Strength of immune selection in tumors varies with sex and age

Andrea Castro1,2,3,13, Rachel Marty Pyke1,2,13, Xinlian Zhang4, Wesley Kurt Thompson4, Chi-Ping Day5, Ludmil B. Alexandrov6,7,8, Maurizio Zanetti8,9,10 & Hannah Carter1,2,8,11,12

Individual MHC genotype constrains the mutational landscape during tumorigenesis. Immune checkpoint inhibition reactivates immunity against tumors that escaped immune surveillance in approximately 30% of cases. Recent studies demonstrated poorer response rates in female and younger patients. Although immune responses differ with sex and age, the role of MHC-based immune selection in this context is unknown. We find that tumors in younger and female individuals accumulate more poorly presented driver mutations than those in older and male patients, despite no differences in MHC genotype. Younger patients show the strongest effects of MHC-based driver mutation selection, with younger females showing compounded effects and nearly twice as much MHC-II based selection. This study presents evidence that strength of immune selection during tumor development varies with sex and age, and may influence the availability of mutant peptides capable of driving effective response to immune checkpoint inhibitor therapy.
The major histocompatibility complex (MHC) exposes protein content on the cell surface to allow detection of antigens by the immune system. This applies to nonself antigens such as viral proteins, and self-proteins that include tumor antigens. Tumor cells harbor oncogenic alterations that can be presented to the immune system by the MHC, causing immune recognition and elimination (immune surveillance)\(^1\). However, in order to grow, invade, and spread, tumors must evade immune surveillance. Common mechanisms of immune evasion include loss of the MHC molecules and the upregulation of immune checkpoint molecules on cell surfaces that normally regulate the amplitude and duration of a T-cell response\(^2\). Immune checkpoint blockade (ICB) uses antibodies to block these immune checkpoint molecules, and can invigorate inactive and/or exhausted T cells to produce antitumor effects that confer long-term survival benefits in certain types of cancer\(^3\). However, ICB is effective in only 10–40\% of patients for reasons that remain unclear. Meta-analyses of clinical trials in multiple cancer types treated with ICB suggest that young and female patients are characterized by low response rates\(^4\)–\(^8\). The reason(s) for the poor response of these two populations remains elusive.

An accumulating body of literature points to sexual dimorphism in immune response\(^9\). Moderated by genetic and hormonal factors, females have twice the antibody response to influenza vaccines\(^10\) and higher CD4\(^+\) T-cell counts than males\(^11\). Moreover, females are far more susceptible to autoimmune diseases\(^12\), demonstrating a stark imbalance in the way the immune response causes diseases in the two sexes. Immunosequencing of over 800 individuals revealed sex-associated differences in the extent to which HLA molecules propagate selection and expansion of CD8\(^+\) T cells\(^13\). Interestingly, a stronger immune response in females has been observed across several species\(^14\)–\(^16\), and sexual dimorphism has been demonstrated in immune selection and restriction of intratumor genetic heterogeneity in a mouse model of B-cell lymphoma\(^17\). In addition, a recent study has found sex-based differences in molecular biomarkers and immune checkpoint expression in multiple tumor types treated with ICB\(^18\). Altogether, these studies suggest that these differences are sex-specific and not lifestyle dependent.

Studies have demonstrated age-related changes in immune response as well. As humans age, there is a decrease of general immune function including production of IL-2, a pivotal growth factor for T cells\(^18\). Reduced thymic output, lower numbers of naive T cells, and overall reshaping of the size and specificity of the T-cell repertoire by microbial pathogens may explain why, for example, about 90\% of excess deaths during flu season occur in patients greater than 65 years of age\(^19\). In addition, elderly people have reduced phagocytic function and HLA-II expression on antigen presenting cells\(^20\). Collectively, these factors render elderly individuals less able to mount a T-cell response to new antigens and respond to vaccination.

Recently, we developed the Patient Harmonic-mean Best Rank (PHBR) score that quantifies patients’ ability to present somatic mutations in their tumor by their specific MHC-I and MHC-II haplotypes\(^21,22\). PHBR-I and PHBR-II scores aggregate predicted peptide/MHC molecule binding affinities from established tools\(^23,24\) to produce a mass spectrometry-validated, residue-centric, and patient-specific presentation score that captures a mutant peptide’s visibility to the immune system. In previous publications we used PHBR scores to assess the role of MHC genotype in shaping mutation accumulation during tumorigenesis\(^21,22\). We found that patients tend to accumulate driver mutations that cannot be effectively presented by their own MHC molecules, likely a consequence of immune-based elimination of tumor cells harboring well-presented driver mutations, a selective process referred to as immunoediting\(^25\). This analysis revealed that thyroid carcinoma and low-grade glioma patients experience the highest MHC-based selective pressure on driver mutations\(^21,22\). Interestingly, these tumor types also had the youngest average age at diagnosis compared to all studied tumor types. In light of these observations, we reasoned that younger and female patients may experience stronger immunoediting early in their tumor history, accumulating mutations that are less favorably presented by their MHC, i.e., mutations more invisible to their immune system, at the time of diagnosis. Predictably, a depletion of potentially immunogenic mutant peptides would cause ICB to be ineffective. At first approximation we ruled out an effect due to sex-specific (MHC-I Pearson R = 0.99, MHC-II Pearson R = 0.99) or age-specific (MHC-I Pearson R = 0.98, MHC-II Pearson R = 0.99) imbalances in MHC genotype frequencies. Therefore, we sought to test the hypothesis that sex- and age-specific differences in driver mutation presentation are the result of differential immunoediting.

In this study we find that female and younger patients exhibit stronger immune selection in their tumors, measured by the affinity of their observed, expressed driver mutations compared to male and older patients. MHC-II appears to have a stronger effect compared to MHC-I. Our findings, based on TCGA samples, are validated in an independent validation cohort.

**Results**

**Fewer presentable drivers in female and younger patients.** We focused on a set of 1018 driver mutations, defined in\(^27\), as driver mutations are more prevalent in the clonal architecture of an individual’s cancer and confer a selective growth advantage. We assigned MHC-I and MHC-II types using PolySolver and HLA- HD, two exome-based calling methods\(^25,26\), and considered only microsatellite-stable TCGA tumors. After excluding 515 patients from class I and 1064 patients from class II analyses due to HLA genotype incompatibility with NetMHCpan affinity prediction software, 9913 patients with MHC-I calls and 7174 patients with MHC-II calls remained. These patients were diverse in sex, with more males than females (Supplementary Fig. 1A), and a broad distribution of age at diagnosis (Supplementary Fig. 1B). PHBR-I and -II scores were calculated for all patients across the 1018 driver events by taking the harmonic mean of each allele’s best NetMHCpan percentile rank affinity score, providing an estimate of each patient’s potential to present each mutation via MHC-I and MHC-II, respectively. Importantly, the PHBR-I and PHBR-II scores aggregate percentile rank scores of mutated peptides relative to large numbers of random peptide provided by NetMHCpan-4.0 and NetMHCIIpan3.2. For single peptide-HLA pairs, percentile rank scores of 0.5\% and 2\% for MHC-I and 2\% and 10\% for MHC-II have been used to represent strong and weak binding cutoffs respectively\(^28,29\).

To rule out other covariates, we performed a series of control analyses. We categorized patients into subgroups according to sex (male versus female) and age (younger versus older based on pan-cancer 30th and 70th percentiles at age of diagnosis for categorical analyses). For sex-specific analyses, we further excluded seven sex-specific tumor types (breast, cervical, ovarian, uterine, prostate, and testicular cancer). First, we established that there were similar average numbers of driver mutations across sex and age patient groups (Supplementary Fig. 2). We previously found that TCGA patients with somatic MHC-I mutations had altered mutational landscapes, with a higher fraction of binding mutant peptides than patients without MHC-I mutations\(^30\). To ensure that somatic MHC-I mutations would not skew the driver mutation PHBR-I score distributions, we compared scores for patients with and without MHC-I mutations grouped by sex and age and found no significant differences (Supplementary Fig. 3). We then compared the distributions of patient PHBR-I and PHBR-II scores across the 1018 driver mutations (Supplementary Fig. 4A–D) and found
significant \( p \) values, but very small effect sizes between groups. To ensure that the potential to present driver mutations was consistent across sex and age, we compared the fraction of presented drivers at various score thresholds, and found no significant differences (Supplementary Fig. 4E, F). The overall similarity of MHC presentation suggests that patients of both sexes and various ages at diagnosis present driver mutations with roughly equivalent efficacy, implying that specificity of MHC presentation resulting from specific allele combinations is not a mechanism causing differences in ICB response rate.

We therefore reasoned that the discrepancy might be due to differences in the strength of immune selection, e.g., tumors with stronger immunoeediting should retain fewer driver mutations that are presentable to T cells by the patient’s own MHC molecules. For sex- and age-specific groups in each cohort, we compared the PHBR-I and PHBR-II score distributions for observed, RNA-expressed driver mutations observed in patient tumors, excluding 4782 patients with no drivers from the list of 1018. While the number of observed drivers was not significantly different between sex and age groups (Supplementary Fig. 2), younger female patients were overrepresented in the group with no observed driver mutations (Fisher’s exact test: class I: OR = 1.12, \( p < 0.12 \); class II: OR = 1.28, \( p < 0.05 \)). We note this group had an overrepresentation of thyroid cancer cases, a disease associated with low mutational burden and that typically only has a single driver mutation\(^{31}\). We therefore performed sex-specific analysis for unique 2900 patients and age-specific analysis for 3928 unique patients.

Across pan-cancer cohorts, females were at a significant disadvantage (higher PHBR scores) in presenting their driver mutations by both their MHC-I and MHC-II molecules (Fig. 1a, b, \( p < 2.6e^{-04} \) and \( p < 1.2e^{-07} \), respectively). Younger patients also tended to have worse presentation of driver mutations by both MHC-I and MHC-II molecules (Fig. 1c, d, \( p < 2.4e^{-5} \) and \( p < 7.3e^{-04} \), respectively). Notably, the shift in PHBR score distributions between groups occurs near the threshold for weak binding. Given that a limited number of somatic mutations generate mutant peptides and not all of these are immunogenic, this small shift may translate to significantly less opportunity to generate a host antitumor response upon ICB. Importantly, we found that these observed between-group differences in PHBR scores were far greater (falling outside the 99% confidence interval) than differences when we randomly reassigned mutations across patients and recalculated patient-specific PHBR scores (Methods; Supplementary Fig. 5), and were an order of magnitude greater than the effect sizes observed when comparing score distributions independent of mutation occurrence (Supplementary Fig. S4). We also found differences in affinity independent of the PHBR score, using median NetMHCpan affinity scores across all alleles (Supplementary Fig. 6).

Altogether this suggests that score differences do indeed result from the interaction of inherited MHC genotype with the observed mutations. Interestingly, the mutation-specific fraction of RNA reads mapping to these driver mutations was significantly lower for females and younger patients (Supplementary Fig. 7), further supporting sex- and age-based differential strength in immune selection.

We next examined evidence for sex and age differences in specific tumor types, adjusting age thresholds according to tumor type. There was a general trend for female and younger patients’ tumors to have higher median PHBR-I and II scores across tumor types, although the difference was only statistically significant in melanoma (Supplementary Fig. 8A). We observed more variability in the trends across tumor types by age. Younger individuals tended toward higher median PHBR-I and II scores in tumors where the 30th/70th percentile was associated with a large age gap and the younger age threshold was under 55, with some notable exceptions that included rectal cancer, thyroid cancer, stomach cancer, and liver (Supplementary Fig. 8B). Overall these trends suggest that stronger pan-cancer immune selection in younger and female patients results broadly from effects observed across multiple tumor types.

Next, we explored the effect of age and sex in the context of the immune system’s ability to eliminate effectively-presented mutations by modeling the relationship between mutation occurrence and immune visibility as modeled by PHBR-I and II scores. We constructed sex- and age-specific generalized additive models with random effects to account for variation in mutation rate across individuals, and examined the coefficients corresponding to independent and interaction effects for PHBR-I, PHBR-II, and sex or age to assess their contribution to immune selection for expressed mutations observed ≥2 times in the cohort, excluding patients with no observed, expressed driver mutations. To control for the fact that some driver mutations occurred in the same tumor, and thus are not completely independent events, we included patient ID as a random effect in our linear model. In both models, we found that PHBR-I and PHBR-II scores alone had significant effects on the probability of a mutation to be a...
target of immune selection (Table 1). Positive coefficients for both PHBR scores indicate that the higher the PHBR score (i.e., poorer presentation), the higher the probability of mutation. Furthermore, when we quantified the influence of both scores on probability of mutation using odds ratios between respective 25th and 75th percentiles, we found that PHBR-II (OR: 3.4, CI [3.19, 3.6]) has a much larger impact on probability of mutation than PHBR-I (OR: 1.27, CI [1.26, 1.29]), echoing the larger effect sizes seen in Fig. 1. As expected, sex and age alone did not influence the probability of mutation; however, of particular interest are the interaction terms that indicate the influence of PHBR scores on probability of mutation within the context of sex and age. Both the PHBR-I:sex and PHBR-I:age interactions as well as the PHBR-II:sex and PHBR-II:age interactions were significant. The negative PHBR-I:age estimates indicate stronger effects of PHBR-I as well as PHBR-II contribution to the probability of mutation in younger patients. On the other hand, positive PHBR-sex estimates indicate stronger effects of PHBR-I and PHBR-II contributing to probability of mutation in females according to the model formulation (Methods). Collectively, these results suggest stronger immune selection in females and younger patients.

As females and younger patients both demonstrated stronger immune selection compared to males and older patients, we further partitioned the cohorts simultaneously by sex and age, and investigated the distribution of PHBR-I and -II scores for these groups. We found that sex and age effects are cumulative, with tumors in younger females exhibiting -II scores for these groups. We found that sex and age effects were signiﬁcantly worse in younger and female patients compared to older and male patients (Fig. 4a–d). These differences were an order of magnitude greater than the effect sizes observed when comparing score distributions independent of mutation occurrence (Supplementary Fig. S9E–H).

When we examined the simultaneous effects of sex and age (Fig. 4e, f), younger females once again had significantly worse presentation of their driver mutations than older males across both MHC-I and MHC-II (p < 0.001, p < 0.007). We repeated the sex- and age-specific analyses using the generalized additive models and found that, for both sex and age, PHBR-II scores alone signiﬁcantly inﬂuenced the probability of mutation, with
higher PHBR scores (i.e., worse presentation) leading to higher probability of mutation (Supplementary Table 3). While PHBR-II:sex and PHBR-II:age coefficients trended in the same direction, with stronger effects in females and younger patients, they did not reach significance, likely due to sample size.

Discussion

Here, we present evidence that both sex and age impact the driver mutations that arise and persist during tumorigenesis. We found that younger and female patients accumulate driver mutations in their tumors that are less readily presented by their MHC molecules (Fig. 5), suggesting a stronger toll by immune selection early in tumorigenesis. This finding is consistent with recent meta-analyses across multiple tumors showing sex- and age-dependent differences in response to ICB. We also observed the strongest effects in MHC-II based selection, in agreement with the fact that females have higher CD4+ T-cell counts than males. A prevalent role of MHC-II driven immune selection can be explained by the fact that CD4+ T cells, besides direct effector function comparable to that of CD8+ T cells, also play a deep-rooted regulatory role in cooperating with CD8+ T cells via associative recognition of antigen. Their function in orchestrating T-cell immunity, in general terms, makes them privileged actors, hence targets of immune selection as revealed herein. In older individuals, immune selection effects by MHC-II presentation of driver mutations are mitigated by a reduced CD4+/CD8+ ratio and greater telomere attrition in CD4+ T cells than in CD8+ T cells leading to accelerated senescence. Taken together, the
evidence suggests that tumors developing in younger and female patients are prone to stronger immunoeediting than those in older and male patients.

Our findings based on the TCGA were reproduced in the smaller validation cohort where we once again observed poorer MHC-based presentation of driver mutations in females versus males and younger versus older patients, with presentation being worse in younger and female patients. When modeling the influence of MHC genotype on the probability of observing driver mutations, the estimated effect sizes are modest, although relatively large compared to effects detected by genome wide association studies where odds ratios are often <1.240. Several sources of uncertainty, including errors in patient genotyping, prediction of the peptide–HLA binding affinities used to calculate the PHBR score, and errors in somatic mutation calling could obscure the true effects21. More accurate estimates will likely require larger sample sizes, and ideally availability of expression data as non-expressed mutations should not reflect the effects of immune selection.

In this analysis, we focused on a set of recurrent missense and indel mutations in established driver genes developed in our previous work. This is motivated by the assumption that these are more likely to occur early during tumorigenesis, and may thus provide a view of immune selection before various mechanisms of immune evasion occur22. However it is unlikely that immune selection operates differently on different categories of mutation, and non-driver mutation-derived neoantigens should be equally capable of triggering a T-cell response. Whether tumor cells can evade T-cell responses more easily when they are targeted against nonessential nondriver mutations remains an important question.

It has been suggested that ICB responses are most effective when a clonal driver neoantigen is present41. While we did not observe large sex or age bias in the mutational signatures associated with the 1018 driver mutations, we speculate that it is possible non-driver mutations could show differences in their potential to serve as neoantigens if the underlying mutational processes are active at different times or are biased to generate mutations in expressed protein coding sequences with characteristics that bias their presentation.

Notwithstanding some limitations, our analysis provides a compelling case for the paradigm that immune selection exerts its toll differently with respect to sex and age, with a greater effect in younger females. Of note, the younger female cohort had the poorest driver mutation presentation across both the discovery...
and validation cohorts, suggesting that these effects are strong and complementary. Although our analysis suggests that younger age is associated with stronger antitumor immune responses, we strongly suggest caution in considering whether this trend could generalize to pediatric tumors. The genomic landscape of pediatric tumors is distinct from that of adulthood tumors, with lower mutation burdens, different driver events and more germline factors and the characteristics of the pediatric immune system differ greatly from those of an adult. Furthermore, we are unable to control for other sex- and age-related factors beyond predicted MHC presentation of driver mutation-derived peptides. These possibilities may include (a) differences in the antigen processing machinery preceding surface exposure of MHC-peptide complexes, and (b) genetic and epigenetic factors causing preferential mutation accumulation in the cohorts for reasons other than immunoediting.

In conclusion, this study indicates that immune selection exerts its toll differently with respect to sex and age, with a greater effect in younger females. As such, the response rate to ICB may be dependent on the strength of immune selection occurring early in tumorigenesis. Methods to accurately predict the impact of immunoediting on a patient-specific basis may lead to better predictive algorithms for response to therapy. As a corollary, we posit that ICB treatment is likely to have a reduced effect in younger female patients since this treatment will attempt to reactivate T cells for immunologically invisible neoantigens. Rather, adaptive T-cell therapy against patient-validated neoantigens or therapeutic vaccination against conserved antigens will likely be more beneficial in these patients. Notably prior to treatment with ICB, male sex (and less consistently older age) are associated with higher risk of recurrence and death in melanoma and may stand to benefit more from ICB, thus it is also possible that overall stronger immune surveillance could prove advantageous in the context of ICB despite differences in the quality of neoantigens. Finally, these findings shed light on the role of immune surveillance in cancer progression.

**Methods**

**HLA typing.** HLA genotyping was performed for class I genes HLA-A, HLA-B, HLA-C, and class II genes HLA-DRB1, HLA-DPA1, HLA-DPB1, HLA-DQA1, and HLA-DQB1, which encode three protein determinants of MHC-I peptide binding specificity, HLA-DR, HLA-DP, and HLA-DQ. TCGA samples were typed with Polysolver, with default parameters, for class I and typed with HLA-HD, using default parameters, for class II. Both tools require germline (whole blood or tissue matched) whole exome sequenced samples. Samples with very low coverage on specific genes are left untyped by HLA-HD. Patients were assigned an HLA-DR type if they were successfully typed for HLA-DRB1. Patients were assigned HLA-DP and -DQ types if they had successful typing for HLA-DPA1/HLA-DPB1 and HLA-DQA1/HLA-DQB1, respectively. Class I and class II types were validated by xHLA, run with default parameters, and only patients where all alleles agreed in both classes were included in the analysis.

**Presentation score assignment.** We used patient presentation scores, as defined in (1), to represent a particular patient’s ability to present a residue given their distinct set of HLA types. For class I, 6 HLA alleles were considered (HLA-A, HLA-B, and HLA-C). For class II, 12 HLA-encoded MHC-II molecules (4 combinations of HLA-DPA1/DPB1 and HLA-DQA1/DQB1; 2 alleles of HLA-DRB1 considered twice—since HLA-DRA1 is invariant—for consistency between resulting molecules). NetMHCpan4.0 and NetMHCIIpan3.2 were used to calculate binding affinities. The PHBR score was assigned as the harmonic mean of the best residue presentation scores for each group of MHC-I and MHC-II molecules. A lower patient presentation score indicates that the patient’s MHC molecules are more likely to present a residue on the cell surface.

**Set of driver mutations.** Somatic mutations were considered to be recurrent and oncogenic if they occurred in one of the 100 most highly ranked oncogenes or tumor suppressors described by Davoli et al. and were observed in at least three TCGA samples. Among these, we retained only mutations that would result in predictable protein sequence changes that could generate neoantigens, including missense mutations and inframe indels. A total of 1018 mutations (512 missense mutations from oncogenes, 488 missense mutations from tumor suppressors, 11 indels from oncogenes and 7 indels from tumor suppressors) were obtained.

**Modeling the effects of PHBR score on mutation probability.** We built two matrices, for PHBR-I scores and PHBR-II scores, from the 1018 mutations and the 1912 patients with both PHBR-I and -II calls. Next, we built a binary mutation matrix $y_{ij} \in [0,1]$ indicating whether patient $i$ has a specific mutation $j$. We evaluated the relationship between this binary matrix, the matched 1912 × 1018 matrices with log PHBR-I and -II scores, $x_{1i}$ and $x_{2i}$, respectively, and the variable $y_{ij}$.
of interest (sex or age) for patient i and mutation j. We fit a generalized additive model for the centered log PHBR-I, centered log PHBR-II scores, centered sex (coded 0 for males/females) or centered age, and mutation probability with the GAM function in the mgcv R package. To estimate the effects of PHBR and sex and age on probability of mutation, we considered the following random effects models:

\[
\text{Logit}(P(y_{ij} = 1)) = \beta_1 x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + \beta_4 x_{ij} + \beta_5 x_{ij} + \eta_i + \eta_j,
\]

And PHBR-I and PHBR-II specific models (results in Supplementary Table 1):

\[
\text{Logit}(P(y_{ij} = 1)) = \beta_1 x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + \beta_4 x_{ij} + \beta_5 x_{ij} + \eta_i + \eta_j.
\]

where \(\eta_i \sim \text{N}(0, \sigma_i)\) are random effects capturing different mutation propensities among patients, using patient IDs. In these models \(\beta_m\) measures the effect of the log of the PHBR-I, log PHBR-II, and sex or age. This analysis was repeated for the validation cohort.

### Mutational signature analysis

Mutational signatures analysis was performed using a previously developed computational framework SigProfiler. A detailed description of the workflow of the framework can be found in ref. 48, while the code can be downloaded freely from: https://www.mathworks.com/matlabcentral/fileexchange/58724-sigprofiler.

### Predicting RNA expression from DNA variant allelic fraction

To predict binary RNA expression (≥5 reads at the mutant allele), we used the LogisticRegressionCV function from the Python sklearn v0.20.3 package to train a logistic classifier on the TCGA discovery cohort, using DNA variant allelic fraction (VAF), VAF percentile rank within the patient, and mutated gene as features. We conducted 10-fold cross-validation, achieving a mean 72% area under the receiver operating curve.

### Statistical analysis

All box plots were evaluated using the default one-tailed Mann–Whitney U statistical test, via the scipy.stats Python package. Mutational signature sex-specific distributions were also compared using the one-tailed Mann–Whitney U test, and \(p\) values were adjusted using the Benjamini–Hochberg Procedure. All boxplot figures include the median line, the box denotes the interquartile range (IQR), whiskers denote the rest of the data distribution and outliers are denoted by points determined by ±1.5 x IQR. Effect sizes were calculated using Cliff’s d (Cliff 1993).

### Data availability

Discovery cohort: data were obtained from publicly available sources including The Cancer Genome Atlas (TCGA) Research Network. TCGA normal exome sequences and TCGA clinical data were downloaded from the GDC on June 23–26th, 2018 and April 25th, 2017, respectively, using the gdc-client v3.3.0. Furthermore, TCGA somatic mutations were accessed from the NCI Genomic Data Commons on May 14th, 2017. Validation cohort: all remaining relevant data are available in the article, Supplementary Information, or from the corresponding author upon reasonable request.

### Code availability

Code to reproduce findings and figures can be freely accessed at https://github.com/CarterLab/HLA-immunoediting.
Acknowledgements

We would like to thank T. Cameron Waller, Tina Wang, and Trey Ideker for scientific discussion. This work was supported by an NIH National Library of Medicine Training Grant T5SL001127 to A.C. an NSF graduate fellowship #201520259 to R.M.P., NIH grants DPS-OD17957, an Emerging Leader Award from The Mark Foundation for Cancer Research to R.M.P., and a CIFAR fellowship to H.C. and RON CA220009 to M.Z. and H.C., P41-GM103504 for computing resources provided by the National Resource for Network Biology (NRNB). We would like to thank the TCGA research network for providing data used in the analyses, the ICGC database, as well as the following studies used in the validation cohort. phs01493v1p1c2 and phs00145s1v1p1c1 We would also like to thank the Blavatnik Family Foundation, grants from the Broad Institute SPARC program, the National Institutes of Health (NCI-SROI01CA151080-02, NCI-NBRI-01T035-02, NCI-SPORE-2PS0CA19142-11A1, NCI-R01CA1665536, and GM08042). A comprehensive description of the dataset can be found at PMID:29220474. phs00519v1p1c1. We thank the Ben and Catherine Ivy Foundation, the Blavatnik Family Foundation, the Broad Institute SPARC program, and NIH (NCI-IROCA155010-02 (to C.J.W.). HLBII-5901I013532-03 (to J.W.N.), Francis and Adele Kittredge Family Immuno-Oncology and Melanoma Research Fund (to P.A.O.), Fairchild Family Research Fund (to P.A.O.), NIH/NCI R21CA216772-01A1 (to D.B.K.), NCI-SPORE-2PS0CA19142-11A1 (to D.B.K.), NHLBI-T32HL07627 (to J.R.L). R50CA211482 (to S.A.S.), Zuckerman STEM Leadership Program (to LT); Benzozy Endowment Fund for the Advancement of Science (to LT); P50 CA165962 (SPORE) and P01 CA61365 (to K.L.L.); DFCI Center for Cancer Immunotherapy Research fellowship (to Z.H.); Howard Hughes Medical Institute Medical Research Fellows Program (to A.J.A.); and American Cancer Society PF-17-04-01-LB (to N.D.M.). C.J.W. is a scholar of the Leukemia and Lymphoma Society. We thank the Center for Neuro-Oncology, J. Russell and Dana-Farber Cancer Institute (DFCI) for Immuno-Oncology (CIO) staff; B. Meyers, C. Harvey and S. Bartel (Clinical Pharmacy); M. Severgnini, K. Kleinsteuber, and E. McWilliams (CIO laboratory); M. Coperisino (Regulatory Affairs), T. Bowman (DFHCC Specialized Histopathology Core Laboratory); A. Lako (CIO); M. Seaman and H. Baroosh (BDIMC); the Broad Institute’s Biological Samples, Genetic Analysis and Genome Sequencing Platforms; J. Petricciani and M. Krane for regulatory advice; B. McDonough (CSBio), I. Javeri and K. Nellaiappan (CurRx) for peptide development. phs00156s1v1p1c1 The research reported in this article was supported by BroadBridge, BroadNest10, NIH R08CA186615, the Howard Hughes Medical Institute, the American Cancer Society Grant #18276 (Z.H.) and Stand Up To Cancer—American Cancer Society Lung Cancer Dream Team Translational Research Grant (grant number: SU2C-AACR-DT17-15). Stand Up To Cancer is a program of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the scientific partner of SU2C.

Author contributions

Original concept, R.M.P.; project supervision, H.C. and M.Z.; project planning and experimental design, A.C., R.M.P., C.P.D., M.Z., and H.C.; statistical advising, X.Z., M. Severgnini, M. Copersisno (Regulatory Affairs), T. Bowman (DFHCC Specialized Histopathology Core Laboratory); A. Lako (CIO); M. Seaman and H. Baroosh (BDIMC); the Broad Institute’s Biological Samples, Genetic Analysis and Genome Sequencing Platforms; J. Petricciani and M. Krane for regulatory advice; B. McDonough (CSBio), I. Javeri and K. Nellaiappan (CurRx) for peptide development. phs00156s1v1p1c1 The research reported in this article was supported by BroadBridge, BroadNest10, NIH R08CA186615, the Howard Hughes Medical Institute, the American Cancer Society Grant #18276 (Z.H.) and Stand Up To Cancer—American Cancer Society Lung Cancer Dream Team Translational Research Grant (grant number: SU2C-AACR-DT17-15). Stand Up To Cancer is a program of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the scientific partner of SU2C.

Competing interests

R.M.P. is an employee and holds stock in Personalis. The remaining authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-17981-0.

Correspondence and requests for materials should be addressed to H.C.

Peer review information Nature Communications thanks Joshua Rubin and the other, anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

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/.

© The Author(s) 2020