ley, T. J., Mulé, J. J., … Weber, J. S. (2014). Clonal Architectures and Driver Mutations in Metastatic Melanomas. *PLoS ONE*, *9*(11), e111153. https://doi.org/10.1371/journal.pone.0111153

Ding, Z., Ying, W., He, Y., Chen, X., Jiao, Y., Wang, J., & Zhou, X. (2021). LncRNA-UCA1 in the diagnosis of bladder cancer. *Medicine*, *100*(11), e24805. https://doi.org/10.1097/MD.0000000000024805

Donehower, L. A. (1996). The p53-deficient mouse: A model for basic and applied cancer studies. *Seminars in Cancer Biology*, *7*(5), 269–278. https://doi.org/10.1006/scbi.1996.0035

Dong, C., Reiter, J. L., Dong, E., Wang, Y., Lee, K. P., Lu, X., & Liu, Y. (2022). Intron-Retention Neoantigen Load Predicts Favorable Prognosis in Pancreatic Cancer. *JCO Clinical Cancer Informatics*, *6*, e2100124. https://doi.org/10.1200/CCI.21.00124

Dong, X., & Chen, R. (2019). Understanding aberrant RNA splicing to facilitate cancer diagnosis and therapy. *Oncogene*. https://doi.org/10.1038/s41388-019-1138-2

Dong, X., Yang, Y., Yuan, Q., Hou, J., & Wu, G. (2021). High Expression of CEMIP Correlates Poor Prognosis and the Tumur Microenvironment in Breast Cancer as a Promisingly Prognostic Biomarker. *Frontiers in Genetics*, *12*. https://www.frontiersin.org/articles/10.3389/fgene.2021.768140

Eggertsson, G., & Söll, D. (1988). Transfer ribonucleic acid-mediated suppression of termination codons in Escherichia coli. *Microbiological Reviews*, *52*(3), 354–374.

El-Bchiri, J., Guilloux, A., Dartigues, P., Loire, E., Mercier, D., Buhard, O., Sobhani, I., de la Grange, P., Auboeuf, D., Praz, F., Fléjou, J.-F., & Duval, A. (2008). Nonsense-Mediated mRNA Decay Impacts MSI-Driven Carcinogenesis and Anti-Tumor Immunity in Colorectal Cancers. *PLoS ONE*, *3*(7), e2583. https://doi.org/10.1371/journal.pone.0002583

Faktor, J., Hernychová, L., Vojtěšek, B., & Hupp, T. (2018). Proteogenomic Platform for Identification of Tumor Specific Antigens. *Klinicka Onkologie*, *31*(Suppl 2), 102–107. https://doi.org/10.14735/amko20182s102

Fan, S., Yuan, R., Ma, Y. X., Meng, Q., Goldberg, I. D., & Rosen, E. M. (2001). Mutant BRCA1 genes antagonize phenotype of wild-type BRCA1. *Oncogene*, *20*(57), 57. https://doi.org/10.1038/sj.onc.1205033

Finotello, F., Rieder, D., Hackl, H., & Trajanoski, Z. (2019). Next-generation computational tools for interrogating cancer immunity. *Nature Reviews Genetics*, *20*(12), 12. https://doi.org/10.1038/s41576-019-0166-7

Floquet, C., Deforges, J., Rousset, J.-P., & Bidou, L. (2011). Rescue of non-sense mutated p53 tumor suppressor gene by aminoglycosides. *Nucleic Acids Research*, *39*(8), 3350–3362. https://doi.org/10.1093/nar/gkq1277

Frezza, V., Chellini, L., Del Verme, A., & Paronetto, M. P. (2023). RNA Editing in Cancer Progression. *Cancers*, *15*(21), 5277. https://doi.org/10.3390/cancers15215277

Fujiwara, K., Shao, Y., Niu, N., Zhang, T., Herbst, B., Henderson, M., Muth, S., Zhang, P., & Zheng, L. (2022). Direct identification of HLA class I and class II-restricted T cell epitopes in pancreatic cancer tissues by mass spectrometry. *Journal of Hematology & Oncology*, *15*(1), 154. https://doi.org/10.1186/s13045-022-01373-6

Gencel-Augusto, J., & Lozano, G. (2020). p53 tetramerization: At the center of the dominant-negative effect of mutant p53. *Genes & Development*, *34*(17–18), 1128–1146. https://doi.org/10.1101/gad.340976.120

Gessulat, S., Schmidt, T., Zolg, D. P., Samaras, P., Schnatbaum, K., Zerweck, J., Knaute, T., Rechenberger, J., Delanghe, B., Huhmer, A., Reimer, U., Ehrlich, H.-C., Aiche, S., Kuster, B., & Wilhelm, M. (2019). Prosit: Proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nature Methods*, *16*(6), 6. https://doi.org/10.1038/s41592-019-0426-7

Gilboa, E. (2016). A quantum leap in cancer vaccines. *Journal for ImmunoTherapy of Cancer*, *4*(1), 3–4. https://doi.org/10.1186/s40425-016-0192-3

Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., & Aebersold, R. (2012). Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis*. *Molecular & Cellular Proteomics*, *11*(6), O111.016717. https://doi.org/10.1074/mcp.O111.016717

Gnanapradeepan, K., Basu, S., Barnoud, T., Budina-Kolomets, A., Kung, C.-P., & Murphy, M. E. (2018). The p53 Tumor Suppressor in the Control of Metabolism and Ferroptosis. *Frontiers in Endocrinology*, *9*. https://www.frontiersin.org/article/10.3389/fendo.2018.00124

Gong, X., & Karchin, R. (2023). Clustering by antigen-presenting genes reveals immune landscapes and predicts response to checkpoint immunotherapy. *Scientific Reports*, *13*(1), 1. https://doi.org/10.1038/s41598-023-28167-1

Guallar-Garrido, S., & Julián, E. (2020). Bacillus Calmette-Guérin (BCG) Therapy for Bladder Cancer: An Update. *ImmunoTargets and Therapy*, *9*, 1–11. https://doi.org/10.2147/ITT.S202006

Gudikote, J. P., Cascone, T., Poteete, A., Sitthideatphaiboon, P., Wu, Q., Morikawa, N., Zhang, F., Peng, S., Tong, P., Li, L., Shen, L., Nilsson, M., Jones, P., Sulman, E. P., Wang, J., Bourdon, J.-C., Johnson, F. M., & Heymach, J. V. (2021). Inhibition of nonsense-mediated decay rescues p53β/γ isoform expression and activates the p53 pathway in MDM2-overexpressing and select p53-mutant cancers. *The Journal of Biological Chemistry*, *297*(5), 101163. https://doi.org/10.1016/j.jbc.2021.101163

Haddad, R. I., Seiwert, T. Y., Chow, L. Q. M., Gupta, S., Weiss, J., Gluck, I., Eder, J. P., Burtness, B., Tahara, M., Keam, B., Kang, H., Muro, K., Albright, A., Mogg, R., Ayers, M., Huang, L., Lunceford, J., Cristescu, R., Cheng, J., & Mehra, R. (2022). Influence of tumor mutational burden, inflammatory gene expression profile, and PD-L1 expression on response to pembrolizumab in head and neck squamous cell carcinoma. *Journal for ImmunoTherapy of Cancer*, *10*(2), e003026. https://doi.org/10.1136/jitc-2021-003026

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, *12*(1), 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, *100*(1), 57–70.

He, F., Li, X., Spatrick, P., Casillo, R., Dong, S., & Jacobson, A. (2003). Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5’ to 3’ mRNA decay pathways in yeast. *Molecular Cell*, *12*(6), 1439–1452. https://doi.org/10.1016/s1097-2765(03)00446-5

He, G., Siddik, Z. H., Huang, Z., Wang, R., Koomen, J., Kobayashi, R., Khokhar, A. R., & Kuang, J. (2005). Induction of p21 by p53 following DNA damage inhibits both Cdk4 and Cdk2 activities. *Oncogene*, *24*(18), 2929–2943. https://doi.org/10.1038/sj.onc.1208474

Hilf, N., Kuttruff-Coqui, S., Frenzel, K., Bukur, V., Stevanović, S., Gouttefangeas, C., Platten, M., Tabatabai, G., Dutoit, V., van der Burg, S. H., thor Straten, P., Martínez-Ricarte, F., Ponsati, B., Okada, H., Lassen, U., Admon, A., Ottensmeier, C. H., Ulges, A., Kreiter, S., … Wick, W. (2019). Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature*, *565*(7738), 7738. https://doi.org/10.1038/s41586-018-0810-y

Hu, J., Wang, Z., Yang, S., Lu, Y., & Li, G. (2022). The edited UPF1 is correlated with elevated asparagine synthetase in pancreatic ductal adenocarcinomas. *Molecular Biology Reports*. https://doi.org/10.1007/s11033-022-07211-9

Huang, X., Tian, M., Hernandez, C. C., Hu, N., & Macdonald, R. L. (2012). The GABRG2 nonsense mutation, Q40X, associated with Dravet syndrome activated NMD and generated a truncated subunit that was partially rescued by aminoglycoside-induced stop codon read-through. *Neurobiology of Disease*, *48*(1), 115–123. https://doi.org/10.1016/j.nbd.2012.06.013

Hundal, J., Kiwala, S., McMichael, J., Miller, C. A., Xia, H., Wollam, A. T., Liu, C. J., Zhao, S., Feng, Y.-Y., Graubert, A. P., Wollam, A. Z., Neichin, J., Neveau, M., Walker, J., Gillanders, W. E., Mardis, E. R., Griffith, O. L., & Griffith, M. (2020). pVACtools: A Computational Toolkit to Identify and Visualize Cancer Neoantigens. *Cancer Immunology Research*, *8*(3), 409–420. https://doi.org/10.1158/2326-6066.CIR-19-0401

Hunt, D. F., Henderson, R. A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A. L., Appella, E., & Engelhard, V. H. (1992). Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science (New York, N.Y.)*, *255*(5049), 1261–1263. https://doi.org/10.1126/science.1546328

Huntzinger, E., Kashima, I., Fauser, M., Saulière, J., & Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA*, *14*(12), 2609–2617. https://doi.org/10.1261/rna.1386208

Hupp, T. R., Meek, D. W., Midgley, C. A., & Lane, D. P. (1992). Regulation of the specific DNA binding function of p53. *Cell*, *71*(5), 875–886. https://doi.org/10.1016/0092-8674(92)90562-q

Huth, M., Santini, L., Galimberti, E., Ramesmayer, J., Titz-Teixeira, F., Sehlke, R., Oberhuemer, M., Stummer, S., Herzog, V., Garmhausen, M., Romeike, M., Chugunova, A., Leesch, F., Holcik, L., Weipoltshammer, K., Lackner, A., Schoefer, C., von Haeseler, A., Buecker, C., … Leeb, M. (2022). NMD is required for timely cell fate transitions by fine-tuning gene expression and regulating translation. *Genes & Development*, genesdev;gad.347690.120v1. https://doi.org/10.1101/gad.347690.120

Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J. S., Jackson, S. E., Wills, M. R., & Weissman, J. S. (2014). Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Reports*, *8*(5), 1365–1379. https://doi.org/10.1016/j.celrep.2014.07.045

Inoue, K., Kurabayashi, A., Shuin, T., Ohtsuki, Y., & Furihata, M. (2012). Overexpression of p53 protein in human tumors. *Medical Molecular Morphology*, *45*(3), 115–123. https://doi.org/10.1007/s00795-012-0575-6

Ishizaki, H., Song, G.-Y., Srivastava, T., Carroll, K. D., Shahabi, V., Manuel, E. R., Diamond, D. J., & Ellenhorn, J. D. I. (2010). Heterologous Prime/Boost Immunization with p53-based Vaccines Combined with Toll-Like Receptor Stimulation Enhances Tumor Regression. *Journal of Immunotherapy (Hagerstown, Md. : 1997)*, *33*(6), 609–617. https://doi.org/10.1097/CJI.0b013e3181e032c6

J, M., T, G., L, V., Om, B., T, T., R, B., & L, R. (2019). Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Molecular Omics*, *15*(5). https://doi.org/10.1039/c9mo00082h

Jagannathan, S., & Bradley, R. K. (2016). Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population. *Genome Research*, *26*(12), 1639–1650. https://doi.org/10.1101/gr.205070.116

Jensen, K. K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J. A., Yan, Z., Sette, A., Peters, B., & Nielsen, M. (2018). Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*, *154*(3), 394–406.

Jhunjhunwala, S., Hammer, C., & Delamarre, L. (2021). Antigen presentation in cancer: Insights into tumour immunogenicity and immune evasion. *Nature Reviews Cancer*. https://doi.org/10.1038/s41568-021-00339-z

Jia, J., Werkmeister, E., Gonzalez-Hilarion, S., Leroy, C., Gruenert, D. C., Lafont, F., Tulasne, D., & Lejeune, F. (2017). Premature termination codon readthrough in human cells occurs in novel cytoplasmic foci and requires UPF proteins. *Journal of Cell Science*, *130*(18), 3009–3022. https://doi.org/10.1242/jcs.198176

Johnson, D. B., Estrada, M. V., Salgado, R., Sanchez, V., Doxie, D. B., Opalenik, S. R., Vilgelm, A. E., Feld, E., Johnson, A. S., Greenplate, A. R., Sanders, M. E., Lovly, C. M., Frederick, D. T., Kelley, M. C., Richmond, A., Irish, J. M., Shyr, Y., Sullivan, R. J., Puzanov, I., … Balko, J. M. (2016). Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nature Communications*, *7*(1), 1. https://doi.org/10.1038/ncomms10582

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., & Nielsen, M. (2017a). NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *The Journal of Immunology*, *199*(9). https://doi.org/10.4049/jimmunol.1700893

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., & Nielsen, M. (2017b). NetMHCpan-4.0: Improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *The Journal of Immunology*, *199*(9), 3360–3368.

Kalaora, S., Barnea, E., Merhavi-Shoham, E., Qutob, N., Teer, J. K., Shimony, N., Schachter, J., Rosenberg, S. A., Besser, M. J., Admon, A., & Samuels, Y. (2016). Use of HLA peptidomics and whole exome sequencing to identify human immunogenic neo-antigens. *Oncotarget*, *7*(5), 5110–5117. https://doi.org/10.18632/oncotarget.6960

Kalathiya, Padariya, Pawlicka, Verma, Houston, Hupp, & Alfaro. (2019). Insights into the Effects of Cancer Associated Mutations at the UPF2 and ATP-Binding Sites of NMD Master Regulator: UPF1. *International Journal of Molecular Sciences*, *20*(22), 5644. https://doi.org/10.3390/ijms20225644

Kanaseki, T., Tokita, S., & Torigoe, T. (2019). Proteogenomic discovery of cancer antigens: Neoantigens and beyond. *Pathology International*, pin.12841. https://doi.org/10.1111/pin.12841

Kang, Z.-J., Liu, Y.-F., Xu, L.-Z., Long, Z.-J., Huang, D., Yang, Y., Liu, B., Feng, J.-X., Pan, Y.-J., Yan, J.-S., & Liu, Q. (2016). The Philadelphia chromosome in leukemogenesis. *Chinese Journal of Cancer*, *35*, 48. https://doi.org/10.1186/s40880-016-0108-0

Kantoff, P. W., Higano, C. S., Shore, N. D., Berger, E. R., Small, E. J., Penson, D. F., Redfern, C. H., Ferrari, A. C., Dreicer, R., Sims, R. B., Xu, Y., Frohlich, M. W., & Schellhammer, P. F. (2010). Sipuleucel-T Immunotherapy for Castration-Resistant Prostate Cancer. *New England Journal of Medicine*, *363*(5), 411–422. https://doi.org/10.1056/NEJMoa1001294

Karousis, E. D., Nasif, S., & Mühlemann, O. (2016). Nonsense-mediated mRNA decay: Novel mechanistic insights and biological impact. *Wiley Interdisciplinary Reviews: RNA*, *7*(5), 661–682. https://doi.org/10.1002/wrna.1357

Khurana, E., Fu, Y., Chakravarty, D., Demichelis, F., Rubin, M. A., & Gerstein, M. (2016). Role of non-coding sequence variants in cancer. *Nature Reviews Genetics*, *17*(2), 2. https://doi.org/10.1038/nrg.2015.17

Kichina, J. V., Rauth, S., Das Gupta, T. K., & Gudkov, A. V. (2003). Melanoma cells can tolerate high levels of transcriptionally active endogenous p53 but are sensitive to retrovirus-transduced p53. *Oncogene*, *22*(31), 4911–4917. https://doi.org/10.1038/sj.onc.1206741

Kinniburgh, A. J., Maquat, L. E., Schedl, T., Rachmilewitz, E., & Ross, J. (1982). MRNA-deficient beta o-thalassemia results from a single nucleotide deletion. *Nucleic Acids Research*, *10*(18), 5421–5427. https://doi.org/10.1093/nar/10.18.5421

Knudson, A. G. (1971). Mutation and cancer: Statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America*, *68*(4), 820–823. https://doi.org/10.1073/pnas.68.4.820

Koboldt, D. C., Larson, D. E., & Wilson, R. K. (2013). Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *44*, 15.4.1-15.4.17. https://doi.org/10.1002/0471250953.bi1504s44

Koşaloğlu-Yalçın, Z., Lee, J., Greenbaum, J., Schoenberger, S. P., Miller, A., Kim, Y. J., Sette, A., Nielsen, M., & Peters, B. (2022). Combined assessment of MHC binding and antigen abundance improves T cell epitope predictions. *IScience*, *25*(2), 103850. https://doi.org/10.1016/j.isci.2022.103850

Kote, S., Pirog, A., Bedran, G., Alfaro, J., & Dapic, I. (2020). Mass spectrometry-based identification of MHC-associated peptides. In *Cancers* (Vol. 12). MDPI AG. https://doi.org/10.3390/cancers12030535

Kranz, L. M., Diken, M., Haas, H., Kreiter, S., Loquai, C., Reuter, K. C., Meng, M., Fritz, D., Vascotto, F., Hefesha, H., Grunwitz, C., Vormehr, M., Hüsemann, Y., Selmi, A., Kuhn, A. N., Buck, J., Derhovanessian, E., Rae, R., Attig, S., … Sahin, U. (2016). Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature*, *534*(7607), 7607. https://doi.org/10.1038/nature18300

Krasnov, G. S., Dmitriev, A. A., Kudryavtseva, A. V., Shargunov, A. V., Karpov, D. S., Uroshlev, L. A., Melnikova, N. V., Blinov, V. M., Poverennaya, E. V., Archakov, A. I., Lisitsa, A. V., & Ponomarenko, E. A. (2015). PPLine: An Automated Pipeline for SNP, SAP, and Splice Variant Detection in the Context of Proteogenomics. *Journal of Proteome Research*, *14*(9), 3729–3737. https://doi.org/10.1021/acs.jproteome.5b00490

Kushwah, R., Wu, J., Oliver, J. R., Jiang, G., Zhang, J., Siminovitch, K. A., & Hu, J. (2010). Uptake of apoptotic DC converts immature DC into tolerogenic DC, which induce differentiation of Foxp3+ regulatory T cells. *European Journal of Immunology*, *40*(4), 1022–1035. https://doi.org/10.1002/eji.200939782

Labani-Motlagh, A., Ashja-Mahdavi, M., & Loskog, A. (2020). The Tumor Microenvironment: A Milieu Hindering and Obstructing Antitumor Immune Responses. *Frontiers in Immunology*, *11*. https://www.frontiersin.org/articles/10.3389/fimmu.2020.00940

Labunskyy, V. M., Hatfield, D. L., & Gladyshev, V. N. (2014). Selenoproteins: Molecular Pathways and Physiological Roles. *Physiological Reviews*, *94*(3), 739–777. https://doi.org/10.1152/physrev.00039.2013

Lan, T., Yuan, K., Yan, X., Xu, L., Liao, H., Hao, X., Wang, J., Liu, H., Chen, X., Xie, K., Li, J., Liao, M., Huang, J., Zeng, Y., & Wu, H. (2019). LncRNA SNHG10 Facilitates Hepatocarcinogenesis and Metastasis by Modulating Its Homolog SCARNA13 via a Positive Feedback Loop. *Cancer Research*, *79*(13), 3220–3234. https://doi.org/10.1158/0008-5472.CAN-18-4044

Lane, D. P. (1992). Cancer. P53, guardian of the genome. *Nature*, *358*(6381), 15–16. https://doi.org/10.1038/358015a0

Lane, D. P., & Crawford, L. V. (1979). T antigen is bound to a host protein in SY40-transformed cells. *Nature*, *278*(5701), 5701. https://doi.org/10.1038/278261a0

Laumont, C. M., Daouda, T., Laverdure, J.-P., Bonneil, É., Caron-Lizotte, O., Hardy, M.-P., Granados, D. P., Durette, C., Lemieux, S., Thibault, P., & Perreault, C. (2016). Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. *Nature Communications*, *7*, 10238. https://doi.org/10.1038/ncomms10238

Laumont, C. M., Vincent, K., Hesnard, L., Audemard, E., Bonneil, E., Laverdure, J.-P., Gendron, P., Courcelles, M., Hardy, M.-P., Cote, C., Durette, C., St-Pierre, C., Benhammadi, M., Lanoix, J., Vobecky, S., Haddad, E., Lemieux, S., Thibault, P., & Perreault, C. (2018). Noncoding regions are the main source of targetable tumor-specific antigens. *Science Translational Medicine*, *10*(470). https://doi.org/10.1126/scitranslmed.aau5516

Lauss, M., Donia, M., Harbst, K., Andersen, R., Mitra, S., Rosengren, F., Salim, M., Vallon-Christersson, J., Törngren, T., Kvist, A., Ringnér, M., Svane, I. M., & Jönsson, G. (2017). Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. *Nature Communications*, *8*(1), 1–10. https://doi.org/10.1038/s41467-017-01460-0

Lebrun, L., Milowich, D., Le Mercier, M., Allard, J., Van Eycke, Y.-R., Roumeguere, T., Decaestecker, C., Salmon, I., & Rorive, S. (2018). UCA1 overexpression is associated with

less aggressive subtypes of bladder cancer. *Oncology Reports*, *40*(5), 2497–2506. https://doi.org/10.3892/or.2018.6697

Lee, S., Hwang, Y., Kim, T. H., Jeong, J., Choi, D., & Hwang, J. (2022). UPF1 Inhibits Hepatocellular Carcinoma Growth through DUSP1/p53 Signal Pathway. *Biomedicines*, *10*(4), 4. https://doi.org/10.3390/biomedicines10040793

Lee, S.-Y., Park, J.-H., Jeong, S., Kim, B.-Y., Kang, Y.-K., Xu, Y., & Chung, S.-K. (2019). K120R mutation inactivates p53 by creating an aberrant splice site leading to nonsense-mediated mRNA decay. *Oncogene*, *38*(10), 1597–1610. https://doi.org/10.1038/s41388-018-0542-3

Lee, W. S., Yang, H., Chon, H. J., & Kim, C. (2020). Combination of anti-angiogenic therapy and immune checkpoint blockade normalizes vascular-immune crosstalk to potentiate cancer immunity. *Experimental & Molecular Medicine*, *52*(9), 1475–1485. https://doi.org/10.1038/s12276-020-00500-y

Leeds, P., Peltz, S. W., Jacobson, A., & Culbertson, M. R. (1991). The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes & Development*, *5*(12a), 2303–2314. https://doi.org/10.1101/gad.5.12a.2303

Lejeune, F. (2022). Nonsense-Mediated mRNA Decay, a Finely Regulated Mechanism. *Biomedicines*, *10*(1), 141. https://doi.org/10.3390/biomedicines10010141

Levine, A. J. (2020). P53 and The Immune Response: 40 Years of Exploration—A Plan for the Future. *International Journal of Molecular Sciences*, *21*(2), 541. https://doi.org/10.3390/ijms21020541

Li, M., Johnson, J. R., Truong, B., Kim, G., Weinbren, N., Dittmar, M., Shah, P. S., Von Dollen, J., Newton, B. W., Jang, G. M., Krogan, N. J., Cherry, S., & Ramage, H. (2019). Identification of antiviral roles for the exon–junction complex and nonsense-mediated decay in flaviviral infection. *Nature Microbiology*. https://doi.org/10.1038/s41564-019-0375-z

Littman, D., & Hexner, E. (2017). Cancer Immunotherapy with Chimeric Antigen Receptor (CAR) T Cells. *Journal of Onco-Nephrology*, *1*(3), 151–155. https://doi.org/10.5301/jo-n.5000035

Liu, D.-T., Peng-Zhao, null, Han, J.-Y., Lin, F.-Z., Bu, X.-M., & Xu, Q.-X. (2014). Clinical and prognostic significance of SOX11 in breast cancer. *Asian Pacific Journal of Cancer Prevention: APJCP*, *15*(13), 5483–5486. https://doi.org/10.7314/apjcp.2014.15.13.5483

Liu, J., Yu, Z., Sun, M., Liu, Q., Wei, M., & Gao, H. (2019). Identification of cancer/testis antigen 2 gene as a potential hepatocellular carcinoma therapeutic target by hub gene screening with topological analysis. *Oncology Letters*, *18*(5), 4778–4788. https://doi.org/10.3892/ol.2019.10811

Liu, M., & Guo, F. (2018). Recent updates on cancer immunotherapy. *Precision Clinical Medicine*, *1*(July), 65–74. https://doi.org/10.1093/pcmedi/pby011

Liu, T., Han, C., Wang, S., Fang, P., Ma, Z., Xu, L., & Yin, R. (2019). Cancer-associated fibroblasts: An emerging target of anti-cancer immunotherapy. *Journal of Hematology & Oncology*, *12*(1), 86. https://doi.org/10.1186/s13045-019-0770-1

Losson, R., & Lacroute, F. (1979). Interference of nonsense mutations with eukaryotic messenger RNA stability. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(10), 5134–5137.

Loughran, G., Chou, M.-Y., Ivanov, I. P., Jungreis, I., Kellis, M., Kiran, A. M., Baranov, P. V., & Atkins, J. F. (2014). Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids Research*, *42*(14), 8928–8938. https://doi.org/10.1093/nar/gku608

Lu, J., Plank, T.-D., Su, F., Shi, X., Liu, C., Ji, Y., Li, S., Huynh, A., Shi, C., Zhu, B., Yang, G., Wu, Y., Wilkinson, M. F., & Lu, Y. (2016). The nonsense-mediated RNA decay pathway is disrupted in inflammatory myofibroblastic tumors. *The Journal of Clinical Investigation*, *126*(8), 3058–3062. https://doi.org/10.1172/JCI86508

Lueck, J. D., Yoon, J. S., Perales-Puchalt, A., Mackey, A. L., Infield, D. T., Behlke, M. A., Pope, M. R., Weiner, D. B., Skach, W. R., McCray, P. B., & Ahern, C. A. (2019). Engineered transfer RNAs for suppression of premature termination codons. *Nature Communications*, *10*(1), 822. https://doi.org/10.1038/s41467-019-08329-4

Łuksza, M., Riaz, N., Makarov, V., Balachandran, V. P., Hellmann, M. D., Solovyov, A., Rizvi, N. A., Merghoub, T., Levine, A. J., Chan, T. A., Wolchok, J. D., & Greenbaum, B. D. (2017). A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature*, *551*(7681), 517–520. https://doi.org/10.1038/nature24473

Lykke-Andersen, S., & Jensen, T. H. (2015). Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes. In *Nature Reviews Molecular Cell Biology* (Vol. 16). Nature Publishing Group. https://doi.org/10.1038/nrm4063

Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., & Lajoie, G. (2003). PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, *17*(20), 2337–2342.

Ma, N. J., Hemez, C. F., Barber, K. W., Rinehart, J., & Isaacs, F. J. (2018). Organisms with alternative genetic codes resolve unassigned codons via mistranslation and ribosomal rescue. *ELife*, *7*, e34878. https://doi.org/10.7554/eLife.34878

Malaina, I., Gonzalez-Melero, L., Martínez, L., Salvador, A., Sanchez-Diez, A., Asumendi, A., Margareto, J., Carrasco-Pujante, J., Legarreta, L., García, M. A., Pérez-Pinilla, M. B., Izu, R., Martínez de la Fuente, I., Igartua, M., Alonso, S., Hernandez, R. M., & Boyano, M. D. (2023). Computational and Experimental Evaluation of the Immune Response of Neoantigens for Personalized Vaccine Design. *International Journal of Molecular Sciences*, *24*(10), 9024. https://doi.org/10.3390/ijms24109024

Malakar, P., Shilo, A., Mogilevsky, A., Stein, I., Pikarsky, E., Nevo, Y., Benyamini, H., Elgavish, S., Zong, X., Prasanth, K. V., & Karni, R. (2017). Long noncoding RNA MALAT1 promotes hepatocellular carcinoma development by SRSF1 up-regulation and mTOR activation. *Cancer Research*, *77*(5), 1155–1167. https://doi.org/10.1158/0008-5472.CAN-16-1508

Malik, V., Rodino-Klapac, L. R., Viollet, L., & Mendell, J. R. (2010). Aminoglycoside-induced mutation suppression (stop codon readthrough) as a therapeutic strategy for Duchenne muscular dystrophy. *Therapeutic Advances in Neurological Disorders*, *3*(6), 379–389. https://doi.org/10.1177/1756285610388693

Mangkalaphiban, K., He, F., Ganesan, R., Wu, C., Baker, R., & Jacobson, A. (2021). Transcriptome-wide investigation of stop codon readthrough in Saccharomyces cerevisiae. *PLoS Genetics*, *17*(4), e1009538. https://doi.org/10.1371/journal.pgen.1009538

Marei, H. E., Althani, A., Afifi, N., Hasan, A., Caceci, T., Pozzoli, G., Morrione, A., Giordano, A., & Cenciarelli, C. (2021). P53 signaling in cancer progression and therapy. *Cancer Cell International*, *21*, 703. https://doi.org/10.1186/s12935-021-02396-8

Martin, L., Grigoryan, A., Wang, D., Wang, J., Breda, L., Rivella, S., Cardozo, T., & Gardner, L. B. (2014). Identification and characterization of small molecules that inhibit nonsense-mediated rna decay and suppress nonsense p53 mutations. *Cancer Research*, *74*(11), 3104–3113. https://doi.org/10.1158/0008-5472.CAN-13-2235

McCubrey, J. A., Abrams, S. L., Steelman, L. S., Cocco, L., Ratti, S., Martelli, A. M., Lombardi, P., Gizak, A., & Duda, P. (2022). APR-246—The Mutant TP53 Reactivator—Increases the Effectiveness of Berberine and Modified Berberines to Inhibit the Proliferation of Pancreatic Cancer Cells. *Biomolecules*, *12*(2), 276. https://doi.org/10.3390/biom12020276

McDonald, C. M., Campbell, C., Torricelli, R. E., Finkel, R. S., Flanigan, K. M., Goemans, N., Heydemann, P., Kaminska, A., Kirschner, J., Muntoni, F., Osorio, A. N., Schara, U., Sejersen, T., Shieh, P. B., Sweeney, H. L., Topaloglu, H., Tulinius, M., Vilchez, J. J., Voit, T., … ACT DMD Study Group. (2017). Ataluren in patients with nonsense mutation Duchenne muscular dystrophy (ACT DMD): A multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet (London, England)*, *390*(10101), 1489–1498. https://doi.org/10.1016/S0140-6736(17)31611-2

Midgley, C. A., Fisher, C. J., Bártek, J., Vojtěsek, B., Lane, D., & Barnes, D. M. (1992). Analysis of p53 expression in human tumours: An antibody raised against human p53 expressed in Escherichia coli. *Journal of Cell Science*, *101 (Pt 1)*, 183–189. https://doi.org/10.1242/jcs.101.1.183

Mingeot-Leclercq, M.-P., & Tulkens, P. M. (1999). Aminoglycosides: Nephrotoxicity. *Antimicrobial Agents and Chemotherapy*, *43*(5), 1003–1012.

Moldoveanu, D., Ramsay, L., Lajoie, M., Anderson-Trocme, L., Lingrand, M., Berry, D., Perus, L. J. M., Wei, Y., Moraes, C., Alkallas, R., Rajkumar, S., Zuo, D., Dankner, M., Xu, E. H., Bertos, N. R., Najafabadi, H. S., Gravel, S., Costantino, S., Richer, M. J., … Watson, I. R. (n.d.). Spatially mapping the immune landscape of melanoma using imaging mass cytometry. *Science Immunology*, *7*(70), eabi5072. https://doi.org/10.1126/sciimmunol.abi5072

Morisaki, T., Kubo, M., Umebayashi, M., Yew, P. Y., Yoshimura, S., Park, J.-H., Kiyotani, K., Kai, M., Yamada, M., Oda, Y., Nakamura, Y., Morisaki, T., & Nakamura, M. (2021). Neoantigens elicit T cell responses in breast cancer. *Scientific Reports*, *11*(1), 13590. https://doi.org/10.1038/s41598-021-91358-1

Mort, M., Ivanov, D., Cooper, D. N., & Chuzhanova, N. A. (2008). A meta-analysis of nonsense mutations causing human genetic disease. *Human Mutation*, *29*(8), 1037–1047. https://doi.org/10.1002/humu.20763

Muhlrad, D., & Parker, R. (1994). Premature translational termination triggers mRNA decapping. *Nature*, *370*(6490), 578–581. https://doi.org/10.1038/370578a0

Muramatsu, T., Heckmann, K., Kitanaka, C., & Kuchino, Y. (2001). Molecular mechanism of stop codon recognition by eRF1: A wobble hypothesis for peptide anticodons. *FEBS Letters*, *488*(3), 105–109. https://doi.org/10.1016/s0014-5793(00)02391-7

Naderi, A., Teschendorff, A. E., Beigel, J., Cariati, M., Ellis, I. O., Brenton, J. D., & Caldas, C. (2007). BEX2 is overexpressed in a subset of primary breast cancers and mediates nerve

growth factor/nuclear factor-kappaB inhibition of apoptosis in breast cancer cell lines. *Cancer Research*, *67*(14), 6725–6736. https://doi.org/10.1158/0008-5472.CAN-06-4394

Nakano, K., Karasawa, N., Hashizume, M., Tanaka, Y., Ohsugi, T., Uchimaru, K., & Watanabe, T. (2022). Elucidation of the Mechanism of Host NMD Suppression by HTLV-1 Rex: Dissection of Rex to Identify the NMD Inhibitory Domain. *Viruses*, *14*(2), 344. https://doi.org/10.3390/v14020344

Namgoong, J. H., & Bertoni, C. (2016). Clinical potential of ataluren in the treatment of Duchenne muscular dystrophy. *Degenerative Neurological and Neuromuscular Disease*, *6*, 37–48. https://doi.org/10.2147/DNND.S71808

Nasif, S., Contu, L., & Mühlemann, O. (2018). Beyond quality control: The role of nonsense-mediated mRNA decay (NMD) in regulating gene expression. *Seminars in Cell & Developmental Biology*, *75*, 78–87. https://doi.org/10.1016/j.semcdb.2017.08.053

Nau, H., & Biemann, K. (1976). Amino acid sequencing by gas chromatography—Mass spectrometry using perfluoro-dideuteroalkylated peptide derivatives. A. Gas chromatographic retention indices. *Analytical Biochemistry*, *73*(1), 139–153. https://doi.org/10.1016/0003-2697(76)90150-0

Nielsen, M., Lundegaard, C., Lund, O., & Keşmir, C. (2005). The role of the proteasome in generating cytotoxic T-cell epitopes: Insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*, *57*(1–2), 33–41. https://doi.org/10.1007/s00251-005-0781-7

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. *Annual Review of Biochemistry*, *79*, 321–349. https://doi.org/10.1146/annurev-biochem-060208-105251

O'Donnell, T. J., Rubinsteyn, A., Bonsack, M., Riemer, A. B., Laserson, U., & Hammerbacher, J. (2018). MHCflurry: Open-Source Class I MHC Binding Affinity Prediction. *Cell Systems*, *7*(1), 129-132.e4. https://doi.org/10.1016/j.cels.2018.05.014

Oka, M., Xu, L., Suzuki, T., Yoshikawa, T., Sakamoto, H., Uemura, H., Yoshizawa, A. C., Suzuki, Y., Nakatsura, T., Ishihama, Y., Suzuki, A., & Seki, M. (2021). Aberrant splicing isoforms detected by full-length transcriptome sequencing as transcripts of potential neoantigens in non-small cell lung cancer. *Genome Biology*, *22*(1), 9. https://doi.org/10.1186/s13059-020-02240-8

Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F., & Ohno, S. (2012). N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and

SMG-5:SMG-7 during NMD. *Nucleic Acids Research*, *40*(3), 1251–1266. https://doi.org/10.1093/nar/gkr791

O'Leary, P. C., Chen, H., Doruk, Y. U., Williamson, T., Polacco, B., McNeal, A. S., Shenoy, T., Kale, N., Carnevale, J., Stevenson, E., Quigley, D. A., Chou, J., Feng, F. Y., Swaney, D. L., Krogan, N. J., Kim, M., Diolaiti, M. E., & Ashworth, A. (2022). Resistance to ATR Inhibitors Is Mediated by Loss of the Nonsense-Mediated Decay Factor UPF2. *Cancer Research*, OF1–OF12. https://doi.org/10.1158/0008-5472.CAN-21-4335

O'Neill, R., Mohtar, A., Gomez-Herranz, M., Singh, A., Mayordomo, M. Y., Alfaro, J., Pilch, M., & Hupp, T. (2018). Interferon responsive genes that link adaptive and innate immunity in human cancer. *Acta Biochimica Polonica*, *65*(S2).

Oshi, M., Asaoka, M., Tokumaru, Y., Angarita, F. A., Yan, L., Matsuyama, R., Zsiros, E., Ishikawa, T., Endo, I., & Takabe, K. (2020). Abundance of Regulatory T Cell (Treg) as a Predictive Biomarker for Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *Cancers*, *12*(10), 3038. https://doi.org/10.3390/cancers12103038

Ott, P. A., Hu, Z., Keskin, D. B., Shukla, S. A., Sun, J., Bozym, D. J., Zhang, W., Luoma, A., Giobbie-Hurder, A., Peter, L., Chen, C., Olive, O., Carter, T. A., Li, S., Lieb, D. J., Eisenhaure, T., Gjini, E., Stevens, J., Lane 10, W. J., … Wu, C. J. (2017). An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature Publishing Group*, *547*. https://doi.org/10.1038/nature22991

Pak, H., Michaux, J., Huber, F., Chong, C., Stevenson, B. J., Müller, M., Coukos, G., & Bassani-Sternberg, M. (2021). Sensitive Immunopeptidomics by Leveraging Available Large-Scale Multi-HLA Spectral Libraries, Data-Independent Acquisition, and MS/MS Prediction. *Molecular & Cellular Proteomics: MCP*, *20*, 100080. https://doi.org/10.1016/j.mcpro.2021.100080

Palma, M., & Lejeune, F. (2021). Deciphering the molecular mechanism of stop codon readthrough. *Biological Reviews*, *96*(1), 310–329. https://doi.org/10.1111/brv.12657

Park, J. S. W., Kim, J. J. H., Reagan, J. L., Fast, L. D., Kinkead, H. L., Jaffee, E. M., Zaidi, N., Kinkead, H. L., Hopkins, A., Lutz, E., Wu, A. A., Yarchoan, M., Cruz, K., Woolman, S., Vithayathil, T., Glickman, L. H., Ndubaku, C. O., Shigemizu, D., Momozawa, Y., … Felip, E. (2018). The nonsense-mediated mRNA decay (NMD) pathway differentially regulates COX17, COX19 and COX23 mRNAs. *Cancer Research*, *8*(1), 0. https://doi.org/10.1101/gr.119974.110

Paston, S. J., Brentville, V. A., Symonds, P., & Durrant, L. G. (2021). Cancer Vaccines, Adjuvants, and Delivery Systems. *Frontiers in Immunology*, *12*. https://www.frontiersin.org/articles/10.3389/fimmu.2021.627932

Pastor, F., Kolonias, D., Giangrande, P. H., & Gilboa, E. (2010). *Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay*. https://doi.org/10.1038/nature08999

Pawlicka, K., Kalathiya, U., & Alfaro, J. (2020). Nonsense-Mediated mRNA Decay: Pathologies and the Potential for Novel Therapeutics. *Cancers*, *12*(3), 3. https://doi.org/10.3390/cancers12030765

Peltz, S. W., Brown, A. H., & Jacobson, A. (1993). MRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. *Genes & Development*, *7*(9), 1737–1754. https://doi.org/10.1101/gad.7.9.1737

Perrin-Vidoz, L., Sinilnikova, O. M., Stoppa-Lyonnet, D., Lenoir, G. M., & Mazoyer, S. (2002). The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons. *Human Molecular Genetics*, *11*(23), 2805–2814. https://doi.org/10.1093/hmg/11.23.2805

Peters, B., Bulik, S., Tampe, R., Van Endert, P. M., & Holzhütter, H.-G. (2003). Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *Journal of Immunology (Baltimore, Md.: 1950)*, *171*(4), 1741–1749. https://doi.org/10.4049/jimmunol.171.4.1741

Pfister, N. T., & Prives, C. (2017). Transcriptional Regulation by Wild-Type and Cancer-Related Mutant Forms of p53. *Cold Spring Harbor Perspectives in Medicine*, *7*(2), a026054. https://doi.org/10.1101/cshperspect.a026054

Phan, G. Q., Yang, J. C., Sherry, R. M., Hwu, P., Topalian, S. L., Schwartzentruber, D. J., Restifo, N. P., Haworth, L. R., Seipp, C. A., Freezer, L. J., Morton, K. E., Mavroukakis, S. A., Duray, P. H., Steinberg, S. M., Allison, J. P., Davis, T. A., & Rosenberg, S. A. (2003). Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(14), 8372–8377. https://doi.org/10.1073/pnas.1533209100

Pierson, A., Darrigrand, R., Rouillon, M., Boulpicante, M., Renko, Z. D., Garcia, C., Ghosh, M., Laiguillon, M.-C., Lobry, C., Alami, M., & Apcher, S. (2019). *Splicing inhibition enhances*

the antitumor immune response through increased tumor antigen presentation and altered MHC-I immunopeptidome [Preprint]. Immunology. https://doi.org/10.1101/512681

Popp, M. W., & Maquat, L. E. (2018). Nonsense-mediated mRNA Decay and Cancer. *Current Opinion in Genetics & Development*, *48*, 44–50. https://doi.org/10.1016/j.gde.2017.10.007

Pulak, R., & Anderson, P. (n.d.). *MRNA surveillance by the Caenorhabditis elegans stag genes*. 14.

Rasmussen, M., Fenoy, E., Harndahl, M., Kristensen, A. B., Nielsen, I. K., Nielsen, M., & Buus, S. (2016). Pan-specific prediction of peptide–MHC class I complex stability, a correlate of T cell immunogenicity. *The Journal of Immunology*, *197*(4), 1517–1524.

Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W. H., Peters, B., & Nielsen, M. (2020). Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *Journal of Proteome Research*.

Rezaei, T., Davoudian, E., Khalili, S., Amini, M., Hejazi, M., Guardia, M. de la, & Mokhtarzadeh, A. (n.d.). Strategies in DNA vaccine for melanoma cancer. *Pigment Cell & Melanoma Research*, *n/a*(n/a). https://doi.org/10.1111/pcmr.12933

Rizvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J. J., Lee, W., Yuan, J., Wong, P., Ho, T. S., Miller, M. L., Rekhtman, N., Moreira, A. L., Ibrahim, F., Bruggeman, C., Gasmi, B., Zappasodi, R., Maeda, Y., Sander, C., … Chan, T. A. (2015). Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science (New York, N.Y.)*, *348*(6230), 124–128. https://doi.org/10.1126/science.aaa1348

Rosenberg, S. A., Tong-On, P., Li, Y., Riley, J. P., El-Gamil, M., Parkhurst, M. R., & Robbins, P. F. (2002). Identification of BING-4 Cancer Antigen Translated From an Alternative Open Reading Frame of a Gene in the Extended MHC Class II Region Using Lymphocytes From a Patient With a Durable Complete Regression Following Immunotherapy. *Journal of Immunology (Baltimore, Md. : 1950)*, *168*(5), 2402–2407.

Rosenthal, R., Cadieux, E. L., Salgado, R., Bakir, M. A., Moore, D. A., Hiley, C. T., Lund, T., Tanić, M., Reading, J. L., Joshi, K., Henry, J. Y., Ghorani, E., Wilson, G. A., Birkbak, N. J., Jamal-Hanjani, M., Veeriah, S., Szallasi, Z., Loi, S., Hellmann, M. D., … TRACERx consortium. (2019). Neoantigen-directed immune escape in lung cancer evolution. *Nature*, *567*(7749), 479–485. https://doi.org/10.1038/s41586-019-1032-7

Roudko, V., Greenbaum, B., & Bhardwaj, N. (2020). Computational Prediction and Validation of Tumor-Associated Neoantigens. *Frontiers in Immunology*, *11*, 27. https://doi.org/10.3389/fimmu.2020.00027

Roy, B., Leszyk, J. D., Mangus, D. A., & Jacobson, A. (2015). Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), 3038–3043. https://doi.org/10.1073/pnas.1424127112

S K, J., S P, D., R, S., Sai Surya, N. U., & Chenmala, K. (2021). Guardian of genome on the tract: Wild type p53-mdm2 complex inhibition in healing the breast cancer. *Gene*, *786*, 145616. https://doi.org/10.1016/j.gene.2021.145616

Saeinasab, M., Bahrami, A. R., González, J., Marchese, F. P., Martinez, D., Mowla, S. J., Matin, M. M., & Huarte, M. (2019). SNHG15 is a bifunctional MYC-regulated noncoding locus encoding a lncRNA that promotes cell proliferation, invasion and drug resistance in colorectal cancer by interacting with AIF. *Journal of Experimental & Clinical Cancer Research*, *38*(1), 172. https://doi.org/10.1186/s13046-019-1169-0

Sahin, U., Derhovanessian, E., Miller, M., Kloke, B.-P., Simon, P., Löwer, M., Bukur, V., Tadmor, A. D., Luxemburger, U., Schrörs, B., Omokoko, T., Vormehr, M., Albrecht, C., Paruzynski, A., Kuhn, A. N., Buck, J., Heesch, S., Schreeb, K. H., Müller, F., … Türeci, Ö. (2017). Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature*, *547*(7662), 7662. https://doi.org/10.1038/nature23003

Sarkar, A., Panati, K., & Narala, V. R. (2022). Code inside the codon: The role of synonymous mutations in regulating splicing machinery and its impact on disease. *Mutation Research/Reviews in Mutation Research*, *790*, 108444. https://doi.org/10.1016/j.mrrev.2022.108444

Sato, H., & Singer, R. H. (2021). Cellular variability of nonsense-mediated mRNA decay. *Nature Communications*, *12*, 7203. https://doi.org/10.1038/s41467-021-27423-0

Schilff, M., Sargsyan, Y., Hofhuis, J., & Thoms, S. (2021). Stop Codon Context-Specific Induction of Translational Readthrough. *Biomolecules*, *11*(7), 1006. https://doi.org/10.3390/biom11071006

Schlautmann, L. P., & Gehring, N. H. (2020). A Day in the Life of the Exon Junction Complex. *Biomolecules*, *10*(6), 6. https://doi.org/10.3390/biom10060866

Schueren, F., Lingner, T., George, R., Hofhuis, J., Dickel, C., Gärtner, J., & Thoms, S. (2014). Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. *ELife*, *3*, e03640. https://doi.org/10.7554/eLife.03640

Schumacher, T., Bunse, L., Pusch, S., Sahm, F., Wiestler, B., Quandt, J., Menn, O., Osswald, M., Oezen, I., Ott, M., Keil, M., Balß, J., Rauschenbach, K., Grabowska, A. K., Vogler, I., Diekmann, J., Trautwein, N., Eichmüller, S. B., Okun, J., … Platten, M. (2014). A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature*, *512*(7514), 324–327. https://doi.org/10.1038/nature13387

Shunji Sugawara, Toru Abo, & Katsuo Kumagai. (1987). A simple method to eliminate the antigenicity of surface class I MHC molecules from the membrane of viable cells by acid treatment at pH 3. *Journal of Immunological Methods*, *100*(1), 83–90. https://doi.org/10.1016/0022-1759(87)90175-X

Smart, A. C., Margolis, C. A., Pimentel, H., He, M. X., Miao, D., Adeegbe, D., Fugmann, T., Wong, K.-K., & Van Allen, E. M. (2018). Intron retention is a source of neoepitopes in cancer. *Nature Biotechnology*, *36*(11), 1056–1058. https://doi.org/10.1038/nbt.4239

Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J. M., Desrichard, A., Walsh, L. A., Postow, M. A., Wong, P., Ho, T. S., Hollmann, T. J., Bruggeman, C., Kannan, K., Li, Y., Elipenahli, C., Liu, C., Harbison, C. T., Wang, L., Ribas, A., … Chan, T. A. (2014). Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *New England Journal of Medicine*, *371*(23), 2189–2199. https://doi.org/10.1056/NEJMoa1406498

Sturm, T., Sautter, B., Wörner, T. P., Stevanović, S., Rammensee, H.-G., Planz, O., Heck, A. J. R., & Aebersold, R. (2021). Mild Acid Elution and MHC Immunoaffinity Chromatography Reveal Similar Albeit Not Identical Profiles of the HLA Class I Immunopeptidome. *Journal of Proteome Research*, *20*(1), 289–304. https://doi.org/10.1021/acs.jproteome.0c00386

Svane, I. M., & Verdegaal, E. M. (2014). Achievements and challenges of adoptive T cell therapy with tumor-infiltrating or blood-derived lymphocytes for metastatic melanoma: What is needed to achieve standard of care? *Cancer Immunology, Immunotherapy: CII*, *63*(10), 1081–1091. https://doi.org/10.1007/s00262-014-1580-5

Tang, S. J., Shen, H., An, O., Hong, H., Li, J., Song, Y., Han, J., Tay, D. J. T., Ng, V. H. E., Bellido Molias, F., Leong, K. W., Pitcheshwar, P., Yang, H., & Chen, L. (2020). Cis- and trans-regulations of pre-mRNA splicing by RNA editing enzymes influence cancer development. *Nature Communications*, *11*(1), 1. https://doi.org/10.1038/s41467-020-14621-5

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., Powderly, J. D., Carvajal, R. D., Sosman, J. A., Atkins, M. B., Leming, P. D., Spigel, D. R., Antonia, S. J., Horn, L., Drake, C. G., Pardoll, D. M., Chen, L., Sharfman, W. H., Anders, R. A., … Sznol, M. (2012). Safety, Activity, and Immune Correlates of Anti–PD-1 Antibody in Cancer. *New England Journal of Medicine*, *366*(26), 2443–2454. https://doi.org/10.1056/NEJMoa1200690

Tsalikis, J., Abdel-Nour, M., Farahvash, A., Sorbara, M. T., Poon, S., Philpott, D. J., & Girardin, S. E. (2019). Isoginkgetin, a Natural Biflavonoid Proteasome Inhibitor, Sensitizes Cancer Cells to Apoptosis via Disruption of Lysosomal Homeostasis and Impaired Protein Clearance. *Molecular and Cellular Biology*, *39*(10). https://doi.org/10.1128/mcb.00489-18

Van Allen, E. M., Miao, D., Schilling, B., Shukla, S. A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Geukes Foppen, M. H., Goldinger, S. M., Utikal, J., Hassel, J. C., Weide, B., Kaehler, K. C., Loquai, C., Mohr, P., Gutzmer, R., Dummer, R., Gabriel, S., … Garraway, L. A. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*, *350*(6257), 207–211. https://doi.org/10.1126/science.aad0095

Vansteenkiste, J. F., Cho, B. C., Vanakesa, T., De Pas, T., Zielinski, M., Kim, M. S., Jassem, J., Yoshimura, M., Dahabreh, J., Nakayama, H., Havel, L., Kondo, H., Mitsudomi, T., Zarogoulidis, K., Gladkov, O. A., Udud, K., Tada, H., Hoffman, H., Bugge, A., … Altorki, N. (2016). Efficacy of the MAGE-A3 cancer immunotherapeutic as adjuvant therapy in patients with resected MAGE-A3-positive non-small-cell lung cancer (MAGRIT): A randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet Oncology*, *17*(6), 822–835. https://doi.org/10.1016/S1470-2045(16)00099-1

Verbruggen, S., Ndah, E., Van Criekinge, W., Gessulat, S., Kuster, B., Wilhelm, M., Van Damme, P., & Menschaert, G. (2019). PROTEOFORMER 2.0: Further Developments in the Ribosome Profiling-assisted Proteogenomic Hunt for New Proteoforms. *Molecular & Cellular Proteomics: MCP*, *18*(8 suppl 1), S126–S140. https://doi.org/10.1074/mcp.RA118.001218

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., & Kinzler, K. W. (2013). Cancer Genome Landscapes. *Science (New York, N.Y.)*, *339*(6127), 1546–1558. https://doi.org/10.1126/science.1235122

Vyasamneni, R., Kohler, V., Karki, B., Mahimkar, G., Esaulova, E., McGee, J., Kallin, D., Sheen, J. H., Harjanto, D., Kirsch, M., Poran, A., Dong, J., Srinivasan, L., Gaynor, R. B., Bushway, M. E., & Srouji, J. R. (2023). A universal MHCII technology platform to characterize antigen-specific CD4+ T cells. *Cell Reports Methods*, *3*(1), 100388. https://doi.org/10.1016/j.crmeth.2022.100388

Wang, L., Geng, H., Liu, Y., Liu, L., Chen, Y., Wu, F., Liu, Z., Ling, S., Wang, Y., & Zhou, L. (2023). Hot and cold tumors: Immunological features and the therapeutic strategies. *MedComm*, *4*(5), e343. https://doi.org/10.1002/mco2.343

Wang, R. F., & Wang, H. Y. (2017). Immune targets and neoantigens for cancer immunotherapy and precision medicine. *Cell Research*, *27*(1), 11–37. https://doi.org/10.1038/cr.2016.155

Wangen, J. R., & Green, R. (2020). Stop codon context influences genome-wide stimulation of termination codon readthrough by aminoglycosides. *ELife*, *9*, e52611. https://doi.org/10.7554/eLife.52611

Warburg, O. (1956). On the origin of cancer cells. *Science (New York, N.Y.)*, *123*(3191), 309–314. https://doi.org/10.1126/science.123.3191.309

Wei, H., Qu, L., Dai, S., Li, Y., Wang, H., Feng, Y., Chen, X., Jiang, L., Guo, M., Li, J., Chen, Z., Chen, L., Zhang, Y., & Chen, Y. (2021). Structural insight into the molecular mechanism of p53-mediated mitochondrial apoptosis. *Nature Communications*, *12*(1), 2280. https://doi.org/10.1038/s41467-021-22655-6

Wei, J., Kishton, R. J., Angel, M., Conn, C. S., Dalla-Venezia, N., Marcel, V., Vincent, A., Catez, F., Ferré, S., Ayadi, L., Marchand, V., Dersh, D., Gibbs, J. S., Ivanov, I. P., Fridlyand, N., Couté, Y., Diaz, J.-J., Qian, S.-B., Staudt, L. M., … Yewdell, J. W. (2019). Ribosomal Proteins Regulate MHC Class I Peptide Generation for Immunosurveillance. *Molecular Cell*, *73*(6), 1162-1173.e5. https://doi.org/10.1016/j.molcel.2018.12.020

Weischenfeldt, J., Damgaard, I., Bryder, D., Theilgaard-Monch, K., Thoren, L. A., Nielsen, F. C., Jacobsen, S. E. W., Nerlov, C., & Porse, B. T. (2008). NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes & Development*, *22*(10), 1381–1396. https://doi.org/10.1101/gad.468808

Wen, B., Xu, S., Sheynkman, G. M., Feng, Q., Lin, L., Wang, Q., Xu, X., Wang, J., & Liu, S. (2014). sapFinder: An R/Bioconductor package for detection of variant peptides in shotgun proteomics experiments. *Bioinformatics (Oxford, England)*, *30*(21), 3136–3138. https://doi.org/10.1093/bioinformatics/btu397

Wicks, E. E., & Semenza, G. L. (n.d.). Hypoxia-inducible factors: Cancer progression and clinical translation. *The Journal of Clinical Investigation*, *132*(11), e159839. https://doi.org/10.1172/JCI159839

Wu, C.-C., Beird, H. C., Andrew Livingston, J., Advani, S., Mitra, A., Cao, S., Reuben, A., Ingram, D., Wang, W.-L., Ju, Z., Hong Leung, C., Lin, H., Zheng, Y., Roszik, J., Wang, W., Patel, S., Benjamin, R. S., Somaiah, N., Conley, A. P., … Futreal, P. A. (2020). Immuno-genomic landscape of osteosarcoma. *Nature Communications*, *11*(1), 1. https://doi.org/10.1038/s41467-020-14646-w

XCaroline Jansen, X. S., XPhD, X., XNataliya Prokhnevska, X., & XHaydn Kissick, X. T. (2018). The requirement for immune infiltration and organization in the tumor microenvironment for successful immunotherapy in prostate cancer. *Urologic Oncology: Seminars and Original Investigations*. https://doi.org/10.1016/j.urolonc.2018.10.011

Xiang, R., Ma, L., Yang, M., Zheng, Z., Chen, X., Jia, F., Xie, F., Zhou, Y., Li, F., Wu, K., & Zhu, Y. (2021). Increased expression of peptides from non-coding genes in cancer proteomics datasets suggests potential tumor neoantigens. *Communications Biology*, *4*(1), 496. https://doi.org/10.1038/s42003-021-02007-2

Xie, Y., Xie, F., Zhang, L., Zhou, X., Huang, J., Wang, F., Jin, J., Zhang, L., Zeng, L., & Zhou, F. (2021). Targeted Anti-Tumor Immunotherapy Using Tumor Infiltrating Cells. *Advanced Science*, *8*. https://doi.org/10.1002/advs.202101672

Yadav, M., Jhunjhunwala, S., Phung, Q. T., Lupardus, P., Tanguay, J., Bumbaca, S., Franci, C., Cheung, T. K., Fritsche, J., Weinschenk, T., Modrusan, Z., Mellman, I., Lill, J. R., & Delamarre, L. (2014). Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature*, *515*(7528), 7528. https://doi.org/10.1038/nature14001

Yamashita, A., Izumi, N., Kashima, I., Ohnishi, T., Saari, B., Katsuhata, Y., Muramatsu, R., Morita, T., Iwamatsu, A., Hachiya, T., Kurata, R., Hirano, H., Anderson, P., & Ohno, S. (2009). SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. *Genes and Development*, *23*(9), 1091–1105. https://doi.org/10.1101/gad.1767209

Yang, Y., & Cao, Y. (2022). The impact of VEGF on cancer metastasis and systemic disease. *Seminars in Cancer Biology*, *86*, 251–261. https://doi.org/10.1016/j.semcancer.2022.03.011

Yewdell, J. W., Antón, L. C., & Bennink, J. R. (1996). Defective ribosomal products (DRiPs): A major source of antigenic peptides for MHC class I molecules? *Journal of Immunology (Baltimore, Md.: 1950)*, *157*(5), 1823–1826.

Zavileyskiy, L., & Bunik, V. (2022). Regulation of p53 Function by Formation of Non-Nuclear Heterologous Protein Complexes. *Biomolecules*, *12*(2), 327. https://doi.org/10.3390/biom12020327

Zhang, H., Liu, T., Zhang, Z., Payne, S. H., Zhang, B., McDermott, J. E., Zhou, J.-Y., Petyuk, V. A., Chen, L., Ray, D., Sun, S., Yang, F., Chen, L., Wang, J., Shah, P., Cha, S. W., Aiyetan, P., Woo, S., Tian, Y., … Townsend, R. R. (2016). Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell*, *166*(3), 755–765. https://doi.org/10.1016/j.cell.2016.05.069

Zhang, H., Lyu, Z., Fan, Y., Evans, C. R., Barber, K. W., Banerjee, K., Igoshin, O. A., Rinehart, J., & Ling, J. (2020). Metabolic stress promotes stop-codon readthrough and phenotypic heterogeneity. *Proceedings of the National Academy of Sciences*, *117*(36), 22167–22172. https://doi.org/10.1073/pnas.2013543117

Zhang, M., Fritsche, J., Roszik, J., Williams, L. J., Peng, X., Chiu, Y., Tsou, C.-C., Hoffgaard, F., Goldfinger, V., Schoor, O., Talukder, A., Forget, M. A., Haymaker, C., Bernatchez, C., Han, L., Tsang, Y.-H., Kong, K., Xu, X., Scott, K. L., … Hwu, P. (2018). RNA editing derived epitopes function as cancer antigens to elicit immune responses. *Nature Communications*, *9*(1), 3919. https://doi.org/10.1038/s41467-018-06405-9

Zhang, M., Heldin, A., Palomar-Siles, M., Öhlin, S., Bykov, V. J. N., & Wiman, K. G. (2018a). Synergistic Rescue of Nonsense Mutant Tumor Suppressor p53 by Combination Treatment with Aminoglycosides and Mdm2 Inhibitors. *Frontiers in Oncology*, *7*, 323. https://doi.org/10.3389/fonc.2017.00323

Zhang, M., Heldin, A., Palomar-Siles, M., Öhlin, S., Bykov, V. J. N., & Wiman, K. G. (2018b). Synergistic Rescue of Nonsense Mutant Tumor Suppressor p53 by Combination Treatment with Aminoglycosides and Mdm2 Inhibitors. *Frontiers in Oncology*, *7*. https://doi.org/10.3389/fonc.2017.00323

Zhang, Q., Bykov, V. J. N., Wiman, K. G., & Zawacka-Pankau, J. (2018). APR-246 reactivates mutant p53 by targeting cysteines 124 and 277. *Cell Death & Disease*, *9*(5), 439. https://doi.org/10.1038/s41419-018-0463-7

Zhang, X., Wu, J., Baeza, J., Gu, K., & Zhou, Z. (2023). *DeepTAP: An RNN-based method of TAP-binding peptide prediction in the selection of tumor neoantigens* [Preprint]. Bioinformatics. https://doi.org/10.1101/2023.02.13.528393

Zhao, Y., Ye, X., Shehata, M., Dunker, W., Xie, Z., & Karijolich, J. (2020). The RNA quality control pathway nonsense-mediated mRNA decay targets cellular and viral RNAs to restrict KSHV. *Nature Communications*, *11*(1), 3345. https://doi.org/10.1038/s41467-020-17151-2

Zhou, W., Xiang, W., Yu, J., Ruan, Z., Pan, Y., Wang, K., & Liu, J. (2023). *NeoTCR: An immunoinformatic database of experimentally-supported functional neoantigen-specific TCR sequences* [Preprint]. Bioinformatics. https://doi.org/10.1101/2023.02.13.528383

Zhu, H., Zhang, L., Yan, S., Li, W., Cui, J., Zhu, M., Xia, N., Yang, Y., Yuan, J., Chen, X., Luo, J., Chen, R., Xing, R., Lu, Y., & Wu, N. (2017). LncRNA16 is a potential biomarker for diagnosis of early-stage lung cancer that promotes cell proliferation by regulating the cell cycle. *Oncotarget*, *8*(5), 7867–7877. https://doi.org/10.18632/oncotarget.13980

Zickmann, F., & Renard, B. Y. (2015). MSProGene: Integrative proteogenomics beyond six-frames and single nucleotide polymorphisms. *Bioinformatics (Oxford, England)*, *31*(12), i106-115. https://doi.org/10.1093/bioinformatics/btv236

Ziegler, I., Ma, B., Nie, E., Bischl, B., Rügamer, D., Schubert, B., & Dorigatti, E. (2022). *What cleaves? Is proteasomal cleavage prediction reaching a ceiling?* (arXiv:2210.12991). arXiv. http://arxiv.org/abs/2210.12991