ARTICLE

Check for updates

Escape from nonsense-mediated decay associates with anti-tumor immunogenicity

Kevin Litchfield [1,2,12✉], James L. Reading [2,3,12], Emilia L. Lim[1], Hang Xu[1], Po Liu [1], Maise Al-Bakir[1], Yien Ning Sophia Wong[3], Andrew Rowan[1], Samuel A. Funt [4], Taha Merghoub[4], David Perkins[5], Martin Lauss[6], Inge Marie Svane[7], Göran Jönsson[6], Javier Herrero [8], James Larkin[9], Sergio A. Quezada [3], Matthew D. Hellmann [4], Samra Turajlic [9,10✉] & Charles Swanton [1,2,11✉]

Frameshift insertion/deletions (fs-indels) are an infrequent but highly immunogenic mutation subtype. Although fs-indels are degraded through the nonsense-mediated decay (NMD) pathway, we hypothesise that some fs-indels escape degradation and elicit anti-tumor immune responses. Using allele-specific expression analysis, expressed fs-indels are enriched in genomic positions predicted to escape NMD, and associated with higher protein expression, consistent with degradation escape (NMD-escape). Across four independent melanoma cohorts, NMD-escape mutations are significantly associated with clinical-benefit to checkpoint inhibitor (CPI) therapy ($P_{meta} = 0.0039$). NMD-escape mutations are additionally found to associate with clinical-benefit in the low-TMB setting. Furthermore, in an adoptive cell therapy treated melanoma cohort, NMD-escape mutation count is the most significant biomarker associated with clinical-benefit. Analysis of functional T cell reactivity screens from personalized vaccine studies shows direct evidence of fs-indel derived neoantigens eliciting immune response, particularly those with highly elongated neo open reading frames. NMD-escape fs-indels represent an attractive target for biomarker optimisation and immunotherapy design.

[1] Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, 1 Midland Rd, London NW1 1AT, UK. [2] Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London WC1E 6DD, UK. [3] Cancer Immunology Unit, Research Department of Haematology, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London WC1E 6BT, UK. [4] Memorial Sloan Kettering Cancer Center, Division of Solid Tumor Oncology, Department of Medicine, Weill Cornell Medical College, and Parker Center for Cancer Immunotherapy, 885 2nd Avenue, New York, NY 10017, USA. [5] Mass Spectrometry Proteomics, The Francis Crick Institute, London NW1 1AT, UK. [6] Faculty of Medicine, Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, Scheelegatan 2, Medicon Village, 22185 Lund, Sweden. [7] Center for Cancer Immune Therapy, Department of Oncology, Copenhagen University Hospital Herlev, Borgmester Ib Juuls Vej 1, 2730 Herlev, Denmark. [8] Bill Lyons Informatics Centre, University College London Cancer Institute, London WC1E 6DD, UK. [9] Renal and Skin Units, The Royal Marsden Hospital, London SW3 6JJ, UK. [10] Cancer Dynamics Laboratory, The Francis Crick Institute, 1 Midland Rd, London NW1 1AT, UK. [11] Department of Medical Oncology, University College London Hospitals, 235 Euston Rd, Fitzrovia, London NW1 2BU, UK. [12]These authors contributed equally: Kevin Litchfield, James L. Reading. ✉email: k.litchfield@ucl.ac.uk; samra.turajlic@crick.ac.uk; charles.swanton@crick.ac.uk

Tumor mutation burden (TMB) is associated with response to immunotherapy across multiple tumor types, and therapeutic modalities, including checkpoint inhibitors (CPIs) and cellular based therapy[1–10]. Although TMB is a clinically relevant biomarker, there are clear opportunities to refine the molecular features associated with response to immunotherapy. In particular, the primary hypothesis that derives from TMB as an immunotherapy biomarker relates to the fact that somatic variants are able to generate tumor specific neoantigens. However, the vast majority of mutations appear to have no immunogenic effect. For example, although hundreds of high affinity neoantigens are predicted in a typical tumor sample, peptide screens routinely detect T cell reactivity against only a few neoantigens per tumor[11]. Additionally, the oligoclonal T cell expansions commonly reported in CPI responders favour the hypothesis that a restricted number of neoantigens mediate anti-tumor immune responses[6]. Finally, immunopeptidome profiling via mass spectrometry has similarly identified only a few neoantigens effectively presented on human leucocyte antigen (HLA) molecules per tumor sample[12]. Detailed analysis of TMB to identify the true underlying subsets of mutations driving immunogenicity may substantially optimise biomarker accuracy and improve therapeutic targeting of neoantigens.

We have previously shown that frameshift insertion/deletions (fs-indels) are an infrequent (pan-cancer median = 4 per tumor) but highly immunogenic subset of somatic variants[13]. Fs-indels can produce an increased abundance of tumor specific neoantigens with greater mutant-binding specificity. We found that fs-indels are associated with improved response to CPI therapy[13] and may be attractive candidates for therapeutic personalised tumor vaccines. However, fs-indels cause premature termination codons (PTCs) and are susceptible to degradation at the messenger RNA level through the process of nonsense-mediated decay (NMD). NMD normally functions as a surveillance pathway to protect eukaryotic cells from the toxic accumulation of truncated proteins. We hypothesized that a subset of fs-indels may escape NMD degradation, and which when translated contribute substantially to directing anti-tumor immunity.

The NMD process is only partially efficient. The canonical NMD model dictates that a mutation triggering a PTC, will escape degradation if located downstream of the last exon junction complex. Therefore, NMD efficiency is intimately linked to sequence position, with reduced efficiency found in the: (i) last gene exon, (ii) penultimate exon within 50 nucleotides of the 3′ exon junction, and (iii) first exon within the first 200 nucleotides of coding sequence (CDS). These rules only partially explain the variance in NMD efficiency however, and an estimated 27% remains unexplained across all genes, increasing to 71% for dosage compensated genes (i.e. genes where copy number deletion is compensated with upregulation of the remaining allele)[14].

Based on these rules, ~30% of fs-indels across cancers are predicted to escape NMD[15]. Fs-indel mutations escaping NMD have been shown to be an abundant source of expressed neoantigen protein in microsatellite instable (MSI) tumors and to correlate with high levels of CD8 infiltration[16]. In addition, targeted inhibition of NMD has been shown to strongly suppress tumor growth[17]. Taken together these data suggest fs-indels escaping NMD are rare but may be disproportionally immunogenic. Indeed, recent work in parallel with our own from Lindeboom et al.[18] has demonstrated that NMD-escaping fs-indels strongly associate with improved response to CPI therapy. To test this hypothesis further and provide independent validation, here we quantify NMD efficiency via allele-specific fs-indel detection in paired DNA and RNA sequencing data. We apply this pipeline to four independent cohorts of melanomas treated with CPI, one melanoma adoptive cell therapy cohort, and conduct further NMD analysis in personalized tumor vaccine studies. For further comparison, we also examined non-immunotherapy treated cases from the cancer genome atlas (TCGA). Collectively these results highlight a subset of fs-indels which escape NMD, and associate with anti-tumor immune response.

Results

Detection of NMD-escape mutations. Expressed frameshift indels (fs-indels) were detected using paired DNA and RNA sequencing, with data processed through an allele-specific bioinformatics pipeline (Fig. 1). Across all processed TCGA samples ($n = 453$, see "Methods" for cohort details) a median of 4 fs-indels were detected per tumor (range 0–552), of which mutant allele expression was detected in a median of 1 per tumor (range 0–92). Thus, expressed fs-indel mutations were present at relatively low frequency and abundance. In fact, 47.9% of samples profiled had zero expressed fs-indel mutations detected. Exon positions were annotated for expressed fs-indels ($n = 2267$), and compared to non-expressed fs-indels (i.e. mutant allele present in DNA, but not in RNA) ($n = 11153$). Expressed fs-indels were enriched for mutations in last exon positions (odds ratio versus non expressed fs-indels = 1.92, 95% confidence interval [1.69–2.19], $P < 2.2 \times 10^{-16}$, Fisher's exact test), as well as penultimate exons within 50 nucleotides of the 3′ exon junction complex (EJC) (OR = 1.90 [1.54–2.34], $P = 4.4 \times 10^{-9}$, Fisher's exact test). By contrast, expressed fs-indels were depleted in middle exon locations (OR = 0.54 [0.48–0.60], $P < 2.2 \times 10^{-16}$, Fisher's exact test), and no significant change was detected either way for penultimate exon, >50 nucleotides from the 3′ EJC, mutations (Fig. 2a). These exon positions are highly consistent with known patterns of NMD-escape[14]. No depletion/enrichment was detected for first exon (within first 200 bp of CDS) position mutations, likely due to the small absolute numbers. Next we considered RNA variant allele frequency (VAF) estimates for expressed fs-indels, and found them to be highest for last (median = 0.32), penultimate ≤50 bp of the 3′ EJC (0.31), penultimate >50 bp of the 3′ EJC (0.27) and first (0.26) exon positions, with middle exon alterations having the lowest value (0.19) (Fig. 2b, $P < 2.2 \times 10^{-16}$, Kruskal–Wallis test). Finally, we obtained protein array expression data from the cancer proteome atlas[19], for a panel of 223 proteins across 453 tumors, which overlapped with the DNA/RNAseq processed cohort. Intersecting samples with both an fs-indel gene mutation(s), and matched protein expression data, we compared the protein levels of expressed ($n = 40$) versus non expressed fs-indels ($n = 96$). Protein abundance was found to be significantly higher for expressed fs-indels (Fig. 2c, $P = 0.018$, Mann–Whitney U test). Taken collectively, these results suggest that expressed fs-indels are (at least partially) escaping NMD and being translated to the protein level (further allele-specific fs-indel protein expression data is also presented below). Expressed fs-indels are here after also referred to as NMD-escape mutations.

NMD-escape associates with clinical benefit to CPI. To assess the impact of NMD-escape mutations on anti-tumor immune response, we assessed the association between NMD-escape mutation count and CPI clinical benefit in four independent melanoma cohorts with matched DNA and RNA sequencing data: Van Allen et al.[8] ($n = 33$, anti-CTLA-4 treated), Snyder et al.[7] ($n = 21$, anti-CTLA-4 treated), Hugo et al.[4] ($n = 25$, anti-PD-1 treated) and Riaz et al.[20] ($n = 24$, anti-PD-1). For each sample, mutation burden was quantified based on the following classifications: (i) TMB: all non-synonymous SNVs (nsSNVs), (ii) expressed nsSNVs, (iii) fs-indels, and (iv) NMD-escape expressed fs-indels. Each mutation class was tested for an association with
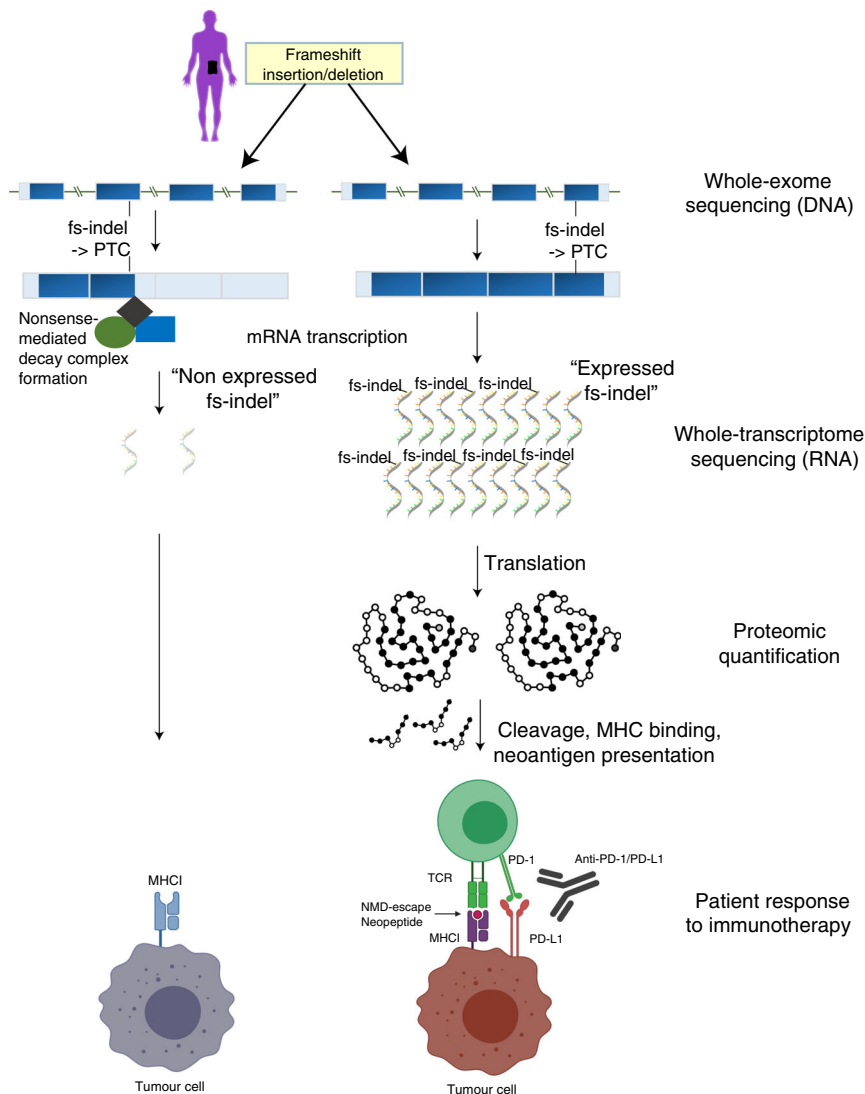
Fig. 1 Allele-specific expression analysis using matched DNA and RNA sequencing data. Shows an overview of study design and methodological approach. The left-hand side of the panel shows a fs-indel triggered premature termination codon, which falls in a middle exon of the gene, a position associated with efficient nonsense-mediated decay (NMD). The right-hand side of the panel shows a fs-indel triggered premature termination codon, which falls in the last exon of the gene, a position associated with bypassing NMD.

clinical benefit (Fig. 3a). In meta-analysis of the four melanoma cohorts with both WES and RNAseq (total $n = 103$), nsSNV, expressed nsSNV and fs-indel counts were higher in patients experiencing clinical benefit, but with non-significant p-value (meta-analysis across all cohorts, $P_{meta} = 0.073$, $P_{meta} = 0.19$ and $P_{meta} = 0.064$, respectively, Fisher's combined probability test) (Fig. 3a). NMD-escape mutation count however showed a statistically significant association with clinical benefit ($P_{meta} = 0.0039$, Fisher's combined probability test) (Fig. 3a). For clarity, we note sample sizes utilised here are smaller than previously reported, since only a subset of cases had both matched DNA and RNA sequencing data available, and that nsSNV and fs-indel measures are significant in the full datasets. Patients with one or more NMD-escape mutations had higher rates of clinical benefit to immune checkpoint blockade compared to patients with no NMD-escape mutations: 56% versus 12% (Van Allen et al.[8]), 57% versus 14% (Snyder et al.[7]), 75% versus 35% (Hugo et al.[4]) and 64% versus 30% (Riaz et al.[20]) (Fig. 4a). We additionally assessed for evidence of correlation between TMB and NMD-escape metrics, and found only a weak correlation between the two variables ($r = 0.23$). In multivariable logistic regression analysis,

we tested both variables together in a joint model to assess for independent significance ($n = 103$, study ID was also included as a model term to control for cohort specific factors), and NMD-escape mutation count was found to independently associate with CPI clinical benefit ($P = 0.0087$), whereas TMB did not reach independent significance ($P = 0.20$).

NMD-escape predicts clinical benefit in low-TMB tumors. In a clinical scenario where TMB is implemented to stratify patients for CPI therapy, patients with low TMB tumors may be not recommended for CPI treatment. It is known however that some low-TMB tumors can respond to CPI therapy, and we reasoned that NMD-escape mutation count may offer independent predictive power in the low-TMB setting to rescue patients who may have a higher chance of response. To investigate this we split the population of CPI treated patients to make a low-TMB cohort (nsSNV count ≤ 217, the median value across all cohorts, approximately equivalent to 10 mutations/Mb[21]), which comprised $n = 52$ patients (all four studies combined). In this cohort, NMD-escape mutation count was significantly associated with clinical benefit to CPI ($P = 0.013$, Mann–Whitney U test),
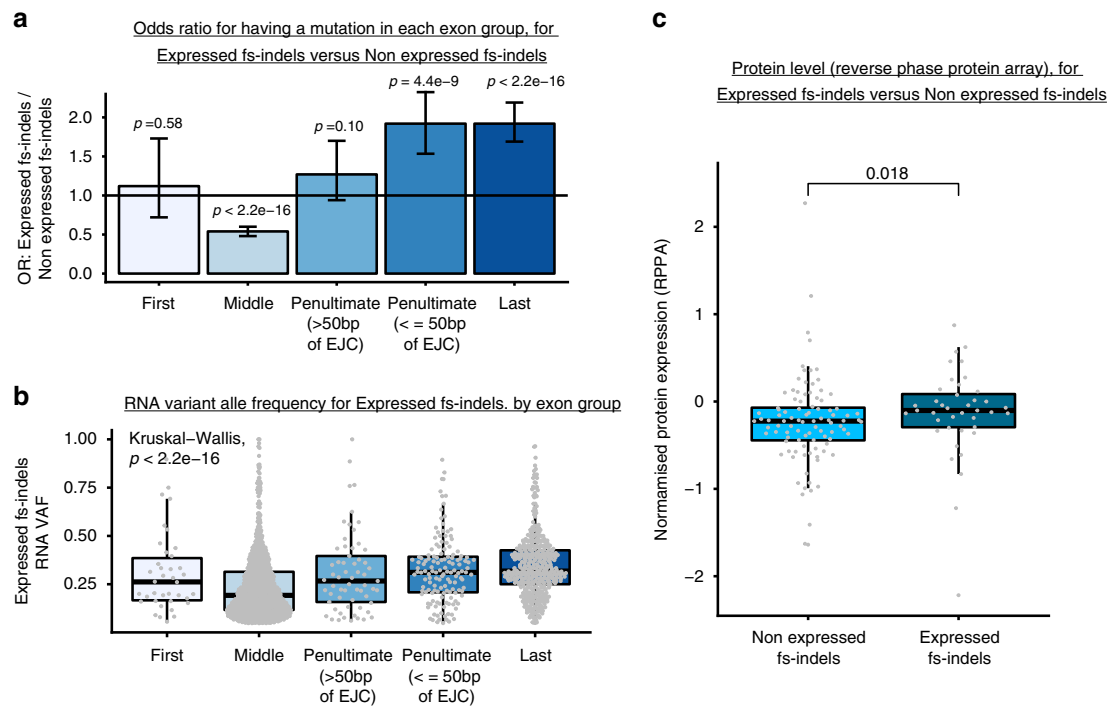
Fig. 2 Expressed fs-indels follow the rules of NMD. a Shows the odds ratio (OR), between expressed fs-indels and non-expressed fs-indels, for falling into either first, middle, penultimate or last exon positions. Odds ratios and associated *p*-values were calculated using Fisher's Exact Test. Data is from $n = 453$ TCGA patients, and statistics are based on $n = 13,420$ mutations. Coloring is used arbitrarily to distinguish groups. The measure of centre is the odds ratio value and error bars denote 95% confidence intervals of OR estimates. **b** Shows variant allele frequencies for expressed fs-indels by exon group position. Kruskal–Wallis test was used to test for a difference in distribution between groups. Data is from $n = 453$ TCGA patients. **c** Shows protein expression levels for not expressed, versus expressed, fs-indel mutations. Two-sided Mann–Whitney *U* test was used to assess for a difference between groups. Data is based on $n = 136$ mutations. In all boxplots in this figure the centre line is the median, the bounds of the box represent the inter-quartile range, the lower whisker $= \max(\min(x),\ Q\_1 - 1.5 \times IQR)$ and upper whisker $= \min(\max(x),\ Q\_3 + 1.5 \times IQR)$.

whereas nsSNV count was not ($P = 0.19$, Mann–Whitney *U* test). Patients with one or more NMD-escape mutation retained a relatively high rate of clinical benefit from CPI at 53%, compared to 15% for patients with zero NMD-escape events (Odds Ratio = 6.0, 95% confidence interval [1.4–28.9], $P = 0.0094$, Fisher's exact test, Fig. 4b). This suggests a potential utility for assessment of NMD-escape mutations in tumors with low overall TMB scores.

NMD-escape predicts clinical benefit from adoptive cell therapy. To further investigate the importance of NMD-escape mutations in directing anti-tumor immune response, we analyzed matched DNA and RNA sequencing data from patients with melanoma ($n = 22$) treated with adoptive cell therapy (ACT)[10]. TMB nsSNVs ($P = 0.027$), fs-indels ($P = 0.025$) and NMD-escape count ($P = 0.021$) were all significantly associated with clinical benefit from therapy (Fig. 4c, Mann–Whitney *U* test). All patients with NMD-escape count ≥ 1 experienced clinical benefit ($n = 4$, 100%), compared to 33% (6/18) of patients who had no NMD-escape mutations, further highlighting the potential strong immunogenic effect from just a single NMD-escape mutation. We acknowledge however the small sample size of this cohort as a limitation and investigation in larger adoptive cell therapy cohorts will be of significant interest.

Allele-specific protein expression of fs-indel mutations. To validate that exact mutated peptides derived from fs-indel mutations could be detected via mass spectrometry, we downloaded raw data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC), available for $n = 81$ samples from the

colon adenocarcinoma TCGA cohort. Fs-indel mutational data was also obtained for these same patients, and all mutated peptides resulting from the frameshift mutations were calculated. Customised library searches were conducted using MASCOT (see "Methods"), which identified a spectral match to a mutated peptide sequence (Fig. 5), thus providing evidence to support mutant allele protein expression. The peptide/spectral match derives from a fs-indel in gene *C15orf39* (sample TCGA-AA-3672), located in the penultimate exon 2 of 3, a region with low predicted NMD efficiency[18]. While only descriptive in nature, this example fits a model of NMD-escape, and interestingly the neo open reading frame (neoORF) triggered from this mutation is highly elongated in nature, generating a total of 131 amino acids of mutated sequence (Fig. 5).

FS-indel neoantigens generate anti-tumor immune response. While of translational relevance and clinical utility, biomarker associations do not directly isolate specific neoantigens driving anti-tumor immune response. Accordingly, we obtained data from two anti-tumor personalised vaccine studies, and one CPI study, in which immune reactivity against specific fs-indel derived neopeptides had been established by functional assay of patient T cells[22–24]. Across these three studies, 15 different fs-indel mutations generated peptides that were functionally validated as eliciting immune reactivity (Table S1); thus at a proof of concept level the ability of fs-indels to elicit anti-tumor immune response has been established. Across these same studies, four fs-indel derived neoantigens had also undergone functional screening, but were found to be non-immunogenic
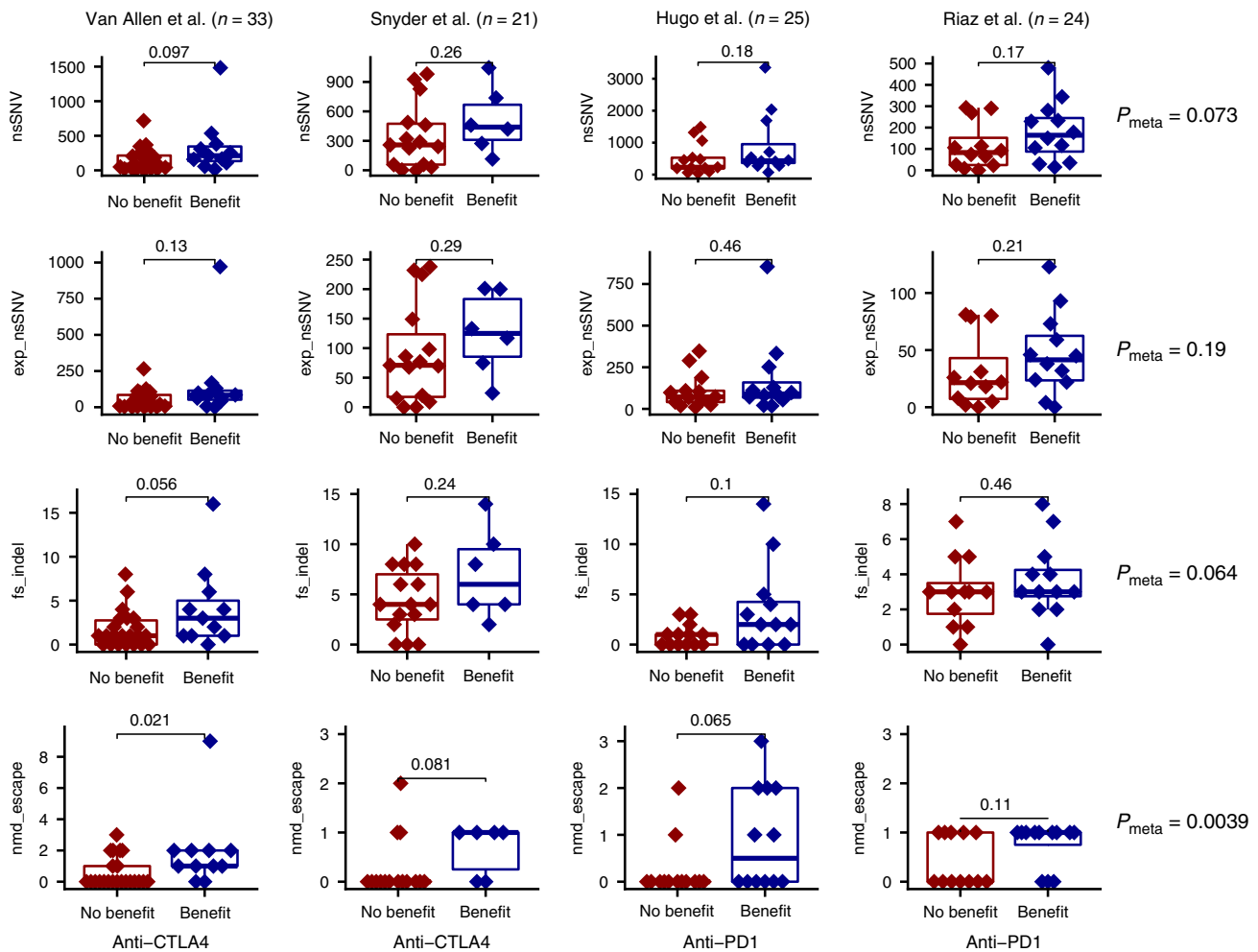
Fig. 3 NMD-escape mutation count predicts immune checkpoint inhibitor response. Shows four melanoma checkpoint inhibitor (CPI) treated cohorts, split into groups based on no-clinical benefit (dark red) or clinical benefit (dark blue) to therapy. Four metrics are displayed per cohort: (top row) TMB non-synonymous SNV count, (second row) expressed non-synonymous SNV count, (third row) frameshift indel count and (fourth row) NMD-escape mutation count. In the first column is the Van Allen et al.[8] anti-CTLA-4 cohort ($n = 33$ patients), second column is the Snyder et al.[7] anti-CTLA4 cohort ($n = 21$ patients), the third column is the Hugo et al.[4] anti-PD1 cohort ($n = 25$ patients) and the fourth column is the Riaz et al.[20] anti-PD1 cohort ($n = 24$ patients). Far right are meta-analysis p-values, for each metric across the four cohorts, showing the association with clinical benefit from CPI treatment. Two-sided Mann–Whitney U test was used to assess for a difference between groups. Meta-analysis of results across cohorts was conducted using the Fisher method of combining P values from independent tests. In all boxplots in this figure the centre line is the median, the bounds of the box represent the inter-quartile range, the lower whisker = $max(min(x), Q_1 - 1.5 \times IQR)$ and upper whisker = $min(max(x), Q_3 + 1.5 \times IQR)$.

(Supplementary Table 1). Although limited by a small sample size, we note that immunogenic fs-indel mutations ($n = 15$) had a significantly longer neoORF length (median = 27 amino acids) than screened, but non-immunogenic, fs-indel mutations ($n = 4$, median = 5 amino acids, $P = 0.0032$, Mann–Whitney U test) (Fig. 6a). We additionally again note several fs-indel mutations with extreme neoORF length (termed super neoORF (SNORF) mutations) were detected (neoORF ≥ 50 amino acids, $n = 5$), and these were restricted to the immunogenic group. The number of peptides screened is likely to be a confounding factor in these comparisons (i.e. a longer neoORF allows more unique peptides to be utilized for immunization), but this also highlights the inherent advantage of SNORF events. In the context of SNORF mutations, we next considered redundancy in HLA allele binding, based on the hypothesis that SNORF events (and indeed fs-indels in general) would generate peptides capable of binding to a broader spectrum of patient HLA alleles. This is likely to be of particular importance in the

context loss of heterozygosity at the HLA locus (LOHHLA), a mechanism used by tumor cells to achieve immune evasion. For example considering class I alleles, LOHHLA is known to occur such that one or two HLA alleles become lost[25]; however loss of all six HLA alleles would be unfavourable to the cancer cell, due to global loss of antigen presentation and resulting attraction of natural killer (NK) cell activity[26]. We analysed this further using neoantigen prediction data from two cohorts[8,20], and note that 54% [95% confidence interval 53–54%] of nsSNV mutations generate peptides that bind to just one HLA allele, with only 46% generating peptides than bind to multiple alleles (Fig. 6b). By contrast, 34% [30–37%] of fs-indel mutations generate peptides which bind against a single HLA allele, and instead the majority (66%) bind to multiple HLA alleles (Fig. 6b). In fact 13% of fs-indel mutations (likely SNORF events) were found to generate peptides binding against ≥4 HLA class I alleles, compared to only 3% of nsSNV mutations (Fig. 6b).
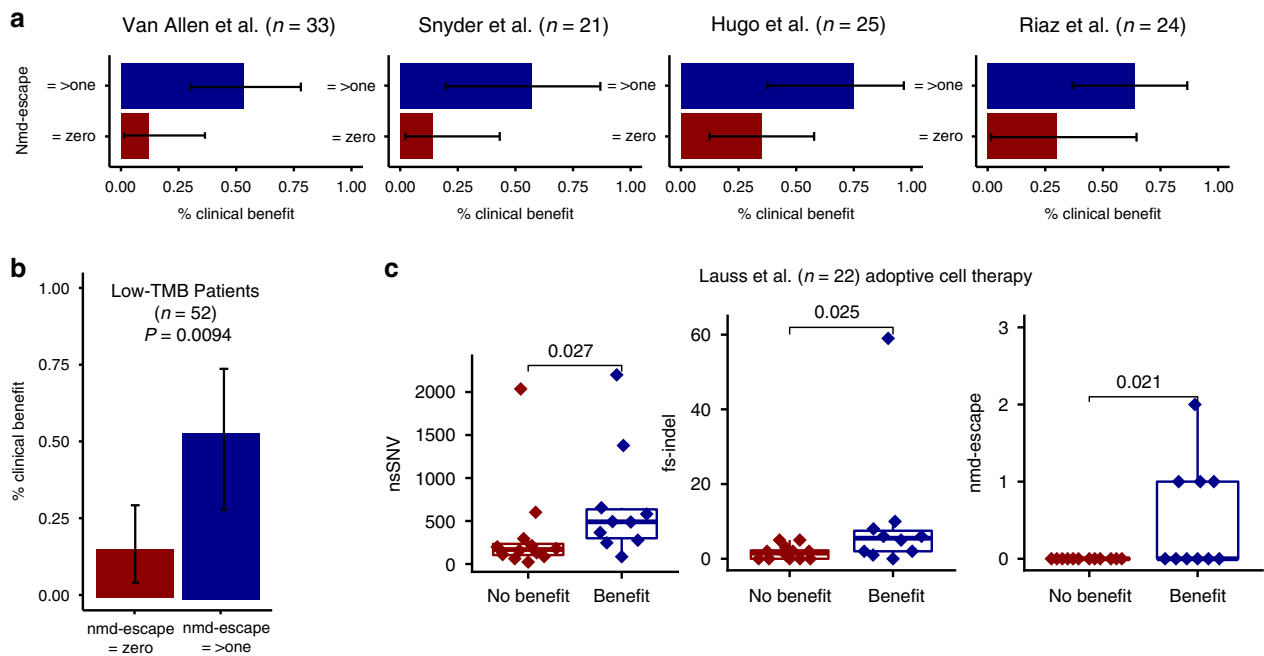
Fig. 4 NMD-escape mutations predict CPI response in low-TMB patients and response to ACT. a Shows the percentage of patient with clinical benefit from CPI therapy, for patients with ≥1 NMD-escape mutation (dark blue) and zero NMD-escape mutations (dark red), error bars denote 95% confidence intervals of the percentage clinical benefit estimates. Sample sizes are listed on the figure and represent number of patients. **b** Shows the combined set of CPI treated patients (across all four studies) split to make a low-TMB cohort (nsSNV count < 217, the median value across all cohorts, approximately equivalent to 10 mutations/Mb). The total patient number is $n = 52$. The percentage clinical benefit rates are shown for patients with zero and ≥one NMD-escape mutation, error bars denote 95% confidence interval and significance determined using Fisher's exact test. **c** Shows non-synonymous SNV count, frameshift indel count and NMD-escape mutation count, compared in an adoptive cell therapy treated cohort of $n = 22$ patients. In all boxplots in this figure the centre line is the median, the bounds of the box represent the inter-quartile range, the lower whisker $= \max(\min(x), Q\_1 - 1.5 \times IQR)$ and upper whisker $= \min(\max(x), Q\_3 + 1.5 \times IQR)$. Two-sided Mann–Whitney U test was used to assess for a difference between groups.
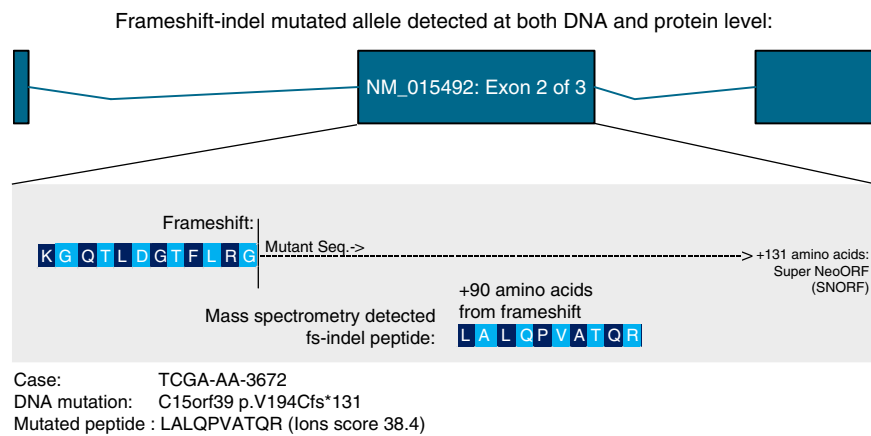

Fig. 5 Fs-indel mutated alleles detected using mass spectrometry. Shows mass spectrometry results, for spectral searches against mutant peptide sequences deriving from fs-indels. A single peptide-spectrum match (PSM) was identified from this analysis and is shown as a descriptive example. The exon structure of the fs-indel mutated gene is depicted, along with the position of the frameshift mutation, the matching peptide sequence before the frameshift, and the mutant peptide detected from the the neo open reading frame (neoORF) window.

NMD-escape mutations show evidence of negative selection. Next, we assessed for evidence of selective pressure against NMD-escape mutations, which may reflect the potential to generate native anti-tumor immunogenicity. In addition to potential immunogenic selective pressure, fs-indels have also previously been reported to be under functional selection[15] due to their loss of protein function effect. To account for this, we used stop-gain

SNV mutations as a benchmark comparator, as these variants have equivalent functional impact but no immunogenic potential (i.e. loss of function but no neoantigens generated). Furthermore, the rules of NMD apply equally to both stop-gain SNVs and fs-indels, as both trigger premature termination codons. Using the skin cutaneous melanoma (SKCM) TCGA cohort, we annotated all fs-indels ($n = 1527$) and stop-gain SNVs ($n = 9439$) for exonic
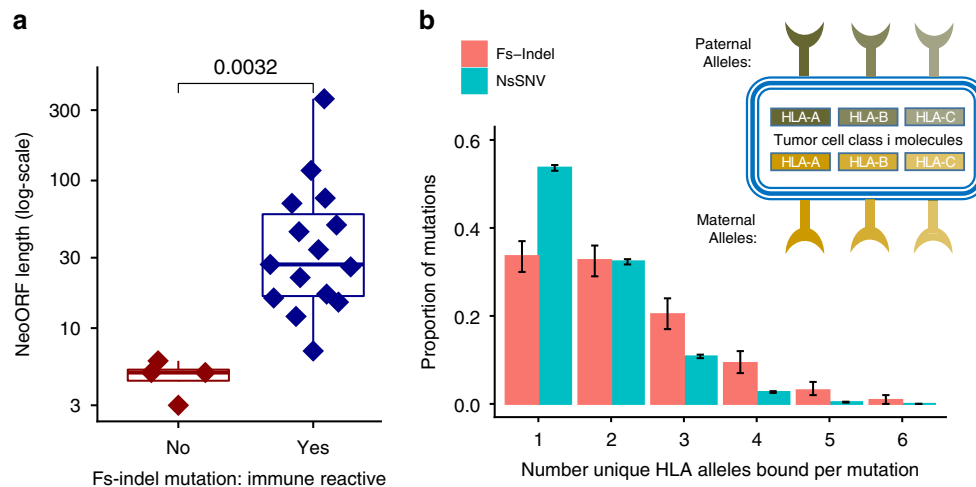
Fig. 6 Direct evidence of T cell reactivity to fs-indel peptides. a Shows the neo open reading frame (neoORF) length, from functional T cell reactivity screening data from recent personalized vaccine and CPI studies. In dark red are lengths from fs-indel peptides which were non-reactive, in dark blue are lengths from fs-indel peptides which were T cell reactive. Two-sided Mann–Whitney U test was used to assess for a difference between groups, with $n = 19$ mutations in total. **b** Shows the number of unique class I HLA alleles that individual nsSNV and fs-indel mutations were found to bind against in pan-cancer TCGA data. In all boxplots in this figure the centre line is the median, the bounds of the box represent the inter-quartile range, the lower whisker = max (min(x), $Q\_1 - 1.5 \times IQR$) and upper whisker = min(max(x), $Q\_3 + 1.5 \times IQR$).

position. Penultimate and last exon alterations were found to be significantly depleted in fs-indels compared to stop-gain events (OR = 0.58 [0.45–0.73], $P = 1.2 \times 10^{-6}$ and OR = 0.65 [0.55–0.76], $P = 9.4 \times 10^{-8}$, respectively, Fisher's exact test) (Supplementary Fig. 1). By contrast, fs-indel mutations were more likely to occur in middle exon positions (OR = 1.57 [1.38–1.79], $P = 2.9 \times 10^{-12}$, Fisher's exact test). First exon mutations were not enriched either way, possibly due to small absolute numbers (only $n = 69$ fs-indels were first exon). These data suggest negative selective immune pressure may act against fs-indel mutations in exonic positions likely to escape NMD (e.g. penultimate and last), leading to cancer cells with middle exon fs-indels being more likely to survive immunoediting. As an additional control to rule out any potential bias in variant calling between fs-indels and stop-gain SNV groups, we repeated the above analysis for germline variants from the Exome Aggregation Consortium (ExAC) database[27]. Due to self-tolerance, no immunogenicity would be expected against either stop-gain or fs-indels, and in accordance with this no depletion in fs-indel mutations was detected in penultimate or last exon positions (all ORs were between 0.8 and 1.2) (Supplementary Fig. 1).

NMD-escape predicts CPI response in pan-cancer cases. Finally, to investigate a potential association in other tumor types, NMD-escape analysis was conducted in a pan-cancer CPI dataset of $n = 542$ cases including lung, renal, colorectal, bladder, head and neck[5,9,28–31]. As RNA-seq data was not available in over half the cohorts NMD-escape mutations were instead predicted from DNA mutation exon position using the recently published model from Lindeboom et al.[18]. Across the pan-cancer cohort predicted NMD-escape mutation count was again associated with improved response to CPI treatment ($P_{meta} = 0.00043$, Fisher's combined probability test) (Supplementary Fig. 2), suggesting a potential broader role for NMD-escape beyond melanoma.

Discussion
In this study, we analysed expressed fs-indels, in the context of NMD and anti-tumor immunogenicity. We show that expressed fs-indels are highly enriched in genomic positions predicted to escape NMD, and have higher protein-level expression (relative

to non-expressed fs-indels). Expressed fs-indels (a.k.a. NMD-escape mutations) also significantly associated with clinical benefit from immunotherapy. The small and heterogeneous individual cohorts (~20–30 cases/cohort) utilized in this study should be acknowledged however.

The primary mechanism controlling NMD in mammalian cells is proposed to be the exon junction complex (EJC) model, whereby transcripts bearing a premature termination codon (PTC) in the last exon or the penultimate exon ≤50 bp of the 3′ EJC escape degradation. Indeed, we observe strong consistency with this model in comparing expressed versus not expressed fs-indels, with the former being highly enriched in penultimate and last exon position mutations (both OR > 1.8, $P < 5 \times 10^{-12}$). In addition, protein abundance was found to be significantly higher in the expressed fs-indel group ($P = 0.018$). The unexplained determinants of NMD should also be recognised however, with recent studies[14] estimating that over a quarter of NMD variance remains unexplained across all genes. Similarly, these exceptions are visible in our data, with an appreciable number of mutations in middle exon position detected as expressed. As well as novel instances of NMD-escape, it is also possible a subset of these middle exon mutations are in fact undergoing partial (or full) NMD degradation, but remain at sufficiently high transcript abundance to be detected. Translational plasticity has also been described as an additional feature impacting NMD efficiency, with a diverse range of mechanisms such as stop codon read-through, alternative translation initiation and alternative splicing, being reported as driving NMD-escape in the germline setting[32]. Furthermore, the highly dysregulated nature of cancer cell transcriptomes may further explain the partial leakiness in NMD patterns we observe in this study. These exceptions, and the currently incomplete understanding of NMD, highlights the importance of establishing fs-indel expression using RNA sequencing data and the need for further mechanistic research in this area.

NMD-escape mutation count was found to significantly associate with clinical benefit from immunotherapy, across both CPI and ACT modalities, and with a stronger association than either nsSNVs or fs-indels. CPI clinical benefit rates for patients with ≥1 NMD-escape mutation were elevated (range across the cohorts analysed = 0.56–0.75) compared to patients with zero

such events (range 0.12–0.35). Furthermore, NMD-escape mutation count was shown to remain significantly associated with clinical benefit to CPI in the low-TMB setting ($P = 0.013$), whereas nsSNV count was not ($P = 0.19$). This raises the prospect of rescuing patients who may fall below the overall 10 mutations per megabase TMB threshold, but have a higher chance of CPI response based on harbouring an elevated number of NMD-escape mutation events. Several potential sources of antigenic peptide material for human leucocyte antigen presentation have been proposed, ranging from classical degradation of previously functional proteins, to alternative sources including the pioneer round of mRNA translation and defective ribosomal products[33,34]. The enrichment of NMD-escape features, in the expressed fs-indels observed in this study, would favour a classical route of translational as at least one source of peptide material in these cohorts. However, the appreciable number of expressed fs-indels deriving from middle exon positions, suggests additional possible sources may be present, such as those from the pioneer round of translation.

Experimental evidence, analyzed from anti-tumor vaccine and CPI studies, demonstrates T cell reactivity against frameshifted neoepitopes directly in human patients ($n = 15$). T cell reactive fs-indel neoantigens were also enriched for longer neoORF length (median = 27 amino acids), versus experimentally screened, but T cell non-reactive fs-indels (median = 5 amino acids) ($P = 0.0032$). These elongated neoORF mutations create the additional benefit of increased redundancy in HLA allele binding, based on the intuitive result that a greater number of peptides will be capable of binding to a broader spectrum of patient HLA alleles. Selection analysis demonstrated a depletion of fs-indels in penultimate and last exon positions, as compared to functionally equivalent stop-gain SNVs, suggesting potential negative immune selection against NMD-escape events during tumor evolution. As a negative control, we demonstrate this same association is not present in germline mutations from the ExAC database. Checkpoint control of immune response is a likely compensatory mechanism used by tumor cells to manage remaining NMD-escape events, a notion in keeping with the elevated CPI clinical benefit rates we see in patients with even a small number of NMD-escape alterations. We also note recent work highlighting the technical challenges in detecting neoantigen depletion signal in cancer datasets[35], and further modelling work is likely needed. In terms of study limitations, we acknowledge that escape from NMD has not been functionally demonstrated in this work, and instead the enrichment of penultimate/last exon position mutations, together with higher protein expression levels, is suggestive of NMD-escape but not direct proof. This limitation is in keeping with the translational biomarker scope of this study, but nevertheless further functional investigation of NMD-escape mechanisms will be of significant interest. Furthermore, we acknowledge that neoantigen presentation is an inefficient process, and that NMD-escape mutations identified by DNA/RNA sequencing are unlikely to be directly causative in all tumors. Here we present evidence to support NMD-mutations as a rare mutation type, with enhanced potential to elicit an anti-tumor immune response. The bioinformatics challenges in accurate indel calling are a further limitation, meaning a reduced sensitivity to detect all fs-indels. However, we demonstrate that using both DNA and RNA sequencing assays improves calling accuracy and leads to a high confidence call set, due to alteration detection at both DNA and RNA levels (see "Methods"). In summary, here we highlight NMD-escape mutations as a highly immunogenic mutational subset, rare in frequency but found to significantly associate with clinical benefit to immunotherapy. These mutations may represent attractive targets for personalized immunotherapy design, as well as contributing to the refinement of genomic biomarkers to predict CPI response.

Methods

Study cohorts. Matched DNA/RNA sequencing analysis was conducted in the following cohorts all treated with immunotherapy:

- Van Allen et al.[8], an advanced melanoma CPI (anti-CTLA-4) treated cohort. Cases with both RNA sequencing and whole-exome (DNA) sequencing data were utilised ($n = 33$).
- Snyder et al.[7], an advanced melanoma CPI (anti-CTLA-4) treated cohort. Cases with both RNA sequencing and whole-exome (DNA) sequencing data were utilised ($n = 21$).
- Hugo et al.[4], an advanced melanoma CPI (anti-PD-1) treated cohort. Cases with both RNA sequencing and whole-exome (DNA) sequencing data were utilised ($n = 25$).
- Riaz et al.[20], an advanced melanoma CPI (anti-PD-1) treated cohort. Cases with both RNA sequencing and whole-exome (DNA) sequencing data, from the ipilimumab-naive cohort, were utilised ($n = 24$). In keeping with the original publication, we found the other patient cohort in this study (cases pre-treated and progressive on ipilimumab therapy (Ipi-P)), to have no association between mutation load metrics (nsSNVs, fs-indels, NMD-escape mutations) and subsequent benefit from anti-PD1 therapy.
- Lauss et al.[10], an advanced melanoma adoptive cell therapy treated cohort. Cases with both RNA sequencing and whole-exome (DNA) sequencing data were utilised ($n = 22$).

Matched DNA/RNA sequencing analysis was conducted in the following cohorts (not specifically treated with immunotherapy):

- Skin cutaneous melanoma (SKCM) tumors, obtained from the cancer genome atlas (TCGA) project. Cases with paired end RNA sequencing data and curated variant calls from TCGA GDAC Firehose (2016_01_28 release) were utilised ($n = 364$). Melanoma was selected as a cohort in order to match with the immunotherapy treated samples.
- Microsatellite instable (MSI) tumors, across all histological subtypes from TCGA project. MSI case IDs were identified based on classification from Cortes-Ciriano et al.[36]. Cases with paired end RNA sequencing data (see below) and curated variant calls from TCGA GDAC Firehose (2016_01_28 release) were utilised ($n = 89$). MSI tumors were selected on account of the high numbers of fs-indels per sample, and hence greater power to measure NMD-escape.

T cell reactivity response analysis was conducted in the following immunotherapy treated cohorts:

- Ott et al.[22], an advanced melanoma personalized vaccine treated cohort ($n = 6$ cases).
- Rahma et al.[23], a metastatic renal cell carcinoma personalized vaccine treated cohort ($n = 6$ cases).
- Le et al.[24], an advanced mismatch repair-deficient cohort, across cancers across 12 different tumor types, treated with anti-PD-1 blockade ($n = 86$ cases, functional neoantigen reactivity T cell work only conducted in $n = 1$ case).

Whole-exome sequencing (DNA) variant calling. For Van Allen et al.[8], Snyder et al.[7], and Riaz et al.[20] cohorts, we obtained germline/tumor BAM files from the original authors and reverted these back to FASTQ format using Picard tools (version 1.107) SamToFastq. Raw paired-end reads in FastQ format were aligned to the full hg19 genomic assembly (including unknown contigs) obtained from GATK bundle (version 2.8), using bwa mem (bwa-0.7.7). We used Picard tools to clean, sort and to remove duplicate reads. GATK (version 2.8) was used for local indel realignment. We used Picard tools, GATK (version 2.8), and FastQC (version 0.10.1) to produce quality control metrics. SAMtools mpileup (version 0.1.19) was used to locate non-reference positions in tumor and germline samples. Bases with a Phred score of less than 20 or reads with a mapping quality <20 were omitted. VarScan2 somatic (version 2.3.6) used output from SAMtools mpileup to identify somatic variants between tumor and matched germline samples. Default parameters were used with the exception of minimum coverage for the germline sample, which was set to 10, and minimum variant frequency was changed to 0.01. VarScan2 processSomatic was used to extract the somatic variants. Single nucleotide variant (SNV) calls were filtered for false positives with the associated fpfilter.pl script in Varscan2, initially with default settings then repeated with min-var-frac=0.02, having first run the data through bam-readcount (version 0.5.1). MuTect (version 1.1.4) was also used to detect SNVs, and results were filtered according to the filter parameter PASS. In final QC filtering, an SNV was considered a true positive if the variant allele frequency (VAF) was >2% and the mutation was called by both VarScan2, with a somatic p-value ≤ 0.01, and MuTect. Alternatively, a frequency of 5% was required if only called in VarScan2, again with a somatic p-value ≤ 0.01. For small scale insertion/deletions (INDELs), only calls classed as high confidence by VarScan2 processSomatic were kept for further analysis, with somatic_p_value scores $< 5 \times 10^{-4}$. Variant annotation was performed using Annovar (version 2016Feb01). For the Hugo et al.[4] cohort, we obtained final post-quality control mutation annotation files generated as from[4]. Briefly, SNVs were detected using MuTect, VarScan2 and the GATK Unified Genotyper, while INDELs were detected using VarScan2, IndelLocator and GATK-

UGF. Mutations that were called by at least two of the three SNV/INDEL callers were retained as high confidence calls. For the Lauss et al.[10] cohort, SNVs and INDELs were detected as outlined in ref. [10]. Briefly, SNVs were detected using the intersection of MuTect and VarScan2 variants, while INDELs were detected using VarScan2 only. For VarScan2, high confidence calls at a VAF greater than 10% were retained.

Whole-transcriptome sequencing (RNA) variant calling. RNAseq data was obtained in BAM format for all studies, and reverted back to FASTQ format using bam2fastq (v1.1.0), and only samples with successfully recovered paired end R1 and R2 fastqs were utilised. Insertion/deletion mutations were called from raw paired end FASTQ files, using mapsplice (v2.2.0), with sequence reads aligned to hg19 genomic assembly (using bowtie pre-built index). Minimum QC thresholds were set to retain variants with ≥5 alternative reads, and variant allele frequency ≥0.05. Insertions and deletions called in both RNA and DNA sequencing assays were intersected, and designated as expressed indels, with a ±10 bp padding interval included to allow for minor alignment mismatches. SNVs in RNA sequencing data were called directly from the hg19 realigned BAM files, using Rsamtools to extract read counts per allele for each genomic position where a SNV had already called in DNA sequencing analysis. Similarly, minimum QC thresholds of ≥5 alternative reads, and variant allele frequency ≥0.05, were utilised and variants passing these thresholds were designated as expressed SNVs.

Consensus indel variant calling accuracy. As an additional methodological check, indels were re-called from datasets[7,8], using two additional DNA variant callers, Mutect2 and Scalpel, in addition to Varscan2. The aim was to assess if the joint DNA and RNA calling approach used in this study lead to higher consensus between variant callers, and hence a reduction in the risk of caller specific artefacts. Using a DNA calling only approach, we observed the consensus between variant callers ranging from 67% [called in all three tools Varscan2/Mutect2/Scalpel] to 82% [called in Varscan2 and one of Mutect2 or Scalpel]. These same values for indels called in both DNA and RNA sequencing data increased to 81% and 100%, respectively. Thus, we find 100% of the reported NMD-escape indel mutations (detected in both DNA and RNA) were called in two or more different DNA indel calling algorithms.

Isoform annotation. For analysis in Fig. 2a, b, variants were annotated on a tumor specific isoform basis. Specifically, level 3 isoform data was obtained from the broad firehose repository (https://gdac.broadinstitute.org/). For each mutated gene in each tumor, the corresponding isoform expression values were extracted (for the gene in question), and the isoform with highest abundance was selected for annotation purposes. Isoform annotation was conducted using Annovar, with frameshift indel mutations grouped into five categories, based on the position of the premature termination codon following the frameshift: (i) first exon, (ii) middle exon, (iii) penultimate exon more than 50 bp of the last exon junction complex, (iv) penultimate exon less than or equal to 50 bp of the last exon junction complex, (v) last exon.

Protein expression analysis. We retrieved Level 4 (L4) normalized protein expression data for 223 proteins, across $n = 453$ TCGA melanoma/MSI tumors (which overlapped with the TCGA cohorts also analysed via DNA/RNA sequencing) from the cancer proteome atlas (http://tcpaportal.org/tcpa/index.html). We filtered the data to sample/protein combinations which also contained an fs-indel mutation ($n = 136$), as called by DNA sequencing. The dataset was then split into two groups, based on the fs-indel being expressed or not (as measured by RNAseq, using the method detailed above). The two groups were compared using a two-sided Mann–Whitney test. We note that a limitation of this analysis is the fact that frameshift indel mutations themselves may alter antibody binding efficiency to the mutated proteins, however this potential bias applies equally to both groups being compared.

Mass spectrometry analysis. We downloaded raw mass spectrometry files from the CPTAC data portal (https://cptc-xfer.uis.georgetown.edu/publicData/Phase_II_Data/TCGA_Colorectal_Cancer/), in.raw and.mzML formats, for $n = 96$ samples from the colon adenocarcinoma TCGA cohort. For matching cases, patient level curated mutation annotation files were obtained from TCGA GDAC Firehose (2016_01_28 release) (https://gdac.broadinstitute.org/). Fs-indel mutations were filtered from mutation annotation files, and in total $n = 81$ patients had both mass spectrometry data coupled with fs-indel mutational events. All mutated peptide sequences, resulting from frameshift mutation events, were calculated and used as a custom search library. Mass spectral searches were conducted using the Mascot search engine (v2.3.1)[37], with mutated peptide sequences appended to the known human proteome, in FASTA file format. The following settings were used: Fixed modifications: none; Variable modifications: Oxidation (M),Carbamido-methyl (C),Acetyl (N-term); MSMS tolerance: 0.5 Da; MS tolerance: 2 Da and Enzyme: non-specific.

Tumor specific neoantigen analysis. For each patient, 4-digit HLA types were derived from germline exome data using POLYSOLVER[38]. Next, peptide sequences were generated computationally, for all 9, 10, and 11-mer mutated and wildtype epitopes, based on the non-synonymous single nucleotide variants and frameshift insertion/deletions called in each sample as described above. For each mutant peptide, binding affinity to each class I HLA allele (from POLYSOLVER) was predicted using NetMHCpan (v3.0) and NetMHC (v4.0)[39]. Peptides with biding prediction <0.5 from the rank score were considered neoantigen binders.

Outcome analysis. Across all immunotherapy treated cohorts, measures of patient clinical benefit/no-clinical benefit were kept as consistent with original author's criteria/definitions.

Selection analysis. To test for evidence of selection, fs-indel mutations were compared to stop-gain SNV mutations, in the SKCM TCGA cohort. Stop-gain SNV mutations were utilised a benchmark comparator, due to their likely equivalent functional impact (i.e. loss of function), equivalent treatment by the NMD pathway (i.e. last exon stop-gain SNVs will still escape NMD and cause truncated protein accumulation) but lack of immunogenic potential (i.e. no mutated peptides are generated). All alterations in each group were annotated for exon position (i.e. first, middle, penultimate or last exon). The odds of having an fs-indel in first, middle, penultimate or last exon positions was then benchmarked against the equivalent odds for a stop-gain SNV. As a negative control this same method was additionally applied to germline mutations from the ExAC database, downloaded from: https://console.cloud.google.com/storage/browser/gnomad-public/legacy/exacv1_downloads/.

Pan-cancer CPI response analysis. Pan-cancer data ($n = 542$) from nine additional studies treated with CPI therapy was utilised, taken from the following studies: Cristescu et al.[28] an advanced melanoma anti-PD-1 treated cohort, Cristescu et al.[28] an advanced head and neck cancer anti-PD-1 treated cohort, Cristescu et al.[28] "all other tumor types" cohort (from KEYNOTE-028 and KEYNOTE-012 studies), treated with anti-PD-1, Snyder et al.[29], a metastatic urothelial cancer anti-PD-L1 treated cohort, Mariathasan et al.[9], a metastatic urothelial cancer anti-PD-L1 treated cohort, Mcdermot et al.[30], a metastatic renal cell carcinoma anti-PD-L1 treated cohort, Rizvi et al.[5], a non-small cell lung cancer anti-PD-1 treated cohort, Hellman et al., a cohort of non-small cell lung cancer samples treated with anti-PD-1, and Le et al.[31], a colorectal cancer cohort treated with anti-PD-1 therapy. Raw exome was data downloaded and processed as described above. CPI response was defined based on radiological response (responder = CR/PR, non-responder = SD/PD). NMD-escape predictions were calculated using the recently published model from Lindeboom et al.[18], with NMD efficiency scores <0.25 considered to be predicted NMD-escape (the 0.25 threshold is taken directly from the Lindeboom et al.[18] paper).

Statistical methods. Odds ratios were calculated using Fisher's Exact Test for Count Data, with each exon position group compared to all others. Kruskal–Wallis test was used to test for a difference in distribution between three or more independent groups. Two-sided Mann–Whitney U test was used to assess for a difference in distributions between two population groups. Meta-analysis of results across cohorts was conducted using the Fisher method of combining P values from independent tests. Logistic regression was used to assess multiple variables jointly for independent association with binary outcomes. Statistical analysis were carried out using R3.4.4 (http://www.r-project.org/). We considered a P value of 0.05 as being statistically significant, and all tests were two-sided.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data for the Van Allen et al.[8] cohort is available in dbGap under accession number phs000452.v2.p1. Data for the Snyder et al.[7] melanoma cohort is available in dbGap under accession number phs001041.v1.p1. The transcriptome data for Hugo et al.[4] is available through GEO accession number GSE78220. Data for the Riaz et al.[20] cohort is available in SRA through accessions SRP094781 (RNAseq) and SRP095809 (exome data). Data for the Lauss et al.[10] cohort was obtained via email request from the corresponding author(s). Data for the pan-cancer CPI analysis came from accession numbers: phs001572.v1.p1, EGAS00001002556, EGAS00001002928 and phs000980.v1.p1. Data from Snyder et al.[37], Hellman and Le et al.[39] cohorts was obtained via email request from the corresponding author(s). TCGA data was obtained from [https://portal.gdc.cancer.gov/] and CPTAC mass spectrometry data from [https://cptc-xfer.uis.georgetown.edu/publicData/]. T cell reactivity results were taken directly from papers published by Ott et al.[22], Rahma et al.[23], and Le et al.[24]. Cancer proteome expression data was obtained from [http://tcpaportal.org/tcpa/index.html]. Germline ExAC data was downloaded from: [https://console.cloud.google.com/storage/browser/gnomad-public/legacy/exacv1_downloads/].

Code availability

References

1. Forde, P. M. et al. Neoadjuvant PD-1 blockade in resectable lung cancer. N. Engl. J. Med. 378, 1976–1986 (2018).
2. Hellmann, M. D. et al. Tumor mutational burden and efficacy of nivolumab monotherapy and in combination with ipilimumab in small-cell lung cancer. Cancer cell. 33, 853–861 e4 (2018).
3. Hellmann, M. D. et al. Nivolumab plus Ipilimumab in lung cancer with a high tumor mutational burden. N. Engl. J. Med. 378, 2093–2104 (2018).
4. Hugo, W. et al. Genomic and transcriptomic features of response to Anti-PD-1 therapy in metastatic melanoma. Cell 165, 35–44 (2016).
5. Rizvi, N. A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 348, 124–128 (2015).
6. Roh, W. et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. Sci. Transl. Med. 9(Suppl. 379):eaah3560 (2017).
7. Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N. Engl. J. Med. 371, 2189–2199 (2014).
8. Van Allen, E. M. et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science 350, 207–211 (2015).
9. Mariathasan, S. et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature 554, 544–548 (2018).
10. Lauss, M. et al. Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. Nat. Commun. 8, 1738 (2017).
11. Robbins, P. F. et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. Nat. Med. 19, 747–752 (2013).
12. Bassani-Sternberg, M. et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. Nat. Commun. 7, 13404 (2016).
13. Turajlic, S. et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. Lancet Oncol. 18, 1009–1021 (2017).
14. Lindeboom, R. G., Supek, F. & Lehner, B. The rules and impact of nonsense-mediated mRNA decay in human cancers. Nat. Genet. 48, 1112–1118 (2016).
15. Hu, Z., Yau, C. & Ahmed, A. A. A pan-cancer genome-wide analysis reveals tumour dependencies by induction of nonsense-mediated decay. Nat. Commun. 8, 15943 (2017).
16. Maby, P., Galon, J. & Latouche, J. B. Frameshift mutations, neoantigens and tumor-specific CD8(+) T cells in microsatellite unstable colorectal cancers. Oncoimmunology 5, e1115943 (2016).
17. Pastor, F., Kolonias, D., Giangrande, P. H. & Gilboa, E. Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay. Nature 465, 227–230 (2010).
18. Lindeboom R. G. H., Vermeulen M., Lehner B., Supek F. The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. Nature Genet. 51, 1645–1651 (2019).
19. Li, J. et al. TCPA: a resource for cancer functional proteomics data. Nat. methods 10, 1046–1047 (2013).
20. Riaz, N. et al. Tumor and microenvironment evolution during immunotherapy with Nivolumab. Cell 171, 934–949 e16 (2017).
21. Budczies, J. et al. Optimizing panel-based tumor mutational burden (TMB) measurement. Ann. Oncol. 30, 1496–1506 (2019).
22. Ott, P. A. et al. An immunogenic personal neoantigen vaccine for patients with melanoma. Nature 547, 217–221 (2017).
23. Rahma, O. E. et al. A pilot clinical trial testing mutant von Hippel-Lindau peptide as a novel immune therapy in metastatic renal cell carcinoma. J. Transl. Med. 8, 8 (2010).
24. Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science 357, 409–413 (2017).
25. McGranahan, N. et al. Allele-specific HLA loss and immune escape in lung cancer evolution. Cell 171, 1259–1271 e11 (2017).
26. Karre, K., Ljunggren, H. G., Piontek, G. & Kiessling, R. Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. Nature 319, 675–678 (1986).
27. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291 (2016).
28. Cristescu, R. et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. Science. 362(Suppl. 6411):eaar3593 (2018).
29. Snyder, A. et al. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis. PLoS Med. 14, e1002309 (2017).
30. McDermott, D. F. et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. Nat. Med. 24, 749–757 (2018).
31. Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. N. Engl. J. Med. 372, 2509–2520 (2015).
32. Jagannathan, S. & Bradley, R. K. Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population. Genome Res. 26, 1639–1650 (2016).
33. Apcher, S. et al. Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. Proc. Natl Acad. Sci. USA 108, 11572–11577 (2011).
34. Rock, K. L., York, I. A., Saric, T. & Goldberg, A. L. Protein degradation and the generation of MHC class I-presented peptides. Adv. Immunol. 80, 1–70 (2002).
35. Van den Eynden, J., Jimenez-Sanchez, A., Miller, M. L. & Larsson, E. Lack of detectable neoantigen depletion signals in the untreated cancer genome. Nat. Genet. 51, 1741–1748 (2019).
36. Cortes-Ciriano, I., Lee, S., Park, W. Y., Kim, T. M. & Park, P. J. A molecular portrait of microsatellite instability across multiple cancers. Nat. Commun. 8, 15180 (2017).
37. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567 (1999).
38. Shukla, S. A. et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. 33, 1152–1158 (2015).
39. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 32, 511–517 (2016).

Author contributions

Study design: K.L., S.T., C.S., data analysis: K.L., J.R., E.L., H.X., P.L., M.A.-B., D.P., S.T. Data contribution and interpretation: S.W., A.R., S.A.F., T.M., M.L., I.M.-S., G.J., J.H., J.L., S.A.Q. Paper writing: K.L., M.D.H., S.T., C.S.

Competing interests

K.L., S.T., and C.S. have a patent on indel burden and checkpoint inhibitor response pending, and a patent on targeting of frameshift neoantigens for personalised immunotherapy pending. K.L. reports speaker fees from Roche Tissue Diagnostics. S.T. reports grants from Ventana, outside the submitted work. J.R. is a consultant for Achilles Therapeutics, and S.A.Q. is CSO and co-founder of Achilles Therapeutics. J.L. reports institutional research support from: BMS, MSD, Novartis, Pfizer, Achilles Therapeutics,

Additional information