

Systematic discovery of neoepitope–HLA pairs for neoantigens shared among patients and tumor types

Received: 22 November 2022

Accepted: 14 August 2023

Published online: 19 October 2023

 Check for updates

Hem R. Gurung^{1,3}, Amy J. Heidersbach^{1,3}, Martine Darwish^{1,3}, Pamela Pui Fung Chan¹, Jenny Li¹, Maureen Beresini¹, Oliver A. Zill¹, Andrew Wallace¹, Ann-Jay Tong¹, Dan Hascall¹, Eric Torres¹, Andy Chang¹, Kenny ‘Hei-Wai’ Lou¹, Yassan Abdolazimi¹, Christian Hammer¹, Ana Xavier-Magalhães¹, Ana Marcu¹, Samir Vaidya¹, Daniel D. Le¹, Ilseyar Akhmetzyanova¹, Soyung A. Oh¹, Amanda J. Moore², Uzodinma N. Uche², Melanie B. Laur², Richard J. Notturmo², Peter J. R. Ebert², Craig Blanchette¹✉, Benjamin Haley¹✉ & Christopher M. Rose¹✉

The broad application of precision cancer immunotherapies is limited by the number of validated neoepitopes that are common among patients or tumor types. To expand the known repertoire of shared neoantigen–human leukocyte antigen (HLA) complexes, we developed a high-throughput platform that coupled an *in vitro* peptide–HLA binding assay with engineered cellular models expressing individual HLA alleles in combination with a concatenated transgene harboring 47 common cancer neoantigens. From more than 24,000 possible neoepitope–HLA combinations, biochemical and computational assessment yielded 844 unique candidates, of which 86 were verified after immunoprecipitation mass spectrometry analyses of engineered, monoallelic cell lines. To evaluate the potential for immunogenicity, we identified T cell receptors that recognized select neoepitope–HLA pairs and elicited a response after introduction into human T cells. These cellular systems and our data on therapeutically relevant neoepitopes in their HLA contexts will aid researchers studying antigen processing as well as neoepitope targeting therapies.

T cells play a critical role in eliminating cancer cells^{1,2}, and immunotherapies that enhance endogenous tumor-specific T cell activity (for example, cancer vaccines) or introduce T cells that target neoantigens have shown clinical efficacy^{3,4}. Several neoantigen-directed therapies require the presentation of neoepitopes—8–11 amino acid peptides derived from mutated proteins—by polymorphic human leukocyte antigen class I (HLA-I, hereafter HLA) molecules on the surface of tumor cells. T cell receptors (TCRs) interact with a cognate peptide in the

context of an HLA complex such that the therapeutic target comprises both the neoepitope sequence and the HLA subtype. Despite promise in the clinic, the application of neoantigen-specific therapeutics is limited, at least in part, by the scope of verified neoepitope–HLA combinations, particularly those that may be common across tumor subtypes or patient populations.

Precision T cell therapies target two broad categories of neoantigens: private and shared⁵. Private neoantigens are somatic

¹Genentech, South San Francisco, CA, USA. ²Adaptive Biotechnologies, Seattle, WA, USA. ³These authors contributed equally: Hem R. Gurung, Amy J. Heidersbach, Martine Darwish. ✉e-mail: cdblanchette@gmail.com; ben.haley@gmail.com; rose.christopher@gene.com

variants unique to an individual's tumor and represent the majority of mutations that arise during cancer progression¹. Shared neoantigens recur across many patients due to common oncogenic mutations in proteins such as KRAS, EGFR, TP53 and BRAF^{2,3}. Prior knowledge of shared neoantigens could enable prioritization of target epitopes and a path toward off-the-shelf therapeutics for patients with an appropriate tumor mutation and HLA haplotype.

Discovery of shared neoepitopes presented in their native context is challenging due to the tremendous number of possible neoepitope–HLA combinations. For each coding variant, there are 38 possible 8–11 amino acid epitopes with the potential to bind to thousands of distinct HLA alleles depending on the amino acid composition of the neoepitope. If considering only 15 HLA alleles and 50 shared cancer neoantigens, more than 28,000 neoepitope–HLA pairs could be formed. Although progress has been made toward the development of computational methods to predict neoepitope–HLA binding events⁴, they are not yet fully able to identify which peptides are processed and presented in a cellular context.

We present here a scalable pipeline for neoepitope–HLA pair discovery. For this, we selected 47 cancer mutations and 15 prevalent HLA alleles to define the neoepitope landscape and, by extension, putative clinical targets. We then employed a high-throughput HLA binding assay^{5,6} and NetMHCpan-4.0 (ref. 7) to experimentally and computationally identify neoepitope–HLA combinations for follow-up. Neoepitope–HLA pairs observed through both methods were assayed for presentation using untargeted and targeted mass spectrometry (MS) analyses of HLA monoallelic cell lines modified to express ~25 amino acid segments corresponding to each of the 47 cancer neoantigens, resulting in 86 observed neoepitope–HLA pairs. To assess the therapeutic potential for these targets, we used TCRs discovered through Multiplex Identification of T cell Receptor Antigen (MIRA) (Adaptive Biotechnologies)⁸ assays and demonstrated mutant-selective T cell targeting of cells expressing these neoepitopes.

Results

Clinico-genomics analysis of shared cancer neoantigens

To establish a computational and experimental pipeline for neoepitope–HLA discovery, we first identified the most common recurrent point mutations across cancer types within a compendium of sequencing data from tumor and normal tissue samples⁹, filtering at a per-indication case prevalence of 2% (Fig. 1a). This led to a list of 36 shared cancer neoantigens (Supplementary Table 1). Next, we mined the Allele Frequency Net Database (AFND) and The Cancer Genome Atlas (TCGA) to catalog common haplotypes, narrowing to those with a carrier frequency of at least 10% in TCGA and an allele frequency of at least 5% in AFND. This analysis led to a list of 16 HLA alleles that combined with the 36 selected neoantigens to provide the foundation for development of our platform (Supplementary Table 1).

High-throughput TR-FRET analysis of neoepitope–HLA stability

To survey all potential neoepitopes between candidate cancer neoantigens and selected HLA alleles, a high-throughput time-resolved fluorescence energy transfer (TR-FRET) assay based on peptide-mediated stabilization of conditional HLA complexes was developed (Fig. 1b)⁵. Our neoantigen target set consisted of 36 shared cancer neoantigens identified above along with 11 additional tumor antigens. Separately, 15 of 16 prioritized HLA variants were viable in the conditional HLA complex format (Fig. 1b). Together, this permitted the characterization of 24,149 neoepitope–HLA complexes after eliminating overlapping cancer neoepitopes as well as one allele due to synthesis challenges (Supplementary Table 2).

Conditional HLA complexes, pre-loaded with ultraviolet (UV)-cleavable peptides, were incubated with a neoepitope of interest at 100-fold molar excess and exposed to UV light for 25 min. This reaction leads to conditional ligand cleavage and conversion of the peptide

from a stable high-affinity 'binder' to an unstable binder that dissociates from the HLA groove. In the presence of a binding neoepitope, peptide exchange would stabilize the HLA complex, whereas a lack of binding results in complex dissociation. Complex stability was monitored using fluorescence of a TR-FRET donor (europium) conjugated to an anti-β2M antibody and a TR-FRET acceptor conjugated to streptavidin, which bound to the biotin tag on the HLA alpha chain, where a TR-FRET signal would be observed only if the complex remained intact. TR-FRET signals were quantified based on the ratio of relative fluorescent units (RFUs), and signals were subjected to a double normalization to generate a robust Z-score (RZ-score). Any neoepitope–HLA combination with an RZ-score ≥5 was considered a 'stable binder', which was a conservative measure based on prior assessment with our positive control CMV-peptide/HLA-A*02:01 complex. A Z-score ≥5 captured 90% of positive control binding events without identifying false-positive binders (Fig. 1c).

NetMHCpan-4.0 (hereafter NetMHC) was employed to better understand how our TR-FRET results compared to computational prediction methods^{10,11}. We considered NetMHC binding affinity (BA) percentile rank (%Rank) relative to TR-FRET results and eluted ligand (EL) %Rank to determine if a neoepitope was predicted to be presented (%Rank ≤2). Representative data for KRAS G12V peptides binding to A*03:01 showed two previously described neoepitopes^{12,13}, VVGAVGVGK and VVVGAVGVGK, as binders with both approaches (Fig. 1d). Further examination of neoantigen–HLA combinations revealed variable concordances between the TR-FRET and NetMHC results (Supplementary Fig. 1a,b). Assessment of KRAS G12D peptides with C*08:02 found a known neoepitope (GADGVGKSAL)¹⁴ to be a binder by both methods (Supplementary Fig. 1c).

When measured as a percentage of all potential neoepitope–HLA complexes, TR-FRET generally identified more stable binders as compared to NetMHC (Fig. 1e and Supplementary Fig. 1d). We found that the percent agreement between NetMHC prediction and TR-FRET when classifying binders was generally less than 30% (Fig. 1f and Supplementary Fig. 1e), whereas much stronger agreement was found for non-binders only (Supplementary Fig. 1f,g). For further comparison, TR-FRET RZ-scores and NetMHC %Ranks were plotted for all candidate neoepitope–HLA pairs, demonstrating that 0.63% of neoepitope–HLA pairs were probable binders by both methods (Fig. 1g and Supplementary Fig. 2a). The different methods identified a similar percentage of additional binding events for neoepitope–HLA pairs, demonstrating that each has the potential to uncover unique binding combinations (Fig. 1g and Supplementary Fig. 2b). These findings highlight the power of our high-throughput TR-FRET assay to identify an expanded and complementary set of neoepitope–HLA pairs relative to computational prediction and suggest that co-deployment of both approaches would be needed for comprehensive neoepitope discovery.

Generation of monoallelic cells co-expressing 47 neoantigens

Despite observed peptide–HLA stabilization *in vitro* or computational prediction of an interaction, mutant protein expression and processing may not result in neoepitope presentation in a cellular context^{15–17}. For this reason, candidate neoepitope validation typically requires evidence of direct physical association with surface-bound HLA via HLA immunoprecipitation (HLA-IP) followed by MS. This process has been enhanced through the use of engineered 'HLA monoallelic' cell lines, although these have largely relied on endogenous mutant protein expression or expression of relatively few mutant transgenes, thus limiting throughput^{13,18,19}.

We anticipated that co-expression of all 47 candidate neoantigen sequences (concatenated ~25 amino acid segments centered on the mutated position) within a single HLA-null cell line would improve throughput of monoallelic cell line generation and subsequent validation of TR-FRET/NetMHC-identified neoepitope–HLA pairs by targeted MS (Fig. 2a). For this, we selected the HMy2.C1R (C1R) lymphoblast cell

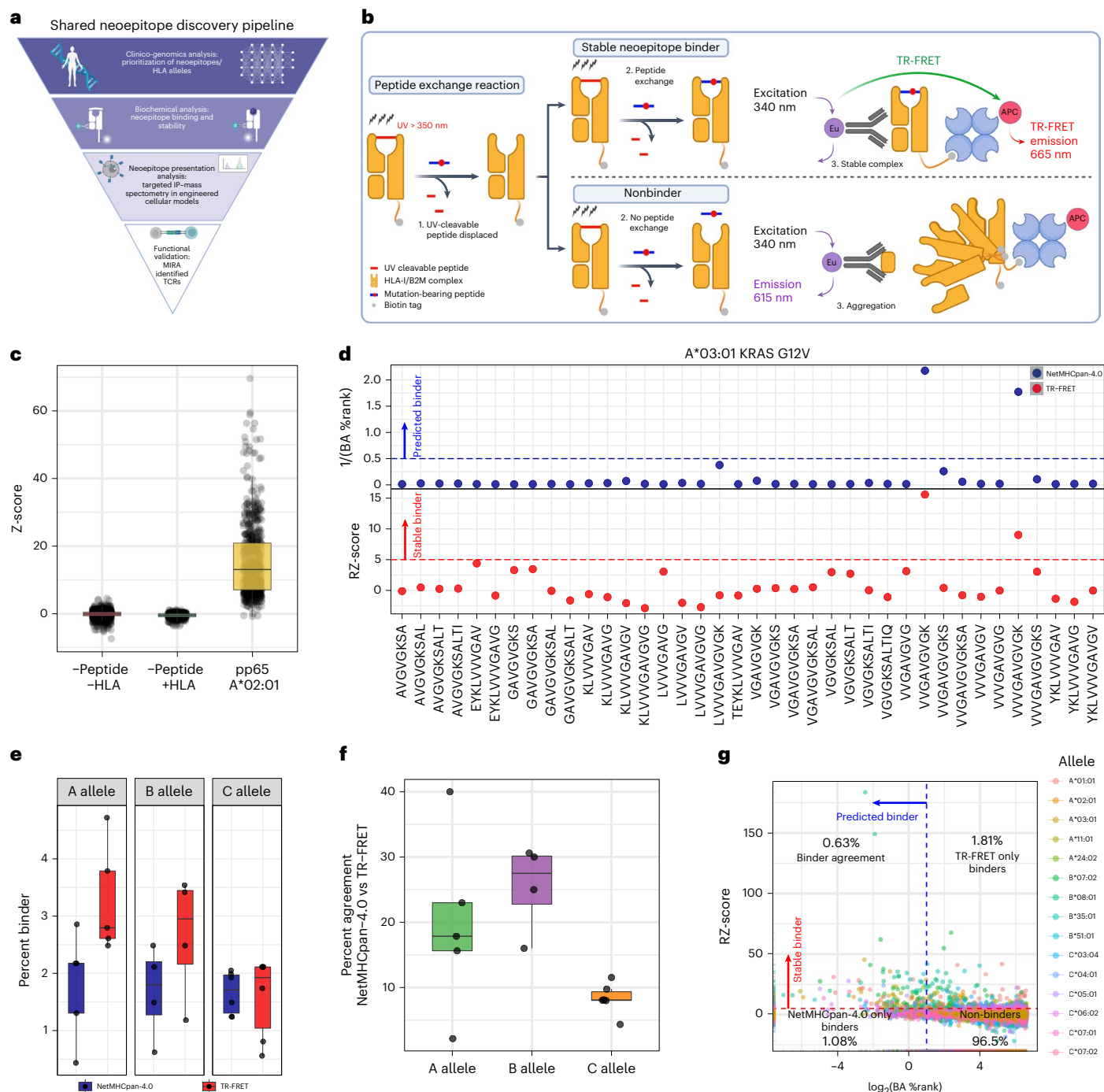


Fig. 1 | A shared neopeptide discovery pipeline featuring characterization of neopeptide-HLA binding through a high-throughput TR-FRET assay and NetMHCpan-4.0 prediction. **a**, Overview of the shared neopeptide discovery pipeline. **b**, Schematic diagram of the TR-FRET assay used to measure stable neopeptide-HLA binding. In brief, HLA monomers bound to UV-cleavable peptides are exposed to UV light in the presence of mutation-bearing candidate neopeptides. Successful exchange of the candidate peptide will lead to complex stabilization and TR-FRET emission (top). Unsuccessful exchange will lead to aggregation and no TR-FRET emission (bottom). **c**, TR-FRET data for all controls measured within the screen. Z-score was calculated as compared to -Peptide/+HLA controls for each 384-well plate. The box represents the interquartile range; the line represents the median value; and the whiskers represent the minimum and maximum values (excluding outliers). Each dot represents an individual well measurement (-Pep, -HLA $n = 1,477$; -Pep, +HLA $n = 368$; pp65, +A*02:01 $n = 623$). **d**, Representative TR-FRET results for KRAS G12V/A*03:01 comparing NetMHC BA percentile rank (%Rank, blue) and RZ-score (red). **e**, Percent of neopeptide-HLA combinations that were determined to be

stable binders by TR-FRET (red) or NetMHC (blue) across the HLA A, B and C alleles. The box represents the interquartile range; the line represents the median value; and the whiskers represent the minimum and maximum values (excluding outliers). Each dot represents the count of binders for a single allele (A allele, NetMHC $n = 5$; A allele, TR-FRET $n = 5$; B allele, NetMHC $n = 4$; B allele, TR-FRET $n = 4$; C allele, NetMHC $n = 6$; C allele, TR-FRET $n = 6$). **f**, Percent of neopeptide-HLA pairs found to be binders by NetMHC and/or TR-FRET across a single allele. The box represents the interquartile range; the line represents the median value; and the whiskers represent the minimum and maximum values (excluding outliers). Each dot represents the percent agreement for each allele (A allele $n = 5$, B allele $n = 4$, C allele $n = 6$). **g**, Scatter plot of TR-FRET RZ-score and NetMHC BA %Rank. The dashed red line represents the cutoff for stable binders as measured by TR-FRET, where values higher than the red line are considered a stable binder. The dashed blue line represents the cutoff for binders based on NetMHC analysis, where values lower than the blue line are considered binders. **a** and **b** were created with BioRender.com.

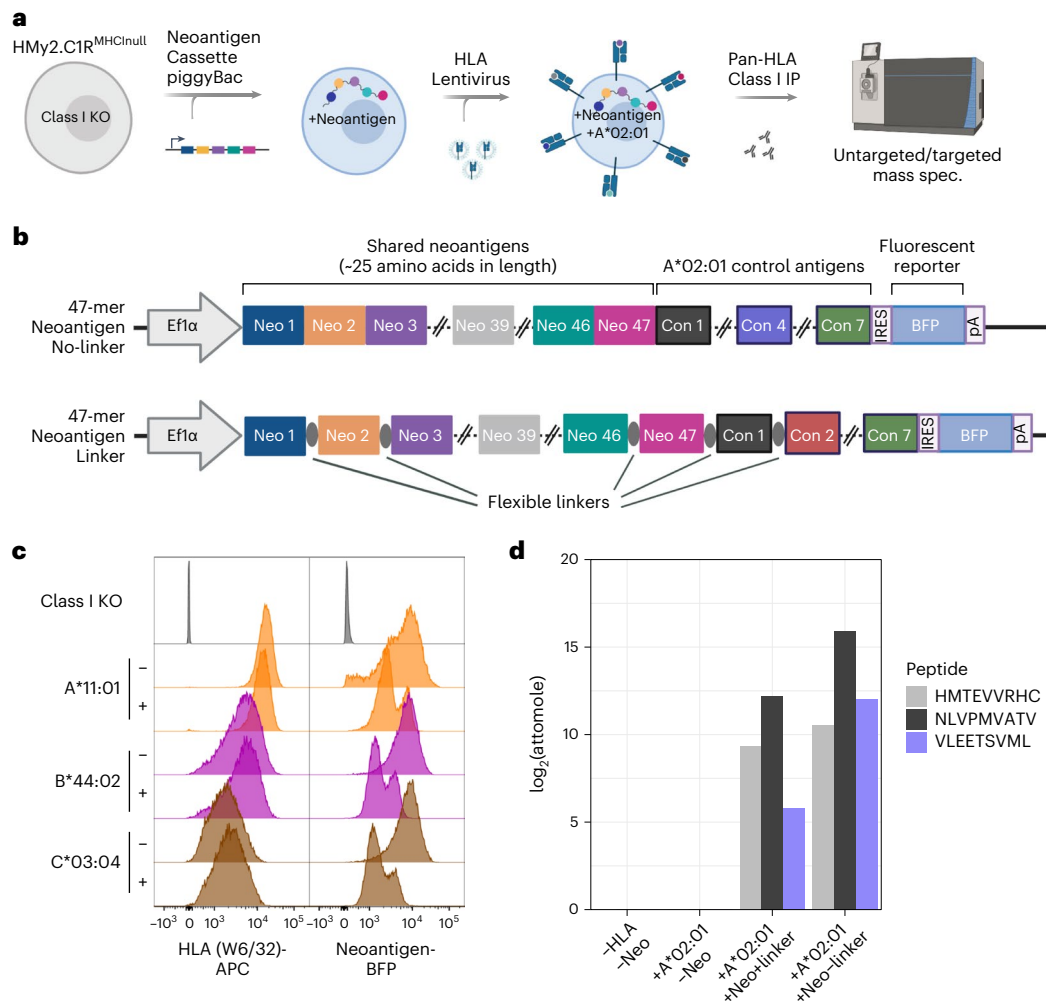


Fig. 2 | Generation of HLA class I monoallelic cell lines that stably express a polyantigen cassette containing 47 shared cancer neoantigens. **a**, Process overview for the cell engineering and MS analysis of peptides presented by polyantigen-expressing HLA monoallelic cell lines. KO, knockout. **b**, Vector map of the piggyBac polyantigen expression constructs used in this study. A single transcript containing 47 tandem neoantigens followed by seven control peptides and an IRES-linked mTagBFP2 (BFP) reporter is driven by an EF1a promoter. Neoantigens were either directly concatenated (no-linker) or interspersed by short flexible linker sequences (linker). **c**, Flow cytometric detection of HLA

expression (W6/32 antibody-APC) and polyantigen cassette reporter (BFP) in selected cell lines or HLA knockout parental line. Here, ‘-’ indicates the absence of linkers and ‘+’ indicates presence of linkers in the polyantigen construct. **d**, Targeted immunopeptidomic detection of a previously described A*02:01-presented TP53 neopeptide (HMTEVVRHC, position 39) as well as two neoantigen control peptides from pp65 (NLVPMVATV, control 1) and IE-1 (VLEETSVML, control 4) known to be presented by the A*02:01 allele. **a** and **b** were created with [BioRender.com](https://www.biorender.com).

line, which lacks HLA-A and HLA-B^{20,21}. To generate a full C1R^{HLA^{null}} cell line, the HLA-C allele (HLA-C*04:01) was disrupted using CRISPR-Cas9, and the HLA-null population was enriched by fluorescence activated cell sorting (FACS) (Supplementary Fig. 3a–c).

Local amino acid sequence context may affect antigen processing²². Accordingly, we engineered unique C1R^{HLA^{null}} lines to stably express concatemers of all 47 prioritized neoantigens that were separated, or not, by short, flexible amino acid linkers (Fig. 2b). Subsequent introduction of the 15 HLA alleles as individual transgenes through stable lentiviral transduction of the linker and no-linker neoantigen-expressing C1R^{HLA^{null}} cell lines resulted in 30 total cell populations (Fig. 2c and Supplementary Fig. 3d,e). To validate functionality of the polyantigen cassettes, the linker and no-linker neoantigen constructs contained an identical set of seven known HLA-A*02:01-presented epitopes. HLA-IP followed by targeted MS analysis confirmed presentation of two control peptides and a previously described TP53 R175H²³ neopeptide in both the linker and no-linker HLA-A*02:01-engineered cells (Fig. 2d).

Detection of neopeptides presented on engineered monoallelic cells

Both targeted and untargeted MS were applied for neopeptide discovery across the panel of monoallelic cell lines. Untargeted MS analysis enabled unbiased identification of peptides from the entire immunopeptidome, whereas targeted analysis facilitated detection of peptides presented at low copies per cell but was constrained to prioritized sequences from our TR-FRET/NetMHC analyses.

Untargeted MS analysis was performed with a semi-automated workflow resulting in 218–6,663 unique 8–11-mer peptides identified from each cell population (Fig. 3a,b). The number of 8–11-mer peptides and general sequence features for each allele overlapped regardless of the polyantigen linker status (Supplementary Fig. 4a) and confirmed that presented peptides fit known motifs (Supplementary Fig. 5). Expression of the polyantigen cassette was confirmed by detection of control viral epitopes from A*02:01 and A*11:01 monoallelic cells (Supplementary Fig. 4b) as well as epitopes from an integrated blue

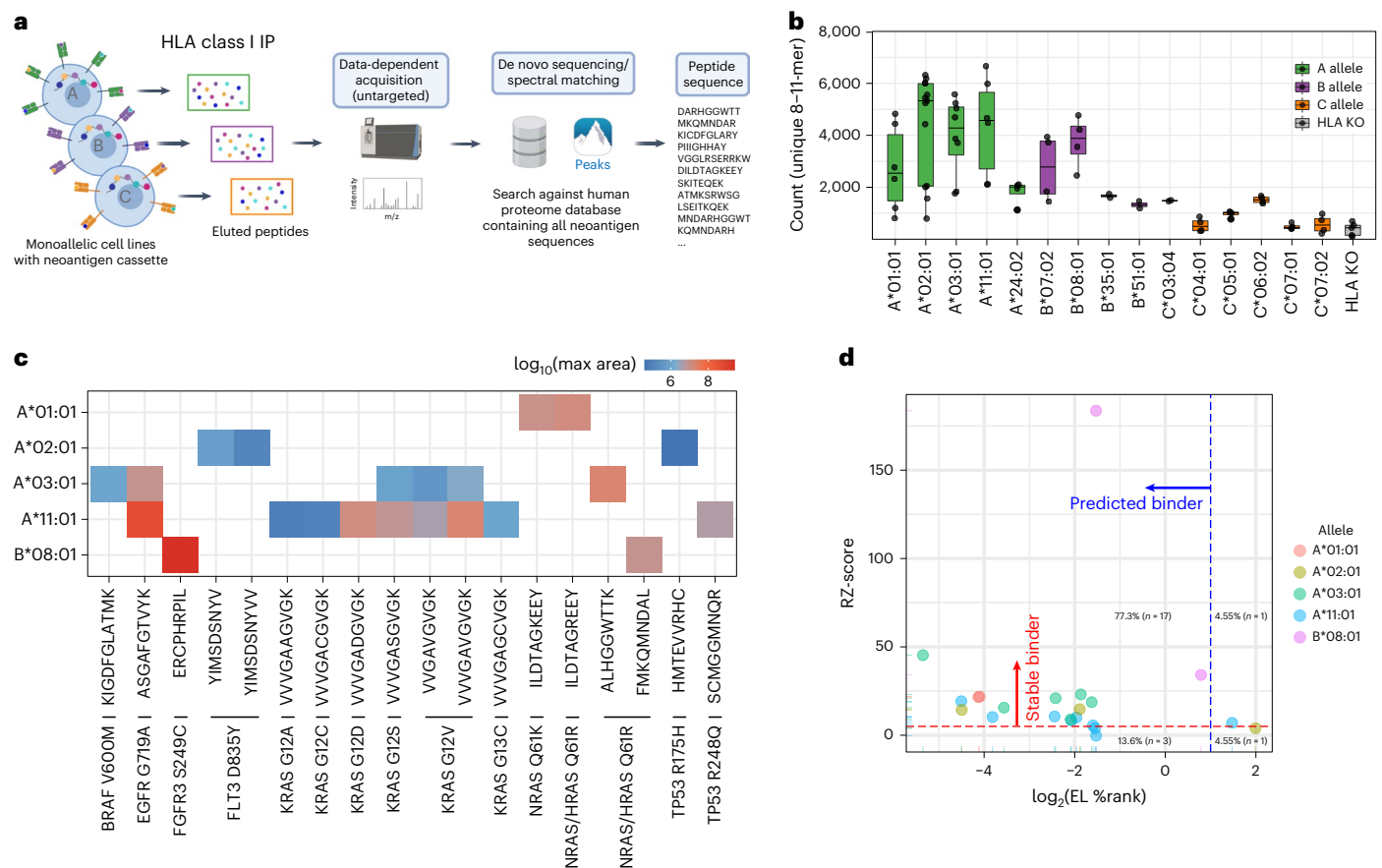


Fig. 3 | Untargeted immunopeptidomic analysis of monoallelic cell lines expressing the polyantigen cassette. **a**, Workflow for untargeted immunopeptidomic analysis of monoallelic cell lines containing the polyantigen cassette. **b**, Number of unique 8–11-mer peptides identified in untargeted immunopeptidomic analysis. The box represents the interquartile range; the line represents the median value; and the whiskers represent the minimum and maximum values (excluding outliers). Each dot represents a separate analysis beginning with a replicate cell pellet (A*01:01 $n = 6$, A*02:01 $n = 13$, A*03:01 $n = 8$,

A*11:01 $n = 6$, A*24:02 $n = 4$, B*07:02 $n = 4$, B*08:01 $n = 4$, B*35:01 $n = 2$, B*51:01 $n = 2$, C*03:04 $n = 2$, C*04:01 $n = 4$, C*05:01 $n = 4$, C*06:02 $n = 4$, C*07:01 $n = 4$, C*07:02 $n = 4$, HLA KO $n = 5$). **c**, Identified shared cancer neoantigen epitopes. The color scale represents the \log_{10} largest area across all analyses. **d**, Comparison of TR-FRET RZ-score and NetMHC EL percentile rank (%Rank) score for each epitope identified through untargeted immunopeptidomic analysis. **a** was created with [BioRender.com](https://www.biorender.com).

fluorescence protein (BFP) selection marker across eight different HLA alleles (Supplementary Fig. 4c).

From our untargeted analyses, we observed 22 neoepitope–HLA pairs and several peptides from non-mutation-bearing regions of the polyantigens. Neoepitopes corresponded to 15 shared neoantigens across five HLAs, representing $\sim 5.4\%$ of neoepitope–HLA pairs predicted by NetMHC and $\sim 3.7\%$ of neoepitope–HLA pairs identified within the TR-FRET assay (Fig. 3c and Supplementary Table 3). Of the 22 neoepitope–HLA pairs, 10 were previously described in the literature, and the remaining 12 were thought to be novel based on a search of T antigen²⁴, CAAtlas²⁵ and NEpdb²⁶ and an extended literature survey (Supplementary Table 3). TR-FRET and NetMHC showed excellent concordance for all 22 identified pairs; 17 were identified as binders by both approaches (Fig. 3d). One and three neoepitope–HLA pairs were uniquely identified as hits by TR-FRET and NetMHC, respectively, demonstrating that each approach can predict distinct neoepitope subsets (Fig. 3d). One neoepitope–HLA pair (TP53 R175H (HMTTEVVRHC)/A*02:01) represented an exception. This pair had a TR-FRET RZ-score of 3.9 and a NetMHC EL %Rank of 3.98 and was not considered a hit by either approach, demonstrating that false negatives remain possible.

Although we surmised that targeted MS analysis would improve detection of presented neoepitopes, this relied on heavy

isotope-labeled standard peptides. As such, a logistically challenging synthesis of 1,786 peptides (47 neoantigens \times 38 possible mutation-bearing candidate neoepitopes) would be needed to screen all potential neoepitopes from our monoallelic cell lines. Therefore, we used the TR-FRET results as a preliminary screen and synthesized all 397 peptides with an RZ-score ≥ 5 . Due to the complementarity of TR-FRET and NetMHC results, an additional 81 peptides were synthesized that had an RZ-score < 5 and NetMHC %Rank ≤ 2 . The 479 peptides were divided into 15 HLA allele-specific pools comprising 21–88 peptides (Fig. 4a,b).

Targeted MS analysis identified 86 neoepitope–HLA pairs across 12 different alleles and 36 neoantigens, representing a ~ 4 -fold improvement compared to untargeted MS analysis (Fig. 4b and Supplementary Table 3). After a search of the literature and relevant databases, we determined that 21 of the neoepitope–HLA pairs were described previously, and 65 were novel (Supplementary Table 3). Twenty of 86 neoepitope–HLA pairs identified across untargeted and targeted analyses were associated with A*11:01. This was likely due to the presence of eight distinct KRAS neoantigen sequences in the polyantigen cassette, as 14 of 20 A*11:01-specific and nine of 14 A*03:01-specific neoepitopes mapped to KRAS G12X or G13X neoantigens.

To assess the value of using TR-FRET and NetMHC results to select peptides for targeted MS, we plotted RZ-score versus NetMHC

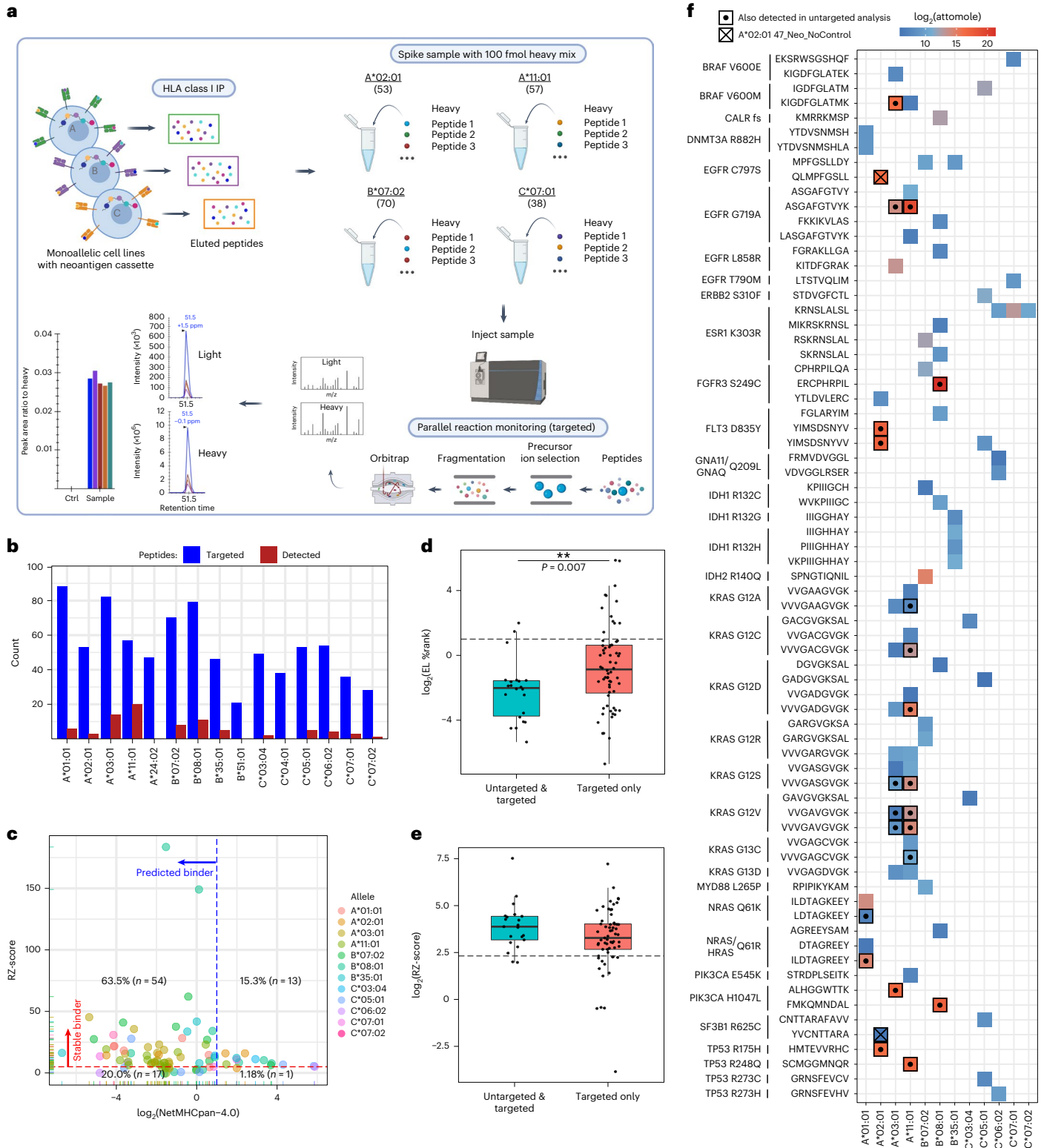


Fig. 4 | Targeted immunopeptidomic analysis of monoallelic cell lines expressing a polyantigen cassette. **a**, Targeted immunopeptidomic workflow for the analysis of candidate neopeptides within monoallelic cell lines expressing a polyantigen cassette. **b**, The number of targeted (blue) and detected (red) shared cancer neoantigen epitopes within each targeted assay. **c**, Comparison of TR-FRET RZ-score and NetMHC EL %Rank score for each epitope identified through targeted immunopeptidome analysis. **d**, NetMHC EL %Rank scores for neopeptides detected in both untargeted and targeted (teal) analysis or targeted analysis alone (red). The box represents the interquartile range; the line represents the median value; and the whiskers represent the minimum

and maximum values (excluding outliers). Each dot represents a neopeptide-HLA pair (untargeted & targeted $n = 22$, targeted only $n = 64$). The P value was calculated using a Wilcoxon test (two-sided). **e**, Same analysis as **d** but for TR-FRET RZ-scores. **f**, Summary of neopeptide-HLA pairs detected from shared cancer neoantigens. Color represents attomole of neopeptides detected on column during analysis. Bolded squares with centered dots represent neopeptides also detected in untargeted analysis. A square with an 'X' indicates A*02:01-specific neopeptides that were detected in a cell line containing a polyantigen construct lacking control neoantigen sequences. **a** was created with [BioRender.com](https://www.biorender.com).

%Rank for each of the observed 86 neoepitope–HLA pairs (Fig. 4c). This revealed that 55 neoepitopes were stable binders by TR-FRET and predicted to be presented by NetMHC. Thirteen neoepitope–HLA pairs were found as hits in TR-FRET only, whereas 18 neoepitope–HLA pairs were hits in NetMHC alone (Fig. 4c). To understand the binding characteristics of neoepitope–HLA pairs identified by targeted analysis alone, we plotted RZ and NetMHC %Rank scores for peptides observed in both untargeted and targeted analysis compared to peptides found only in targeted analysis (Fig. 4d,e). Neoepitope–HLA pairs identified by targeted analysis alone had a broader range of NetMHC %Rank and TR-FRET RZ-scores relative to neoepitopes also detected in untargeted analysis (Fig. 4d,e). This suggests that targeted analysis can identify neoepitopes that are weaker binders compared to those observed by untargeted means.

Targeted MS permits absolute quantification of peptide presentation across neoepitopes. Overall, the measured amount of neoepitope presentation spanned from 60 amol to 2.5 pmol (Fig. 4f) and was consistent across independent replicates of cell line growth and sample preparation (Supplementary Fig. 6). Two peptides detected by untargeted MS, EGFR G719A (ASGAFGTVYK) and FGFR3 S249C (ERCPHPIL), exhibited the highest absolute quantities (Fig. 4f). When the absolute amounts of neoepitopes detected were compared to RZ-score, NetMHC EL %Rank or NetMHC BA %Rank for each allele, no clear correlation could be found (Supplementary Fig. 7a–c). This suggests that each score has predictive value for neoantigen presentation but also that these cannot be used to estimate the absolute amount presented.

Polyantigen cassette design impacts neoepitope presentation

The polyantigen sequence included neoantigens with known A*02:01-binding epitopes to confirm translation, processing and presentation of the cassette. It was possible that the controls could compete with experimental neoepitopes, thus creating an avenue for false negatives. To evaluate this, a separate A*02:01 cell line was created that stably expressed a no-linker polyantigen cassette lacking control sequences. Upon analysis of the no-control line, two additional neoepitopes were detected: YVCNTTARA (SF3B1 R625C; RZ-score = 16; EL %Rank = 5.3) and QLMPFGSL (EGFR C797S; RZ-score = 7; EL %Rank = 0.21) (Fig. 4f, squares with 'X'). These results suggest that strong binding peptides could inhibit presentation of certain neoepitopes, and a revised workflow may omit control sequences from the polyantigen cassette.

Polyantigen cassette length is an important consideration when designing cancer vaccines, and a concern that translation of neoantigens at the C-terminal/3' end of the cassette will be decreased may have factored into the use of shorter cassettes in clinical settings (for example, 10-mer or 34-mer)²⁷. To characterize the translation efficiency of our 47-mer polyantigen transgene, we performed ribosome profiling (Ribo-Seq) on A*02:01 monoallelic cells containing either the linker or no-linker cassette with A*02:01 controls (Supplementary Fig. 8a,b). These analyses demonstrated consistent translation across the no-linker polyantigen cassette, whereas the cassette containing linkers had a substantial decrease in translation after ~20 neoantigen sequences (Supplementary Fig. 8a,b).

We next sought to understand if the difference in translation between cassette designs was reflected within our targeted immunopeptidomics results. For this, we plotted the highest attomole abundance of presented peptide for each neoantigen (irrespective of HLA) versus neoantigen position within the linker and no-linker polyantigen cassettes (Supplementary Fig. 8c). This revealed a potential bias toward presentation of peptides derived from the first ~20 neoantigen sequences regardless of format. Within the portion of the polyantigen cassette that exhibits lower translation, we detected six additional neoepitope–HLA pairs from cells expressing the no-linker cassette, suggesting that the no-linker format may be advantageous for assaying ≥ 20 target sequences (Supplementary Fig. 8c).

For neoepitopes detected in both the linker and no-linker cell lines, there was not a clear difference in the maximum presentation, suggesting that positional effects detected in the Ribo-Seq data could be buffered at the level of presentation (Supplementary Fig. 8c). This was further supported by roughly equivalent presentation of KRAS G12X and G13X neoepitopes (which are identical except for the mutated residue) across the polyantigen cassette (Supplementary Fig. 8d). To evaluate the impact of linkers more broadly, we plotted the highest absolute amount of neoepitope presented and found that presentation of some neoepitopes increased in the presence of linkers while presentation of other neoepitopes decreased (Supplementary Figs. 9a,b and 10). Together, these data demonstrate that the no-linker polyantigen cassette enabled detection of a greater number of neoepitope–HLA pairs. However, if a neoepitope was detected in linker and no-linker cells, the presence of linkers did not impact abundance of presentation in a consistent manner.

Validation of neoepitope presentation from full-length protein

Neoepitopes derived from a polyantigen construct may not reflect peptides processed from a full-length mutant protein. To address this, we developed four HLA-A*11:01 monoallelic C1R lines expressing an inducible, full-length wild-type, G12C, G12D or G12V mutant KRAS transgene and compared neoepitope presentation from these cell lines with a cell line expressing the same HLA and a no-linker polyantigen cassette. Expression of full-length variant proteins was confirmed using a whole-cell targeted proteomic assay comprising a peptide that can detect total KRAS as well as three unique peptides that measured individual KRAS mutants (Fig. 5a). Little to no mutant peptide signal was observed in total protein samples from the polyantigen cassette-expressing cell line (Fig. 5a).

We then performed HLA-IP and targeted MS to quantify presentation of previously identified 9-mer and 10-mer KRAS epitopes associated with HLA-A*11:01 (Fig. 5b)^{12,13}. In cell lines expressing full-length mutant transgenes, clear induction of neoepitope presentation was observed for both G12V epitopes as well as the 10-mer epitope of G12D (Fig. 5b). From cells expressing the polyantigen cassette, all targeted mutant KRAS epitopes were detected and measured at higher absolute copies per cell compared to lines expressing full-length mutant proteins (Fig. 5b). Detection of KRAS peptides after HLA-IP but not from total cell protein suggested that the polyantigen concatemer was likely unstable and efficiently degraded, resulting in enhanced epitope presentation^{28,29}. Therefore, monoallelic cells containing the polyantigen cassette provided a reliable, higher throughput and more sensitive system for discovery of neoepitopes from shared cancer neoantigens relative to cell lines expressing a full-length antigen.

Lastly, we sought to demonstrate that neoepitopes discovered by our pipeline can be identified within cells that endogenously co-express relevant proteins and HLA alleles. Targeted MS assays were used to quantify four neoepitopes—9-mer and 10-mer from KRAS G12C and G12D—within cell lines that express A*11:01 as well as KRAS G12C (HOP62 and NCIH2030), KRAS G12D (HuCCT1 and SNU601) or KRAS G12V (SW527) (Fig. 5c). One of these neoepitopes (KRAS G12C (VVGACG-VGK)) has not previously been described, whereas the remaining three neoepitopes have been confirmed only within cellular systems that exogenously express the neoantigen¹². We confirmed presentation of the four target neoepitopes within cell lines that harbor the target neoantigens (KRAS G12C and G12D), whereas there was no observed presentation in a control cell line that contained KRAS G12V (Fig. 5c). In both HOP62 and NCIH2030 cells, KRAS G12C 9-mer neoepitopes appeared to have higher absolute presentation as compared to the previously described 10-mer (Fig. 5c). Furthermore, whereas the presentation of the 10-mer KRAS G12D epitope was similar across HuCCT1 and SNU601 cells, presentation of the 9-mer KRAS G12D neoepitope was much higher within HuCCT1 (Fig. 5c). This suggests that presentation

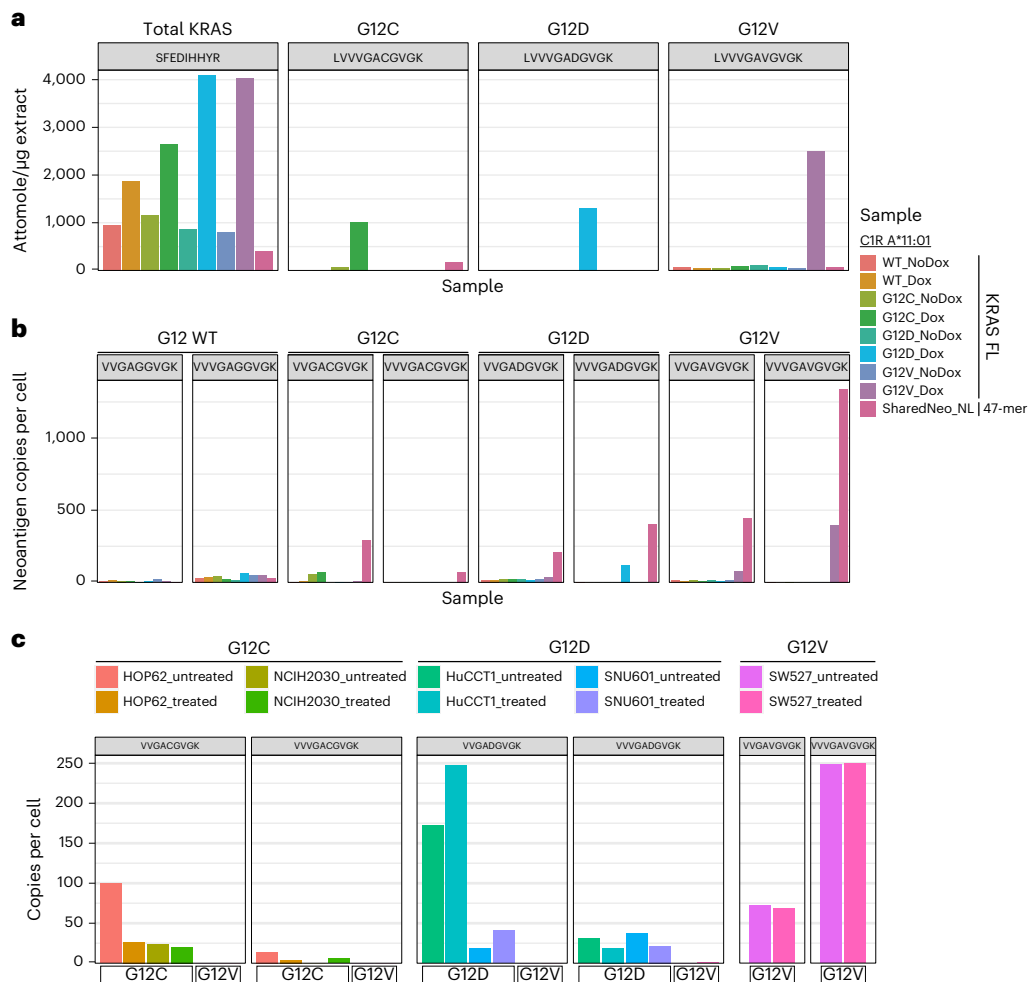


Fig. 5 | Presentation of KRAS neoepitopes derived from exogenous and endogenous expression of full-length mutant protein. a, b. A*11:01 monoallelic cells were engineered to express doxycycline (dox)-inducible full-length (FL) KRAS mutant proteins (G12C, G12D and G12V). These were compared against an A*11:01 monoallelic cell line containing the no-linker polyantigen cassette. **a**, Absolute amount of KRAS wild-type (WT) and mutant proteins in the cell lysate by targeted MS. **b**, Copies per cell of presented KRAS 9-mer (VVGAGGVGK) and 10-mer (VVGAGXGVGK) neoepitopes as measured by A*11:01 monomers

containing heavy synthetic neoepitope peptides spiked in before affinity purification and targeted MS. **c**, Targeted immunopeptidomic analysis of neoepitopes in cell lines that endogenously express both KRAS and A*11:01. Two neoantigens for KRAS G12C (VVGACGVGK and VVGAGGVGK) and KRAS G12D (VVGADGVGK and VVGADGVGK) were analyzed in cell lines that harbor KRAS G12C (HOP62 and NCIH2030), G12D (HuCCT1 and SNU601) or G12V (SW527). These neoepitopes are either novel or were described previously in non-endogenous systems. ‘Treated’ samples were treated with interferon-gamma.

of slightly varying neoepitopes can differ substantially based on the cell line from which they are derived. In total, these data demonstrate that neoepitopes discovered through our pipeline can be found from both exogenously expressed full-length proteins and within systems that endogenously express both the HLA and neoantigen.

Functional validation of tumor-specific antigen–HLA pairs

To determine whether neoepitopes identified through our workflow could be recognized by human T cells, we employed a modified multiplexed TCR discovery method⁸. Using two of the identified neoepitope–HLA pairs (FLT3-p.D835Y/A*02:01, PIK3CA-p.E545K/A*11:01) as examples, neoepitopes were first allocated to peptide pools in unique combinations before healthy human donor CD8⁺ T cells were expanded using autologous monocyte-derived dendritic cells, restimulated with the neoepitope peptide pools, sorted for activation marker upregulation and subjected to TCR β sequencing. This method was used for donors spanning a range of HLA genotypes, enabling the association of TCRs with a variety of peptide–HLA pairs. However, owing to the multiallelic nature of donor cells, the HLA restriction of identified

neoepitopes was not initially disambiguated among the 3–6 donor HLA alleles.

For neoepitopes that elicited a T cell response, associated TCR β and TCR α sequences were determined using a parallel multiplexed assay³⁰ that enabled construction of paired TCR expression vectors and the selection of candidate neoepitope-specific TCRs. The specificity and potential efficacy of each TCR were then assessed through cellular assays. TCR encoding in vitro transcribed mRNA was introduced via electroporation into primary human T cells, which were then incubated with either an increasing concentration of the candidate neoepitope in the presence of A*02:01⁺ T2 cells or monoallelic K562 cells that co-expressed an HLA allele and neoantigen of interest.

We found dose-dependent upregulation of CD137 after 12-h co-culture of primary human CD8⁺ T cells transfected with predicted FLT3-p.D835Y/A*02:01-specific TCRs in response to T2 cells incubated with exogenously delivered YIMSDSNVY peptide (Fig. 6a). Furthermore, these T cells were activated by and specifically killed monoallelic A*02:01-K562 cells expressing a mutant FLT3-p.D835Y transgene (minigene encoding 21 amino acids) but were not activated by and did

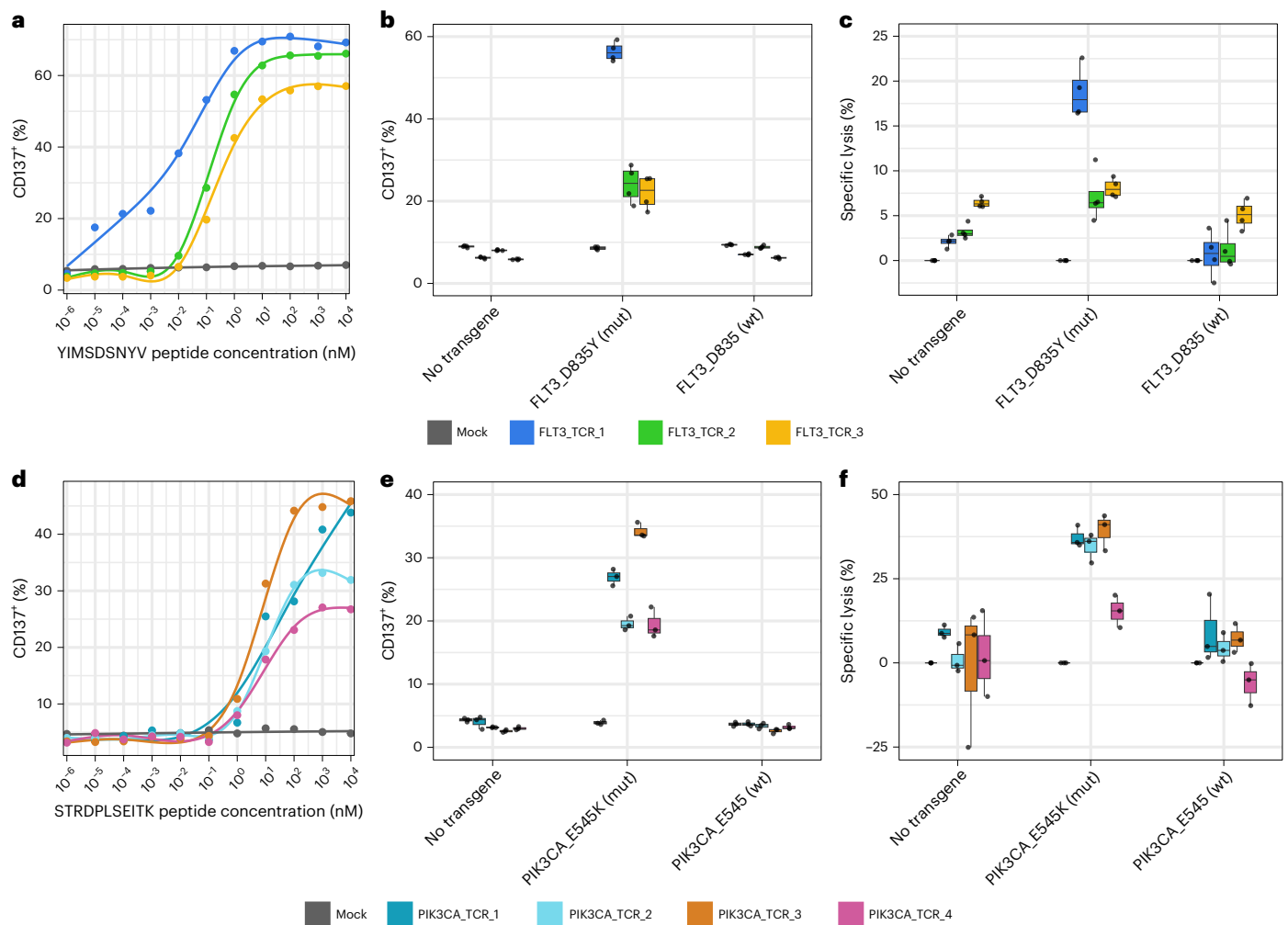


Fig. 6 | Discovery of neopeptide-specific TCRs demonstrates immunogenic potential of discovered neopeptide–HLA pairs. Human CD8⁺ T cells were transfected with either FLT3-p.D835Y-specific or PIK3CA-p.E545K-specific TCR RNA. **a–c**, CD137 expression was assessed after T cells transfected with FLT3-p.D835Y-specific TCRs were co-cultured overnight with YIMSDSNVY peptide-pulsed A*02:01⁺-T2 cells (**a**). CD137 expression (**b**) and specific lysis (**c**) were determined after co-culture with A*02:01⁺-K562 cells transfected with no RNA, transgene containing mutant FLT3-p.D835Y or transgene containing FLT3-D835 wild-type sequence. The box represents the interquartile range; the line represents the median value; and the whiskers represent the minimum and maximum values (excluding outliers) ($n = 4$). **d–f**, CD137 expression was assessed after T cells transfected with PIK3CA-p.E545K-specific TCRs were co-cultured overnight with STRDPLSEITK peptide-pulsed A*11:01⁺-K562 cells (**d**). CD137 expression (**e**) and specific lysis (**f**) were determined after co-culture with A*11:01⁺-K562 cells transfected with no RNA, transgene containing mutant PIK3CA-E545K or transgene containing PIK3CA-E545 wild-type sequence. The box represents the interquartile range; the line represents the median value; and the whiskers represent the minimum and maximum values (excluding outliers) ($n = 3$). mut, mutant; wt, wild-type.

not kill monoallelic A*02:01-K562 cells expressing a wild-type FLT3 transgene (Fig. 6b,c). These TCRs appear to be exquisitely specific for the mutant neopeptide, which is an important characteristic because a similar non-mutant epitope IMSDSNVV was identified by untargeted analysis in A*02:01 monoallelic cells.

As a second proof of concept, T cells were transfected with predicted PIK3CA-p.E545K/HLA-A*11:01 TCRs and mixed with monoallelic A*11:01-expressing K562 cells incubated with an increasing concentration of the predicted neopeptide STRDPLSEITK (Fig. 6d). Here, TCR-transfected T cells demonstrated dose-dependent activation as measured by CD137 expression. Furthermore, these T cells demonstrated higher levels of activation and cell killing when mixed with A*11:01 K562 cells expressing a PIK3CA-p.E545K transgene (minigene encoding 21 amino acids) as compared to cells that expressed a wild-type PIK3CA transgene (Fig. 6e,f). Mutations that introduce anchor residues are thought to have high immunogenic potential because the immune system has not built tolerance to a similar

wild-type epitope. For PIK3CA-p.E545K/A*11:01, the E → K mutation introduces an anchor residue within the context of A*11:01, and the wild-type STRDPLSEITE epitope was not detected in untargeted MS analyses of A*11:01 monoallelic cells. Although false negatives are anticipated in our MS workflow, the wild-type epitope was also not predicted to bind A*11:01 by NetMHC (12.8). Taken together, these data provide a clear mechanism for the specificity of PIK3CA-p.E545K TCRs for recognition of mutant PIK3CA as compared to wild-type and lend support for these TCRs as potential therapeutic candidates.

wild-type epitope. For PIK3CA-p.E545K/A*11:01, the E → K mutation introduces an anchor residue within the context of A*11:01, and the wild-type STRDPLSEITE epitope was not detected in untargeted MS analyses of A*11:01 monoallelic cells. Although false negatives are anticipated in our MS workflow, the wild-type epitope was also not predicted to bind A*11:01 by NetMHC (12.8). Taken together, these data provide a clear mechanism for the specificity of PIK3CA-p.E545K TCRs for recognition of mutant PIK3CA as compared to wild-type and lend support for these TCRs as potential therapeutic candidates.

Discussion

Most neopeptide discovery efforts have focused on a limited number of neoantigens and HLA alleles in the search for presented tumor-associated peptides^{12,31,32}. We, therefore, developed a scalable, multiplexed platform that integrates a high-throughput binding assay, computational neopeptide binding prediction, complex cellular engineering of monoallelic cell lines and targeted MS to identify dozens of

unique tumor-associated neoepitopes in context with specific HLA alleles, representing putative targets for neoantigen-based cancer immunotherapies.

In total, 24,149 potential neoepitope–HLA pairs were surveyed from 47 shared cancer neoantigens across 15 common HLA alleles, resulting in 844 stable combinations. From this, subsequent proteomic assessment using monoallelic cell lines identified 86 unique neoepitope–HLA pairs derived from 36 neoantigens across 12 HLA alleles. We selected two example combinations (FLT3-p.D835Y/A*02:01 and PIK3CA-p.E545K/A*11:01) for cell-based assays to validate a cohort of TCRs identified in a separate MIRA workflow, which demonstrated T cell activation or target cell killing and mutant peptide selectivity.

Despite a high rate of rediscovery for known peptide–HLA combinations with our platform, only a fraction of those found here were evaluated by HLA-IP-MS using cells that natively express the neoantigens and HLAs or express full-length mutant cDNAs. Also, our T cell/target cell co-culture assays relied on peptide pulsing or expression of the neoantigen from minigenes. A demonstration that T cells can be modified to target cells with endogenous expression of the newly observed neoepitope–HLA pairs would further substantiate our findings^{13,23,33}. However, a paucity of appropriate cell lines poses a challenge to the study of endogenous neoepitope presentation, which may explain why only six of 21 previously reported neoepitope–HLA pairs described in the literature have been validated in a native context. This includes the KRAS G12V¹³, PIK3CA H104L³³ and TP53 R175H²³ neoepitope–HLA pairs validated in T cell targeting assays; neoepitope–HLA pairs from NRAS/HRAS Q61R¹⁹, NRAS Q61K³⁴ and KRAS G12V^{12,13,35} that were detected through MS alone and a neoepitope–HLA pair from MYD88 L265P³⁶ that was detected through ELISpot (Supplementary Table 3).

We extended this list by validating additional 9-mer and 10-mer neoepitope–HLA pairs from cells endogenously expressing either KRAS G12C or G12V and A*11:01 and found that copy/cell levels of neoepitope presentation as well as the relative ratios of 9-mer to 10-mer presentation varied across cell lines. This could have been due to differences in KRAS abundance in the cell and/or expression of genes involved in antigen processing, but a broader study of presentation across endogenous cell lines could reveal important insights into KRAS neoepitope presentation.

As described in previous studies, detection of neoepitopes by MS may be impacted by the amino acid composition of the peptides³⁷. Thirty-four of 86 unique neoepitope–HLA pairs that we observed were associated with either A*03:01 or A*11:01. This was potentially due to the overrepresentation of KRAS variants in the polyantigen cassette but may also be due to a basic residue (lysine or arginine) at the C-terminus. Additional charges, either through additional basic residues or labeling with a chemical tag^{34,38}, generally improve ionization, fragmentation and identification of peptides. The cysteine-containing HMTEVVRHC (TP53 R175H, A*02:01) was the only peptide that failed to reach significance by NetMHC or TR-FRET but was found by untargeted analysis. At least two details may explain this: cysteines have been underrepresented in MS data used to train prediction algorithms, and these residues can cause peptide dimerization in solution.

The datasets and tools that we developed represent a valuable and expandable resource for future studies of neoepitope presentation. For example, the TR-FRET dataset could be used for training or benchmarking neoepitope prediction algorithms that factor in neoepitope–HLA complex formation. Additionally, we provide raw data for untargeted and targeted MS analysis, enabling re-analysis with improved search algorithms³⁴, peptide false discovery rate (FDR) determination³⁸ or specific workflows that detect rare events within the antigen presentation pathway³⁹. Monoallelic cell lines expressing the polyantigen cassette also represent a versatile system for characterizing the processing and presentation of private, shared and unconventional cancer antigens³⁹.

The workflow that we describe provided insight into targets for future precision immunotherapies. In particular, few (86 total out

of 24,149 initially screened neoepitope–HLA combinations (0.36%) neoepitope–HLA pairs were detected as presented peptides. Neoepitopes for 14 of 36 cancer neoantigens were detected in the context of only one HLA allele, and, of the cancer neoantigens that presented epitopes across multiple alleles, nine were KRAS G12X or G13X mutations. Given this narrow spectrum of bona fide neoepitope–HLA targets, a broadened use of this platform and additional neoepitope–HLA discovery efforts will be needed to increase the coverage of patient populations most likely to benefit from shared neoantigen-specific immunotherapies.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-023-01945-y>.

References

- Jhunjhunwala, S., Hammer, C. & Delamarre, L. Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. *Nat. Rev. Cancer* **21**, 298–312 (2021).
- Klebanoff, C. A. & Wolchok, J. D. Shared cancer neoantigens: making private matters public. *J. Exp. Med.* **215**, 5–7 (2018).
- Zhang, Z. et al. Neoantigen: a new breakthrough in tumor immunotherapy. *Front. Immunol.* **12**, 672356 (2021).
- Mei, S. et al. A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction. *Brief. Bioinform.* **21**, 1119–1135 (2020).
- Darwish, M. et al. High-throughput identification of conditional MHC ligands and scaled-up production of conditional MHC complexes. *Protein Sci.* **30**, 1169–1183 (2021).
- Rodenko, B. et al. Generation of peptide–MHC class I complexes through UV-mediated ligand exchange. *Nat. Protoc.* **1**, 1120–1132 (2006).
- Jurtz, V. et al. NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* **199**, 3360–3368 (2017).
- Klinger, M. et al. Multiplex identification of antigen-specific T cell receptors using a combination of immune assays and immune receptor sequencing. *PLoS ONE* **10**, e0141561 (2015).
- Chang, M. T. et al. Accelerating discovery of functional mutant alleles in cancer. *Cancer Discov.* **8**, 174–183 (2018).
- Ott, P. A. et al. A phase Ib trial of personalized neoantigen therapy plus anti-PD-1 in patients with advanced melanoma, non-small cell lung cancer, or bladder cancer. *Cell* **183**, 347–362 (2020).
- Sahin, U. et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547**, 222–226 (2017).
- Choi, J. et al. Systematic discovery and validation of T cell targets directed against oncogenic KRAS mutations. *Cell Rep. Methods* **1**, 100084 (2021).
- Bear, A. S. et al. Biochemical and functional characterization of mutant KRAS epitopes validates this oncoprotein for immunological targeting. *Nat. Commun.* **12**, 4365 (2021).
- Tran, E. et al. T-cell transfer therapy targeting mutant KRAS in cancer. *N. Engl. J. Med.* **375**, 2255–2262 (2016).
- Garstka, M. A. et al. The first step of peptide selection in antigen presentation by MHC class I molecules. *Proc. Natl Acad. Sci. USA* **112**, 1505–1510 (2015).
- Jappe, E. C., Kringelum, J., Trolle, T. & Nielsen, M. Predicted MHC peptide binding promiscuity explains MHC class I ‘hotspots’ of antigen presentation defined by mass spectrometry eluted ligand data. *Immunology* **154**, 407–417 (2018).

17. Skora, A. D. et al. Generation of MANAbodies specific to HLA-restricted epitopes encoded by somatically mutated genes. *Proc. Natl Acad. Sci. USA* **112**, 9967–9972 (2015).
18. Abelin, J. G. et al. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity* **46**, 315–326 (2017).
19. Wang, Q. et al. Direct detection and quantification of neoantigens. *Cancer Immunol. Res.* **7**, 1748–1754 (2019).
20. Zemmour, J., Little, A. M., Schendel, D. J. & Parham, P. The HLA-A,B ‘negative’ mutant cell line C1R expresses a novel HLA-B35 allele, which also has a point mutation in the translation initiation codon. *J. Immunol.* **148**, 1941–1948 (1992).
21. Creary, L. E. et al. Next-generation HLA typing of 382 International Histocompatibility Working Group reference B-lymphoblastoid cell lines: report from the 17th International HLA and Immunogenetics Workshop. *Hum. Immunol.* **80**, 449–460 (2019).
22. Gomez-Perosanz, M., Ras-Carmona, A. & Reche, P. A. PCPS: a web server to predict proteasomal cleavage sites. *Methods Mol. Biol.* **2131**, 399–406 (2020).
23. Lo, W. et al. Immunologic recognition of a shared p53 mutated neoantigen in a patient with metastatic colorectal cancer. *Cancer Immunol. Res.* **7**, 534–543 (2019).
24. Olsen, L. R. et al. TANTIGEN: a comprehensive database of tumor T cell antigens. *Cancer Immunol. Immunother.* **66**, 731–735 (2017).
25. Yi, X. et al. caAtlas: an immunopeptidome atlas of human cancer. *iScience* **24**, 103107 (2021).
26. Xia, J. et al. NEPdb: a database of T-cell experimentally-validated neoantigens and pan-cancer predicted neoepitopes for cancer immunotherapy. *Front. Immunol.* **12**, 644637 (2021).
27. Liu, J. et al. Cancer vaccines as promising immuno-therapeutics: platforms and current progress. *J. Hematol. Oncol.* **15**, 28 (2022).
28. Jensen, S. M., Potts, G. K., Ready, D. B. & Patterson, M. J. Specific MHC-I peptides are induced using PROTACs. *Front. Immunol.* **9**, 2697 (2018).
29. Moser, S. C., Voerman, J. S. A., Buckley, D. L., Winter, G. E. & Schliehe, C. Acute pharmacologic degradation of a stable antigen enhances its direct presentation on MHC class I molecules. *Front. Immunol.* **8**, 1920 (2018).
30. Howie, B. et al. High-throughput pairing of T cell receptor α and β sequences. *Sci. Transl. Med.* **7**, 301ra131 (2015).
31. Leidner, R. et al. Neoantigen T-cell receptor gene therapy in pancreatic cancer. *N. Engl. J. Med.* **386**, 2112–2119 (2022).
32. Yang, J. C. & Rosenberg, S. A. Adoptive T-cell therapy for cancer. *Adv. Immunol.* **130**, 279–294 (2016).
33. Chandran, S. S. et al. Immunogenicity and therapeutic targeting of a public neoantigen derived from mutated *PIK3CA*. *Nat. Med.* **28**, 946–957 (2022).
34. Pfammatter, S. et al. Extending the comprehensiveness of immunopeptidome analyses using isobaric peptide labeling. *Anal. Chem.* **92**, 9194–9204 (2020).
35. Douglass, J. et al. Bispecific antibodies targeting mutant *RAS* neoantigens. *Sci. Immunol.* **6**, eabd5515 (2021).
36. Nelde, A. et al. HLA class I-restricted MYD88 L265P-derived peptides as specific targets for lymphoma immunotherapy. *Oncoimmunology* **6**, e1219825 (2016).
37. Sachs, A. et al. Impact of cysteine residues on MHC binding predictions and recognition by tumor-reactive T cells. *J. Immunol.* **205**, 539–549 (2020).
38. Stopfer, L. E., Mesfin, J. M., Joughin, B. A., Lauffenburger, D. A. & White, F. M. Multiplexed relative and absolute quantitative immunopeptidomics reveals MHC I repertoire alterations induced by CDK4/6 inhibition. *Nat. Commun.* **11**, 2760 (2020).
39. Oreper, D., Klaeger, S., Jhunjhunwala, S. & Delamarre, L. The peptide woods are lovely, dark and deep: hunting for novel cancer antigens. *Semin. Immunol.* **67**, 101758 (2023).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© Genentech, Inc. and Adaptive Biotechnologies Corp. 2023

Methods

Engineering of monoallelic polyantigen cassette-expressing HMy2.C1R cell lines

An HLA class-I null cell population was generated by CRISPR–Cas9-mediated gene disruption of the endogenous HLA-C locus in HMy2.C1R cells. Wild-type HMy2.C1R cells were electroporated with Cas9/RNP (Invitrogen) containing an HLA-C-specific sgRNA (Synthego, sequence: TTCATCGCAGTGGGCTACG) (Supplementary Fig. 3a) using an Amaxa V system (program D-023). After an expansion period, cells were stained with anti-pan-HLA (W6/32), and antigen-negative cells were enriched by FACS (Supplementary Fig. 3d,e). For flow cytometric data collection, experiments were performed on BD Celesta, BD Fortessa or BD Symphony machines using FACSDiva version8/version9 acquisition software.

The HLA null HMy2.C1R cells were stably engineered with a piggyBac neoantigen expression plasmid system designed to co-express 47 shared cancer neoantigens and seven A*02:01 control antigens. In brief, neoantigen segments (~25 amino acids each) were concatenated and converted to codon-optimized DNA segments (Integrated DNA Technologies) with or without a flexible linker separating most neoantigen sequences. A version of the neoantigen cassette without linkers and lacking control antigens was also generated for use in the A*02:01 monoallelic context. The polyantigen cassettes were synthesized and cloned into a piggyBac transposon plasmid downstream of a constitutive human EF1a and transcriptionally linked to an IRES-TagBFP2 reporter element. A separate hPGK promoter-driven puromycin resistance gene was included on the same vector for selection purposes. The polyantigen expression plasmid was co-electroporated with pBO (piggyBac transposase, Hera BioLabs) using a NEON system (Invitrogen) and a 100- μ l kit (buffer R, 1,230 V, 20 ms, three pulses). Polyantigen-expressing cells were selected by culture in 1 μ g ml⁻¹ puromycin (Gibco) and further purified by FACS enrichment of the TagBFP2-positive population.

Unique HLA allele open reading frames (ORFs), each with a distinct 19-base pair (bp) DNA barcode, were cloned downstream of the human EF1a promoter (GenScript) in a custom-modified pLenti6.3 backbone (Thermo Fisher Scientific). Lentivirus was generated by Lipofectamine 2000 (Invitrogen)-mediated co-transfection of HEK293T cells with individual lenti-HLA expression constructs and packaging plasmids. Seventy-two hours after transfection, viral supernatant was harvested, filtered through a 0.45- μ m filter and concentrated by LentiX concentrator reagent (Takara) following the manufacturer's recommended protocol. Linker or no-linker polyantigen-expressing HLA-null HMy2.C1R cells were transduced with HLA expression vectors via spin infection (800g for 30 min at room temperature with 8 μ g ml⁻¹ polybrene). Transgenic HLA-expressing cells subsequently were purified by magnetic bead-based enrichment (biotin-W6/32, BioLegend, SA-MACS). HLA allele identification was confirmed by barcode sequencing (amplicon primers: Fwd-TCCAGAGCCACCGTTACAC, Rev-GACTTAACGCGTCTGTTGC; sequencing primer: CTGGTTGCAGCGCTTACGCT), and uniform expression of both the HLA allele and polyantigen cassette was confirmed by flow cytometry (Fig. 3c and Supplementary Fig. 2c,d) before analysis by MS.

For studies evaluating neoantigen presentation in the context of full-length neoantigen-containing proteins, A*11:01 monoallelic cell lines were stably engineered with a doxycycline (dox)-inducible piggyBac vector expressing wild-type or mutant alleles (G12D, G12C and G12V) of human KRAS. The KRAS allele of interest and an IRES-linked mCherry reporter were driven by a dox-responsive TRE3G promoter. A puromycin resistance gene and the Tet-on3G element were encoded on the same vector downstream of a constitutive hPGK promoter. After puro selection and expansion, KRAS expression was induced by treating cells with 1 μ g ml⁻¹ dox for 5 d before subsequent analysis.

Clinico-genomics analysis

Prevalence data for common cancer mutations (single-nucleotide variants (SNVs) and insertions and deletions (indels)) were obtained from the Cancer Hotspots database (<http://cancerhotspots.org>)⁶ and cross-referenced with TCGA data obtained from the cBioPortal for Cancer Genomics (<http://cbioportal.org>). Prevalence data for common HLA alleles were obtained by tabulating HLA types from the AFND (<http://allelefrequencies.net>) and from TCGA normal samples. Allele frequency data for the HLA-A, HLA-B and HLA-C genes across seven selected populations were downloaded from the AFND in May 2020 (Supplementary Table 2). We focused on large datasets ($n > 24,000$ for each population) from the National Marrow Donor Program. HLA alleles with allele frequency below 1% in all populations were removed. We then calculated the overall allele frequency for each allele as the mean across all populations and used this overall frequency in filtering and ranking alleles. We also analyzed HLA typing data from TCGA that were generated by running PolySolver⁴⁰ on the whole-exome data from 9,741 matched normal samples (Amir Horowitz, Icahn School of Medicine at Mount Sinai; Supplementary Table 1). We tabulated and ranked the most prevalent HLA alleles in TCGA and overlapped them with the list of prevalent alleles in the AFND, which allowed us to confirm that the HLA alleles we selected were generally present with similar frequencies in cancer and non-cancer settings.

From these datasets, the 47 most common cancer mutations were determined based on prevalence per cancer type; the 47 most common HLA alleles were also determined. Additional ranking of these mutations was performed that considered the overall prevalence of each cancer type and whether a neoantigen-specific therapy could be readily developed in a clinical setting.

Predicted neoepitope landscape analysis

After translating mutations to peptide sequences, neoepitope–HLA binding predictions were generated using NetMHCpan-4.0 (ref. 4) on all combinations of 8-mer, 9-mer, 10-mer and 11-mer peptides derived from the 47 cancer neoantigens combined with 15 prevalent HLA alleles. Both BA and EL predictions were obtained, which were then used for downstream analysis. Predicted neoepitopes were defined as neoepitope–HLA combinations with mutant EL percentile rank <2.

Automated high-throughput neoepitope exchange

See Supplementary Methods for protein expression, peptide synthesis and HLA–peptide refolding information. Peptides were diluted to 10 μ M in 25 mM Tris pH 8.0, 150 mM NaCl, 4 mM EDTA and 4.35% ethylene glycol in 96-deep-well plates (VWR) using a Biomek i5 automated liquid handler (Beckman Coulter). The peptide–buffer mixtures were dispensed and reformatted into 384-well plates (Lab-cyte) at a volume of 47.5 μ l per well, resulting in identical plates of up to 352 unique neoepitopes for screening against each of the 15 HLA alleles. The first two columns of the plate were reserved for controls. A*02:01 with and without exchange peptide was included on each plate as positive and negative controls for exchange, respectively. The well-characterized A*02:01-specific viral epitope, CMV pp65 peptide (NLVPMVATV, Elim Biopharmaceuticals), was plated in quadruplicate as a positive control for peptide exchange. Negative controls for exchange included wells to which no peptide was added and, instead, received ethylene glycol only during the peptide dilution step. Negative control wells for the HLA allele being screened were plated in octuplicate.

Using a MANTIS Liquid Handler (Formulatrix), 2.5 μ l of 0.1 mg ml⁻¹ UV peptide–HLA complexes was added to each well, with one HLA allele screened for binding per plate. Positive control wells received A*02:01, and negative control wells received either HLA A*02:01 or the HLA allele specific to the plate. The resultant peptide exchange reaction mixtures contained 10 μ M peptide, 0.1 μ M UV–HLA complex and 5% ethylene glycol (v/v).

The peptide exchange protocol was adapted from a previously described method² by decreasing the UV exposure time and adding an incubation step after UV exposure. Plates containing the peptide exchange reaction mixtures were incubated under UV lamps (UVP 3UV Lamp, Analytik Jena) for 25 min using one lamp per plate. Plates were then sealed and incubated for 18 h at 37 °C.

TR-FRET assay

The homogenous TR-FRET assay was carried out in MAKO 1,536-well white solid-bottom plates (Aurora Microplates). The total assay volume was 4 µl per well, including 2 µl of diluted samples and 2 µl of reagent mix. In brief, 1.8 µl per well of assay diluent (PBS, 0.5% BSA + 0.05% Tween 20 + 10 ppm proclin, Genentech) was added to the 1,536-well destination plate by a Multidrop Combi nL dispenser (Thermo Fisher Scientific). Then, 200 nl of 5 µg ml⁻¹ HLA-I complex sample was dispensed from the Echo-qualified 384-well source plate (Beckman Coulter) into the destination plate by an Echo 550 acoustic liquid dispenser (Beckman Coulter). After centrifugation for 3 min, 2 µl of master mix donor at 2 nM (europium mouse anti-human β2-microglobulin (β2M), BioLegend, custom labeled by PerkinElmer) and acceptor at 40 nM (SureLight Allophycocyanin-Streptavidin (APC-SA), PerkinElmer) in assay diluent were dispensed into each well of the destination plate with the Multidrop Combi nL dispenser. After incubation at room temperature for 1 h, the destination plates were read on the PHERAstar FS plate reader (BMG Labtech) with donor excitation at 337 nm, donor emission at 615 nm and acceptor emission at 665 nm.

The signal was expressed as the ratio of RFUs in each well (RFU ratio = (665 nm/615 nm) × 10⁴). For ranking the binders, a double normalization was applied to obtain %DeltaF. $\Delta F(\%) = \frac{(\text{RFU}[\text{sample}] - \text{mean RFU}[\text{negative}])}{\text{mean RFU}[\text{negative}]} \times 100$. The RZ-score was calculated on the sample plate basis. For screening quality control, large-scale prepared positive control (A*02:01 with pp65) and a negative control (A*02:01 only) were added to designated wells in each sample plate. The acceptance of the screen was determined by the Z-factor calculated from the assay controls (Z-factor = $1 - \frac{3 \text{ s.d.}[\text{positive}] - 3 \text{ s.d.}[\text{negative}]}{(\text{mean}[\text{positive}] - \text{mean}[\text{negative}])}$). Sample plates with a Z-factor >0.4 were qualified for data processing.

Untargeted MS and database search

See Supplementary Information for HLA-IP information. One-third of each sample was loaded into a 25 cm × 75 µm ID, 1.6 µm C18 IonOpticks Aurora Series column (IonOpticks, AUR2-25075C18A) on a Thermo Ultimate 3000 high-performance liquid chromatography (HPLC) system (Thermo Fisher Scientific) at a flow rate of 400 nl min⁻¹. Peptides were separated with a 90-min gradient of 2% to 35% or 40% buffer B (98% ACN, 2% water and 0.1% FA) at a flow rate of 300 nl min⁻¹. The gradient was further raised to 75% buffer B for 5 min and to 90% buffer B for 4 min at the same flow rate before final equilibration with 98% buffer A (98% water, 2% ACN and 0.1% FA) and 2% buffer B for 10 min at a flow rate of 400 nl min⁻¹.

Peptide mass spectra were acquired using either an Orbitrap Fusion Lumos or an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific) with MS¹ Orbitrap resolution of 240,000 and MS/MS fragmentation of the precursor ions by collision-induced dissociation (CID), followed by spectra acquisition at MS² Orbitrap resolution of 15,000. All data-dependent acquisition (DDA) spectral raw files were searched in PEAKSOnline (Bioinformatics Solutions, PEAKS Online X build 1.7) against a UniProt-derived *Homo sapiens* proteome (downloaded on 3 October 2019) that contained appended concatenated sequences of the 47 most common mutations flanked by -13-mer sequences on either end of each mutation with or without stretches of glycine and serine residue (GS) linkers along with sequences of BFP. Within PEAKSOnline, because HLA peptides are non-tryptic, the enzyme specificity was set as none; CID was selected as an activation method; and Orbitrap (Orbi-Orbi) was chosen as an

instrument parameter. In-depth de novo assisted database search and quantification were performed with precursor mass error tolerance of 15 ppm, fragment mass error tolerance of 0.02 Da and missed cleavage allowance of 3. Carbamidomethylation (Cys+57.02) was set as a fixed modification, whereas deamidation (Asn+0.98 and Gln+0.98) and oxidation (Met+15.99) were set as variable post-translational modifications (PTMs), allowing a maximum of three variable PTMs per peptide. Additional report filters included peptide spectral match FDR of 1%, proteins $-\log_{10}P \geq 20$ and de novo only amino acid residue average local confidence of 50%. For label-free analysis, a new group was created for each sample, and match between runs was performed with default parameters, except that retention time shift tolerance was set to 4 min and base sample was selected as 'Average'. Output CSV files were exported and further analyzed in R.

Targeted MS

See Supplementary Information for HLA-IP information. Absolute quantification (AQUA) synthetic heavy peptides (8–11-mer) (Elim Biopharmaceuticals) for all 47 mutation-derived neoantigens with TR-FRET RZ-score ≥ 5 or predicted NetMHC %Rank ≤ 2 (for a subset of mutations) were reconstituted in 30% ACN/0.1% formic acid (FA). DMSO was added for peptides that were not readily soluble in 30% ACN/0.1% FA. A working solution of 25 µM was made for each AQUA peptide from which allele-specific mastermix was made at 25 pmol per peptide. The peptides were reduced/alkylated and cleaned up with C18 cartridges on AssayMAP Bravo. After drying, the peptides were reconstituted in 0.1% FA/0.05% HFBA at 100 fmol per peptide. For each allele-specific assay, the intact modified mass was calculated for each peptide in that assay using TomahaqCompanion software⁴⁰, which was then used to build an inclusion list MS method for a scouting run to get the retention time and mass-to-charge (*m/z*) of each target peptide. Then, 1 µl of each assay was injected into the IonOpticks C18 column and sprayed into the mass spectrometer for a 125-min run as described above, and the raw files were imported and analyzed in Skyline (64-bit, 19.1.0.193) to select appropriate charge for each peptide. A mass list table was built for each assay where a 4-min retention time window was created on both sides of the retention time for each target peptide, which was then imported into the Xcalibur instrument method application and saved as an allele-specific parallel reaction monitoring (PRM) method. For both Fusion Lumos and Eclipse instruments, MS¹ was acquired at Orbitrap resolution of 240,000 with a maximum injection time of 50 ms, followed by a quadrupole isolation window of 1.2 *m/z*, CID fragmentation of parent ions, maximum injection time of 300 ms and MS² acquisition at Orbitrap resolution of 60,000. For Eclipse acquisition, MS¹ and MS² AGC targets were set at 250% and 400%, respectively. One-third of each monoallelic sample was spiked with 100 fmol of corresponding AQUA mastermix and injected into the mass spectrometer using the same HPLC setup as described above. Raw PRM data were imported and analyzed in Skyline in an allele-specific manner. The ratios of the light peptides to their heavy counterparts across samples were exported as CSV files and further analyzed in R. For each neoepitope, background signal detected in the synthetic peptide-only analysis was subtracted from endogenous peptide signal before calculation of a final attomole amount. See Supplementary Information for methods relating to targeted MS quantification of full-length KRAS protein and neoepitopes.

Ribo-Seq

Ribo-Seq was performed as previously described^{41,42}. In brief, 8 million linker and no-linker HLA-A*02:01-engineered cells were lysed in poly-some lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM MgCl₂, 1 mM DTT, 1% Triton X-100, 25 U ml⁻¹ Turbo DNase and 0.1 mg ml⁻¹ cycloheximide). The lysate volume equivalent to 30 µg of RNA was digested with 7.5 U of RNase I (LGC Biosearch Technologies) for 15 min at room temperature. Monosomes were purified using MicroSpin S-400 HR columns (Cytiva) according to the manufacturer's instructions. The

flow-throughs were mixed with TRI Reagent (Zymo Research), and the RNA was purified using the Direct-zol RNA Miniprep Kit (Zymo Research). Ribosome footprints were purified and size selected by electrophoresis in a 15% polyacrylamide TBE-Ureagel (Invitrogen), and footprints between 25 nucleotides (nt) and 32 nt were collected. rRNAs were depleted using the riboPOOL kit (siTOOLS Biotech) according to the manufacturer's instructions. TruSeq Small RNA Library Preparation Kit (Illumina) was used for library preparation with modifications as previously described⁴¹. Libraries were purified by electrophoresis using 6% polyacrylamide TBE gels (Invitrogen). The quality of the libraries was confirmed on a D1000 ScreenTape (Agilent) on the 4200 TapeStation system (Agilent), and libraries were quantified using the Qubit dsDNA High-Sensitivity Assay (Invitrogen) on a Qubit 3.0 Fluorometer (Invitrogen). Sequencing was performed in a NovaSeq 6000 sequencer (Illumina) with a read length of 50 bp, single end, and depth of 100 million reads per sample.

Software used for the analysis of Ribo-Seq data was sourced from <https://anaconda.org/> unless otherwise stated. Ribo-Seq reads were trimmed with Cutadapt (version 3.4_py38h4a8c8d9_1) with the following parameters: `-j 8 -u 3 -u -5 -m 10 -a <adapter sequence file>`. For RNA PCR primer sequences, see Supplementary Methods.

Next, reads aligning to rRNA and tRNA were removed using Bowtie (version 1.3.0_py38hcf49a77_2) with the following parameters: `-p8 -v3 -un`. This procedure outputs a FASTQ file containing reads that map outside of rRNA and tRNA loci. These reads were aligned using STAR (version 2.7.10b) with two sets of parameters:

1. `--outFilterType BySJout, outFilterMismatchNmax 2, outSAMtype BAM SortedByCoordinate, quantMode TranscriptomeSAM GeneCounts, outFilterMultimapNmax 1, outFilterMatchNmin 16, alignEndsType EndToEnd, runThreadN 16`
2. `--runMode alignReads--alignIntronMax 1--outFilterMismatchNmax 20--outFilterScoreMinOverLread 0.25--outFilterMatchNminOverLread 0.25--outSAMmode NoQS--outSAMtype BAM SortedByCoordinate--alignEndsType Extend5pOfReads12--outSAMattributes nM MD NH`

SAMtools (version 1.13_h8c37831_0) was used for indexing and sorting the BAM files. De-duplication of the resulting BAM file was performed with Picard MarkDuplicates (version 2.25.7_hdfd78af_0). The de-duplicated BAM files were then indexed using SAMtools as mentioned previously. Lastly, BAM files were processed with three previously published Ribo-Seq ORF-calling and quality control programs: Price (version https://github.com/erhard-lab/gedi/releases/tag/Price_1.0.3b), RiboCode (version 1.2.11_pyh145b6a8_1) and RibORF (version <https://github.com/zhejilab/RibORF/tree/master/RibORF.2.0>). RiboCode accepts BAM files generated using the STAR alignment method (1), whereas Price and RibORF accept BAM files generated using the STAR alignment method (2).

TCR discovery

A total of 376 predicted and MS-identified neoantigen-derived peptides were synthesized (GenScript), and each was added to six of 11 peptide pools such that each neoepitope (or group of similar neoepitopes) occupied a unique combination of six pools⁵. CD8⁺ T cells were isolated (STEMCELL Technologies) from healthy human donor leukopaks and expanded either on anti-CD3 coated plates (+anti-CD28/IL-2, BioLegend) or in the presence of matched donor-derived monocyte-derived dendritic cells⁴⁰ and a pool of all 376 neoepitopes. At day 10–15, T cells were recovered, supplemented with one of the 11 neoepitope pools, incubated 8–14 h, enriched (Miltenyi Biotec) and then sorted using an anti-CD137 antibody (stained at 1/20: 5 µl of antibody was added to 100 µl of FACS buffer) (BioLegend). Sorted cells were then subjected either to immunoSEQ or pairSEQ (Adaptive Biotechnologies) to identify TCRB sequences displaying neoepitope-specific responsiveness and to associate TCRB with TCRA sequences in parallel, respectively.

TCR sequences were encoded in pcDNA vectors as a single ORF, in the form of the full TCRB sequence followed by an RAKR motif and porcine teschovirus 2A cleavage peptide with the full TCRA sequence after in frame. TCR-encoding pcDNA vectors were then used as templates to generate TCR-encoding in vitro transcribed RNA (mMESSAGE mMACHINE, Thermo Fisher Scientific) for electroporation of primary human T cells.

TCR reactivity assays

CD8⁺ cells were enriched from human peripheral blood mononuclear cells with EasySep Human CD8⁺ T Cell Isolation Kit (STEMCELL Technologies) and stimulated with 5 µg ml⁻¹ Ultra-LEAF anti-human CD3 (BioLegend) and 2.5 µg ml⁻¹ Ultra-LEAF anti-human CD28 (BioLegend). Cells were cultured in the presence of 20 ng ml⁻¹ recombinant human IL-2 for 6 d. Human expanded CD8⁺ T cells were transfected with FLT3-p.D835Y-specific or PIK3CA-p.E545K-specific TCR RNA using a Lonza 4D-Nucleofector, P3 primary cell 4D-Nucleofector kit, program EO-115 (Lonza). RNA was purchased from TriLink BioTechnologies or in vitro transcribed. FLT3-p.D835Y-specific TCRs were co-cultured overnight with HLA-A*02:01-expressing T2 cells pulsed with YIMSDSNYV or HLA-A*02:01-expressing K562 cells transfected with a construct encoding the mutant or wild-type sequence. K562 cells were transfected using a Lonza 4D-Nucleofector, SF cell line 4D-Nucleofector kit, program FF-120 (Lonza). To determine specific cell lysis, an equal mixture of transfected HLA-A*02:01 + K562 cells and untransfected CellTrace Far Red (Thermo Fisher Scientific)-labeled HLA-A*02:01 + K562 cells was co-cultured overnight with T cells at a 2:1 effector-to-target ratio. Percent (%) specific cell lysis = $(P_{\text{mock-transfected T cells}} - P_{\text{TCR-transfected T cells}}) / (P_{\text{mock-transfected T cells}}) \times 100$, where P is the proportion of transfected K562 targets relative to untransfected K562 cells, as measured by flow cytometry. CD137 expression on CD8⁺ T cells was assessed after an overnight co-culture with an anti-CD137 PE antibody (1/20: 5 µl of antibody was added to 100 µl of FACS buffer) (BioLegend).

PIK3CA-p.E545K-specific TCRs were co-cultured overnight with HLA-A*11:01-expressing K562s pulsed with STRDPLSEITK or transfected with a construct encoding the mutant or wild-type sequence. Equal mixtures of CellTrace Far Red-labeled HLA-A*11:01 + K562 cells were added to each well. T cell response to PIK3CA-presenting K562 cells was assessed as above.

Resource availability

Further information and requests for resources and reagents should be directed to, and will be fulfilled by, the lead contact, Chris Rose (rose.christopher@gene.com). Plasmids generated in this study are the property of Genentech but can be made available under a material transfer agreement (MTA). Cell lines generated in this study are the property of Genentech but can be made available under an MTA. Recombinant HLA complexes generated in this study are the property of Genentech but can be made available under an MTA. MTAs can be requested at <https://www.gene.com/scientists/mta>.

Reporting summary

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

Data availability

Prevalence data for common cancer mutations (SNVs and indels) were obtained from the Cancer Hotspots database (<http://cancerhotspots.org>) and cross-referenced with TCGA data obtained from the cBioPortal for Cancer Genomics (<http://cbioportal.org>). Prevalence data for common HLA alleles were obtained by tabulating HLA types from the AFND (<http://allelefrequencies.net>) and from TCGA normal samples. All MS data have been deposited in the MASSIVE repository⁴³ and are publicly available as of the date of publication under the identifier [MSV000090323](https://doi.org/10.1038/s41587-023-01945-y).

References

40. Shukla, S. A. et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152–1158 (2015).
41. Aeschmann, F., Xiong, J., Arnold, A., Dieterich, C. & Großhans, H. Transcriptome-wide measurement of ribosomal occupancy by ribosome profiling. *Methods* **85**, 75–89 (2015).
42. McGlincy, N. J. & Ingolia, N. T. Transcriptome-wide measurement of translation by ribosome profiling. *Methods* **126**, 112–129 (2017).
43. Wang, M. et al. Assembling the community-scale discoverable human proteome. *Cell Syst.* <https://doi.org/10.1016/j.cels.2018.08.004> (2018).

Acknowledgements

We thank the Genentech internal reviewers, J. Lill, S. Klaeger and H. Melichar, for helpful discussions and comments on the manuscript.

Author contributions

H.R.G. performed HLA-IP and immunopeptidomic data collection. H.R.G. and C.M.R. analyzed immunopeptidomic data. A.J.H., Y.A. and B.H. conceived of cell and/or performed cell engineering, cell growth and flow cytometry experiments. M.D., P.P.F.C., J.L., M.B., A.C., K.H.-W.L. and C.B. conceived of and/or performed the HT-TR-FRET assay. D.H. and E.T. performed analysis of HT-TR-FRET assay. O.A.Z., C.H. and A.W. performed clinicogenomic analysis as well as NetMHCpan-4.0 predictions. A.X.-M. performed Ribo-Seq experiments. S.V. and D.D.L. performed Ribo-Seq data analysis. A.M., I.A. and S.A.O. performed or directed cell culture for sample generation. A.J.M., U.N.U., M.B.L., R.J.N. and P.J.R.E. conceived and/or performed TCR discovery and validation studies. H.R.G., A.J.H., M.D.,

C.B., B.H. and C.M.R. wrote the manuscript and generated the figures. C.B., B.H. and C.M.R. directed the research.

Competing interests

H.R.G., A.J.H., M.D., P.P.F.C., J.L., M.B., O.A.Z., A.W., A.-J.T., D.H., E.T., A.C., K.H.W.L., Y.A., C.H., A.X.-M., A.M., S.V., D.D.L., I.A., S.A.O., C.B., B.H. and C.M.R. were employees or contract workers at Genentech, Inc. at the time of performing the research and writing the manuscript. A.J.M., U.N.U., M.B.L., R.J.N. and P.J.R.E. were employees of Adaptive Biotechnologies at the time of performing the research and writing the manuscript. The described workflow of a high-throughput TR-FRET binding assay combined with untargeted and targeted immunopeptidomic analysis of monoallelic cell lines expressing a large number of candidate neoantigens relates to a patent application filed by Genentech, Inc. with H.R.G., B.H., A.J.H., J.L., C.M.R., A.-J.T., C.B., P.P.F.C. and M.D. as inventors (PCT/US2022/078831). The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41587-023-01945-y>.

Correspondence and requests for materials should be addressed to Craig Blanchette, Benjamin Haley or Christopher M. Rose.

Peer review information *Nature Biotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection For flow cytometric data collection, experiments were performed on BD Celesta, BD Fortessa or BD Symphony machines using FACSDIVA v8/v9 acquisition software.

Data analysis Flow cytometry data was analyzed initially by FlowJo 2 v10.7/v10.8. Untargeted Mass Spectrometry was analyzed by PEAKS Online X build 1.7 (Bioinformatics Solutions Inc.). Targeted Mass spectrometry data were analyze with Skyline (64-bit, 19.1.0.193). TomahaqCompanion was used for retention time determination (<https://github.com/CMRose3355/TomahaqCompanionProgram>). Ribo-Seq data analysis utilized cutadapt (Version:3.4_py38h4a8c8d9_1), using bowtie (Version:1.3.0_py38hcf49a77_2), STAR (Version:2.7.10b), Samtools (Version:1.13_h8c37831_0), Picard MarkDuplicates (Version:2.25.7_hdfd78af_0), PRICE (Version:https://github.com/erhard-lab/gedi/releases/tag/Price_1.0.3b), RiboCode(Version:1.2.11_pyh145b6a8_1), and RibORF (Version:<https://github.com/zhejilab/RibORF/tree/master/RibORF.2.0>).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data Availability: Prevalence data for common cancer mutations (SNVs and indels) were obtained from the Cancer Hotspots database (<http://cancerhotspots.org>) and cross-referenced with TCGA data obtained from the cBioPortal for Cancer Genomics (<http://cbiportal.org>). Prevalence data for common HLA alleles were obtained by tabulating HLA types from the Allele Frequency Net Database (AFND, <http://allelefrequency.net>) and from TCGA normal samples. All mass spec data has been deposited in the MASSIVE repository (Wang et al., 2018) and are publicly available as of the date of publication under the identifier MSV000090323.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	<input type="text" value="N/A"/>
Population characteristics	<input type="text" value="N/A"/>
Recruitment	<input type="text" value="N/A"/>
Ethics oversight	<input type="text" value="N/A"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<input type="text" value="Sample size for each experiment indicated in the figure legend, methods or associated text. Sample sizes were determined by the number of replicates needed to provide confidence in the data or what was feasibly possible within the limitations of resources."/>
Data exclusions	<input type="text" value="A technical replicate may be missing due to technical problems/sample loss during data collection."/>
Replication	<input type="text" value="All comparative studies were performed a minimum of 2 times with representative experiments depicted. Biological and technical replicates for each experiment indicated in figure legend, methods or associated text."/>
Randomization	<input type="text" value="Randomization was applied when appropriate (e.g., order in which MS runs were performed). However, for most experiments randomization does not apply."/>
Blinding	<input type="text" value="Blinding was not possible as the researcher performing the experiment needed to know the components of the sample in order to perform sample preparation and analysis."/>

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Antibody clone, company and staining protocol listed in methods antibodies.

Validation

-W6/332-
Anti-human HLA-A,B,C mouse monoclonal;
Biolegend 311434 (biotin);311410 (APC):

Validated by Biolegend and:
Darrow TL, et al. 1989. J. Immunol. 142:3329.
Stern P, et al. 1987. J. Immunol. 138:1088.
Tran TM, et al. 2001. Immunogenetics 53:440.
Barbatis C, et al. 1981. Gut 22:985.
Ayyoub M, et al. 2004. Cancer Immunity 4:7.
DeFelice M, et al. 1990. Cell. Immunol. 126:420.
Fayen J, et al. 1998. Int. Immunol. 10:1347.
Turco MC, et al. 1988. J. Immunol. 141:2275.
Geppert TD, et al. 1989. J. Immunol. 142:3763.
Wooden SL, et al. 2005. J. Immunol. 175:1383.
Nagano M, et al. 2007. Blood 110:151.
McLoughlin RM, et al. 2008. J. Immunol. 181:1323.
<https://www.biolegend.com/en-us/products/biotin-anti-human-hla-a-b-c-antibody-12535>

-Anti-CD137-
Anti-CD137 PE: Biolegend cat# 309804, Clone 4B4-1
Validation by Biolegend and:
1. Liu T, et al. 2021. Gut. 70:1965.
2. Shitaoka K, et al. 2018. Cancer Immunol Res. 6:378.
3. Arce Vargas F et al. 2018. Cancer cell. 33(4):649-663 .
4. Yamaguchi S, et al. 2021. Eur J Immunol. 51:2306.
5. Lamichhane R, et al. 2020. Eur J Immunol. 50:178.
6. Goncharov MM, et al. 2022. Elife.
7. Shin JJ, et al. 2022. J Clin Immunol.
8. Jutz S, et al. 2016. J Immunol Methods. 430:10-20.
9. Pavlicek D, et al. 2017. Immun Ageing. 14:22.
10. Gordon-Alonso M, et al. 2017. Nat. Commun. 10.1038/s41467-017-00925-6.
11. Rajamanickam V, et al. 2021. Cancer Immunol Res. 9:602.
12. Jung JH, et al. 2021. Nat Commun. 12:4043.

Anti-CD8 FITC: Biolegend cat# 300906, Clone HIT8a
Validation by Biolegend and:
1. Zhang H, et al. 2021. Front Immunol. 12:644520.
2. Bilich T, et al. 2021. Cancer Discov. 11:1982.
3. Galindo-Albarrán A, et al. 2016. Cell Rep. 17:2151-2160.
4. Thaker YR, et al. 2022. Front Oncol. 12:884196.
5. Sun L, et al. 2020. J Diabetes Res. 2020:2583257.
6. Zhou Y, et al. 2017. Front Cell Infect Microbiol. 7:457.
7. Mender I, et al. 2020. Cancer Cell. 38(3):400-411.e6.
8. Warmuth S, et al. 2022. Oncoimmunology. 10:2004661.
9. Shen X, et al. 2021. Front Immunol. 12:710750.
10. Bratke K, et al. 2017. Clin Exp Allergy. . 10.1111/cea.13017.
11. Durning S, et al. 2016. J Immunol. 197: 3076 - 3085.
12. Chulpanova DS, et al. 2021. Biology (Basel). 10

Biotin anti-human CD137 (4-1BB) Antibody, Biolegend, cat# 309806, clone 4B4-1
Validation by Biolegend and:
1. Garni-Wagner B, et al. 1996. Cell. Immunol. 169:91.
2. Salih HR, et al. 2000. J. Immunol. 165:2903.
3. Kienzle G, et al. 2000. Int. Immunol. 12:73.
4. Langstein J, et al. 1998. J. Immunol. 160:2488.

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Cell lines sourced from the Genentech cell bank (gCell): HMy2.C1R HOP62 NCIH2030 HuCCT1 SNU601 SW527 HEK293T Cell lines sourced from ATCC: (CONFIRM WITH ADAPTIVE) T2 K562
Authentication	gCell sourced cell lines authenticated by STR profiling as described doi.org/10.1038/nature14397 . T2 cells and K562 cells were used from ATCC without further authentication.
Mycoplasma contamination	Cells tested negative for mycoplasma contamination were maintained under mycoplasma-free conditions.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in these studies.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	w6/32 Biologend 311410 (APC) Pre-wash cells 1 X with staining buffer. Stain: 5ul antibody in 100ul staining buffer. Incubation 20-30min @ 4deg. Wash: 2 X with staining buffer.
Instrument	BD Celesta, BD Fortessa or BD Symphony
Software	FACSDIVA v8/v9 was utilized as acquisition software and FlowJo 2 v10.7/v10.8 was used for subsequent analysis
Cell population abundance	Live cell were defined by FSC-A/SSC-A profiles as indicated Supplementary Fig. 3b.
Gating strategy	Gating strategy indicated in Supplementary Fig. 3b. Live cell populations were gated using fsc and ssc profiles. HLA positive cells were stained with pan-HLA (w6/32) antibody staining and evaluated for APC signal. Polyneoeantigen expressing cells specifically were identified by detection of the transcriptionally linked mTagBFP2.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.