



OPEN ACCESS

EDITED BY

Hang Xu,
Genentech Inc., United States

REVIEWED BY

Susan Klaeger,
Genentech Inc., United States
Michael Volkmar,
German Cancer Research Center (DKFZ),
Germany
Ruping Sun,
University of Minnesota Twin Cities,
United States

*CORRESPONDENCE

Kevin Litchfield
✉ k.litchfield@ucl.ac.uk

RECEIVED 01 December 2023

ACCEPTED 19 February 2024

PUBLISHED 15 March 2024

CITATION

Sng CCT, Kallor AA, Simpson BS, Bedran G, Alfaro J and Litchfield K (2024) Untranslated regions (UTRs) are a potential novel source of neoantigens for personalised immunotherapy. *Front. Immunol.* 15:1347542. doi: 10.3389/fimmu.2024.1347542

COPYRIGHT

© 2024 Sng, Kallor, Simpson, Bedran, Alfaro and Litchfield. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Untranslated regions (UTRs) are a potential novel source of neoantigens for personalised immunotherapy

Christopher C. T. Sng¹, Ashwin Adrian Kallor², Benjamin S. Simpson¹, Georges Bedran², Javier Alfaro^{2,3,4} and Kevin Litchfield^{1*}

¹Cancer Research UK Lung Cancer Centre of Excellence, University College London (UCL) Cancer Institute, London, United Kingdom, ²International Center for Cancer Vaccine Science, University of Gdansk, Gdansk, Poland, ³Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada, ⁴Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, Edinburgh, United Kingdom

Background: Neoantigens, mutated tumour-specific antigens, are key targets of anti-tumour immunity during checkpoint inhibitor (CPI) treatment. Their identification is fundamental to designing neoantigen-directed therapy. Non-canonical neoantigens arising from the untranslated regions (UTR) of the genome are an overlooked source of immunogenic neoantigens. Here, we describe the landscape of UTR-derived neoantigens and release a computational tool, PrimeCUTR, to predict UTR neoantigens generated by start-gain and stop-loss mutations.

Methods: We applied PrimeCUTR to a whole genome sequencing dataset of pre-treatment tumour samples from CPI-treated patients (n = 341). Cancer immunopeptidomic datasets were interrogated to identify MHC class I presentation of UTR neoantigens.

Results: Start-gain neoantigens were predicted in 72.7% of patients, while stop-loss mutations were found in 19.3% of patients. While UTR neoantigens only accounted 2.6% of total predicted neoantigen burden, they contributed 12.4% of neoantigens with high dissimilarity to self-proteome. More start-gain neoantigens were found in CPI responders, but this relationship was not significant when correcting for tumour mutational burden. While most UTR neoantigens are private, we identified two recurrent start-gain mutations in melanoma. Using immunopeptidomic datasets, we identify two distinct MHC class I-presented UTR neoantigens: one from a recurrent start-gain mutation in melanoma, and one private to Jurkat cells.

Conclusion: PrimeCUTR is a novel tool which complements existing neoantigen discovery approaches and has potential to increase the detection yield of neoantigens in personalised therapeutics, particularly for neoantigens with high dissimilarity to self. Further studies are warranted to confirm the expression and immunogenicity of UTR neoantigens.

KEYWORDS

UTR, untranslated region, neoantigen, checkpoint inhibitor, personalised vaccine, PrimeCUTR, immunopeptidomics

1 Introduction

Neoantigens arise from mutated proteins which can be processed and expressed on the surface of cancer cells, forming key targets in anti-tumour immunity. The success of checkpoint inhibitor (CPI) immunotherapy, particularly in tumours with a high mutational burden (as a proxy of neoantigen load), has spurred interest in the identification of the underlying neoantigens (1, 2). Early studies found neoantigens predicted from somatic mutations could stimulate patient-derived CD8⁺ T-cells, and were associated with response to CPI treatment (3–6). Furthermore, small studies demonstrating the ability of neoantigen-specific T cells to induce tumour regression hinted at the promise of neoantigen-directed therapy (7–9). More recently, neoantigen vaccine trials have demonstrated vaccine-induced T cell expansion, and evidence of durable disease response in some patients (10–13). Thus, these cancer-specific antigens represent an important target in development of personalised immunotherapy.

Traditional approaches to neoantigen identification have typically involved sequencing the protein-coding regions of the cancer genome for missense or insertion-deletion mutations, followed by HLA-binding prediction and neoantigen prioritisation (14). This may yield hundreds to thousands of putative neoantigens, but only a small fraction appear to contribute to immune responses (6, 7, 15). In a combined effort, the Tumour Neoantigen Selection Alliance (TESLA) global consortium identified 608 top-ranked neoantigens in 6 solid cancer samples, of which only 37 (6%) could be recognised by matched patient T cells (16). Likewise, approaches to identify predicted neoantigens in cancer immunopeptidomes have had limited yield (6, 17). Additionally, certain tumour types such as neuroblastoma and pancreatic adenocarcinoma bear an inherently low mutational burden, reducing the pool of candidate neoantigens (18). Further obstacles to immune recognition include immune-exclusion, immunosuppressive microenvironment, variable gene expression, mRNA quality control pathways (e.g. nonsense-mediated decay), and intra-tumoural heterogeneity (14, 19). To date, no personalised neoantigen-directed therapies have emulated the clinical response rates or widespread regulatory approval of CPI treatment.

Given the attrition of candidate neoantigens through the discovery process, expanded neoantigen search strategies are essential to capture the breadth of neoantigens to direct therapeutic design. Various studies have demonstrated the presentation on MHC class I of non-canonical/cryptic peptides arising from ostensibly non-coding regions or alternative reading frames (20–23). The majority of these studies focus on non-mutated peptides which are not necessarily cancer-specific, increasing the likelihood self-tolerance. In this study, we present a novel R package, PrimeCUTR, which identifies candidate neopeptides in the 5' and 3' untranslated region (UTR) of genes generated by premature start-gain and stop-loss mutations respectively. Start-gain mutations create novel open reading frames (neoORFs) through the generation of novel upstream start-codons (uAUG) within the 5'UTR region of an mRNA transcript (Figure 1A). Meanwhile, stop-loss mutations which convert the stop codon into a sense codon, theoretically result in read-through of the 3'UTR following the canonical peptide sequence. We describe how these neoantigens contribute to the immune landscape of cancer. To our knowledge, this is the first publicly available tool to predict these UTR neoantigens.

2 Methods

2.1 Cohort

355 patients with metastatic cancer who received CPI treatment were selected from the Hartwig Medical Cohort for analysis of pre-treatment somatic tumour mutation calls in conjunction with clinical response data. Of these, 341 patients had HLA typing data available and were included in this study. This cohort consisted of patients with melanoma (n = 153), lung cancer (n = 69), bladder cancer (n = 58), renal cancer (n = 20) and other cancers (n = 41). Corresponding whole genome sequencing (WGS) somatic mutation data was obtained in Variant Call Format (VCF) via Hartwig Medical Foundation data access request (license agreement DR-087). These VCFs files were generated by the Hartwig Medical Foundation and received aligned to GRCh37 (25). Variants were filtered to include only those with a PASS flag. Based on RECIST1.1

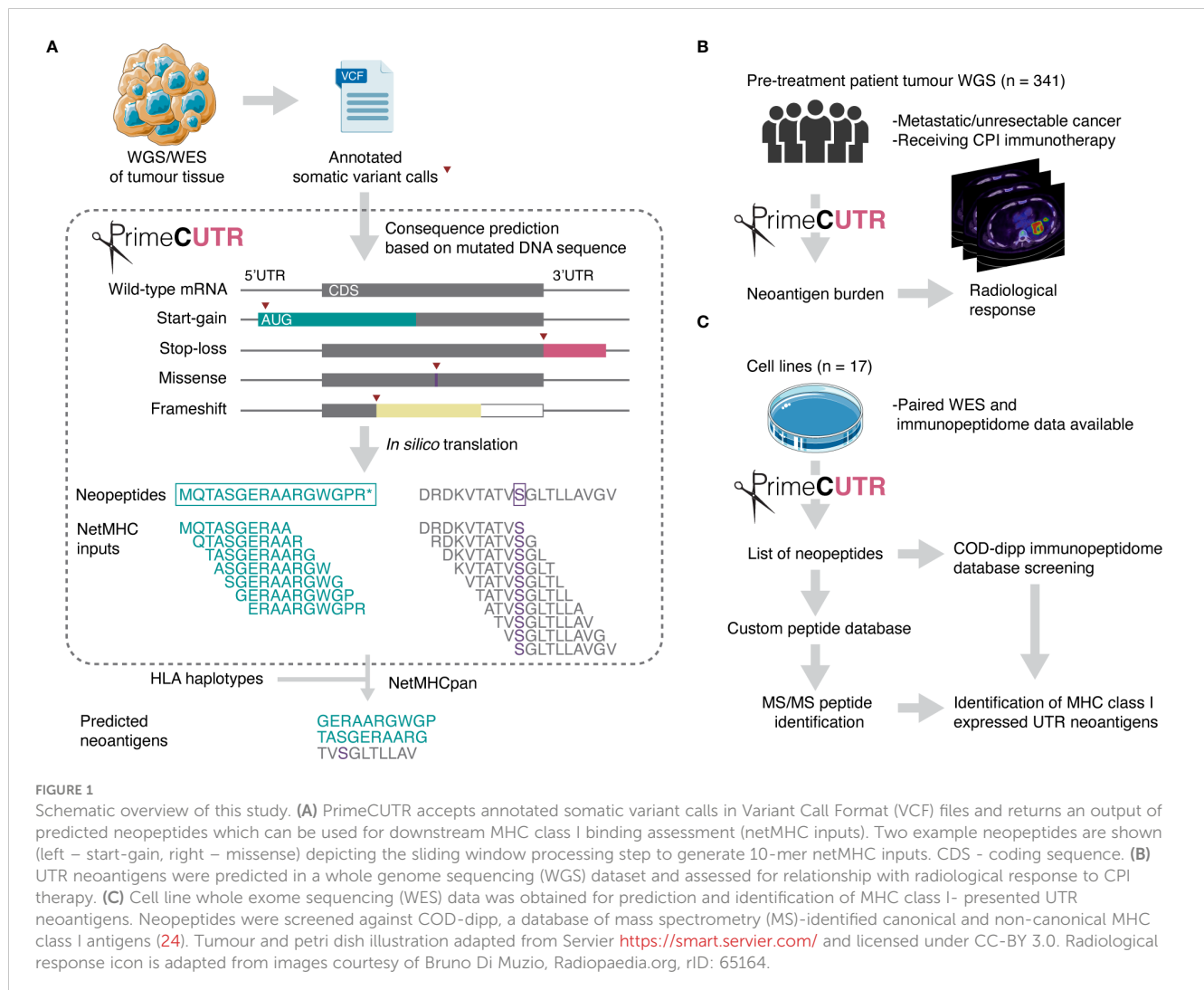


FIGURE 1

Schematic overview of this study. (A) PrimeCUTR accepts annotated somatic variant calls in Variant Call Format (VCF) files and returns an output of predicted neopeptides which can be used for downstream MHC class I binding assessment (netMHC inputs). Two example neopeptides are shown (left – start-gain, right – missense) depicting the sliding window processing step to generate 10-mer netMHC inputs. CDS - coding sequence. (B) UTR neoantigens were predicted in a whole genome sequencing (WGS) dataset and assessed for relationship with radiological response to CPI therapy. (C) Cell line whole exome sequencing (WES) data was obtained for prediction and identification of MHC class I-presented UTR neoantigens. Neopeptides were screened against COD-dipp, a database of mass spectrometry (MS)-identified canonical and non-canonical MHC class I antigens (24). Tumour and petri dish illustration adapted from Servier <https://smart.servier.com/> and licensed under CC-BY 3.0. Radiological response icon is adapted from images courtesy of Bruno Di Muzio, Radiopaedia.org, rID: 65164.

criteria treatment response data, patients with a complete or partial response were classified as CPI responders while those with stable or progressive disease were classified as CPI non-responders. While stringent, this grouping is concordant with published immunotherapy biomarker validation studies (26, 27). Median neoantigen burden was compared between responding and non-responding groups using the Wilcoxon signed-rank test. Multivariable logistic regression was used to assess this relationship while correcting for tumour mutational burden (TMB). TMB was defined as number of mutations per megabase (mut/MB).

2.2 Identifying start-gain and stop-loss neoantigens

Start-gain mutations were defined as any single-nucleotide variant (SNV) or short insertion-deletion which resulted in a new ATG codon in the upstream 5'UTR region of a transcript (uAUG). All uAUG-forming mutations in 5'UTR sequences from all Ensembl-annotated protein coding transcripts were included in

the neoantigen prediction. The relevant reference genome (GRCh37 or GRCh38) was used depending on the prior alignment of the somatic mutation calls. Reading 5' to 3', *in silico* translation of the cDNA sequence, beginning from the uAUG was performed until a stop codon (TAA, TAG or TGA) was reached. Stop-loss mutations were defined as any SNV which altered the annotated stop codon of a transcript into a sense codon. In this case, *in silico* translation is continued from the new sense codon until a stop codon is reached. Open reading frames from insertion-deletion mutations were obtained similarly according to the preceding reading frame. In rare cases where no stop codon is reached within the transcript, the alternate reading frame (from a start-gain, stop-loss or insertion-deletion mutation) is read through to the mRNA poly-A tail, resulting in a poly-lysine sequence (28). Protein-coding SNVs, dinucleotide variants (DNVs), and small in-frame insertion-deletions are grouped as missense mutations. In this report, missense and frameshift mutations refer specifically to mutations occurring in the protein-coding regions of the genome.

All neopeptides were processed using a sliding window to generate 9-, 10- and 11-mers which included at least one mutated/frameshifted residue (Figure 1A). These peptides were

then assessed for predicted MHC class I binding strength, using the pVACtools suite (version 3.1.1) to run pVACbind with the netMHCpan algorithm (29, 30). Peptides binding with an IC50 of less than 500 nM were reported as neoantigens. This widely adopted threshold is consistent with previous work showing that most MHC class I ligands bind below this affinity (31, 32). Start-gain neoORFs overlapping in-frame with coding sequences of other isoforms were excluded by removing any reported neoantigens with exact matches in the canonical human proteome.

2.3 Neoantigen dissimilarity

Neoantigen dissimilarity from the self-proteome (dissimilarity score) and neoantigen homology to known immunogenic epitopes from Immune Epitope Database (foreignness score) were calculated using the *foreignness_score* and *dissimilarity_score* functions in antigen.garnish 2 (<https://github.com/andrewrech/antigen.garnish> accessed 16th October 2023, (33)) with default parameters. Neoantigens were considered highly dissimilar based on a dissimilarity score of >0.7, and highly foreign with a cut-off foreignness score of >0.75. These cutoffs were selected based on natural breaks in the distribution of scores (Supplementary Figures 1B, C).

2.4 Mutational signature extraction

Mutational signatures were extracted for each tumour sample using DeconstructSigs (34) yielding the relative contribution of mutational processes per tumour sample (COSMIC v1.0) (35). To obtain a score per cancer group, samples were grouped by type and the relative contribution of each mutational signature was averaged across samples. COSMIC mutational profiles (v1.0) were downloaded from <https://cancer.sanger.ac.uk/signatures/downloads/> (accessed 16th September 2023). To assess the probability of start-gain formation for a given mutational signature, all non-overlapping 5'UTR sequences in the human genome were obtained from Ensembl (GRCh38 Ensembl release 109). For each given COSMIC mutational signature profile, the probability of each unique single base substitution (SBS) was multiplied by the number of 5'UTR sites where such a substitution lead to uAUG formation, divided by the total number of 5'UTR sites where the substitution could occur. The probabilities of start-gain formation for each of the 96 SBS were then summed to give a probability of start-gain formation for each mutational signature.

TABLE 1 Predicted neoantigens by mutation class.

Class	Mutations (n)	Neoantigens (n)	Neoantigens per mutation	High foreignness neoantigens (n)*	High dissimilarity neoantigens (n)*
Missense	64424	177670	2.76	16438 (9.3%)	1137 (0.6%)
Frameshift	2694	17254	6.40	2155 (12.5%)	1302 (7.5%)
Start-gain	772	4701	6.09	542 (11.5%)	310 (6.6%)
Stop-loss	95	563	5.93	56 (10.0%)	34 (6.0%)

*Percentages expressed as a proportion of total neoantigens in each given mutation class.

2.5 Mass spectrometry (MS) validation

We identified 13 cell lines with whole exome sequencing (WES) data in the Cancer Cell Line Encyclopedia (CCLE, <https://depmap.org/portal/download/all/> accessed 24th May 2023) which had paired immunopeptidomic sequencing (Supplementary Table 1A). CCLE somatic mutations were downloaded aligned to GRCh38. Somatic mutation calls, aligned to GRCh37, from 4 melanoma patient samples were obtained from Bassani-Sternberg et al. (6). PrimeCUTR was used to predict start-gain, stop-loss and frameshift neopeptides from the somatic mutation calls of each of these 17 tumour/cell line samples. In the first validation step, neoORF peptides were compared to a non-canonical MHC-associated peptide database (Closed Open *De novo* – deep immunopeptidomics pipeline (COD-dipp)) generated from MS analysis (24). As a second validation step, Fragpipe version 19.1 (36) was used to performed independent proteogenomic MS database search by appending the neopeptides to the normal protein database as described previously (24). Ion, PSM and peptide-level false discovery rates were set at 1%.

2.6 Translation initiation site prediction

Web-based TIS prediction algorithms, TISrover (37, <http://bioit2.irc.ugent.be/rover/tisrover>) and TIS Transformer (38, <https://jdcla.ugent.be/>) were used to assess the likelihood of translation initiation of start-gain neoORFs. The cDNA sequence for the mutant transcripts were uploaded in FASTA format for analysis using default settings.

3 Results

3.1 Inferring UTR neoantigens from cancer mutation data

PrimeCUTR accepts somatic mutation VCF data annotated by Ensembl Variant Effects Predictor (VEP) (39) or the Hartwig Medical Foundation variant calling pipeline (25) (Figure 1A). VEP annotates mutation consequence per given gene transcript, providing necessary information for PrimeCUTR to classify mutations into missense, frameshift, 5'UTR variant and stop-loss. All 5'UTR variants are checked for start-gain formation (as of v111, VEP does not annotate start-gain variants), while all SNVs arising

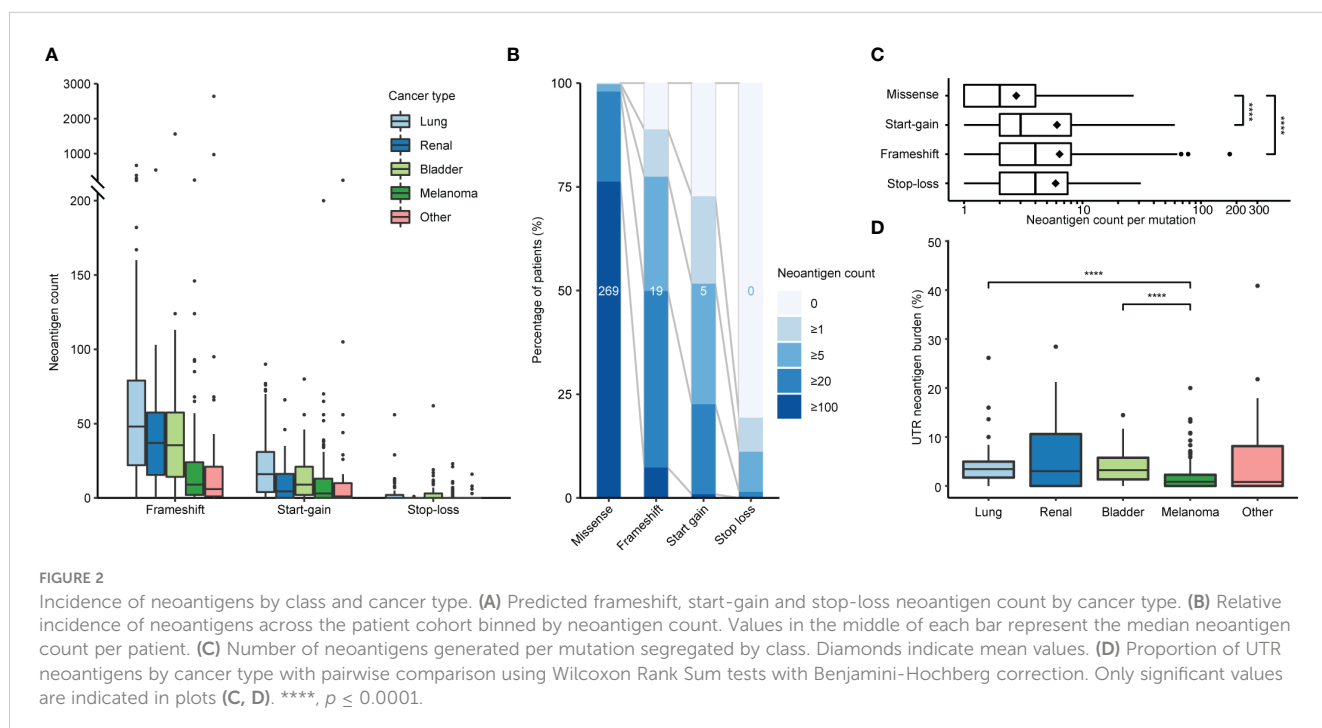
in the final codon of protein coding transcripts are additionally checked for stop-loss regardless of VEP annotation. The *get.peptide* function can be used interactively in R to predict the resulting neopeptide from pairs of Ensembl transcript ID and Human Genome Variation Society (HGVS) coding DNA sequence variant nomenclature provided by VEP. The *get.orfs* function accepts whole VCF files producing three outputs: (1) tab-separated files containing neopeptides per mutation class per sample (see [Supplementary Tables 1B–E](#) for example output), (2) FASTA-format text files containing 9, 10 and 11-mers which include at least one mutated/frameshifted residue, (3) relevant log files. *get.orfs* also returns neopeptides with normal flanking residues extending to the next up- or downstream trypsin cleavage site with or without a missed cleavage, allowing seamless integration of PrimeCUTR output into a proteomic search pipeline. The 9,10 and 11-mer FASTA files can be fed directly to MHC-binding prediction algorithms for neoantigen prediction. Additionally, for start-gain mutations, *get.peptide* scores start-gain Kozak consensus sequence strength (weak, moderate or strong – see Whiffin et al. (40)), estimates overlap with wild-type open reading frames, and flags potential in-frame overlap with protein coding sequences, allowing rapid screening of the most promising neopeptides. Further details on installation, usage and output of the PrimeCUTR R package, including a tutorial and example datasets can be found at <https://github.com/christophersng/primeCUTR>.

3.2 Incidence of UTR neoantigens

We applied PrimeCUTR and netMHCpan to pre-treatment cancer WGS samples from the Hartwig Medical Foundation to identify the contribution of the different mutation classes to the

neoantigen landscape in patients who received CPI treatment ($n = 341$) ([Figure 1B](#)). Across the cancer types, the majority of predicted neoantigens with MHC binding ($IC_{50} < 500$ nM) arose from missense mutations (SNVs, DNVs, in-frame indels, total 177670, 88.8%), while frameshift, start-gain and stop loss mutations contributed 17254 (8.6%), 4701 (2.3%) and 563 (0.3%) respectively ([Table 1](#)). Among cancer types, lung cancer had the highest burden of mutations and neoantigens in all mutation classes ([Figure 2A](#)). Start-gain neoantigens were predicted in 72.7% of patients across cancer types, with a median of 5 unique neoantigens per patient (range 0–237) while stop-loss neoantigens were predicted in only 19.3% of patients ([Figure 2B](#)). By comparison, frameshift mutations were predicted in 88.9% of patients. Overall, frameshift, start-gain and stop-loss mutations generated neoORFs of similar lengths (median: 19 versus 20 versus 17 amino acid residues respectively). Although rarer, start-gain mutations generated significantly more neoantigens per mutation than missense mutations (median: 3 versus 2, mean: 6.09 versus 2.76, adjusted p -value $< 2 \times 10^{-16}$) ([Table 1](#) and [Figure 2C](#)).

Given the translation of neoORFs, we hypothesised start-gain and stop-loss mutations would be more distinct from the self-proteome. Previously, Richman et al. (33) demonstrated that neoantigens with high dissimilarity from the self-proteome, as well as high homology to known immunogenic peptides from Immune Epitope Database (foreignness score) were correlated with measures of immunogenicity. Using the same approach (see Methods), we found that frameshift, start-gain and stop-loss neoantigens were approximately 10 times more likely to have high dissimilarity compared to missense mutations ([Table 1](#)). Therefore, across the cohort, while UTR neoantigens only accounted for 2.6% of predicted neoantigens, they comprised 12.4% of high dissimilarity neoantigens ([Supplementary Figure 1A](#)).



Most UTR neoORFs were private: Only 2 start-gain mutations and no stop-loss mutations were shared by 3 or more patients. The two recurrent start-gain mutations occurred exclusively in melanoma samples: RPL8; ENST00000262584:c.-94G>A (7 patients, 4.6%) and DCAF7; ENST00000310827:c.-207G>A (3 patients, 2.0%). They respectively produced neoORFs 54 and 58 amino acid residues long, and were predicted to generate multiple patient-specific HLA binding neoantigens (Supplementary Tables 2A–C). These recurrent mutations were identified in an independent melanoma WGS cohort (41), where 6 patient samples (3.3%) had RPL8; ENST00000262584:c.-94G>A and 1 patient sample had DCAF7; ENST00000310827:c.-207G>A. Neither variant was reported in dbSNP, The Cancer Genome Atlas or gnomAD.

3.3 Start-gain incidence by mutational signature

All neoantigen classes were correlated with TMB (Supplementary Figure 2A). Interestingly, despite having a median TMB comparable to lung cancer (16.2 versus 16.4 mut/MB, Supplementary Figure 2B), melanoma showed significantly lower relative UTR neoantigen burden (0.8%) compared to lung (3.5%; corrected p -value = 4.9×10^{-8}) and bladder (3.2%, corrected p -value = 4.8×10^{-6}) malignancies (Figure 2D). This primarily reflected the lower relative incidence of start-gain mutations in melanoma.

Given the majority of start-gain mutations arose from SNVs generating a new AUG codon in the 5'UTR, we hypothesised that the underlying single base substitution (SBS) mutational signature could explain the differences in relative start-gain mutation frequency between cancer types. We aggregated the probabilities of uAUG formation in all unique 5'UTR sequences of the human genome for every given 96 SBS mutational signature (COSMIC v1, 35) (Figure 3). This showed that mutational signatures linked to aging (1A/B), smoking (4), DNA mismatch repair (6, 14, 15, 20, 21) and POLE mutation (10, 14) favoured the formation of start-gain mutations (42). Meanwhile, ultraviolet (UV)-related mutational signature 7 strongly suppressed the likelihood of start-gain formation, explaining the relative sparsity of start-gain neoantigens in melanoma.

3.4 Response to checkpoint inhibitor immunotherapy

As described above, UTR neoantigens generate proportionally more neoantigens with a high dissimilarity from the self-proteome and thus may be a more potent immune target. We therefore assessed whether UTR neoantigen load was associated with response to CPI treatment. In univariate analysis, CPI responders had significantly higher missense and start-gain neoantigens (Supplementary Figure 3). Missense mutations are closely linked to TMB, an established marker of CPI response (2). The significant association of start-gain neoantigens and CPI response was not maintained in multivariable logistic regression when accounting for TMB (p -value = 0.8).

3.5 Immunopeptidomic discovery

In order to demonstrate the expression of UTR neoantigens on MHC class I, we screened the UTR neopeptides against COD-dipp, a database of MS-identified canonical and non-canonical MHC class I antigens (24). 48 UTR neoORFs were predicted in 12 out of 17 studied cell lines from somatic mutation data (Supplementary Table 1A). Among these, one candidate UTR neoantigen, peptide ILLNFSTTTK, was identified in COD-dipp, matching a neoORF from a private start-gain mutation (OAT; ENST00000539214:c.-61C>T) in the Jurkat cell line (Supplementary Table 1D, row 40). This was further confirmed with high-confidence using independent proteogenomic MS database search of two Jurkat immunopeptidome replicates (Figure 4A). ILLNFSTTTK was identified in 7 and 8 different spectra in each replicate respectively, but not in any other cell line nor in the normal immunopeptidome. Of the known HLA alleles for Jurkat (43), MHC-binding prediction with netMHCpan showed that ILLNFSTTTK binds strongly to HLA-A*03:01 (17.13 nM). However, we noted that ILLNFSTTTK is also overlapped by a wild-type upstream open reading frame (uORF) beginning at position -46 (Figure 4B). While the wild-type uORF is not previously described in a repository of wild-type RIBO-Seq identified uORFs (www.sorfs.org, accessed 26th April 2023) (44), the origin of ILLNFSTTTK from this uORF could not be excluded. To investigate the origin of translation of ILLNFSTTTK, we used TISRover to compare the likelihood of translation of the wild-type and mutant uORF (37). TISRover appeared to favour the mutant start-gain uORF (score: 9.2×10^{-5}) over the wild-type uORF (score: 1×10^{-6}) as a translation initiation site (TIS), although both scored lower than the canonical start codon (score: 2.8×10^{-2}) (Figure 4C and Supplementary Table 1F). One study used a TISRover score cut-off of 0.1 to annotate translation initiation sites (45). Another TIS prediction algorithm, TIS Transformer, did not annotate either the wild-type or mutant uORFs as potential TIS (38).

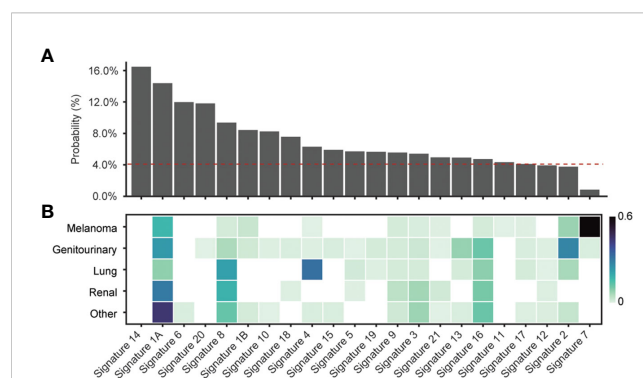


FIGURE 3 (A) Probability of start-gain formation based on COSMIC v1 mutational signature. The red dashed line indicates the probability of start-gain formation given a neutral mutational signature. (B) Heatmap of relative composition of mutations attributable to a given mutational signature averaged within each cancer group. Proposed aetiologies for the mutational signatures include: aging (1A/B), smoking (4), DNA mismatch repair (6, 14, 15, 20, 21), POLE mutation (10,14) and UV (7) (35, 42).

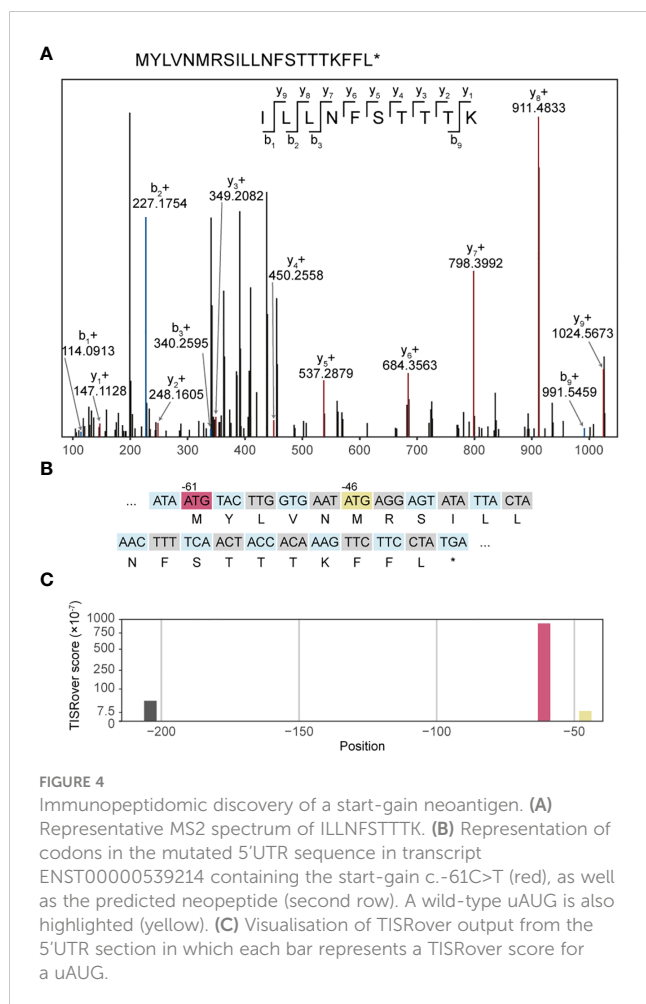


FIGURE 4

Immunopeptidomic discovery of a start-gain neoantigen. (A) Representative MS2 spectrum of ILLNFSTTTK. (B) Representation of codons in the mutated 5'UTR sequence in transcript ENST00000539214 containing the start-gain c.-61C>T (red), as well as the predicted neopeptide (second row). A wild-type uAUG is also highlighted (yellow). (C) Visualisation of TISRover output from the 5'UTR section in which each bar represents a TISRover score for a uAUG.

Separately, we searched for the presence of the recurrent start-gain neoORFs RPL8; ENST00000262584:c.-94G>A and DCAF7; ENST00000310827:c.-207G>A within melanoma immunopeptidome datasets in the COD-dipp database. This identified peptide SAALVNRTR, which matched the RPL8; ENST00000262584:c.-94G>A neoORF, exclusively in the immunopeptidome of one patient-derived melanoma within all three replicate samples. SAALVNRTR was found to have strong patient-specific HLA binding to HLA-A*68:01 (Supplementary Table 2D). No peptides corresponding to RPL8; ENST00000262584:c.-94G>A were found in 10 healthy skin immunopeptidome datasets. SAALVNRTR had also been predicted from genomic data in our primary patient cohort (Supplementary Table 2B). In contrast to ILLNFSTTTK, this neoORF had no overlap with wild-type uORFs or coding regions. TISRover and TIS Transformer both verified the mutant uORF as a viable translation initiation site (Supplementary Tables 2E, F).

Taken together, this evidence supports the translation and expression of the start-gain neoORFs in a tumour-specific manner.

4 Discussion

In this study, we present PrimeCUTR, an open-source R package to identify UTR start-gain and stop-loss neopeptides

from tumour somatic mutation calls. PrimeCUTR is applicable to WGS data as well as WES data (albeit limited by UTR coverage). PrimeCUTR is easily incorporated into any bioinformatic neoantigen discovery workflow and is scalable to the processing of large datasets via a high-performance computing cluster.

Using PrimeCUTR, we show that UTR neoORFs occur frequently across different subtypes of cancer, yielding a previously overlooked source of neoantigens. Like frameshift mutations, when compared to missense mutations, start-gain and stop-loss mutations yield more than double the neoantigens per given mutation. We show that start-gain mutation frequency is influenced by background mutational signature, being favoured in MMR deficiency (Signatures 6, 14, 15, 20 and 21) or POLE mutations (Signatures 10 and 14) which can be found in colorectal and endometrial cancers, as well as age (Signature 1A/B) and tobacco smoking (Signature 4) (42). Signature 8, of unconfirmed aetiology but previously observed in cancers including breast cancer, is also associated with start-gain mutation formation. Meanwhile, UV exposure (Signature 7) suppresses the formation of start-gain mutations in melanoma, although this is offset by the higher overall mutational burden in melanoma. Prevailing mutational signatures for a given cancer type may therefore guide whether personalised neoantigen profiling strategies should use extended UTR coverage sequencing approaches.

Previous studies have demonstrated that neoantigen dissimilarity from the self-proteome is an important predictor of immunogenicity and immunoediting (33, 46). In our analysis, frameshift and UTR neoantigens were found to be more dissimilar to the self-proteome than missense neoantigens and thus enriched amongst the pool of highly dissimilar neoantigens. As the cognate T cells of UTR neoantigens are less likely to have been subject to central mechanisms of tolerance, they represent a promising target for boosting anti-tumour immunity, while minimising off-target effects.

Studying 17 cell lines with paired somatic mutation and immunopeptidomic data, we identified one MHC class I-presented UTR neoantigen, ILLNFSTTTK, which matched a predicted private start-gain neoORF (OAT; ENST00000539214:c.-61C>T) in the Jurkat cell line. This peptide, along with its associated start-gain mutation was exclusive to Jurkat cells. This paired-discovery approach was limited by the fact that the cell line mutation data was derived from WES. Commonly used WES kits only cover up to 20% of UTR bases (47). Taking a more general approach, we searched for expression of recurrent UTR neoantigens within large immunopeptidomic datasets. From our primary WGS patient cohort, we identified a recurrent start-gain neoORF (RPL8; ENST00000262584:c.-94G>A) in 7 patients with melanoma, which was validated in 6 patient samples from an independent melanoma WGS cohort (41). This identified a further UTR neoantigen, SAALVNRTR, in one patient which exactly matched the RPL8; ENST00000262584:c.-94G>A neoORF (Supplementary Table 2D).

Current MS approaches detect only a small fraction of expressed peptides, compounding our limited identification of UTR neoantigens. Cuevas et al. (48) found that only 0.44% of non-canonical translation events (including uORFs within the 5'UTR) were detected by MS. Nevertheless, this is, to our knowledge, the first immunopeptidomic discovery of start-gain UTR neoantigens.

In the patient cohort, we found that UTR neoantigen burden was not significantly associated with CPI response when correcting for TMB. However, the significant disparity between predicted neoantigens and those able to elicit immune responses (6, 7, 15–17) suggests CPI response is driven by a small but important fraction of neoantigens. Prioritisation of these neoantigens for therapy remains a technical and biological challenge: beyond MHC class I binding affinity and dissimilarity, factors such as intratumoural heterogeneity, RNA expression, and location within the protein sequence all influence expression and immunogenicity (14, 49). The expression of uORFs (and start-gain mutations) is also governed by the sequence context of the translation initiation site. Given the current limitations of immunopeptidomic validation, incorporation of translation initiation prediction algorithms will be critical to support the prioritisation of UTR neoantigens (37, 38). While our computational approach has yielded two promising candidates for expressed UTR neoantigens, validation of MS spectra with synthetic peptides was not done due to limited access to the identical MS instrumentation utilised across the diverse source datasets. Further studies with paired whole genome and immunopeptidomic analysis of patient tumour samples, as well as T cell reactivity assays are ultimately needed to confirm the expression and immunogenic potential of UTR neoantigens. In summary, we describe a computational tool to study the contribution of UTR neoantigens to the immune landscape of cancer with the potential to boost neoantigen search strategies for personalised immunotherapy.

Data availability statement

The data analysed in this study is subject to the following licenses/restrictions: Patient level data from the Hartwig Medical Foundation is considered private identifiable information and therefore access-controlled. Requests to access these datasets should be directed to <https://www.hartwigmedicalfoundation.nl/en/data/data-access-request/>.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements. All patients provided explicit consent to the Hartwig Medical Foundation for data sharing for cancer research in accordance with the license agreement.

Author contributions

CS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software,

Validation, Visualization, Writing – original draft, Writing – review & editing. AK: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing – review & editing, Visualization. BS: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing – review & editing, Data curation, Writing – original draft. GB: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – review & editing. JA: Conceptualization, Resources, Supervision, Writing – review & editing. KL: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. CS is funded by a National Institute for Health and Care Research (NIHR) Academic Clinical Fellowship. KL is funded by the UK Medical Research Council (MR/P014712/1 and MR/V033077/1), the Rosetrees Trust and Cotswold Trust (A2437), and CRUK (C69256/A30194). This work was also supported by the "International Centre for Cancer Vaccine Science" project (MAB/2017/3) carried out within the International Research Agendas programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund. The work was also supported by the KATY project that has received funding from the European Union's Horizon 2020 research and innovation programme (grant agreement number 101017453).

Acknowledgments

The authors would like to thank Krupa Thakkar, Hongui Cha, Alexander Coulton and Maria Litovchenko for their thoughtful review and advice on code used in analyses, and Fong Chun Chan, Emilia Lim and Ashley Wong for help with testing PrimeCUTR. We thank CI-TASK, Gdansk, and the PLGrid Infrastructure, Poland (grant numbers: PLG/2023/016653 and PLG/2023/016406) for providing their hardware and software resources. This publication and the underlying study have been made possible partly based on data that Hartwig Medical Foundation has made available to the study through the Hartwig Medical Database. We also thank Joris van der Haar for assistance with HLA and clinical data. Finally, we thank the patients and families who have contributed to this study.

Conflict of interest

KL reports personal fees from Kynos Therapeutics, MonoPTeros Therapeutics, Ellipses Pharma, Tempus Labs and Roche Tissue Diagnostics, and grants from Genesis Therapeutics and Cancer

Research UK/Ono Pharmaceutical Co., Ltd./LifeArc IO Alliance, and is an employee of Isomorphic Labs, all outside the submitted work. BS is an employee of DIOSynVax Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Marabelle Aurélien, Fakih M, Lopez J, Shah M, Shapira-Frommer R, Nakagawa K, et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol.* (2020) 21:1353–65. doi: 10.1016/S1470-2045(20)30445-9
- Yarchoan M, Hopkins A, Jaffe EM. Tumor mutational burden and response rate to PD-1 inhibition. *N Engl J Med.* (2017) 377:2500–2505. doi: 10.1056/NEJMc1713444
- Gubin MM, Zhang X, Schuster H, Caron E, Ward JP, Noguchi T, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature.* (2014) 515:577–81. doi: 10.1038/nature13988
- Snyder A, Makarov V, Merghoub T, Jianda Y, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med.* (2014) 371:2189–99. doi: 10.1056/NEJMoa1406498
- Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* (2015) 348:124–28. doi: 10.1126/science.aaa1348
- Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun.* (2016) 7:13404. doi: 10.1038/ncomms13404
- Robbins PF, Lu Y-C, El-Gamil M, Li YF, Gross C, Gartner J, et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Med.* (2013) 19:747–52. doi: 10.1038/nm.3161
- Zacharakis N, Chinnasamy H, Black M, Xu H, Lu Y-C, Zheng Z, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nat Med.* (2018) 24:724–30. doi: 10.1038/s41591-018-0040-8
- Kristensen NP, Heeke C, Tvingsholm SA, Borch A, Draghi A, Crowther MD, et al. Neoantigen-reactive CD8⁺ T cells affect clinical outcome of adoptive cell therapy with tumor-infiltrating lymphocytes in melanoma. *J Clin Invest.* (2022) 132(2):e150535. doi: 10.1172/JCI150535
- Li T, Zhao L, Yonghao Y, Yao W, Yong Z, Jindong G, et al. T cells expanded from PD-1+ Peripheral blood lymphocytes share more clones with paired tumor-infiltrating lymphocytes. *Cancer Res.* (2021) 81:2184–94. doi: 10.1158/0008-5472.CAN-20-2300
- Leidner R, Silva NS, Huang H, Sprout D, Zheng C, Shih Y-P, et al. Neoantigen T-cell receptor gene therapy in pancreatic cancer. *N Engl J Med.* (2022) 386:2112–19. doi: 10.1056/NEJMoa2119662
- Rojas LA, Sethna Z, Soares KC, Olcese C, Pang N, Patterson E, et al. Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer. *Nature.* (2023) 618:144–50. doi: 10.1038/s41586-023-06063-y
- Weber JS, Carlino MS, Khattak A, Meniawy T, Ansstas G, Taylor MH, et al. Individualised Neoantigen Therapy mRNA-4157 (V940) plus Pembrolizumab versus Pembrolizumab Monotherapy in Resected Melanoma (KEYNOTE-942): A Randomised, Phase 2b Study. *Lancet.* (2024) 403(10427):632–44. doi: 10.1016/S0140-6736(23)02268-7
- Lybaert L, Lefever S, Fant B, Smits E, De Geest B, Breckpot K, et al. Challenges in neoantigen-directed therapeutics. *Cancer Cell.* (2023) 41:15–40. doi: 10.1016/j.ccell.2022.10.013

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2024.1347542/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

(A) Proportion of neoantigens originating from different mutation classes, stratified by neoantigen quality metrics. High binding neoantigens were those with predicted IC50 <50nM. Histogram of (B) foreignness scores and (C) dissimilarity scores for all neoantigens. High foreignness (>0.75) and high dissimilarity thresholds (>0.7) indicated by red bars.

SUPPLEMENTARY FIGURE 2

(A) Correlation matrix of neoantigen count and tumour mutational burden (TMB). (B) TMB values by cancer type.

SUPPLEMENTARY FIGURE 3

Univariate analysis of CPI response based on (A) missense, (B) frameshift, (C) start-gain or (D) stop-loss neoantigen count.

- Parkhurst MR, Robbins PF, Tran E, Prickett TD, Gartner JJ, Jia Li, et al. Unique neoantigens arise from somatic mutations in patients with gastrointestinal cancers. *Cancer Discov.* (2019) 9:1022–35. doi: 10.1158/2159-8290.CD-18-1494
- Wells DK, van Buuren MM, Dang KK, Hubbard-Lucey VM, Sheehan KCF, Campbell KM, et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell.* (2020) 183:818–834.e13. doi: 10.1016/j.cell.2020.09.015
- Löffler MW, Mohr C, Bichmann L, Freudenmann LK, Walzer M, Schroeder CM, et al. Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. *Genome Med.* (2019) 11:28. doi: 10.1186/s13073-019-0636-8
- Vareki SM. High and low mutational burden tumors versus immunologically hot and cold tumors and response to immune checkpoint inhibitors. *J ImmunoTher Cancer.* (2018) 6:157. doi: 10.1186/s40425-018-0479-7
- Litchfield K, Reading JL, Lim EL, Xu H, Liu Po, Al-Bakir M, et al. Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nat Commun.* (2020) 11:3800. doi: 10.1038/s41467-020-17526-5
- Starck SR, Shastri N. Non-conventional sources of peptides presented by MHC class I. *Cell Mol Life Sci.* (2011) 68:1471–795. doi: 10.1007/s00018-011-0655-0
- Laumont CélineM, Vincent K, Hesnard L, Audemard Éric, Bonnel Éric, Laverdure J-P, et al. Noncoding regions are the main source of targetable tumor-specific antigens. *Sci Transl Med.* (2018) 10:eaa5516. doi: 10.1126/scitranslmed.aau5516
- Smart AC, Margolis CA, Pimentel H, He MX, Miao D, Adeegbe D, et al. Intron retention is a source of neoepitopes in cancer. *Nat Biotechnol.* (2018) 36:1056–585. doi: 10.1038/nbt.4239
- Ouspenskaia T, Law T, Clauser KR, Klaeger S, Sarkizova S, Aguet François, et al. Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat Biotechnol.* (2021) 40(2):209–17. doi: 10.1038/s41587-021-01021-3
- Bedran G, Gasser H-C, Weke K, Wang T, Bedran D, Laird A, et al. The immunopeptidome from a genomic perspective: establishing the noncanonical landscape of MHC class I-associated peptides. *Cancer Immunol Res.* (2023) 11:747–62. doi: 10.1158/2326-6066.CIR-22-0621
- Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature.* (2019) 575:210–16. doi: 10.1038/s41586-019-1689-y
- Litchfield K, Reading JL, Puttick C, Thakkar K, Abbosh C, Bentham R, et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell.* (2021) 184:596–614.e14. doi: 10.1016/j.cell.2021.01.002
- So WV, Dejardin D, Rossmann E, Charo J. Predictive biomarkers for PD-1/PD-L1 checkpoint inhibitor response in NSCLC: an analysis of clinical trial and real-world data. *J ImmunoTher Cancer.* (2023) 11:e0064645. doi: 10.1136/jitc-2022-006464
- Chandrasekaran V, Juszkiewicz S, Choi J, Puglisi JD, Brown A, Shao S, et al. Mechanism of ribosome stalling during translation of a poly(A) tail. *Nat Struct Mol Biol.* (2019) 26:1132–405. doi: 10.1038/s41594-019-0331-x
- Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics.* (2009) 61:1–135. doi: 10.1007/s00251-008-0341-z

30. Hundal J, Kiwala S, McMichael J, Miller CA, Xia H, Wollam AT, et al. pVACtools: A computational toolkit to identify and visualize cancer neoantigens. *Cancer Immunol Res.* (2020) 8:409–20. doi: 10.1158/2326-6066.CIR-19-0401
31. Sette A, Vitiello A, Reherman B, Fowler P, Nayarsina R, Kast WM, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol.* (1994) 153:5586–92. doi: 10.4049/jimmunol.153.12.5586
32. Bulik-Sullivan B, Busby J, Palmer CD, Davis MJ, Murphy T, Clark A, et al. Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. *Nat Biotechnol.* (2019) 37:55–63. doi: 10.1038/nbt.4313
33. Richman LP, Vonderheide RH, Rech AJ. Neoantigen dissimilarity to the self-proteome predicts immunogenicity and response to immune checkpoint blockade. *Cell Syst.* (2019) 9:375–3825.e4. doi: 10.1016/j.cels.2019.08.009
34. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* (2016) 17:315. doi: 10.1186/s13059-016-0893-4
35. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature.* (2013) 500:415–21. doi: 10.1038/nature12477
36. Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods.* (2017) 14:513–205. doi: 10.1038/nmeth.4256
37. Zuallaert J, Kim M, Soete A, Saeys Y, De Neve W. TISRover: convNets learn biologically relevant features for effective translation initiation site prediction. *Int J Data Min Bioinf.* (2018) 20:267–845. doi: 10.1504/IJDMB.2018.094781
38. Clauwaert J, McVey Z, Gupta R, Menschaert G. TIS transformer: remapping the human proteome using deep learning. *NAR Genomics Bioinf.* (2023) 5:lqad021. doi: 10.1093/nargab/lqad021
39. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol.* (2016) 17:1225. doi: 10.1186/s13059-016-0974-4
40. Whiffin N, Karczewski KJ, Zhang X, Chothani S, Smith MJ, Evans DG, et al. Characterising the loss-of-function impact of 5' Untranslated region variants in 15,708 individuals. *Nat Commun.* (2020) 11:2523. doi: 10.1038/s41467-019-10717-9
41. Hayward NK, Wilmott JS, Waddell N, Johansson PA, Field MA, Nones K, et al. Whole-genome landscapes of major melanoma subtypes. *Nature.* (2017) 545:175–80. doi: 10.1038/nature22071
42. Alexandrov LB, Kim J, Haradhvala NJ, Huang MiNi, Ng AWT, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature.* (2020) 578:94–101. doi: 10.1038/s41586-020-1943-3
43. Andreatta M, Nicastrì A, Peng Xu, Hancock G, Dorrell L, Ternette N, et al. MS-rescue: A computational pipeline to increase the quality and yield of immunopeptidomics experiments. *PROTEOMICS.* (2019) 19:18003575. doi: 10.1002/pmic.201800357
44. Olexiouk V, Van Criekinge W, Menschaert G. An update on sORFs.Org: A repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res.* (2018) 46:D497–502. doi: 10.1093/nar/gkx1130
45. Bartas M, Volná A, Beaudoin CA, Poulsen ET, Červeň Jiří, Brázda Václav, et al. Unheeded SARS-coV-2 proteins? A deep look into negative-sense RNA. *Briefings Bioinf.* (2022) 23:bbac045. doi: 10.1093/bib/bbac045
46. Łuksza M, Sethna ZM, Rojas LA, Lihm J, Bravi B, Elhanati Y, et al. Neoantigen quality predicts immunoeediting in survivors of pancreatic cancer. *Nature.* (2022) 606:389–95. doi: 10.1038/s41586-022-04735-9
47. Barbitoff YA, Poley DE, Glotov AS, Serebryakova EA, Shcherbakova IV, Kiselev AM, Kostareva AA, et al. Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. *Sci Rep.* (2020) 10(1):2057.
48. Cuevas MVR, Hardy M-P, Holly J, Bonnell Éric, Durette C, Courcelles M, et al. Most non-canonical proteins uniquely populate the proteome or immunopeptidome. *Cell Rep.* (2021) 34(10):108815. doi: 10.1016/j.celrep.2021.108815
49. Müller M, Huber F, Arnaud M, Kraemer AI, Altimiras ER, Michaux J, et al. Machine learning methods and harmonized datasets improve immunogenic neoantigen prediction. *Immunity.* (2023) 56:2650–2663.e6. doi: 10.1016/j.immuni.2023.09.002