

Persistent mutation burden drives sustained anti-tumor immune responses

Received: 1 December 2021

Accepted: 30 November 2022

Published online: 26 January 2023

 Check for updates

Noushin Niknafs¹, Archana Balan¹, Christopher Cherry¹, Karlijn Hummelink², Kim Monkhorst², Xiaoshan M. Shao¹, Zineb Belcaid¹, Kristen A. Marrone¹, Joseph Murray¹, Kellie N. Smith^{1,3}, Benjamin Levy¹, Josephine Feliciano¹, Christine L. Hann¹, Vincent Lam¹, Drew M. Pardoll^{1,3}, Rachel Karchin^{1,4}, Tanguy Y. Seiwert¹, Julie R. Brahmer^{1,3}, Patrick M. Forde^{1,3}, Victor E. Velculescu^{1,3} & Valsamo Anagnostou^{1,3} ✉

Tumor mutation burden is an imperfect proxy of tumor foreignness and has therefore failed to consistently demonstrate clinical utility in predicting responses in the context of immunotherapy. We evaluated mutations in regions of the genome that are unlikely to undergo loss in a pan-cancer analysis across 31 tumor types ($n = 9,242$) and eight immunotherapy-treated cohorts of patients with non-small-cell lung cancer, melanoma, mesothelioma, and head and neck cancer ($n = 524$). We discovered that mutations in single-copy regions and those present in multiple copies per cell constitute a persistent tumor mutation burden (pTMB) which is linked with therapeutic response to immune checkpoint blockade. Persistent mutations were retained in the context of tumor evolution under selective pressure of immunotherapy and tumors with a high pTMB content were characterized by a more inflamed tumor microenvironment. pTMB imposes an evolutionary bottleneck that cancer cells cannot overcome and may thus drive sustained immunologic tumor control in the context of immunotherapy.

The current working hypothesis for tumor-intrinsic features that determine the magnitude of anti-tumor immune responses relies on the assumption that each mutation contributes equally to a composite measure of tumor foreignness, reflected in the number of sequence alterations per coding DNA sequence or tumor mutation burden (TMB). However, with the exception of mismatch repair-deficient tumors, TMB has failed to consistently demonstrate clinical utility in predicting responses to cancer immunotherapy. Efforts to separate subsets of alterations that may predominantly drive an effective anti-tumor immune response have yet to reveal a universal genomic predictive biomarker^{1,2}. We have previously shown that heterozygous mutations and neoantigens can be selectively eliminated through chromosomal deletions and loss of heterozygosity (LOH) conferring acquired resistance to

immune checkpoint blockade (ICB)³. In line with these findings, we have discovered that a higher number of sequence alterations contained in single-copy regions of the genome differentiate responding from nonresponding tumors in the context of immunotherapy⁴. Together, these findings suggested that mutations and associated neoantigens contained in regions of the genome present in a single copy per cancer cell are less likely to be eliminated by chromosomal loss under the selective pressure of therapy and therefore may mediate sustained neoantigen-driven immune responses and long-term clinical benefit⁴.

To extend these findings beyond single-copy genomic regions, we hypothesized that tumors with a higher frequency of sequence alterations in either haploid regions or multiple copies would have a fitness disadvantage in the context of immunotherapy, as these alterations

¹The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ²Netherlands Cancer Institute, Amsterdam, the Netherlands. ³The Bloomberg-Kimmel Institute for Cancer Immunotherapy, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ⁴Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD, USA. ✉e-mail: vanagno1@jhmi.edu

would continuously render them visible to the immune system, resulting in sustained immunologic tumor control. Deletions of single-copy alleles through chromosomal loss are typically not tolerated unless they are relatively small homozygous deletions⁵, as larger chromosomal deletions could contain essential genes in linkage with the mutation and thus be lethal. Similarly, mutation loss by chromosomal deletions³ is evolutionarily unlikely when mutations are contained in multiple copies. Therefore, these ‘persistent’ mutations (which we hereafter refer to as pTMB) may function as an intrinsic driver of tumor rejection in the tumor microenvironment (TME) (Extended Data Fig. 1).

Results

To investigate these hypotheses in a pan-cancer manner, we first evaluated the rate of loss in regions of the genome with a single copy per cell (haploid) versus euploid regions (two copies per cell) using copy number profiles of 5,244 tumors across 31 tumor types from The Cancer Genome Atlas (TCGA) (Supplementary Table 1). These analyses revealed that the rate of loss in haploid regions was consistently lower than that in euploid regions (Fig. 1a), supporting the notion that mutations contained in these regions would be difficult to eliminate. We then examined the frequency of haploid and polyploid regions across the genome and quantified the fraction of the genome in single-copy versus multi-copy states ($N = 9,991$; Fig. 1b and Supplementary Table 2). Some tumor types, including endometrial carcinosarcomas, bladder cancers, adrenocortical carcinomas, lung squamous carcinomas, lung adenocarcinomas, ovarian cancers and cutaneous melanomas, were enriched for genomic regions in the multi-copy state, while cholangiocarcinomas, pancreatic adenocarcinomas, mesotheliomas and kidney chromophobe tumors showed a higher genome fraction in the single-copy state (Fig. 1b). Integration of sequence alterations in only-copy and multi-copy states revealed a cancer lineage-dependent distribution of persistent mutations (Fig. 1c,d and Supplementary Table 2). Next, we characterized the distribution of persistent mutation load in the background of the overall TMB within each tumor type, and found that TMB does not fully explain the abundance of multi-copy and only-copy mutations, as tumor types with similar TMB exhibited differences in multi-copy and only-copy mutation content ($N = 9,242$; Fig. 1d and Supplementary Table 2). Notably, a wide range of prevalence of mutations in only-copy or multi-copy states was observed across the range of overall TMB, suggesting that persistent mutations provide a measure of alterations that is distinct from TMB (Extended Data Fig. 2). We further evaluated the degree of correlation between TMB and pTMB and found a substantial degree of variation in their association across the 31 tumor types analyzed (Spearman ρ : median 0.49, range: 0.02–0.89; Fig. 2a and Supplementary Table 3). Similar patterns were observed when multi-copy (Spearman ρ median: 0.42, range: 0.02–0.76) and single-copy mutations (Spearman ρ median: 0.21, range: –0.12 to 0.48) were considered separately. To understand the potential reclassification of tumors based on pTMB, we employed a series of quantile values ranging from 5% to 95% to define high/low groups for TMB and pTMB (Methods). These analyses revealed reclassification rates as high as 53% (range: 15–53%), with a median reclassification rate of 33% across all tumor types (Fig. 2b and Extended Data Fig. 3).

Next, we explored the relationship between persistent mutation content and TMB in seven published ICB cohorts across three tumor types ($n = 485$; melanoma^{6–9}, non-small-cell lung cancer (NSCLC)^{2,10} and mesothelioma¹¹) and a new cohort of patients with HPV-negative head and neck cancer (head and neck squamous cell carcinoma (HNSCC)) who received ICB ($n = 39$; Supplementary Tables 4–8). Similar to the TCGA analyses, we did not detect a significant enrichment for a higher persistent mutation fraction in tumors harboring a higher TMB in the HNSCC (Spearman $\rho = -0.083$, $P = 0.61$), melanoma (Spearman $\rho = 0.066$, $P = 0.35$) and mesothelioma cohorts (Spearman $\rho = 0.065$, $P = 0.69$), while a weak correlation between TMB and persistent mutation fraction was observed in the NSCLC cohorts (NSCLC-Anagnostou:

Spearman $\rho = 0.26$, $P = 0.03$; NSCLC-Shim: Spearman $\rho = 0.41$, $P = 2.3 \times 10^{-8}$; Fig. 2c and Supplementary Table 9). Collectively, these findings further support the notion that pTMB is distinct from TMB and that tumors are differentially ranked by their pTMB content in a cancer lineage-dependent manner.

We theorized that the impact of pTMB would be exemplified in the context of ICB treatment, where inherent anti-tumor immune responses would be enhanced and sustained in the presence of continued persistent mutation-associated neoantigen (pMANA) stimulation. As our analyses pointed towards subsets of mutations within TMB that may carry differential weights in demarcating tumor foreignness, we evaluated persistent mutations in comparison with mutations that are more likely to be lost in the context of tumor evolution. We refer to the latter as ‘loss-prone’ mutations and these account for the majority of coding alterations that constitute a tumor’s TMB. We next asked the question of whether there are differential clonal compositions between the persistent and loss-prone mutation subsets. In the HNSCC, mesothelioma and NSCLC cohorts, we did not detect a difference in the fraction of clonal alterations between persistent mutations and loss-prone mutations, while in the melanoma cohort, persistent mutations tended to be more clonal (Extended Data Fig. 4). When multi-copy mutations were considered separately, higher cellular fractions were noted for the multi-copy subset compared with loss-prone mutations in melanoma and HNSCC, in line with the notion that multi-copy mutations may be acquired before somatic copy number gains. To further study the clonal architecture of persistent mutations, we evaluated the correlation between pTMB and the fraction of clonal mutations in 31 TCGA tumor types and in the ICB cohorts. In the TCGA dataset, we observed a wide range of correlations between pTMB and fraction of clonal mutations (Spearman ρ range: –0.11 to 0.59; Extended Data Fig. 4 and Supplementary Table 3). We found a moderate degree of anti-correlation between clonal mutation fraction and the number of only-copy mutations in most tumor types (Spearman ρ range: –0.57 to 0.07), while a moderate degree of correlation was detected between the fraction of clonal mutations and multi-copy mutations (Spearman ρ range: 0.01–0.60; Extended Data Fig. 4). In the ICB cohorts, we did not detect strong correlations between tumor clonal heterogeneity and persistent mutations (Spearman ρ range: –0.25 to 0.33) (Supplementary Fig. 1 and Supplementary Table 9). These findings suggest that persistent mutations are encountered at the full spectrum of tumor clonal heterogeneity.

Conceptually, persistent mutations, which by definition reside in aneuploid regions of the genome, are integrally linked with tumor aneuploidy and, next, we assessed the relationship between persistent mutations and fraction of the genome with allelic imbalance. In the ICB cohorts, a moderate degree of correlation was observed between tumor aneuploidy and pTMB (Spearman ρ range: 0.39–0.60; Extended Data Fig. 5). In parallel, we evaluated tumors for whole-genome doubling (WGD) events, which would enable the acquisition of additional mutant copies of the mutations present before the doubling event, and may therefore be a critical contributor to pTMB. Indeed, in all ICB cohorts analyzed, tumors with WGD also harbored a higher number of multi-copy mutations (HNSCC, Mann–Whitney $P = 1.6 \times 10^{-5}$; melanoma, $P = 3.14 \times 10^{-14}$; mesothelioma, $P = 8.24 \times 10^{-6}$; NSCLC-Anagnostou, $P = 6.82 \times 10^{-9}$; NSCLC-Shim, $P = 9.3 \times 10^{-17}$; Supplementary Fig. 2). Notably, tumors that have undergone WGD are by definition expected to have a very small fraction of the genome at total copy number of 1, and, consistent with this notion, we observed a much lower prevalence of only-copy mutations in genomes with WGD.

We investigated potential bias related to different timing of acquisition of persistent mutations, background mutation rates and accuracy of mutation calls in these loci and evaluated the distribution of sequence properties such as GC content and replication timing as well as mutation call quality in persistent versus loss-prone mutations in 9,242 tumors from TCGA (Extended Data Fig. 6). We found a similar GC

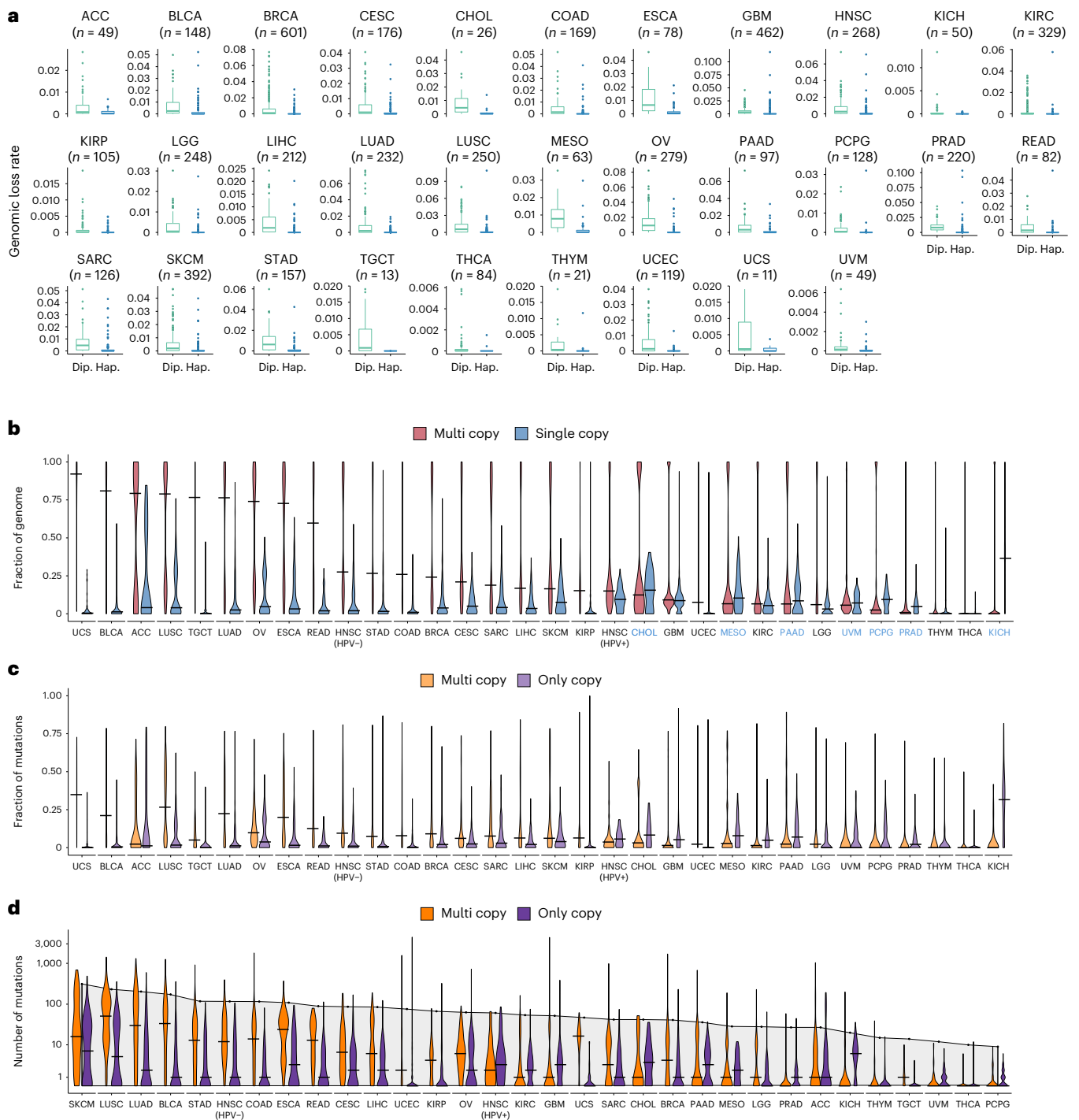


Fig. 1 | Pan-cancer distribution of persistent mutation load. a, The background rate of genomic loss was quantified in 31 tumor types from TCGA ($N = 5,244$). In all tumor types, the rate of loss was significantly higher in diploid (allele-specific copy numbers 1–1) versus haploid (single-copy; allele-specific copy numbers 1–0) regions of the genome (two-sided Mann–Whitney U -test, $P < 0.05$ for all tumor types except for UCS for which Mann–Whitney $P = 0.067$), supporting the notion that mutations that reside in haploid regions of the genome would less likely be lost. Box plots depict the median value and hinges correspond to the first and third quartiles. The whiskers extend from the corresponding hinge to the furthest value within $1.5 \times$ interquartile range from the hinge. **b**, An analysis of somatic copy number aberrations in 9,991 tumors across 31 tumor types from TCGA identified the fraction of the genome with a single copy present (total copy number of 1) and with multiple copies of a parental haplotype. A differential enrichment pattern was noted, whereby cancers including UCS, BLCA, ACC, LUSC, LUAD, OV and SKCM had a higher fraction of the genome with multiple

copies compared with CHOL, PAAD, MESO and KICH which showed a higher genome fraction in the only-copy state. Violin plots depict the distribution of genome fractions in each state, and the horizontal black segments indicate median values. Tumor types with predominance of single-copy genome fraction are marked in blue font. **c**, The prevalence of somatic mutations present in multiple copies per cell (multi-copy), and those present in haploid regions of the genome (only-copy), is depicted for 9,242 tumors, highlighting similar differential distributions per cancer type. **d**, The number of multi-copy and only-copy somatic mutations is shown against a background of the median TMB of the corresponding tumor type. Notably, the median TMB within a tumor type does not fully predict the abundance of multi-copy and only-copy mutations, and tumor types with very similar TMB may exhibit differences in multi-copy and only-copy mutation load (for instance, UCS versus GBM and SARC). Dip., diploid; Hap., haploid; HPV+, human papillomavirus positive; HPV-, human papillomavirus negative.

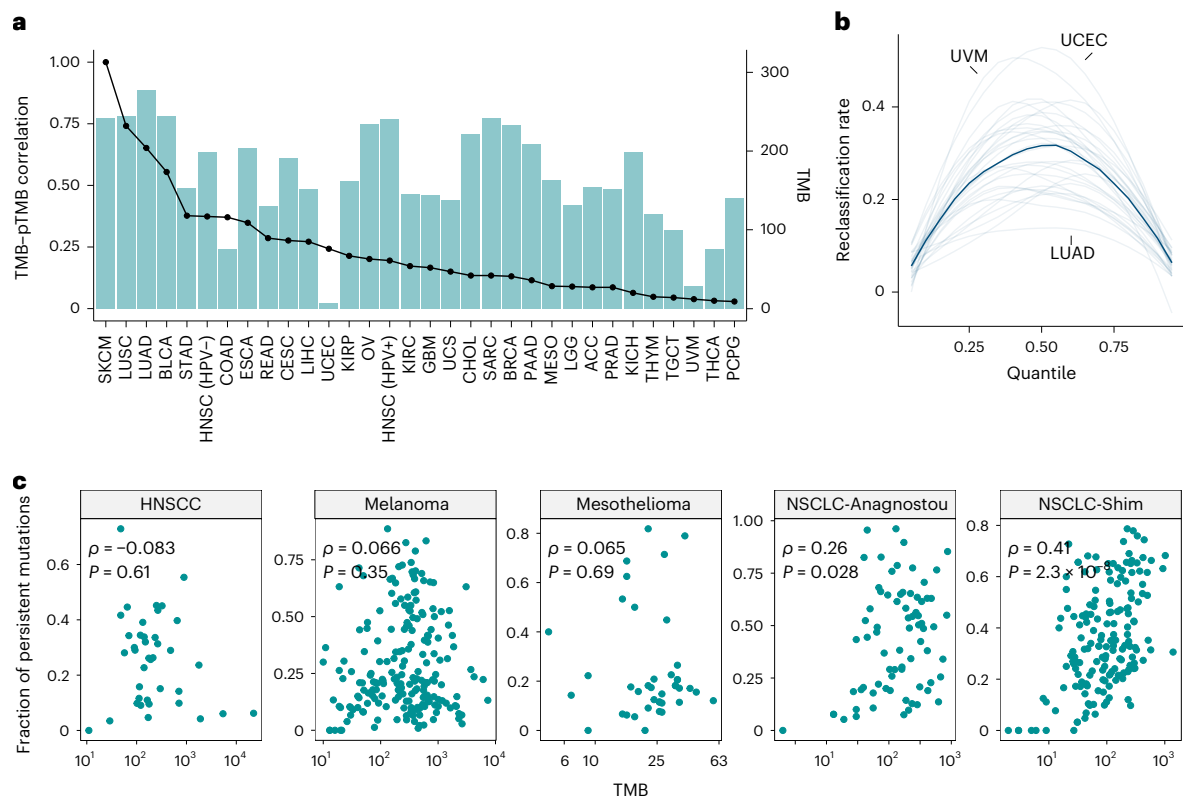


Fig. 2 | Evaluation of the association between persistent mutation content and TMB. a, Analysis of 9,242 tumors across 31 tumor types revealed a large variation in the correlations between TMB and pTMB (blue bars), which is not entirely explained by TMB alone (median TMB values within each tumor type are shown in solid trace). **b**, The tumor reclassification rate was calculated by applying a given quantile of TMB and pTMB and is shown for 31 tumor types within the TCGA cohort (semi-transparent traces), with the median reclassification rate depicted in a solid line. In tumors such as UVM and UCEC, up to 40% of the samples could be differentially classified as pTMB-high versus pTMB-low using persistent mutations rather than the overall TMB value.

c, The fraction of persistent mutations exhibited a variable degree of correlation with TMB in five ICB-treated cohorts ($n = 524$) across four tumor types. In HNSCC ($n = 39$), melanoma ($n = 202$) and mesothelioma ($n = 40$), no significant correlation was observed (HNSCC: Spearman $\rho = -0.083$ and $P = 0.61$; melanoma: $\rho = 0.066$ and $P = 0.35$; mesothelioma: $\rho = 0.065$ and $P = 0.69$; two-sided P values were calculated assuming asymptotic t approximation), while a weak/moderate correlation was observed in NSCLC (NSCLC-Anagnostou, $n = 74$, Spearman $\rho = 0.26$, $P = 0.028$; and NSCLC-Shim, $n = 169$, Spearman $\rho = 0.41$, $P = 2.3 \times 10^{-8}$). Spearman's rank correlation coefficients and corresponding P values are shown in inserts in panel **c**.

composition surrounding loci with persistent and loss-prone mutations (Cohen's $d = 0.08$, persistent mean = 0.54, loss-prone mean = 0.52). We then performed a cancer lineage-specific evaluation of replication timing in the melanoma and NSCLC subsets and found a similar distribution in persistent and loss-prone mutations (melanoma Cohen's $d = -0.035$, NSCLC Cohen's $d = -0.032$). Furthermore, we quantified the fraction of mutations in each category that could theoretically be affected by limitations of next-generation sequencing (NGS) analysis^{12,13} and found similar distributions in persistent and loss-prone mutations (Extended Data Fig. 6). These findings suggest that persistent mutation calling is not confounded by background mutation rates, replication timing or technical artifacts.

We then evaluated whether a higher pTMB was linked with clinical outcome in patients with previously untreated tumors from TCGA (Methods). Our analyses showed that the association between pTMB and clinical outcome was context-dependent, whereby a significant association with prolonged overall survival was noted for lung squamous cell carcinoma (pTMB: 56.27 versus 43.86 months, log-rank $P = 0.085$; clonal pTMB: 60.48 versus 35.32 months, log-rank $P = 0.028$), melanoma (pTMB: 65.83 versus 23.69 months, log-rank $P = 0.036$; clonal pTMB: 65.83 versus 23.69 months, log-rank $P = 0.013$) and uterine carcinosarcoma (pTMB: 27.53 versus 17.15 months, log-rank $P = 0.021$; clonal pTMB: 50.13 versus 14.68 months, log-rank $P = 2.66 \times 10^{-3}$), but not for any other cancer type studied (Extended Data Fig. 7 and Supplementary Table 10). TMB was more weakly

associated with overall survival in the lung squamous cell carcinoma (log-rank $P = 0.50$), melanoma (log-rank $P = 0.98$) and uterine carcinosarcoma (log-rank $P = 0.17$) sets.

Importantly, we hypothesized that tumors with a high persistent mutation content would be the most 'visible' to the immune system and would therefore regress in the context of immunotherapy, a phenomenon that would be reflected in sustained clinical responses to therapy. To this end, we evaluated the potential of pTMB, multi-copy and only-copy mutations in predicting ICB response in 524 patients with melanoma, NSCLC, mesothelioma and HNSCC (Methods and Supplementary Table 9). We discovered that tumors with a high pTMB attained higher rates of therapeutic response with ICB, while TMB alone or the number of loss-prone mutations less optimally distinguished responding from nonresponding tumors (Fig. 3 and Supplementary Table 11). As a representative example that illustrates the difference between pTMB and TMB, patient 44 with metastatic melanoma harboring a pTMB in the 81% quantile but in the 59% quantile by TMB attained a prolonged progression-free survival on ICB (Fig. 3a). High pTMB more accurately differentiated responding from nonresponding tumors in the melanoma cohort ($n = 202$, Mann-Whitney U -test $P = 2.3 \times 10^{-6}$, $P = 6.0 \times 10^{-7}$, $P = 1.92 \times 10^{-3}$ and $P = 2.6 \times 10^{-5}$ for pTMB, clonal pTMB, loss-prone mutation load and TMB, respectively; Fig. 3b). Similarly, in the HNSCC ICB cohort, pTMB was associated with therapeutic response ($n = 39$, Mann-Whitney U -test $P = 0.05$, $P = 0.06$, $P = 0.16$ and $P = 0.09$ for pTMB, clonal pTMB, loss-prone mutations and TMB, respectively;

Fig. 3c). In the mesothelioma cohort, pTMB outperformed TMB in predicting response to durvalumab plus platinum-pemetrexed chemotherapy ($n = 40$, Mann–Whitney U -test $P = 0.03$, $P = 0.05$, $P = 0.09$ and $P = 0.12$ for pTMB, clonal pTMB, loss-prone mutations and TMB, respectively; Fig. 3d). A higher pTMB differentiated responding from nonresponding NSCLC (NSCLC-Anagnostou, $n = 74$, Mann–Whitney U -test $P = 1.3 \times 10^{-4}$, $P = 1.0 \times 10^{-4}$, $P = 0.01$ and $P = 4.3 \times 10^{-4}$ for pTMB, clonal pTMB, loss-prone mutations and TMB, respectively; NSCLC-Shim, $n = 169$, $P = 1.9 \times 10^{-3}$, $P = 1.6 \times 10^{-3}$, $P = 0.03$ and $P = 8.0 \times 10^{-3}$ for pTMB, clonal pTMB, loss-prone mutations and TMB, respectively; Fig. 3e,f).

Next, we evaluated the effect size of persistent mutations, loss-prone mutations and TMB on clinical outcome (Supplementary Table 11). In the melanoma, HNSCC and mesothelioma cohorts, the effect size for pTMB was larger than TMB or loss-prone mutations (HNSCC: pTMB Cohen's $d = -0.96$, TMB $d = -0.64$, loss-prone $d = -0.61$; melanoma: pTMB $d = -0.57$, TMB $d = -0.44$, loss-prone $d = -0.35$; mesothelioma: pTMB $d = -0.74$, TMB $d = -0.51$, loss-prone $d = -0.58$). In the NSCLC cohorts, the effect size for pTMB, while very close to that of TMB, exceeded the effect size of loss-prone mutations (NSCLC-Anagnostou: pTMB $d = -0.89$, TMB $d = -0.93$, loss-prone $d = -0.58$; NSCLC-Shim: pTMB $d = -0.53$, TMB $d = -0.54$, loss-prone $d = -0.44$). These findings suggest that the power of TMB to distinguish between responding and nonresponding tumors in the context of ICB is largely driven by their persistent mutation content. Notably, in the NSCLC cohort, clonal pTMB more optimally distinguished responding from nonresponding tumors (Mann–Whitney U -test $P = 1.03 \times 10^{-4}$ and $P = 1.60 \times 10^{-3}$ for NSCLC-Anagnostou and NSCLC-Shim, respectively). In the melanoma cohort, the number of multi-copy mutations was tightly correlated with therapeutic response (Mann–Whitney U -test $P = 5.42 \times 10^{-7}$), while in the mesothelioma cohort, the number of only-copy mutations better distinguished responding and nonresponding tumors (Mann–Whitney U -test $P = 3.15 \times 10^{-2}$; Fig. 3g). Importantly, pTMB outperformed loss-prone mutation content in all ICB cohorts, despite the latter representing a larger fraction of the overall TMB (Mann–Whitney U -test $P = 1.92 \times 10^{-3}$ versus 2.25×10^{-6} , $P = 0.16$ versus 0.05 , $P = 0.09$ versus 0.03 , 1.03×10^{-2} versus 1.26×10^{-4} , $P = 3.20 \times 10^{-2}$ versus 1.87×10^{-3} for loss-prone versus pTMB in melanoma, HNSCC, mesothelioma, NSCLC-Anagnostou and NSCLC-Shim, respectively; Fig. 3g). We next asked the question of whether the tumors differentially classified by pTMB compared with TMB have different responses to ICB. To this end, we evaluated the number of cases with differential pTMB/TMB classification and compared the therapeutic response rates between pTMB-low/TMB-high and pTMB-high/TMB-low tumors. In the melanoma cohort, 23 tumors fell in the TMB-high/pTMB-low category and 23 tumors fell in the TMB-low/pTMB-high category. We found a higher frequency of responding tumors in the pTMB-high/TMB-low

category compared with the pTMB-low/TMB-high group (Fisher's exact $P = 0.04$, pTMB-high/TMB-low group: 16 responders, 7 nonresponders; pTMB-low/TMB-high group: 8 responders, 15 nonresponders). These findings tie into the reclassification analyses from the TCGA cohorts, and together support a differential classification of tumors based on their persistent mutation content, which is reflective of improved outcomes in the ICB setting.

To further explore the immediate clinical utility of pTMB, we evaluated the feasibility of estimating pTMB from gene panel-targeted NGS, which is widely used in clinical cancer care. Using the genomic intervals from a widely adopted targeted NGS gene panel (309 genes; Methods), we performed in silico simulations utilizing whole exome sequence data from the melanoma and NSCLC ICB cohorts and computed TMB and pTMB in each tumor as captured by the region of interest of the targeted NGS panel. pTMB more accurately distinguished responding from nonresponding tumors (melanoma, $n = 202$, $P = 1.37 \times 10^{-7}$ for pTMB and $P = 1.22 \times 10^{-5}$ for TMB; NSCLC-Shim, $n = 169$, $P = 6.7 \times 10^{-4}$ for pTMB and $P = 0.014$ for TMB; NSCLC-Anagnostou, $n = 74$, $P = 0.02$ for pTMB and $P = 2.0 \times 10^{-3}$ for TMB; Mann–Whitney U -test; Extended Data Fig. 8).

As tumor aneuploidy has been associated with inferior outcomes to ICB³⁴, we investigated whether pTMB has an incremental value over tumor aneuploidy and WGD events in predicting therapeutic response. Tumor aneuploidy (Mann–Whitney U -test $P = 0.35$, $P = 0.73$, $P = 0.35$, $P = 0.07$, $P = 0.50$ for the NSCLC-Anagnostou, NSCLC-Shim, melanoma, mesothelioma and HNSCC cohorts, respectively) or WGD alone (Fisher's exact $P = 0.43$, $P = 0.73$, $P = 0.11$, $P = 0.23$, $P = 0.48$ for the NSCLC-Anagnostou, NSCLC-Shim, melanoma, mesothelioma and HNSCC cohorts, respectively) failed to predict ICB response in all cohorts (Fig. 3g).

To establish the biological plausibility of persistent mutations in the context of tumor evolution, we performed whole exome sequencing analyses of serial tumor samples before and after ICB. We hypothesized that clonal persistent mutations would not be eliminated under the selective pressure of immunotherapy, as they are unlikely to undergo subclonal elimination in the context of therapy and also are unlikely to be lost by chromosomal deletions (potentially lethal in the case of mutations residing in single-copy regions and biologically implausible in the multiple-copy regions). Consistent with our hypothesis, in analyzing pre-treatment and post-acquired resistance NSCLCs from eight ICB-treated patients (Methods and Supplementary Tables 8, 12 and 13), we discovered a marked difference in the frequency of loss between clonal persistent and loss-prone mutations. A total of 363 out of 2,836 clonal mutations that were detected in the baseline tumors were lost in the descendant tumors. Of these, the vast majority were clonal loss-prone mutations (358 out of 363, 98.6%). In six out of

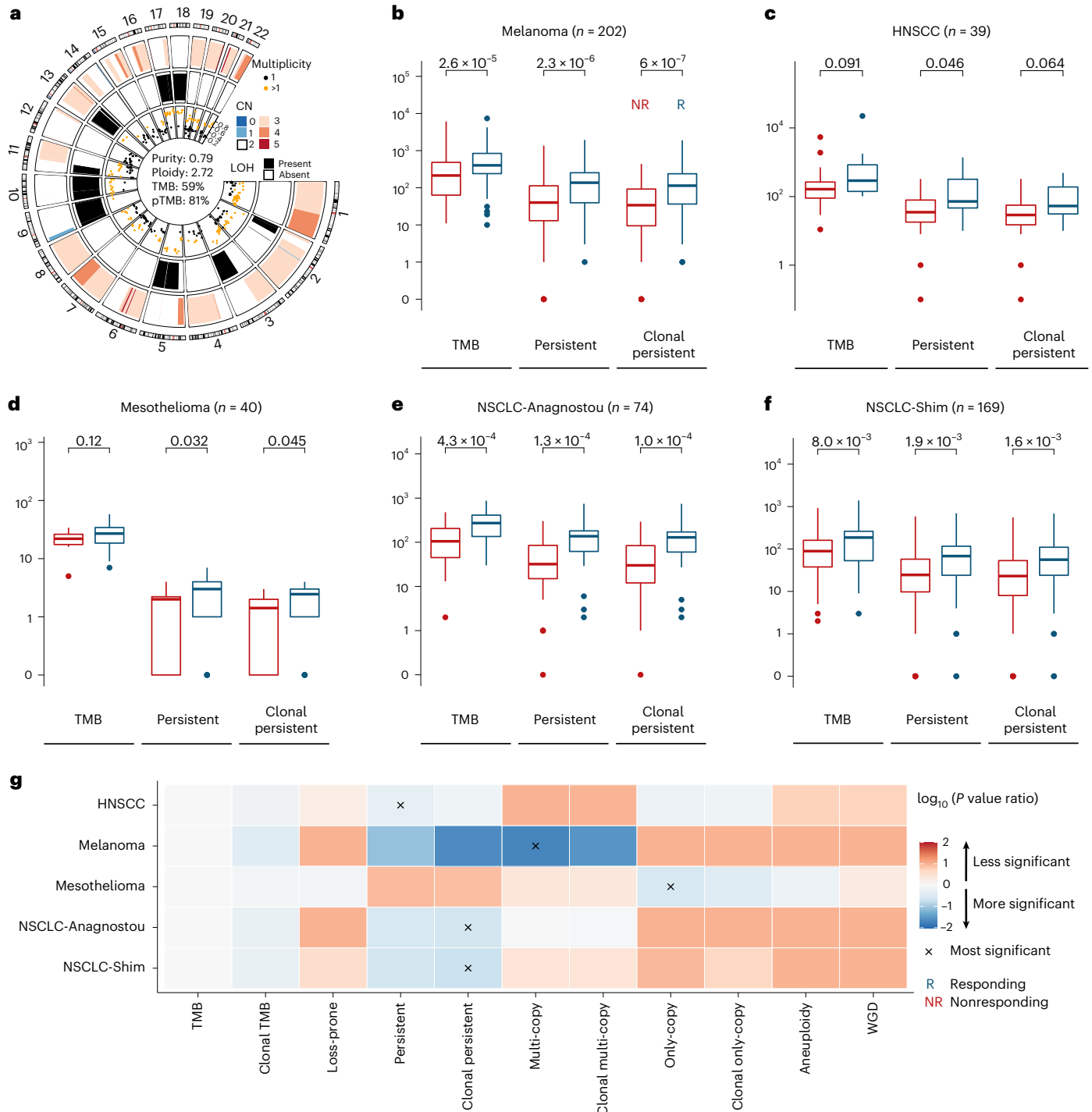
Fig. 3 | pTMB is linked with therapeutic response with ICB. a, A representative example of a responding tumor harboring an intermediate TMB (59% quantile) but high pTMB (81% quantile). The outer and middle ring depict segmental copy number and LOH, respectively. The genomic coordinates of mutations and their mutant allele fraction are shown in the inner ring, colored by their estimated multiplicity. **b**, In melanoma ($n = 202$; NR = 115, R = 87), pTMB distinguished responding from nonresponding tumors (MW $P = 2.3 \times 10^{-6}$ for pTMB, $P = 6.0 \times 10^{-7}$ for clonal pTMB) more optimally compared with TMB (MW $P = 2.6 \times 10^{-5}$). Box plots depict the median value and hinges correspond to the first and third quartiles. The whiskers extend from the corresponding hinge to the furthest value within $1.5 \times$ interquartile range from the hinge. **c**, In HPV-negative HNSCC, we found a greater difference in pTMB of responding versus nonresponding tumors compared with TMB ($n = 39$; NR = 29, R = 10; MW $P = 0.046$ for pTMB, $P = 0.064$ for clonal pTMB, $P = 0.091$ for TMB). **d**, For mesotheliomas, pTMB outperformed TMB in prediction of response to chemo-immunotherapy ($n = 40$; NR = 16, R = 24; MW $P = 0.032$ for pTMB, $P = 0.045$ for clonal pTMB, non-significant for TMB). **e, f**, In NSCLC, a higher pTMB differentiated responding from nonresponding tumors (NSCLC-Anagnostou: $n = 74$; NR = 41, R = 33; MW $P = 1.3 \times 10^{-4}$ for pTMB, $P = 1.0 \times 10^{-4}$ for clonal pTMB,

$P = 4.3 \times 10^{-4}$ for TMB; NSCLC-Shim: $n = 169$; NR = 49, R = 120; MW $P = 1.9 \times 10^{-3}$ for pTMB, $P = 1.6 \times 10^{-3}$ for clonal pTMB, $P = 8.0 \times 10^{-3}$ for TMB). **g**, The significance of association of eight mutation-based and two copy number-based features with therapeutic response was compared with that of TMB (\log_{10} of feature to TMB P value ratios visualized). For each cohort, the feature with the most significant association with therapeutic response is marked with an 'x' mark. For all ICB cohorts, pTMB outperformed TMB in predicting ICB response. Notably, the best-performing feature in NSCLC was clonal pTMB (MW $P = 1.03 \times 10^{-4}$, $P = 1.60 \times 10^{-3}$ for NSCLC-Anagnostou and NSCLC-Shim, respectively), while in melanoma the number of multi-copy persistent mutations (MW $P = 5.42 \times 10^{-7}$) and in mesothelioma the number of only-copy persistent mutations (MW $P = 3.15 \times 10^{-2}$) better distinguished between responding and nonresponding tumors. pTMB outperformed loss-prone mutation load in distinguishing responding from nonresponding tumors (HNSCC: MW $P = 0.16$ versus $P = 0.05$; melanoma: $P = 1.92 \times 10^{-3}$ versus $P = 2.25 \times 10^{-6}$; mesothelioma: $P = 0.09$ versus $P = 0.03$; NSCLC-Anagnostou: $P = 1.03 \times 10^{-2}$ versus $P = 1.26 \times 10^{-4}$; NSCLC-Shim: $P = 3.20 \times 10^{-2}$ versus $P = 1.87 \times 10^{-3}$). Tumor aneuploidy or WGD alone failed to predict ICB response. P values are two-sided. CN, copy number; MW, Mann–Whitney; NR, nonresponding; R, responding.

eight patients analyzed, no clonal persistent mutation was lost in the descendent tumor, and of the two remaining patients, each had two clonal multi-copy mutations that were not detected in the descendent tumor, suggesting an extremely low rate of loss (clonal multi-copy mutations: 4 out of 1,031 lost (0.4%), clonal only-copy mutations: 1 out of 117 lost (0.9%); Fig. 4), and an odds ratio for the loss frequency of clonal loss-prone versus persistent mutations of 61.43 ($P < 2.2 \times 10^{-16}$; Supplementary Table 14). These analyses supported the robustness and biological basis of persistent mutations that are retained in the course of tumor evolution.

Shifting our focus from the tumor to the TME, we explored transcriptomic profiles in the TME of ICB-treated tumors and postulated that a high pTMB would generate an un-interrupted feed of

neoantigens which would in turn trigger interferon- γ signaling and adaptive immunity cascades that may be further enhanced with ICB. Serial RNA sequencing (RNA-seq) analyses of ICB-treated melanomas⁹ revealed a marked enrichment in interferon- γ and inflammatory response-related gene sets before therapy (Fig. 5a–c) which was significantly enhanced during ICB for high-pTMB tumors (Extended Data Fig. 9 and Supplementary Table 15). Notably, the differential enrichment in pro-inflammatory gene sets was lessened in tumors stratified by their TMB content (Fig. 5a, Extended Data Fig. 9 and Supplementary Table 15). Similar trends were observed in transcriptomic analyses comparing the TME of melanomas in the TCGA set that were stratified by pTMB (Supplementary Fig. 3 and Supplementary Table 16). We also found positive correlations between pTMB and CD8 and CD4



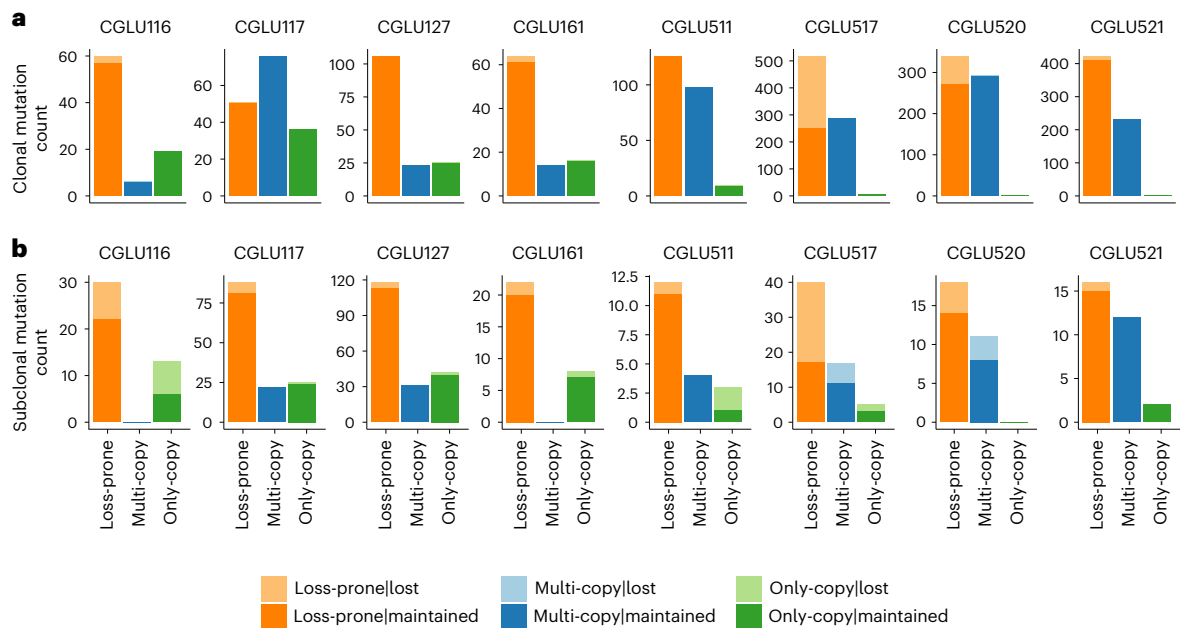


Fig. 4 | Persistent mutations are retained during cancer evolution under selective pressure of ICB. a, The presence of loss-prone, multi-copy and only-copy mutations identified in pre-ICB treatment NSCLC tumors was evaluated post-ICB therapy in tumors from eight patients with NSCLC and durable clinical benefit from ICB. Serial tumor samples were acquired with a minimum time difference of 6 months between biopsies and were analyzed by means of whole exome sequencing. A marked difference in the frequency of loss between clonal loss-prone and persistent mutation sets was observed, with an odds ratio of 61.46 ($P < 2.2 \times 10^{-16}$). Across 16 serially biopsied tumors from 8 patients, a total of 363 out of 2,836 clonal mutations that were detected in the baseline tumors were lost in the descendent tumors; of these 358 (98.6%) were clonal loss-prone mutations.

In six patients analyzed, we did not identify any clonal persistent mutation that was lost in the descendent tumor. In two patients, we detected four clonal multi-copy mutations that were not detected in the descendent tumors, resulting in an extremely low rate of loss in this mutation category (clonal multi-copy mutations: 4 out of 1,031 lost, 0.4% loss frequency; clonal only-copy mutations: 1 out of 117 lost, 0.9% loss frequency). **b**, While the loss frequency was higher for subclonal loss-prone compared to persistent mutations, this did not reach statistical significance (odds ratio = 1.24, $P = 0.44$). Notably, the loss frequency for subclonal multi-copy mutations was 9.3% compared with 14.8% for subclonal loss-prone mutations, potentially indicating differential selection pressures for these alterations.

T cell abundances (Supplementary Fig. 4 and Supplementary Table 17) and the ratio of M1 to M2 macrophages in both baseline (Spearman's $\rho = 0.36$) and on-ICB tumors (Spearman's $\rho = 0.28$). Next, we investigated the relationship of pTMB, tumor aneuploidy and cytolytic activity (Methods). We found a higher expression of cytolytic markers in pTMB-high compared with TMB-high or aneuploidy-low tumors in both the TCGA (TMB, $P > 0.05$ for all genes; pTMB, $P = 0.02$ for GZMK, IFNG and PRF1; $P = 0.04$ for NKG7; aneuploidy, $P > 0.05$ for all genes; Supplementary Fig. 5) and ICB melanoma cohorts (GZMB: TMB $P = 0.02$, pTMB $P = 6.1 \times 10^{-3}$, aneuploidy $P = 0.80$; IFNG: TMB $P = 0.01$, pTMB $P = 5.6 \times 10^{-3}$, aneuploidy $P = 0.88$; PRF1: TMB $P = 0.03$, pTMB $P = 5.1 \times 10^{-3}$, aneuploidy $P = 0.69$; Fig. 5). In modeling the expression of key cytolytic genes based on pTMB and aneuploidy, we found that a high pTMB counteracts the negative (but not statistically significant) impact of aneuploidy on cytolytic activity and ICB response (Fig. 5 and Supplementary Fig. 6).

Taken together, our analyses suggested that a high pTMB, which comprises a biologically relevant measure of tumor foreignness within the overall TMB, may represent an 'uneditable' target set for adaptive immune responses (Extended Data Fig. 1). This hypothesis relies on the basis that pMANAs are less likely to be eliminated by chromosomal loss due to the intrinsic fitness cost to the tumor and therefore may mediate sustained neoantigen-driven immune responses and long-term clinical benefit. Similar to pTMB, pMANA load and, importantly, expressed pMANA load distinguished responding from non-responding tumors (melanoma, $n = 42$, pTMB $P = 2.51 \times 10^{-4}$, pMANA $P = 3.37 \times 10^{-4}$, expressed pMANA $P = 2.91 \times 10^{-4}$; NSCLC, $n = 74$, pTMB $P = 1.26 \times 10^{-4}$, pMANA $P = 1.10 \times 10^{-4}$, expressed pMANA $P = 1.15 \times 10^{-4}$; Mann-Whitney U -test; Supplementary Table 18). We did not observe a further improvement of pMANA performance by restricting our

analyses to the subset of MANAs with computationally inferred high MHC class I binding affinity¹⁵, which is highlighting the limitations with MANA-predicting algorithms in identifying biologically relevant neoepitopes. To this end, we sought to generate additional functional proof that pMANAs are indeed recognized and elicit epitope-specific T cell expansions and pulsed autologous T cells from a patient with NSCLC with peptides derived from persistent and loss-prone mutations. All but one of the peptides that elicited T cell receptor (TCR) clonotypic expansions were encoded by persistent mutations (Extended Data Fig. 10 and Supplementary Tables 19 and 20), suggesting that pMANAs are recognized by CD8⁺ T cells and further supporting the biological importance of persistent mutations.

Discussion

Since the first reports recognizing TMB as a predictor of ICB response for patients with melanoma and NSCLC^{16,17}, it has become clear that TMB as a numeric value or binarized feature can only partially predict therapeutic response. Our findings suggest that a high pTMB, a biologically relevant measure of tumor foreignness within the overall TMB, represents an 'uneditable' target set for adaptive immune responses and may function as an intrinsic driver of sustained immunologic tumor control that cannot be readily bypassed by neoantigen loss via chromosomal deletions during cancer evolution.

Similar to TMB, which is linked with response to immunotherapy in a dose-dependent and cancer lineage-specific manner¹⁸, pTMB has to be considered in the context of the background aneuploidy rate within a specific tumor type. What we have learned from the increasing number of studies evaluating the overall TMB in predicting ICB response is that using a fixed pan-cancer threshold for a biomarker with different distributions and dynamic ranges depending on cancer lineage¹⁹

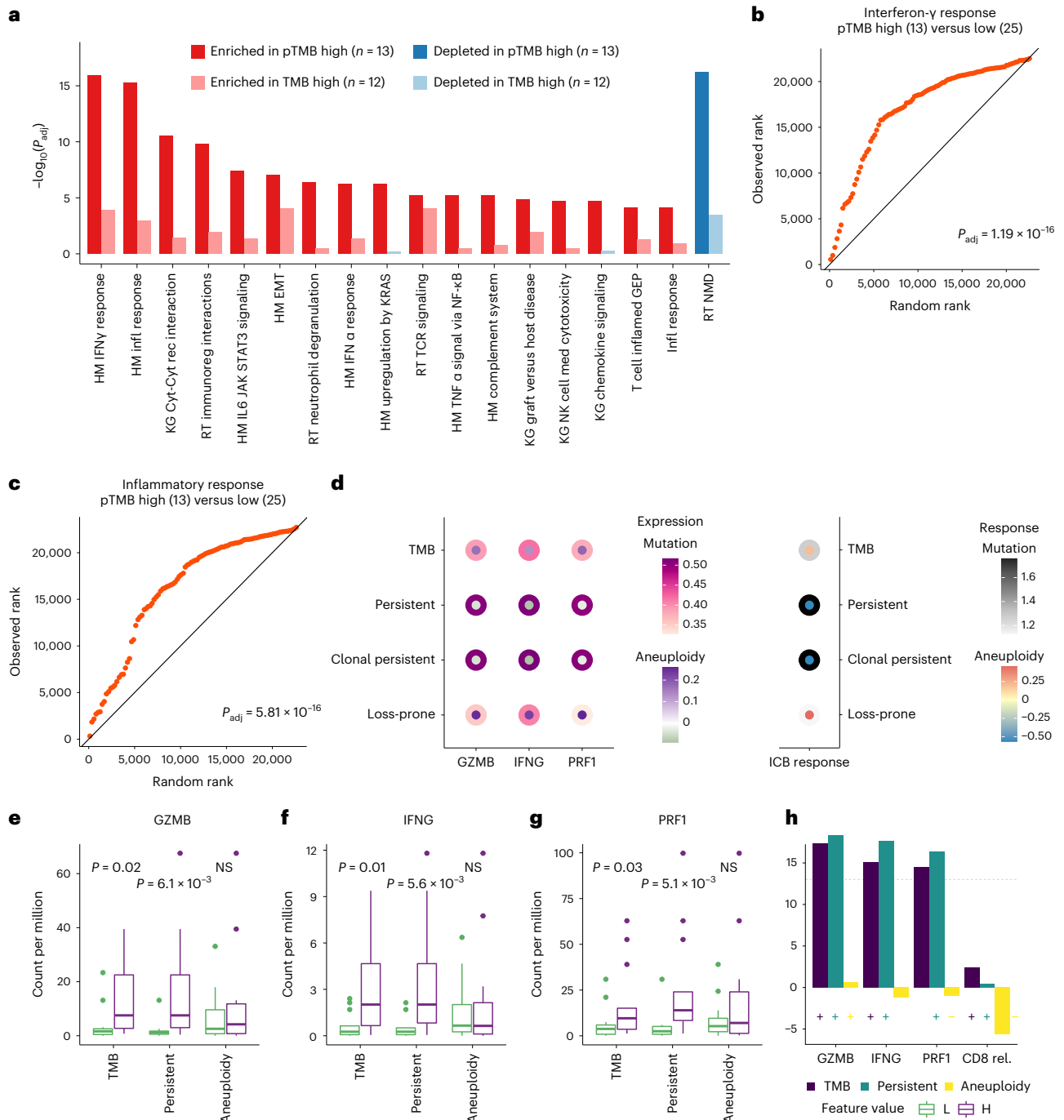


Fig. 5 | pTMB is associated with an inflamed TME in ICB-treated melanomas.

a, Gene set enrichment analysis revealed a marked enrichment in interferon- γ (IFN- γ) response and adaptive immunity gene sets in ICB-treated melanomas with high versus low pTMB, assessed before immunotherapy initiation. In contrast, a considerably lower enrichment in inflammatory gene sets was observed in the TME of tumors stratified by their overall TMB. Nominal two-sided P values are calculated using permutation testing, and FDR-adjusted P values shown for gene set differential expression are provided for comparison of pTMB/TMB-high versus -low groups. **b,c**, A prominent upregulation of IFN- γ (**b**) and inflammatory response (**c**)-related gene expression programs was noted in the TME of pTMB-high melanomas. Quantile–quantile plots were generated to visually compare the ranks of genes in the pathway with ranks that were sampled from a discrete uniform distribution. **d**, pTMB counteracts the negative impact of aneuploidy on cytolytic activity and response to ICB (GZMB, IFNG, PRF1: $\beta_{pTMB} = 0.5$, $P < 3 \times 10^{-3}$; $\beta_{aneuploidy}$ in $[-0.09, -0.03]$, $P > 0.05$; ICB response, $\beta_{pTMB} = 1.8$, $P = 4.0 \times 10^{-3}$; $\beta_{aneuploidy} = -0.56$, $P = 0.21$; β : logistic regression model coefficient, two-sided

P values are calculated assuming normally distributed test statistic). **e**, A greater difference in cytolytic activity was observed between tumors of high ($n = 13$) versus low ($n = 13$) pTMB, compared with TMB and aneuploidy, as evaluated by expression of GZMB (TMB $P = 0.02$, pTMB $P = 6.1 \times 10^{-3}$, aneuploidy $P = 0.80$). **f,g**, Similarly, IFNG (**f**) and PRF1 (**g**) expression was significantly higher in pTMB-high tumors compared to TMB and was not different in tumors based on their aneuploidy content (IFNG: TMB $P = 0.01$, pTMB $P = 5.6 \times 10^{-3}$, aneuploidy $P = 0.88$; PRF1: TMB $P = 0.03$, pTMB $P = 5.1 \times 10^{-3}$, aneuploidy $P = 0.69$; MW U -test; two-sided). **h**, No significant difference in relative abundance of CD8 T cells was observed for TMB, pTMB or aneuploidy. Box plots indicate the median value and hinges correspond to the first and third quartiles. The whiskers extend from the corresponding hinge to the furthest value within $1.5 \times$ interquartile range from the hinge. HM, hallmark; KG, KEGG; RT, reactome; Cyt, Cytokine; Rec, receptor; EMT, epithelial–mesenchymal transition; Med, mediated; FC, fold-change; NS, non-significant; L, low; H, high; rel., relative.

can be challenging and may misevaluate up to 25% of ICB-responsive tumors²⁰. In our current work and to avoid these challenges, we evaluated both persistent mutations and TMB as continuous variables in the context of ICB response; thus, our findings are less susceptible to artefactual associations resulting from application of a threshold. To expand our analyses in supporting a biologically distinct role for pTMB which is reflected in therapeutic response difference compared with overall TMB-based classifications, we evaluated the number of tumors with differential pTMB/TMB classification within the melanoma ICB cohort and found a higher response rate among tumors reclassified by their pTMB content. Notably, the relative contribution of multi- and only-copy mutation components to the overall pTMB varies across cancer lineages. This pattern appears to be driven by the dominant copy number state of the tumor, suggesting that the dominant copy number state has to be considered together with the sequence alteration load affecting these genomic regions.

The premise of pTMB relies on the potential of pMANAs to mediate sustained neoantigen-driven immune responses. Overall MANA burden has failed to demonstrate an incremental value over TMB in predicting clinical outcomes with ICB, as it is the neoantigen quality and not the quantity which may be most informative in predicting therapeutic response^{21–23}. Another feature that may determine the role of MANAs in anti-tumor immune responses is expression, as expressed single-base substitution-derived neopeptides have been shown to more accurately predict response to ICB compared with TMB⁹. In line with this notion, the significance of mutant protein abundance in driving T cell responses may further support the importance of multi-copy persistent mutations, as their presence at higher numbers of copies per cell likely correlates with higher expression of mutant messenger RNAs and proteins. We indeed found a marginal improvement in outcome prognostication of pMANA burden compared with pTMB in ICB-treated NSCLC. Importantly, by testing pMANA-specific TCR clonotypic expansions *in vitro*, we provide proof that pMANAs can elicit memory T cell responses that are likely to drive tumor elimination.

Placing pTMB in the context of other genomic features that have been associated with response to ICB¹, we assessed the clonal architecture of persistent mutations and considered the potential confounding effect of tumor clonal heterogeneity. Overall, in the TCGA and ICB cohorts, the cellular fractions of persistent mutations did not differ from loss-prone mutations; notably, persistent mutations in the multi-copy category tended to be more clonal in a cancer lineage-dependent manner, suggesting that these may have been acquired earlier in tumor evolution before the copy number gain event. Persistent mutations more optimally distinguished responding from nonresponding tumors compared with clonal TMB in all ICB cohorts analyzed. Importantly, clonal persistent mutations were more significantly associated with response in the ICB-treated NSCLC cohorts. These findings suggest that considering pTMB and clonal heterogeneity together may be most informative in predicting response to immunotherapy.

While pTMB is related to tumor aneuploidy and a higher degree of large-scale chromosomal changes has been reported in ICB nonresponsive tumors¹⁴, in our analyses tumor aneuploidy alone failed to differentiate responding from nonresponding tumors. While our analyses did not show a strong association between tumor aneuploidy or WGD and response to ICB as individual predictive biomarkers, the number of persistent mutations was correlated with tumor aneuploidy, and we found an enrichment for multi-copy persistent mutations in tumors with WGD. Our findings highlight the importance of measuring mutational burden in regions of the genome with structural changes rather than considering overall TMB or tumor aneuploidy independently.

Importantly, we studied the evolution of persistent mutations in the evolutionary trajectories shaped by selective pressure of ICB. We hypothesized that multi-copy mutations would inherently be more difficult to lose, as the process of loss would require multiple distinct

genomic events and, similarly, single-copy mutations are unlikely to be lost by chromosomal deletions as these may be detrimental to the cancer cell, rendering persistent mutation loss biologically improbable. Consistent with this notion, we discovered that persistent mutations were retained in the context of tumor evolution while losses predominantly affected loss-prone mutations. While approximately 99% of mutations lost in serial analyses of NSCLC during ICB were loss-prone mutations, four clonal multi-copy persistent mutations were not detected in comparative analyses of baseline/ICB-resistant tumors, which may be explained by the presence of multiple copies of a mutation in tandem or in close proximity on a common chromosomal segment; in these cases, loss of multiple copies could be achieved by a single genomic event.

Taken together, our findings suggest that mutations located in single-copy regions or those present in multiple copies in the cancer genome are unlikely to be lost under the selective pressure of immunotherapy due to the intrinsic fitness cost to the tumor, and therefore may serve as a key driver of sustained immunologic tumor control.

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/s41591-022-02163-w>.

References

- Litchfield, K. et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614.e14 (2021).
- Anagnostou, V. et al. Multimodal genomic features predict outcome of immune checkpoint blockade in non-small-cell lung cancer. *Nat. Cancer* **1**, 99–111 (2020).
- Anagnostou, V. et al. Evolution of neoantigen landscape during immune checkpoint blockade in non-small cell lung cancer. *Cancer Discov.* **7**, 264–276 (2017).
- Forde, P. M. et al. Durvalumab with platinum-pemetrexed for unresectable pleural mesothelioma: survival, genomic and immunologic analyses from the phase 2 PrE0505 trial. *Nat. Med.* **27**, 1910–1920 (2021).
- Leary, R. J. et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc. Natl Acad. Sci. USA* **105**, 16224–16229 (2008).
- Riaz, N. et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* **171**, 934–949.e16 (2017).
- Liu, D. et al. Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat. Med.* **25**, 1916–1927 (2019).
- Hugo, W. et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **165**, 35–44 (2016).
- Anagnostou, V. et al. Integrative tumor and immune cell multi-omic analyses predict response to immune checkpoint blockade in melanoma. *Cell Rep. Med.* **1**, 100139 (2020).
- Shim, J. H. et al. HLA-corrected tumor mutation burden and homologous recombination deficiency for the prediction of response to PD-(L)1 blockade in advanced non-small-cell lung cancer patients. *Ann. Oncol.* **31**, 902–911 (2020).
- Forde, P. M. et al. PrE0505: Phase II multicenter study of anti-PD-L1, durvalumab, in combination with cisplatin and pemetrexed for the first-line treatment of unresectable malignant pleural mesothelioma (MPM)—A PrE0505 study. *J. Clin. Oncol.* **38**, 9003 (2020).

12. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).
13. Zook, J. M. et al. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat. Biotechnol.* **32**, 246–251 (2014).
14. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* <https://doi.org/10.1126/science.aaf8399> (2017).
15. Goodman, A. M. et al. MHC-I genotype and tumor mutational burden predict response to immunotherapy. *Genome Med.* **12**, 45 (2020).
16. Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).
17. Rizvi, N. A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
18. Samstein, R. M. et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206 (2019).
19. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
20. Sha, D. et al. Tumor mutational burden as a predictive biomarker in solid tumors. *Cancer Discov.* **10**, 1808–1825 (2020).
21. McGranahan, N. & Swanton, C. Neoantigen quality, not quantity. *Sci. Transl. Med.* <https://doi.org/10.1126/scitranslmed.aax7918> (2019).
22. Shao, X. M. et al. High-throughput prediction of MHC class I and II neoantigens with MHCnuggets. *Cancer Immunol. Res.* **8**, 396–408 (2020).
23. Shao, X. M. et al. HLA class II immunogenic mutation burden predicts response to immune checkpoint blockade. *Ann. Oncol.* <https://doi.org/10.1016/j.annonc.2022.03.013> (2022).

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) 2023

Methods

Cohorts

We evaluated 10,742 tumor samples from TCGA and 485 NSCLC, melanoma and mesothelioma tumor samples from published cohorts of patients that received ICB^{2,4,6–8,10,24} (Supplementary Table 8). Tumors across 31 tumor types were analyzed from TCGA (ACC: Adrenocortical carcinoma, BLCA: Bladder Urothelial Carcinoma, BRCA: Breast invasive carcinoma, CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma, CHOL: Cholangiocarcinoma, COAD: Colon adenocarcinoma, ESCA: Esophageal carcinoma, GBM: Glioblastoma multiforme, HNSC: Head and Neck squamous cell carcinoma, KICH: Kidney Chromophobe, KIRC: Kidney renal clear cell carcinoma, KIRP: Kidney renal papillary cell carcinoma, LGG: Brain Lower Grade Glioma, LIHC: Liver hepatocellular carcinoma, LUAD: Lung adenocarcinoma, LUSC: Lung squamous cell carcinoma, MESO: Mesothelioma, OV: Ovarian serous cystadenocarcinoma, PAAD: Pancreatic adenocarcinoma, PCPG: Pheochromocytoma and Paraganglioma, PRAD: Prostate adenocarcinoma, READ: Rectum adenocarcinoma, SARC: Sarcoma, SKCM: Skin Cutaneous Melanoma, STAD: Stomach adenocarcinoma, TGCT: Testicular Germ Cell Tumors, THCA: Thyroid carcinoma, THYM: Thymoma, UCEC: Uterine Corpus Endometrial Carcinoma, UCS: Uterine Carcinosarcoma, UVM: Uveal Melanoma). Patients with melanoma across four source studies^{6–8,22} were combined to generate an aggregated melanoma cohort ($n = 202$). Clinical outcomes were retrieved from the original publications (Supplementary Table 8). We further performed whole exome sequencing analyses for a cohort of 39 patients with HPV-negative HNSCC who received ICB at the University of Chicago (HNSCC cohort). We assessed serially sampled NSCLC tumors from four patients from a published study³ as well as from four patients with NSCLC who received ICB at the Nederlands Kanker Instituut (NKI set). The studies were conducted in accordance with the Declaration of Helsinki and were approved by the Nederlands Kanker Instituut Institutional Review Board (IRB) and the University of Chicago Comprehensive Cancer Center IRB (8980). All patients provided written, informed consent for sample acquisition for research purposes; their participation in the protocol was not compensated. Clinical and pathologic characteristics for all patients, as well as self-reported sex and self-reported race, are summarized in Supplementary Table 4.

Whole exome sequencing and sequence data processing

DNA extraction and genomic library preparation were performed following manufacturers' protocols². The coding sequences were captured in solution using the SureSelect XT Human All Exon V6 kit in the HNSCC cohort, and using the SureSelect Human All Exon V4 kit in the NKI cohort. Whole exome sequencing-derived multi-center mutation calls from the TCGA pan-cancer atlas²⁵ were retrieved from the NCI Genomic Data Commons (<https://gdc.cancer.gov/about-data/publications/mc3-2017>) and filtered to keep nonsynonymous alterations. For the ovarian cancer (OV) tumor type, indels were excluded from all downstream analyses to minimize technical artifacts²⁶. Somatic copy number profiles including estimates of tumor purity and ploidy, as well as allele-specific copy number states²⁷, were acquired via the pan-cancer atlas (<https://gdc.cancer.gov/about-data/publications/pancanatlas>). Clinical annotations of tumors were accessed using the TCGA clinical data resource²⁸. For the HNSCC subset in TCGA, HPV status was retrieved from cBioPortal (https://www.cbioportal.org/study/summary?id=hnscc_tcg_pan_atlas_2018). Analyses of copy number profiles to establish the background rate of genomic loss were performed on 10,742 tumor samples where the segmental allele-specific copy numbers were available. For a subset of 9,242 tumor samples from the above, both somatic mutation calls and copy number profiles were available, enabling assessment of persistent mutations. For a subset of 8,925 tumors where clinical data annotations including overall survival and tumor stage assessment were available, survival analyses evaluating the contribution of persistent mutations were performed.

For the publicly available published ICB cohorts, we retrieved allele-specific copy number profile, tumor purity and ploidy estimates, as well as somatic mutation calls, raw gene expression counts and clinical annotations of response to treatment from the original publications. Furthermore, for the Riaz et al. melanoma cohort⁶, allele-specific somatic copy number profile and tumor purity and ploidy estimates were generated by application of FACETS (v.0.6.0) to tumor and matched normal sequence data²⁹. For the Liu et al. melanoma cohort⁷, the short-read archive files were accessioned from the Sequence Read Archive (SRA) and this sample set was filtered to keep only tumors with no previous anti-CTLA4 treatment. Adapter sequences were detected and trimmed using FASTP (v.0.20.0)³⁰. Sequenced reads were aligned to the reference genome assembly hg19 using bowtie2 (v.2.3.5)³¹, and duplicate reads marked by sambamba (v.0.8)³². Tumor purity and ploidy estimates, as well as somatic copy number profiles, were derived by application of FACETS²⁹ to tumor and matched normal pairs. For the Hugo et al. melanoma cohort⁸, fastq files were obtained from the SRA. Sequencing read processing and alignment were performed as described for the Liu et al. cohort, and copy number profiles were similarly obtained by application of FACETS. For the Shim et al. NSCLC cohort¹⁰, somatic mutations were narrowed down to those with mutant allele fraction greater than or equal to 10% to minimize sequencing artifacts. Tumor purity and ploidy estimates, and somatic copy number profiles, were generated by application of FACETS²⁹ to tumor and matched normal pairs. For the HNSCC cohort, somatic mutations were identified using the Strelka mutation calling pipeline (v.2.9.10)³³. Mutations in common single nucleotide polymorphism (SNP) locations (dbSNP v.138) and with greater than one BLAT³⁴ hit were filtered out. The final set of mutations were obtained after filtering for tumor mutant allele fraction $\geq 10\%$, normal mutant allele fraction $\leq 3\%$ and matched normal coverage $\geq 11\times$. For samples from the NKI set, sequence read processing and alignment were performed similarly to the samples from the Liu cohort. Tumor purity and ploidy estimates and somatic copy number profiles were derived by application of FACETS. While we did not have uniform documentation of microsatellite instability (MSI) status in the ICB cohorts analyzed, the very low background prevalence of MSI-high tumors in NSCLC (<1%), melanoma (<1%), mesothelioma (~2%) and HNSCC (<1%)³⁵ renders MSI an unlikely confounder in this study.

Evaluation of mutation multiplicity and cancer cell fraction

Mutation cellular fractions were estimated as follows^{3,36}: considering the tumor sample purity α , tumor copy number n_T and normal copy number n_N , the expected variant allele fraction V_{exp} for a mutation at cellular fraction C with multiplicity m (that is, m mutant copies per cancer cell) can be calculated as:

$$V_{\text{exp}} = \frac{m C \alpha}{\alpha n_T + (1 - \alpha) n_N}$$

The purity and segmental tumor and normal copy numbers were determined via genome-wide analysis of sequencing coverage distribution and b-allele frequency of heterozygous SNPs in each cohort. Assuming a binomial distribution for the number of reads harboring the mutant allele, a 95% confidence interval (CI) is constructed for V_{exp} using the distinct total coverage and mutant read counts for each mutation (that is, coverage and read counts after exclusion of reads marked as duplicates). Since estimates for α , n_T and n_N are available, this yields a 95% CI for the product of mutation at cellular fraction C and multiplicity m . By application of the following rules, one can derive estimates for C and m : (1) If the CI for mC contains an integer, the mutation is deemed clonal and that value is assigned to the multiplicity. (2) If the entire CI is below 1, multiplicity is assumed to be 1 and the mutation is subclonal, except cases where it is within a tolerance threshold of 1 ($C > 0.75$). (3) For a CI that is entirely above 1 and does not include any

integer, m is greater than 1 and is assigned such that the CI falls within the expected range [0,1]. Mutation clonality can now be calculated using the rule in (2).

Assessment of single-copy, multi-copy and persistent TMB

The nonsynonymous somatic mutations in each tumor were intersected with the segmental integer copy number profile to assign minor and major copy number states to the mutated loci. Mutation multiplicity (number of mutated copies per cell) and cancer cell fraction (proportion of cancer cells harboring the mutation) were estimated based on the mutant read count, total coverage, tumor purity, and the major and minor allele-specific copy numbers in the tumor and normal compartments for each mutation. Mutations present in more than one copy per cancer cell constituted the multi-copy category. Those present in regions of the genome with a single copy (total copy number = 1) were included in the single-copy category. The pTMB was defined as the number of mutations in either the multi-copy or single-copy category. For mesotheliomas, given the predominance of copy number losses^{4,37}, the persistent mutation burden was limited to mutations within single-copy regions of the genome. Furthermore, to assess the differential potential of persistent mutation in predicting outcome compared with TMB, we defined the number of loss-prone mutations in each tumor sample as the difference between the total number of mutations assessed (excluding mutations on sex chromosomes or those at loci without copy number assignment) and the number of persistent mutations. Finally, to achieve harmonized comparisons, mutations on sex chromosomes or on loci lacking allele-specific copy number assignment were excluded from analyses.

Characterization of tumor aneuploidy

Aneuploidy metrics were calculated for tumor samples from TCGA and ICB cohorts, and their relationship with persistent mutation burden was characterized. Furthermore, in ICB-treated cohorts, aneuploidy metrics were also considered as independent predictors of outcome. The fraction of genome with allelic imbalance was calculated as a broad metric summarizing the aneuploidy level across the autosomes. We also considered the fraction of genome with single copies (total copy number of 1) and the fraction of genome with multiple copies (that is, major allele-specific copy number greater than 1), given their direct link to persistent mutation burden. Tumor samples with more than 50% of the autosomal length at major allele-specific copy number of 2 or above were marked as having undergone WGD³⁸.

Evaluation of the background rate of genomic loss

To evaluate the background rate of genomic loss, we analyzed the somatic copy number profiles of 10,742 samples from TCGA. In each tumor sample, the chromosome arms in diploid state were defined as those where 75% of the length of segments covering the arm was copy neutral (total copy number of 2) and did not harbor LOH. The chromosome arms in haploid state had 75% of their length covered by segments with a total copy number of 1. The rate of loss in diploid regions of the genome was defined as R_D :

$$R_D = \frac{2 \times l_D^{HD} + l_D^{HM}}{2 \times l_D}$$

Where l_D indicates the total length of segments in arms of diploid state, l_D^{HD} is the total length of the segments with homozygous deletion in diploid arms and l_D^{HM} is the total length of the segments with single-copy loss in diploid arms. Similarly, the rate of loss in haploid regions of the genome was defined as R_H :

$$R_H = \frac{l_H^{HD}}{l_H}$$

Where l_H is the total length of segments in haploid arms and l_H^{HD} is the total length of segments with homozygous deletion in those arms. Comparison of the background rates of genomic loss was

performed on subsets of the TCGA tumor samples where at least one chromosome arm was found in each of diploid and haploid states ($n = 5,244$).

Evaluation of pTMB quantification by gene panel-targeted NGS

To determine the feasibility of using clinical targeted NGS to estimate pTMB, we performed in silico simulations as follows. Given the inherent limitation of targeted NGS in identification of mutations in tumor types with low TMB, we performed a focused analysis in melanoma and NSCLC cohorts. We assumed that allele-specific copy number estimates could be derived by targeted NGS as previously shown^{29,39}. Therefore, we focused our analysis on the subset of mutations that would be captured by the genomic intervals contained in FoundationOne CDx, which is a widely used clinical targeted NGS panel. The list of 309 genes with their full coding sequence included in the FoundationOne CDx panel was retrieved from the Food and Drug Administration website at https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170019S006C.pdf. The RefSeq Select transcript set was used to determine the genomic coordinates of the coding exons for each gene. Mutations in each tumor sample were intersected with panel coordinates to determine simulated estimates of TMB and pTMB as captured by the panel.

Differential expression and gene set enrichment analysis

Expression counts from RNA-seq of pre-treatment melanoma tumors from the CM038 melanoma cohort were retrieved from the original publication²⁴. Differential expression testing was performed using DESeq2 (ref. 40) and the resulting P values were corrected for multiple testing using the Benjamini–Hochberg procedure. For the TCGA tumor type SKCM, the TCGAAbiolinks (v.2.25.0) R package⁴¹ was used to download harmonized raw RNA-seq counts data from the NCI Genomic Data Commons within the target cancer type. This sample set was then narrowed down to the set of samples with persistent mutation estimates and available overall survival data, and comparisons were performed between samples within the top tertile of pTMB/TMB-informed risk (high risk) versus the remaining set (low risk). For gene set enrichment analysis, each gene that passed the count threshold was ranked by ‘ $-\log(p) * \text{sign}(fc)$ ’, where p is P value and fc is fold-change, resulting in ranking where the genes on each flank represent the mostly statistically significantly up- or down-regulated genes and the genes in the middle are the least significant. Gene set enrichment analysis (gsea) was then performed using the fgsea (v.1.20.0) R package⁴² with a curated list of gene sets from the Molecular Signatures Database related to immune responses and cancer hallmarks. Tumors were classified into high or low groups for TMB and pTMB using the second tertile value. The complete list of gene sets can be found in Supplementary Table 15, which contains the gsea results for comparisons based on pTMB and TMB in baseline and on-treatment samples. The P values for gsea were corrected for multiple testing with the Benjamini–Hochberg procedure. Quantile–quantile plots were made to provide a visual comparison of the ranks of pathway genes with a set of ranks sampled from the background distribution.

Modeling of cytolytic activity

Gene level expression values (in counts per million) were used from the CM038 IO melanoma cohort²⁴ and the TCGA melanoma cohort. In each cohort, expression levels for a selected set of gene markers of cytolytic activity were compared between tumors in the top and bottom tertiles of a number of key variables of interest using Mann–Whitney U -test. Furthermore, a multivariable linear regression model defined the combined contribution of a mutation-based marker (that is, pTMB, TMB and so on) and aneuploidy (as measured by the fraction of genome with allelic imbalance) to cytolytic activity. Briefly, both mutation-based marker values and gene expression levels were pseudo log-transformed to control the right skew in the distribution. Next, each variable was scaled to have zero mean and unit variance over

the analyzed cohort. In each regression model, predictor coefficients and the associated *P* values were recorded. In the IO melanoma cohort, multivariable logistic regression was used to model the contribution of mutation-based markers and aneuploidy. In addition, estimates for the relative abundances of 22 immune cell subpopulations derived by CIBERSORT v.1.06 were retrieved from the earlier publication. For the TCGA melanoma tumors, the relative abundances of CD8 T cells were retrieved from the genomic data commons⁴³.

Longitudinal tracking of persistent mutations

For the eight patients with NSCLC with serially biopsied tumor samples, tumor samples were acquired before ICB and at the time of acquired resistance; for all cases, a minimum of 6 months lapsed between ICB initiation and re-biopsy in the setting of acquired resistance. For each patient, the set of mutations identified in the baseline sample was annotated with distinct total coverage, distinct mutant read count, and minor and major allele-specific copy numbers. These annotations were combined with the estimated purity of the tumor sample to yield estimates of mutation cancer cell fraction and multiplicity. The combination of copy number assignment and multiplicity estimate for each mutation in the baseline sample enabled identification of only-copy, multi-copy and persistent mutations, as well as those prone to loss (loss-prone). Mutations identified in the baseline sample with mutant allele fraction of zero at the time of progression were deemed lost (Supplementary Tables 12 and 13).

Analysis of mutation-associated neoantigens

An integrated analysis of persistent mutation-associated neoantigens was performed in the ICB NSCLC² and melanoma²⁴ cohorts. Briefly, MANA predictions by ImmunoSelect-R pipeline (Personal Genome Diagnostics) were retrieved from the original studies. The set of predicted peptides was restricted to those with predicted MHC class I binding affinity (half-maximum inhibitory concentration (IC_{50})) less than 500 nM, and mutations with at least one associated peptide were marked as MANA-encoding. Mutations in genes with nonzero median expression in the respective TCGA tumor type were marked as expressed.

Assessment of replication timing for persistent mutations

We compared persistent and loss-prone mutations with regard to the replication timing of the mutated loci in melanoma (TCGA-SKCM) and NSCLC (TCGA-LUAD, TCGA-LUSC) tumors from TCGA. We retrieved replication timing scores for NHEK (skin) and IMR90 (lung) cell lines, which were measured by Repli-Seq methodology as part of the ENCODE project, using the UCSC Table Browser. In each cell line, we used scores from the 'wavlet-smoothed signal' track, which is the result of application of a wavelet smoothing transformation to the weighted average of the percentage-normalized signals in 1-kilobase intervals across the genome, where higher values indicate earlier replication timing. Genomic intervals were marked based on their quintile membership, and the frequencies of persistent and loss-prone mutations across the quintiles were visualized. Replication timings of persistent and loss-prone mutations were compared by evaluating Cohen's *d* effect size.

Analysis of NGS technical limitations

To assess the possibility of technical artifacts preferentially impacting the somatic mutation calls in our pan-cancer analysis of 31 tumor types, we determined the prevalence of persistent and loss-prone mutations identified in regions of the genome susceptible to limitations of NGS analysis. The UCSC Table Browser was used to retrieve the 'Problematic Regions' track, including regions marked by ENCODE¹², Genome-In-A-Bottle¹³ and NCBI GeT-RM⁴⁴.

Functional T cell assays

The MANAFEST (Mutation-Associated NeoAntigen Functional Expansion of Specific T Cells) assay³ was employed to determine

MANA-specific T cell clonotypic expansions in the peripheral blood of a patient with NSCLC who attained long progression-free and overall survival with ICB (CGLU310). Briefly, candidate neopeptides (JPT Peptide Technologies; Supplementary Table 19) were synthesized and each used to stimulate and co-culture T cells in vitro. On day 10, cells were collected and the CD8⁺ fraction was isolated using a CD8⁺ negative enrichment kit (EasySep, STEMCELL Technologies), followed by DNA extraction from each CD8-enriched culture condition. TCR $\nu\beta$ CDR3 sequencing was performed by the Sidney Kimmel Comprehensive Cancer Center (SKCCC) FEST and TCR Immunogenomics Core (FTIC) on genomic DNA from each T cell condition using the OncoPrint TCR Beta short-read assay (Illumina)¹¹. Following data pre-processing, alignment and trimming, productive frequencies of TCR clonotypes were calculated. To be considered antigen-specific, a T cell clonotype must have met the following criteria: (1) significant expansion (Fisher's exact test with Benjamini-Hochberg correction for false discovery rate (FDR), $P < 0.05$) compared with T cells cultured without peptide; (2) significant expansion compared with every other peptide-stimulated culture (FDR < 0.05); (3) an odds ratio greater than 5 compared with all other conditions; (4) at least 30 reads in the 'positive' well; and (5) at least 2 \times higher frequency than background clonotypic expansions as detected in the HIV-negative control condition.

Evaluation of differentially classified tumors by TMB and pTMB

Reclassification rates based on pTMB versus TMB were computed as follows: in each tumor type and for each variable (TMB and pTMB), a series of quantile values ranging from 5% to 95% in 5% increments were applied to define samples with high and low values for that variable. Next, at each quantile value, we calculated the rate at which sample classification differed by the two metrics (that is, the combined prevalence of samples that were pTMB-low, TMB-high and samples that were pTMB-high, TMB-low within that tumor type for that quantile threshold) to derive the reclassification rate. In the ICB cohorts, we determined the cases with differential pTMB/TMB classification and compared the therapeutic response rates between pTMB-low/TMB-high and pTMB-high/TMB-low tumors. For each predictor variable (pTMB or TMB), the second tertile was used to determine tumor samples with high and low values. Given the current cohort sizes, sufficient sample size for this comparison was only available in the combined melanoma cohort where 23 samples were labeled as TMB-high/pTMB-low and another 23 samples were marked as TMB-low/pTMB-high.

Survival analysis of TCGA tumors

The relationship of pTMB and TMB with overall survival was assessed in 8,925 patients from 31 cohorts in TCGA. Tumor types with more than 50 informative samples were analyzed and overall survival was selected as an endpoint. Briefly, for each tumor type and stage combination, Cox proportional-hazards (CoxPH) models were constructed for each one of the following features as independent continuous predictors: TMB, pTMB, clonal pTMB, multi-copy mutations, clonal multi-copy mutations, only-copy mutations, clonal only-copy mutations (continuous CoxPH model). In cases where an increase in the predictor variable was associated with better outcome (longer survival), a second CoxPH model was used to assess the difference in overall survival between tumors in the top one-third and bottom two-thirds of predicted risk (categorical CoxPH model).

Statistical analyses

For the TCGA cohort, in each tumor type, a CoxPH model was used to evaluate the contributions of TMB and pTMB to overall survival. The predicted risk values from these models were then used to stratify the tumors into low- and high-risk groups using the second tertile of the predicted risk as the threshold. For the immunotherapy cohorts, clinical response assessments were retrieved from the original publications.

The Mann–Whitney U -test was used to evaluate the differences of continuous variables between groups, including the differences of predictive variables between responding and nonresponding tumors, and the differences of background rate of loss between haploid and diploid regions of the genome. Cohen's d statistic was used to quantify the effect size of each predictor variable in the ICB cohorts. Fisher's exact test was used to assess the association of dichotomous variables (such as WGD) with therapeutic response. The associations of the fraction of mutations in single- and multi-copy regions with TMB ranks were evaluated by the Pearson correlation coefficient, while nonparametric correlations were evaluated by the Spearman correlation coefficient. The Kaplan–Meier method was used to estimate the survival function and the survival curves were compared using the nonparametric log-rank test. All P values were based on two-sided testing and differences were considered significant at $P < 0.05$. The P values reported were not adjusted for multiple hypothesis testing unless explicitly stated. Statistical analyses were done using R v.3.6 and higher (<http://www.R-project.org/>).

Reporting summary

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

Data availability

Source data for the TCGA tumor samples were retrieved from <http://cancergenome.nih.gov>. WES-derived somatic mutation calls from the TCGA PanCancer Atlas MC3 project were retrieved from the NCI Genomic Data Commons (<https://gdc.cancer.gov/about-data/publications/mc3-2017>). Somatic copy number profiles (<https://gdc.cancer.gov/about-data/publications/pancanatlas>) and clinical data (<https://gdc.cancer.gov/about-data/publications/PanCan-Clinical-2018>) were accessed from Genomic Data Commons. Previously published genomic data, re-analyzed here, were obtained from the material of the original publications and from dbGaP under accession code [phs000452.v3.p1](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=phs000452.v3.p1) (ref. 7), and from the Sequence Read Archive (SRA) under accession codes [SRP095809](https://www.ncbi.nlm.nih.gov/sra/acc.cgi?acc=SRP095809) (ref. 6), [SRP067938](https://www.ncbi.nlm.nih.gov/sra/acc.cgi?acc=SRP067938) (ref. 8) and [SRP090294](https://www.ncbi.nlm.nih.gov/sra/acc.cgi?acc=SRP090294) (ref. 8). WES sequence data for the HNSCC and NKI cohorts from patients who consented to data deposition can be retrieved from the European Genome-phenome Archive (EGA accession number [EGAS00001006660](https://www.ebi.ac.uk/ega/EGAS00001006660)) in controlled access mode, subject to approval by the Data Access Committee, for use in not-for-profit organizations, with approval of local IRB/ERB, for approved projects, by approved users, and not-for-profit use. The de-identified clinical and genomic data used for the analyses in this study are available in the supplementary tables and all publicly available data elements have been referenced. Source data are provided with this paper.

Code availability

The study did not involve development of new software packages.

References

- Anagnostou, V. et al. Integrative tumor and immune cell multi-omic analyses predict response to immune checkpoint blockade in melanoma. *Cell Rep. Med.* **1**, 100139 (2020).
- Ellrott, K. et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst.* **6**, 271–281 e277 (2018).
- Buckley, A. R. et al. Pan-cancer analysis reveals technical artifacts in TCGA germline variant calls. *BMC Genomics* **18**, 458 (2017).
- Taylor, A. M. et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689 e673 (2018).
- Liu, J. et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400–416 e411 (2018).
- Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
- Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
- Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
- Bonneville, R. et al. Landscape of microsatellite instability across 39 cancer types. *JCO Precis. Oncol.* <https://doi.org/10.1200/PO.17.00073> (2017).
- Niknafs, N., Beleva-Guthrie, V., Naiman, D. Q. & Karchin, R. Subclonal hierarchy inference from somatic mutations: automatic reconstruction of cancer evolutionary trees from multi-region next generation sequencing. *PLoS Comput. Biol.* **11**, e1004416 (2015).
- Hmeljak, J. et al. Integrative molecular characterization of malignant pleural mesothelioma. *Cancer Discov.* **8**, 1548–1565 (2018).
- Bielski, C. M. et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* **50**, 1189–1195 (2018).
- Riester, M. et al. PureCN: copy number calling and SNV classification using targeted short read sequencing. *Source Code Biol. Med.* **11**, 13 (2016).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Colaprico, A. et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
- Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. Preprint at [bioRxiv](https://www.biorxiv.org/content/10.1101/060012v3) <https://www.biorxiv.org/content/10.1101/060012v3> (2021).
- Thorsson, V. et al. The immune landscape of cancer. *Immunity* **48**, 812–830.e14 (2018).
- Mandelker, D. et al. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet. Med.* **18**, 1282–1289 (2016).

Acknowledgements

This work was supported in part by the US National Institutes of Health grant no. CA121113 (V.A. and V.E.V.), the Bloomberg–Kimmel Institute for Cancer Immunotherapy (V.A. and P.M.F.), the Johns Hopkins SKCCC core grant no. NCI CCSG P30 CA006973, the Department of Defense Congressionally Directed Medical Research Programs grant no. CA190755 (V.A. and P.M.F.), the ECOG-ACRIN Thoracic Malignancies Integrated Translational Science Center grant no. UG1CA233259 (V.A. and V.E.V.), the V Foundation (V.A. and V.E.V.) and the LUNGevity Foundation (V.A. and V.E.V.). The results presented in this study are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. This work was supported by the National Human Genome Research Institute Large Scale Sequencing Program, grant no. U54 HG003067.

Author contributions

V.A. and N.N. conceived the study and contributed to the study design, data analysis, interpretation and writing. A.B., C.C., X.M.S. and Z.B. contributed to data analysis. K.H., K.M., K.A.M., J.M., K.N.S., B.L.,

J.F., C.L.H., V.L., D.M.P., R.K., T.Y.S., J.R.B., P.M.F. and V.E.V. contributed to data collection and interpretation and edited the manuscript.

Competing interests

V.A. receives research funding to Johns Hopkins University from AstraZeneca, Personal Genome Diagnostics and Delfi Diagnostics, and has received research funding to Johns Hopkins University from Bristol Myers Squibb in the past 5 years. V.A. is an inventor on patent applications (63/276,525, 17/779,936, 16/312,152, 16/341,862, 17/047,006 and 17/598,690) submitted by Johns Hopkins University, related to cancer genomic analyses, ctDNA therapeutic response monitoring and immunogenomic features of response to immunotherapy, that have been licensed to one or more entities. Under the terms of these license agreements, the University and inventors are entitled to fees and royalty distributions. N.N. is an inventor on patent application 17/598,690 submitted by Johns Hopkins University related to genomic features of response to immunotherapy. C.C. is the founder of CM Cherry Consulting. K.M. has received research funding from AstraZeneca; speakers' fees from MSD, Roche, AstraZeneca and Benecke; and has served in a consultant role for Pfizer, Bristol Myers Squibb, Roche, MSD, Abbvie, AstraZeneca, Diaceutics, Lilly, Bayer, Boehringer Ingelheim and Merck. K.A.M. has served in a consultant/advisory role for Amgen, Janssen, Mirati, AstraZeneca and Puma, and has received research funding to her institution from Mirati and BMS. J.M. has served in a consulting role to MJH Life Sciences, Johnson & Johnson and Doximity, and has received research funding to his institution from Merck via the Conquer Cancer Young Investigators Award. K.N.S. receives research funding to her institution from Bristol Myers Squibb, AstraZeneca and Enara Bio, and holds founder's equity in manaT Bio. B.L. has served in a consultant/advisory role for Janssen, Daiichi Sankyo, AstraZeneca, Eli Lilly, Genentech, Mirati, Amgen, Pfizer, BMS, Guardant 360 and Foundation Medicine. J.F. has served in a consultant/advisory role for Genentech, Eli Lilly, AstraZeneca, Merck, Takeda, Coherus, Regeneron and Pfizer, and has received research funding (directly to the institution) from AstraZeneca, Pfizer and Bristol Myers Squibb. C.L.H. has served in a consultant/advisory role for AbbVie, Amgen, AstraZeneca, BMS, Genentech/Roche, Janssen and GSK, and has received research funding to her institution from AbbVie, Amgen, AstraZeneca, BMS and GSK. V.L. has served in a consultant/advisory role for Takeda, Seattle Genetics, Bristol Myers Squibb, AstraZeneca and Guardant Health, and has received research funding from GlaxoSmithKline, BMS, Merck and Seattle Genetics. D.M.P. is a consultant for Aduro Biotech, Amgen, Bayer, Dynavax, Enara, FLX Bio, Immunomic Therapeutics, Janssen, Merck, Rock Springs Capitol, Tizona and Trieza; is on the Board of Directors of DNAtrix; is on the Scientific Advisory Boards of Immununica, WindMil, Dracen, Camden Partners and Astellas; receives research support from Bristol Myers Squibb and Compugen; and is a founder of ManaT Bio. T.Y.S. has served in an advisory role/

received honoraria from Innate Pharma, Merck/MSD, AstraZeneca, Nanobiotix, Cue Biopharma, Vir Biotechnology, Coherus, Sanofi, Eisei and iTEOS; and reports research funding to his institution from BMS, Genentech/Roche, Merck/MSD, Nanobiotix, AstraZeneca, Cue Biopharma, Regeneron, Innate and Kura. J.R.B. is on the advisory board of/is a consultant for Amgen, AstraZeneca, BMS, Genentech/Roche, Eli Lilly, GlaxoSmithKline, Merck, Sanofi and Regeneron; receives grant research funding from AstraZeneca, BMS, Genentech/Roche, Merck, RAPT Therapeutics, Inc. and Revolution Medicines; and is on the Data and Safety Monitoring Board/Committees of GlaxoSmithKline, Sanofi and Janssen. P.M.F. has received research funding to Johns Hopkins University from AstraZeneca, Bristol Myers Squibb, Novartis, Corvus and Kyowa. He has also served as a consultant for Amgen, AstraZeneca, Bristol Myers Squibb, Daiichi Sankyo, Genentech, G1 Therapeutics, Iteos, Janssen, Merck, Surface Oncology, Mirati, Novartis and Sanofi, and as a DSMB member for Polaris and Flame Therapeutics. V.E.V. is a founder of Delfi Diagnostics, serves on the Board of Directors and as a consultant for this organization, and owns Delfi Diagnostics stock, which is subject to certain restrictions under university policy. Additionally, Johns Hopkins University owns equity in Delfi Diagnostics. V.E.V. divested his equity in Personal Genome Diagnostics (PGDx) to LabCorp in February 2022. V.E.V. is an inventor on patent applications submitted by Johns Hopkins University, related to cancer genomic analyses and cell-free DNA for cancer detection, that have been licensed to one or more entities, including Delfi Diagnostics, LabCorp, Qiagen, Sysmex, Agios, Genzyme, Esoterix, Ventana and ManaT Bio. Under the terms of these license agreements, the University and inventors are entitled to fees and royalty distributions. V.E.V. is an advisor to Danaher, Takeda Pharmaceuticals and Viron Therapeutics. All other authors declare no conflicts of interest.

Additional information

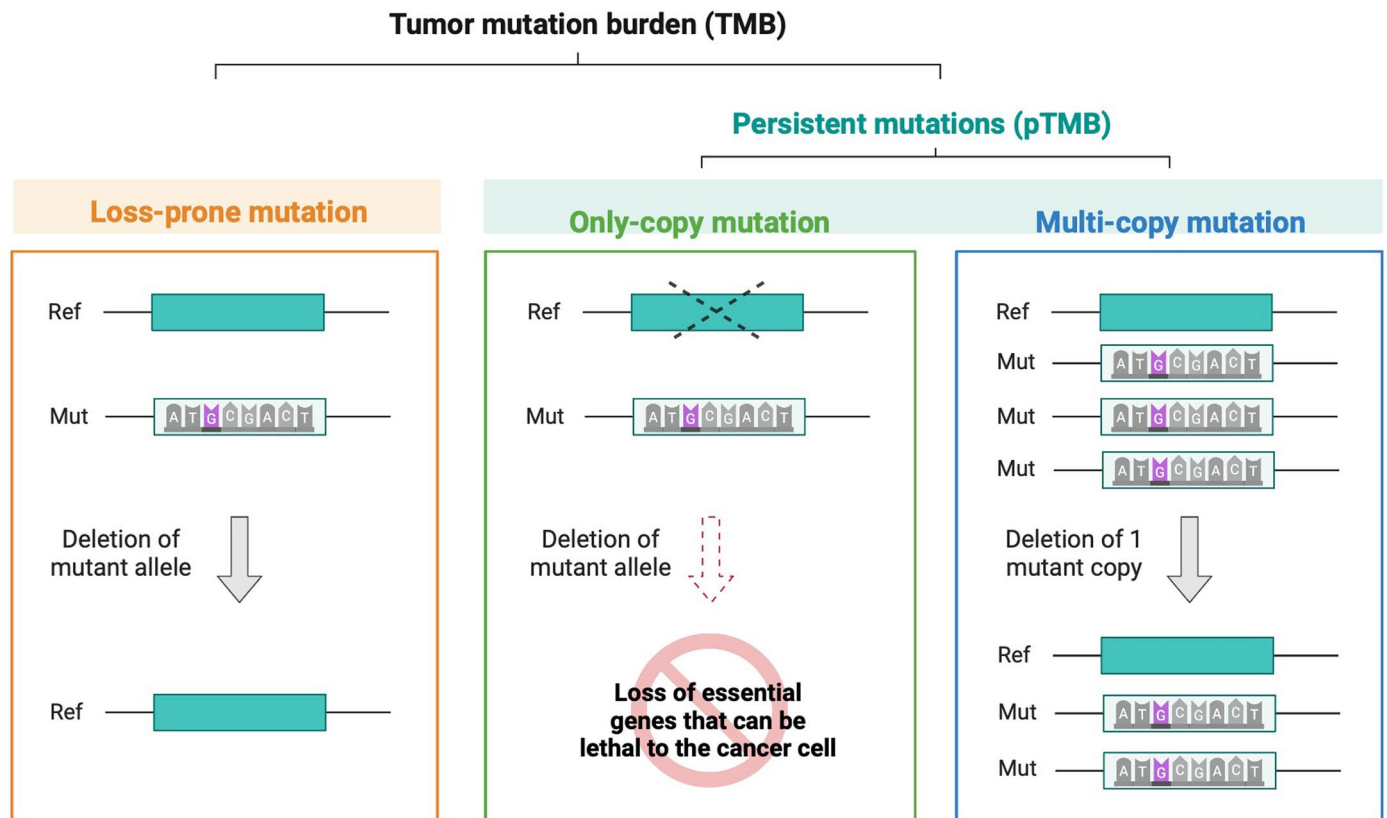
Extended data is available for this paper at <https://doi.org/10.1038/s41591-022-02163-w>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-022-02163-w>.

Correspondence and requests for materials should be addressed to Valsamo Anagnostou.

Peer review information *Nature Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Anna Maria Ranzoni and Saheli Sadanand, in collaboration with the *Nature Medicine* team.

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

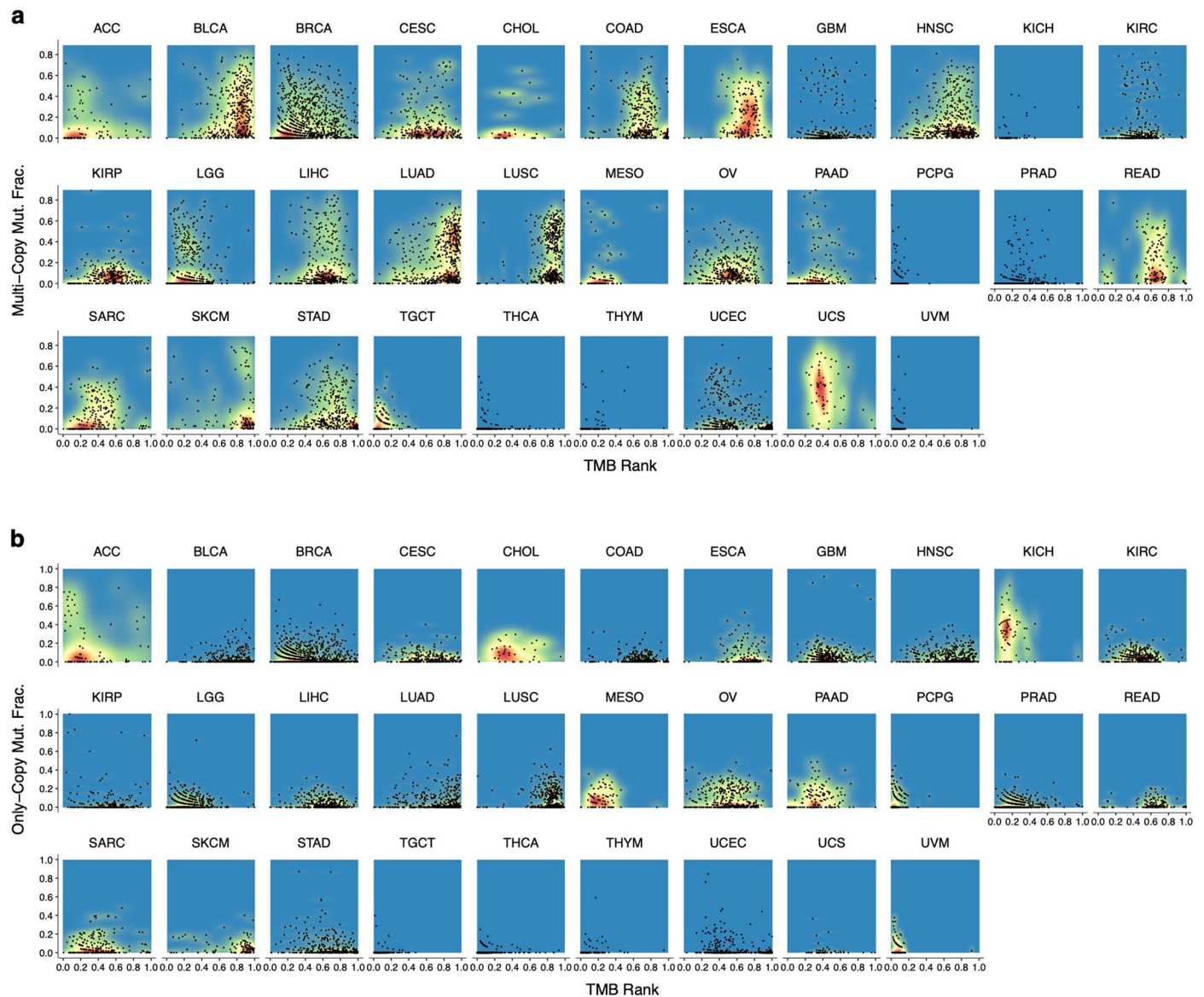


Loss-prone mutations and encoded neoantigens are easily edited under immune pressure.

Extended Data Fig. 1 | The persistent mutation hypothesis. We hypothesized that tumors with a higher frequency of sequence alterations in either haploid regions or in multiple copies would be intrinsically incapable of escaping immune recognition in the setting of immunotherapy, as these alterations

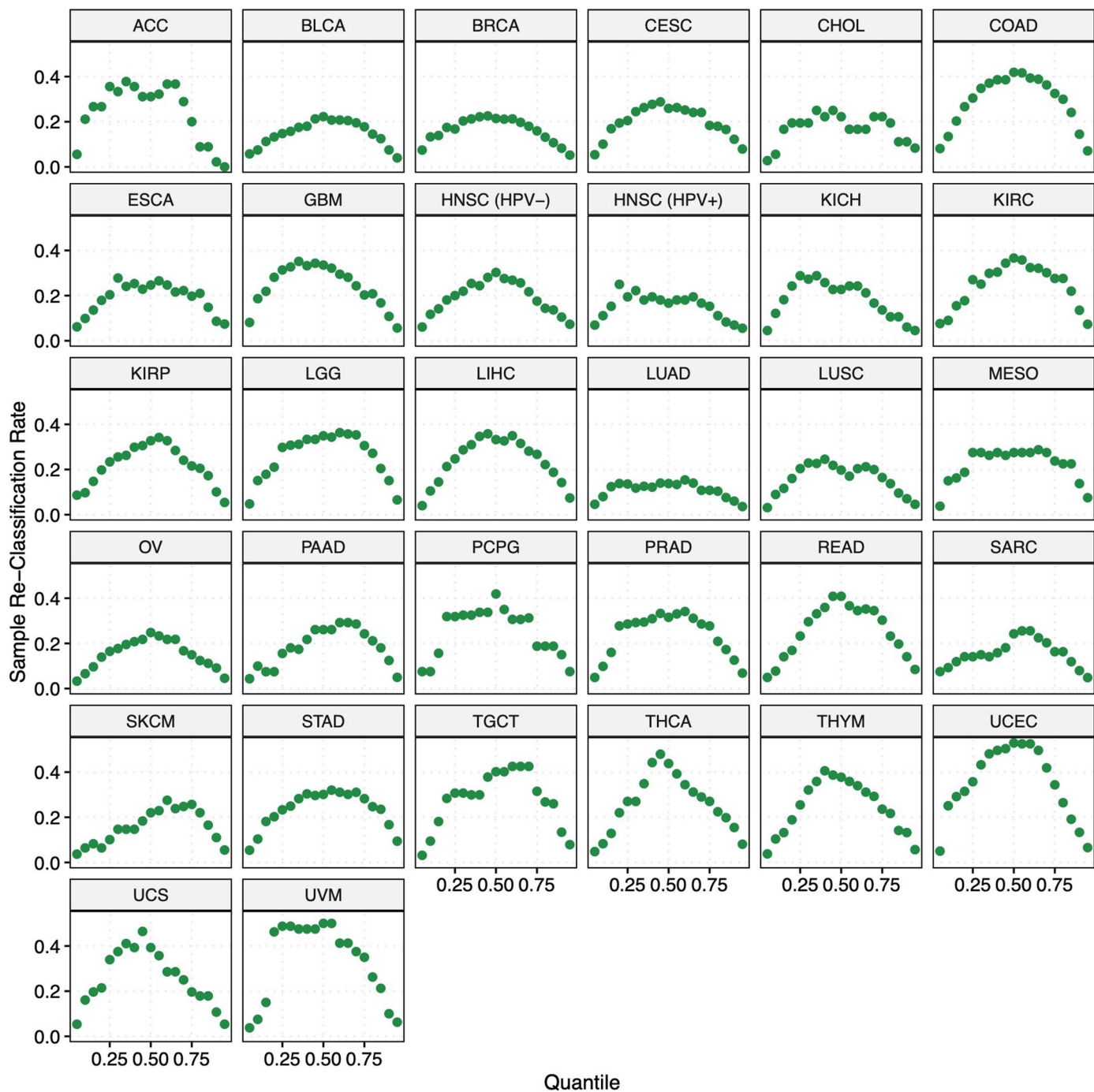
Persistent mutations and encoded neoantigens cannot be edited under immune pressure and drive sustained immunologic tumor control.

would continuously render them visible to the immune system, resulting in sustained tumor rejection. These two distinct genomic mechanisms produce a common feature that we term “persistent mutations”, a key driver of continued immunologic tumor control.



Extended Data Fig. 2 | Interaction of TMB and fraction of mutations in single- and multi-copy regions across 31 cancer types. A combined analysis of TMB ranks in relation to the fraction of mutations in multi-copy (a) and only-copy (b) regions of the genome revealed a complex landscape, where a wide range of multi-copy or single-copy mutation fraction was observed at any given TMB value in 9,256 tumors across 31 cancer types from TCGA. These findings highlight tumors with a low or intermediate TMB that at the same time harbor a large fraction of mutations in multi-copy or single-copy states, therefore suggesting the independent nature of these features. A weak correlation was noted between

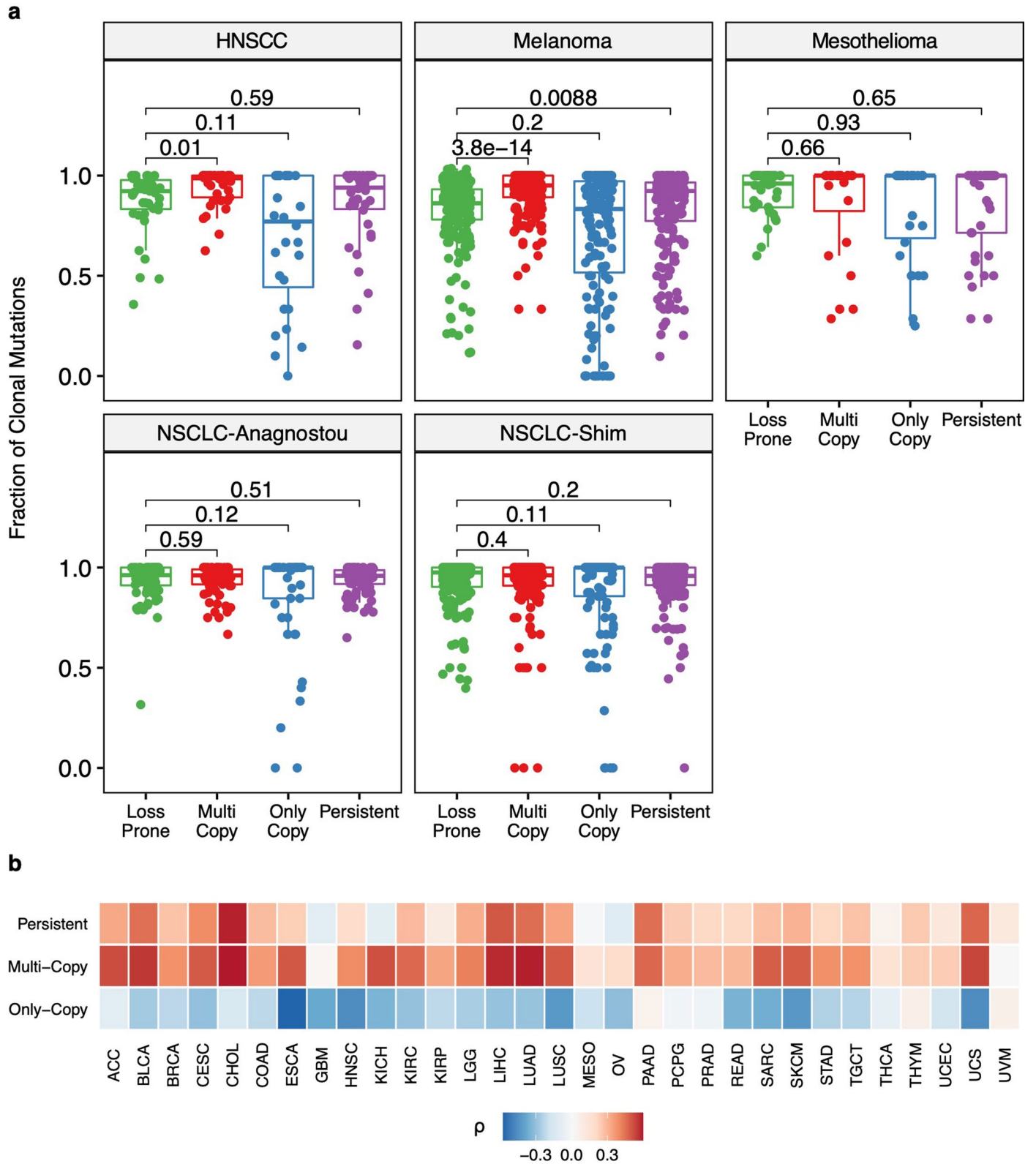
TMB rank and fraction of mutations in multi-copy regions of the genome for BLCA (Pearson's $R = 0.26$), KICH (Pearson's $R = 0.45$), LUAD (Pearson's $R = 0.34$), LUSC (Pearson's $R = 0.21$), SARC (Pearson's $R = 0.24$) while these features were weakly anti-correlated in COAD and UCEC (Pearson's $R = -0.23$ and -0.35 respectively). TMB was weakly anti-correlated with the fraction of mutations in haploid regions for COAD, STAD and UCEC (Pearson's $R = -0.22$, -0.25 and -0.25 respectively). The heatmaps indicate the distribution of tumors within a tumor type, and each black dot represents an analyzed tumor sample.



Extended Data Fig. 3 | Tumor re-classification by pTMB vs TMB. A series of quantile values going from 5% to 95% in 5% increments were used to extract samples with high and low values for each predictor variable in each tumor type. For each quantile value, the re-classification rate per samples within a cancer type group. By definition, sample re-classification rate starts close to 0 at the

Quantile

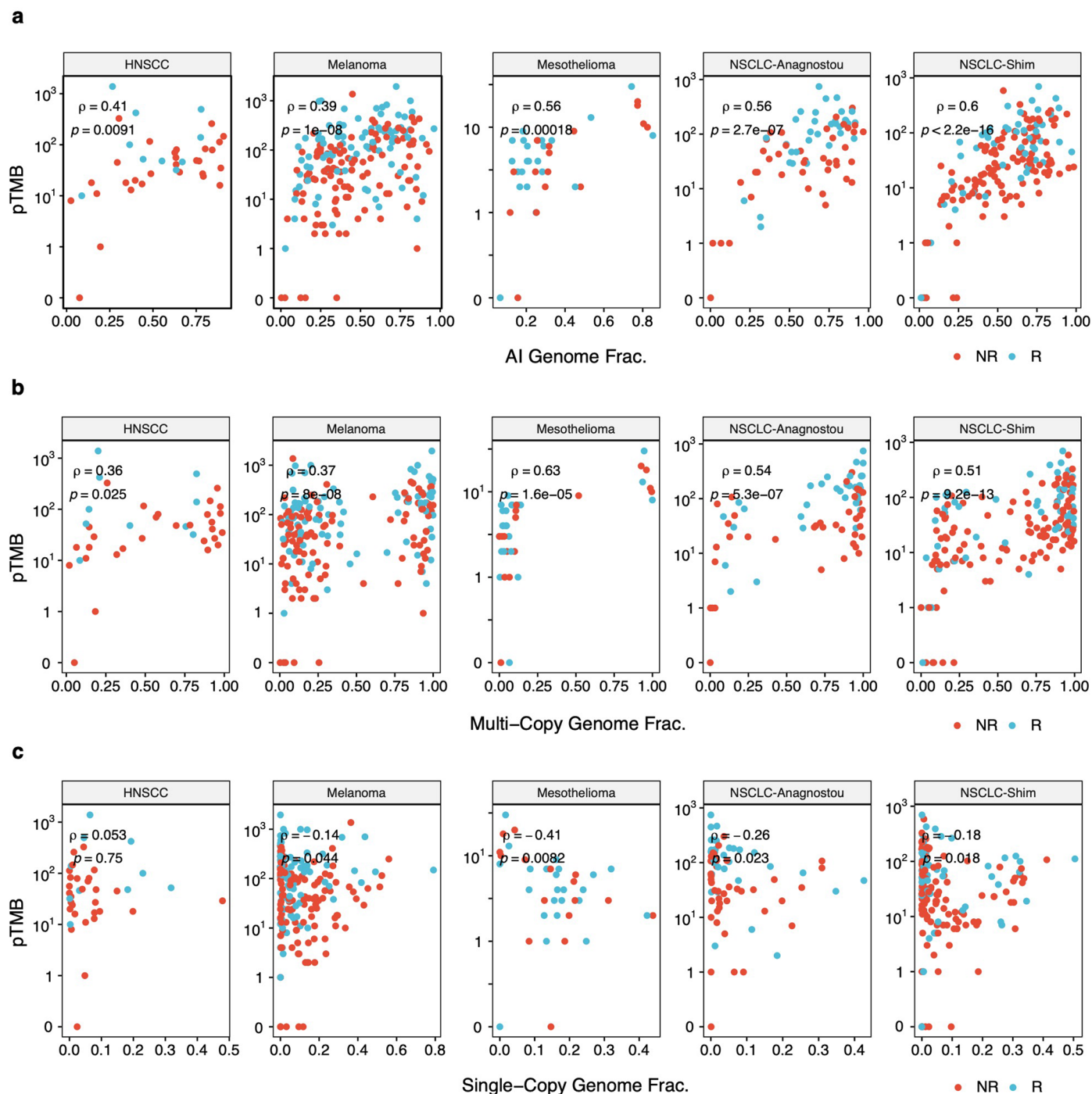
lower end of quantile thresholds (where almost all samples are labeled as high regardless of the metric used), and returns to 0 at the higher end of quantile thresholds (where the great majority of samples are labeled as low irrespective of the metric used). In the intermediate range, re-classification rates as high as 50% were observed.



Extended Data Fig. 4 | See next page for caption.

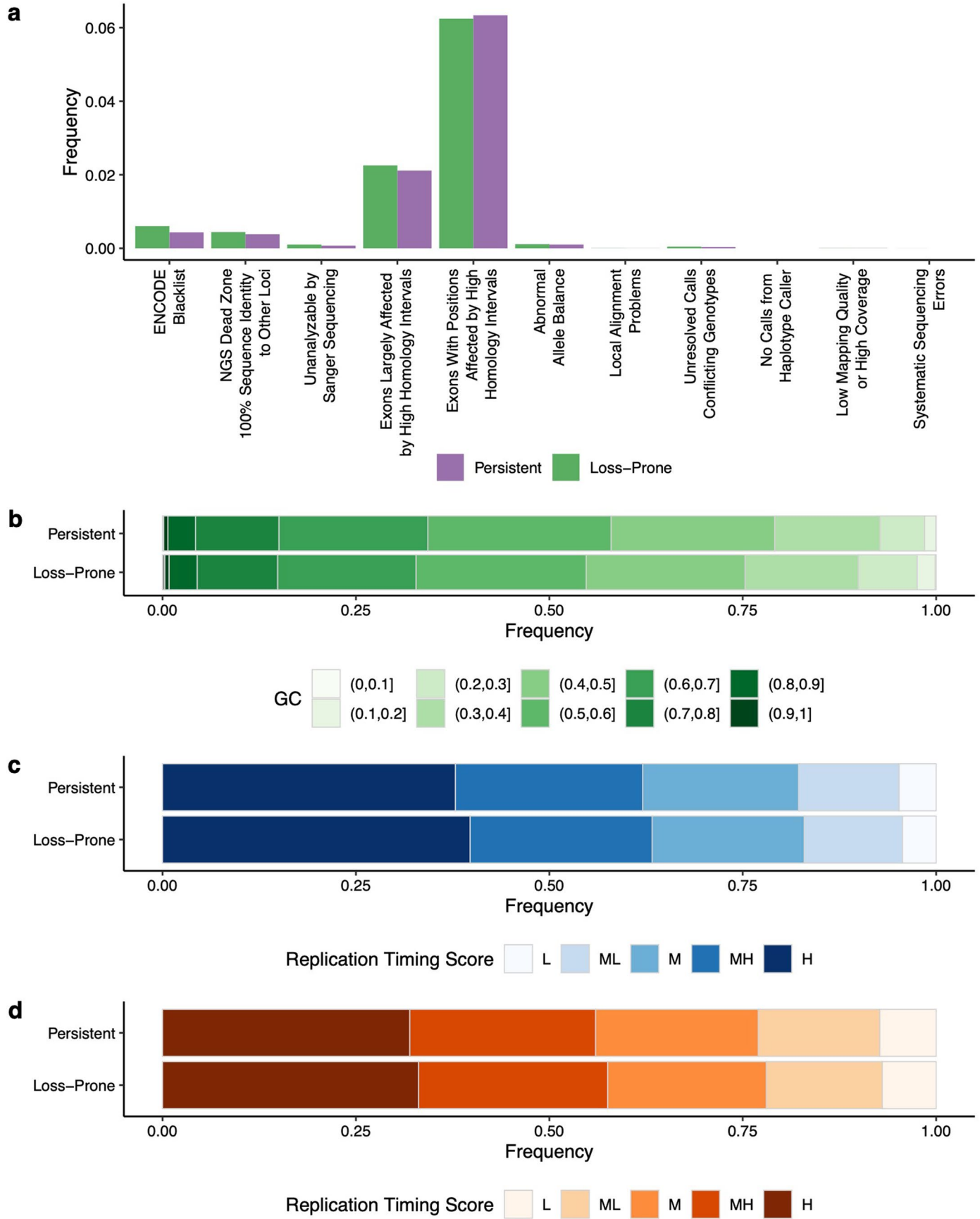
Extended Data Fig. 4 | Clonal architecture of persistent mutations. (a) We computed the fraction of clonal mutations within loss-prone, multi-copy, only-copy, and persistent mutations sets. In the HNSCC (n = 39) and melanoma cohorts (n = 208), multi-copy mutations were more clonal compared to loss-prone mutations (HNSCC; $p = 0.01$, melanoma; $p = 3.8 \times 10^{-14}$), but no such difference was present in the mesothelioma (n = 40, $p = 0.66$) and NSCLC cohorts (NSCLC-Anagnostou; n = 75 and $p = 0.59$, NSCLC-Shim; n = 169 and $p = 0.40$). Only-copy mutations had similar clonal fractions as loss-prone mutations ($p > 0.11$) in all five cohorts. When multi-copy and only-copy mutations were considered together, we did not identify a significant difference in clonal fractions between persistent and loss-prone mutations in the HNSCC ($p = 0.59$), mesothelioma ($p = 0.65$), and NSCLC cohorts (NSCLC-Anagnostou; $p = 0.51$,

NSCLC-Shim; $p = 0.20$). Persistent mutations tended to be more clonal in the melanoma cohort ($p = 8.80 \times 10^{-3}$). Box plots depict the median value and hinges correspond to the first and third quartiles. The whiskers extend from the corresponding hinge to the furthest value within $1.5 \times$ the interquartile range from the hinge. Two-sided Mann-Whitney U-test was used to compare values in across mutation classes. **(b)** A weak to moderate degree of correlation was observed between the fraction of clonal mutations and the number of persistent mutations (Spearman ρ range: $-0.11 - 0.59$) or multi-copy mutations (Spearman ρ range: $0.01 - 0.60$) across the 31 TCGA tumor types analyzed. Notably, the number of only-copy mutations was anti-correlated with the fraction of clonal mutations in the TCGA dataset (Spearman ρ range: $-0.57, 0.07$). Spearman's rank correlation coefficients are depicted in the heatmap.



Extended Data Fig. 5 | Correlations between persistent mutations and tumor aneuploidy. (a) A moderate degree of correlation was observed between pTMB and the fraction of genome with allelic imbalance in the five ICB cohorts analyzed (HNSCC Spearman $\rho = 0.41$, $p = 9e-3$; Melanoma $\rho = 0.39$, $p = 1e-8$; Mesothelioma $\rho = 0.56$, $p = 1.8e-4$; NSCLC-Anagnostou $\rho = 0.56$, $p = 2.7e-7$; NSCLC-Shim $\rho = 0.60$, $p = 2.2e-16$) indicating that tumors with higher levels of aneuploidy tend to have higher pTMB. (b) Similar levels of correlation were observed between pTMB and the fraction of genome with multiple (>2) copies (HNSCC Spearman $\rho = 0.36$,

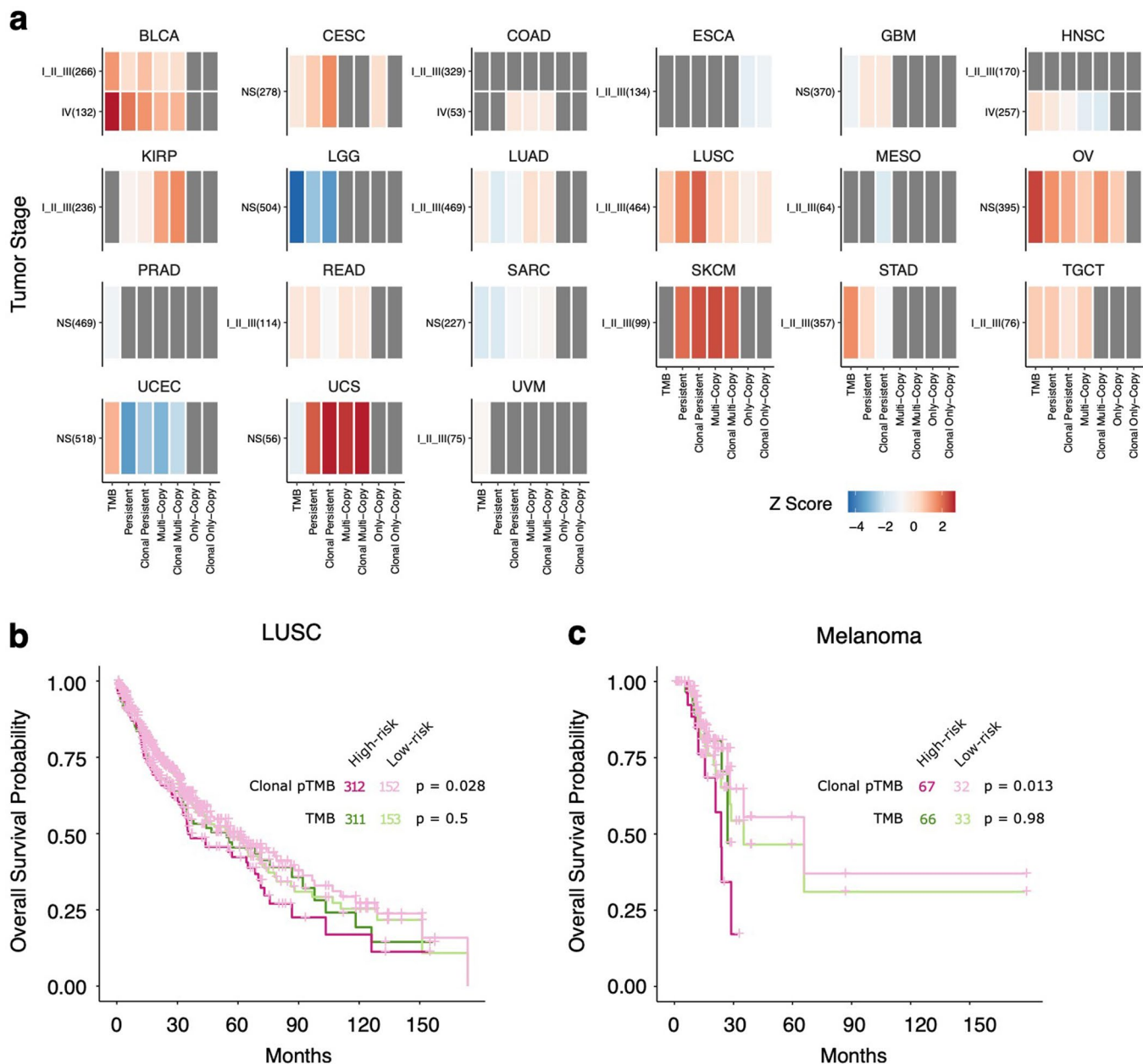
$p = 0.03$; Melanoma $\rho = 0.37$, $p = 8e-8$; Mesothelioma $\rho = 0.63$, $p = 1.6e-5$; NSCLC-Anagnostou $\rho = 0.54$, $p = 5.3e-7$; NSCLC-Shim $\rho = 0.51$, $p = 9.2e-13$). (c) pTMB was weakly anti-correlated with the fraction of genome at single copy number (total CN = 1) in the melanoma ($\rho = -0.14$, $p = 0.04$), mesothelioma ($\rho = -0.41$, $p = 8e-3$), and NSCLC ICB cohorts (NSCLC-Anagnostou $\rho = -0.26$, $p = 0.02$; NSCLC-Shim $\rho = -0.18$, $p = 0.02$). Spearman's rank correlation coefficients are shown, each tumor sample represents a point and points are color coded based on tumor response on ICB. Two-sided p-values are calculated using asymptotic t approximation.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Assessment of genomic characteristics of loci harboring persistent vs loss-prone mutations in TCGA tumors. (a) Genomic regions that are susceptible to limitations of NGS analysis, such as uncertain alignments or variant calls, were infrequent in the somatic mutation call set analyzed (MC3), and had similar representation in the persistent and loss-prone mutation categories ($n = 9,242$). **(b)** Similarly, the difference in GC composition of the immediate bases (9-mers) surrounding persistent and loss-prone mutations was negligible ($n = 9,242$; Cohen's $d = 0.08$, persistent mean = 0.54, loss-prone

mean = 0.52). Persistent and loss-prone mutations were found to have similar replication timing in **(c)** melanoma (TCGA-SKCM, $n = 109$, Cohen's $d = -0.035$) and **(d)** NSCLC (combined set of TCGA-LUAD and TCGA-LUSC, $n = 982$, Cohen's $d = -0.032$). Stacked bar plots depict the frequency of mutations in the five quintiles of replication timing. H, MH, M, ML, and L indicate mutations in the highest to lowest quintiles of replication timing in order; as an example mutations in earliest replicating regions are marked as H.

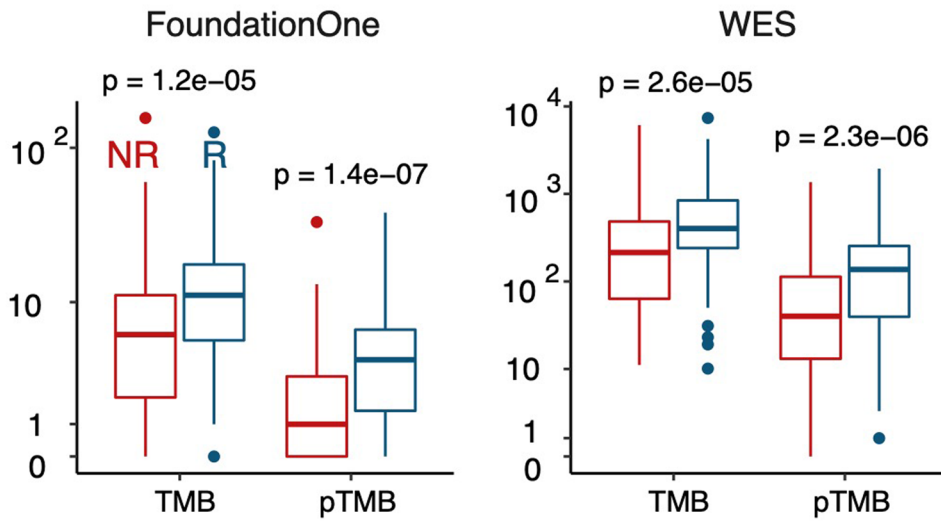


Extended Data Fig. 7 | Context dependence of the association between pTMB and overall survival.

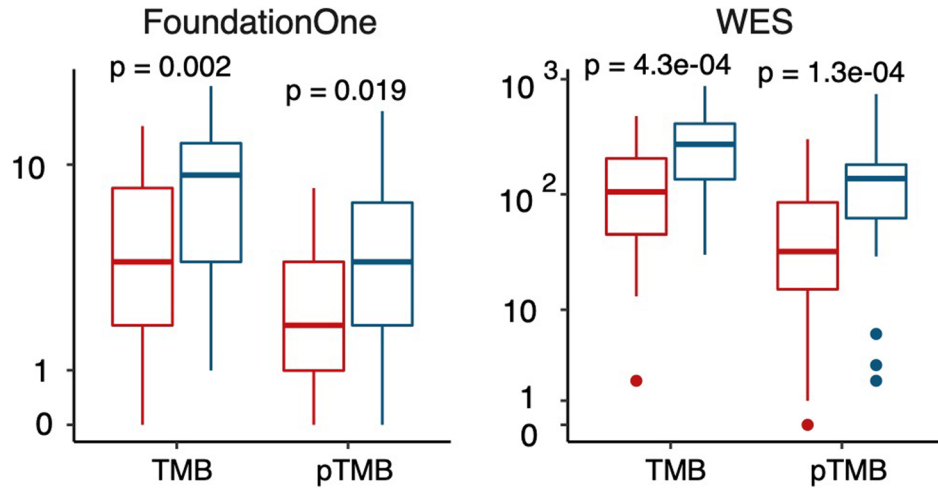
(a) The association between persistent mutations and TMB with overall survival was assessed for 8,925 individuals across 31 cohorts in TCGA. For each tumor type and stage combination, a Cox proportional-hazards (CoxPH) model predicting the overall survival is built for each of seven features (TMB, persistent mutations-pTMB, clonal pTMB, multi-copy mutations, clonal multi-copy mutations, only-copy mutations, clonal only-copy mutations; Continuous CoxPH model). In 21 tumor types shown, an increase in at least one of the seven features listed was associated with longer overall survival and a second CoxPH model was used to assess the difference in overall survival between tumors in the top third and bottom two thirds of predicted risk (Categorical CoxPH model). Heatmap cells depict the Z-score of the model coefficient from categorical CoxPH model. Grey indicates cases where an increase in feature value is associated with shorter survival. A significant association of pTMB

with prolonged overall survival was noted for lung squamous cell carcinoma, melanoma and UCS. (b) Patients with early stage (I, II, and III) squamous lung cancer (n = 464) harboring a high pTMB (low risk) had a significantly longer overall survival compared to patients in the low pTMB (high risk) group especially when clonal pTMB was considered (pTMB: 56.27 vs 43.86 months, log-rank p = 0.085; clonal pTMB: 60.48 vs 35.32 months, log-rank p = 0.028), while no difference was found between high vs. low TMB risk groups (TMB: 55.16 vs 54.37 months, log-rank p = 0.50). (c) Similarly, for patients with early stage melanoma (n = 99), tumors with higher pTMB (low risk) had a significantly longer overall survival compared to those with low pTMB especially when clonal pTMB was considered (pTMB: 65.83 vs 23.69 months, log-rank p = 0.036; clonal pTMB: 65.83 vs 23.69 months, log-rank p = 0.013); while no such difference was observed for TMB-high vs low tumors (35.15 vs 26.97 months, log-rank p = 0.98).

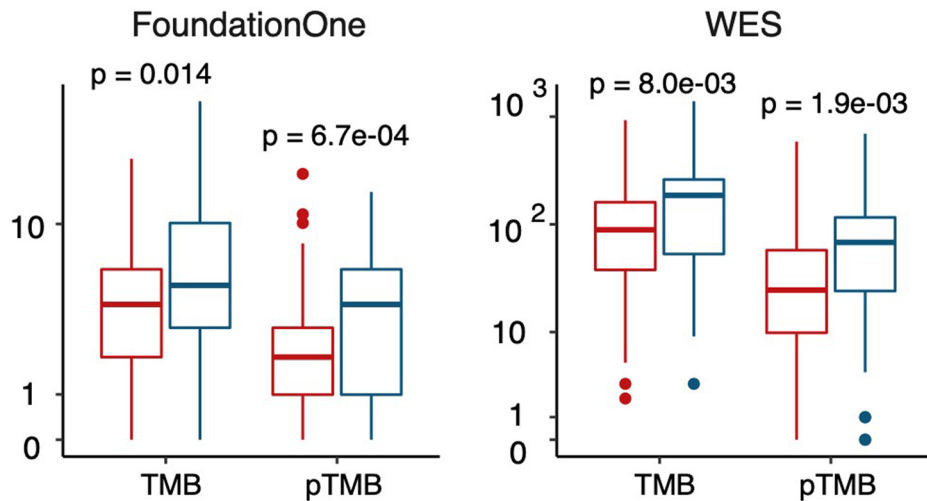
a Melanoma (n=202)



b NSCLC-Anagnostou (n=74)



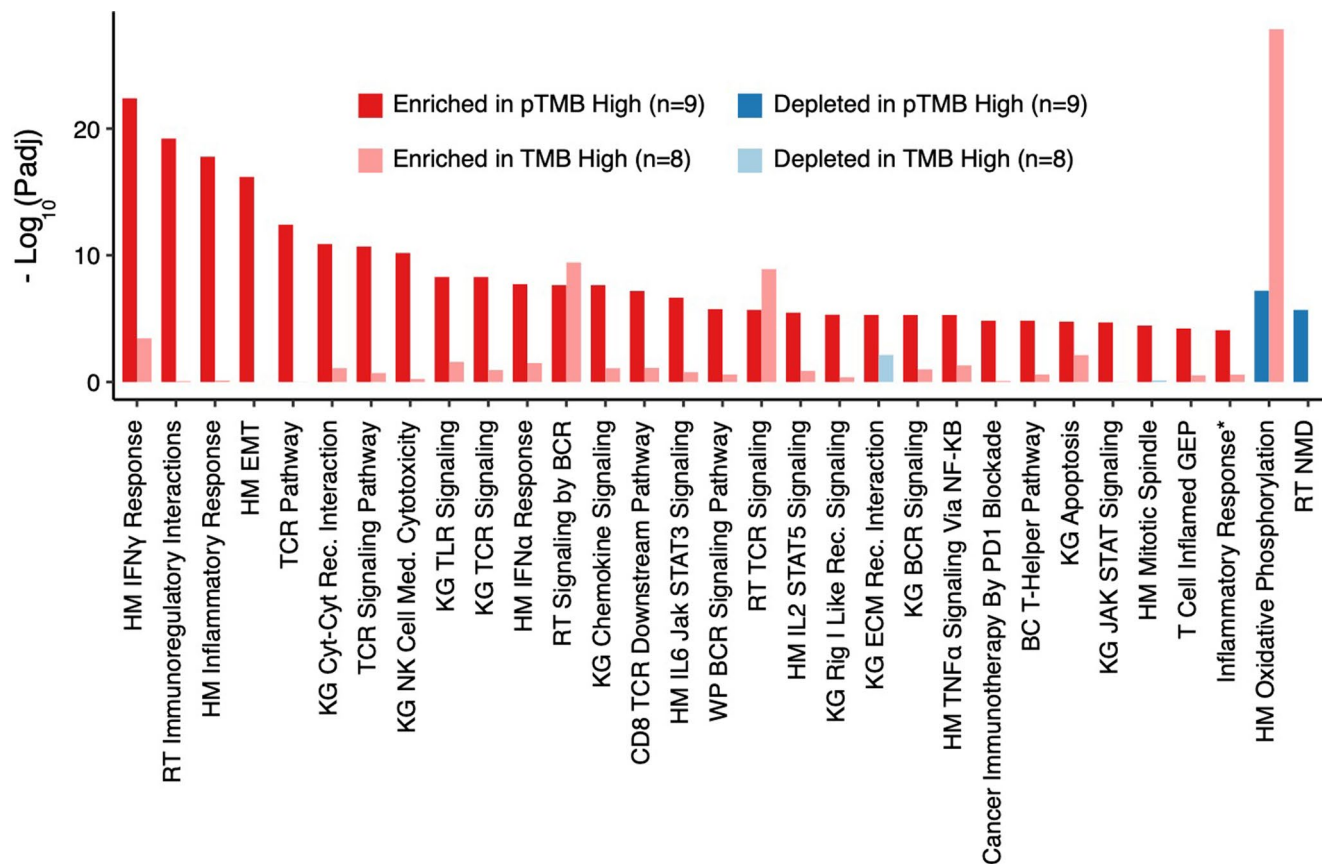
c NSCLC-Shim (n=169)



Extended Data Fig. 8 | See next page for caption.

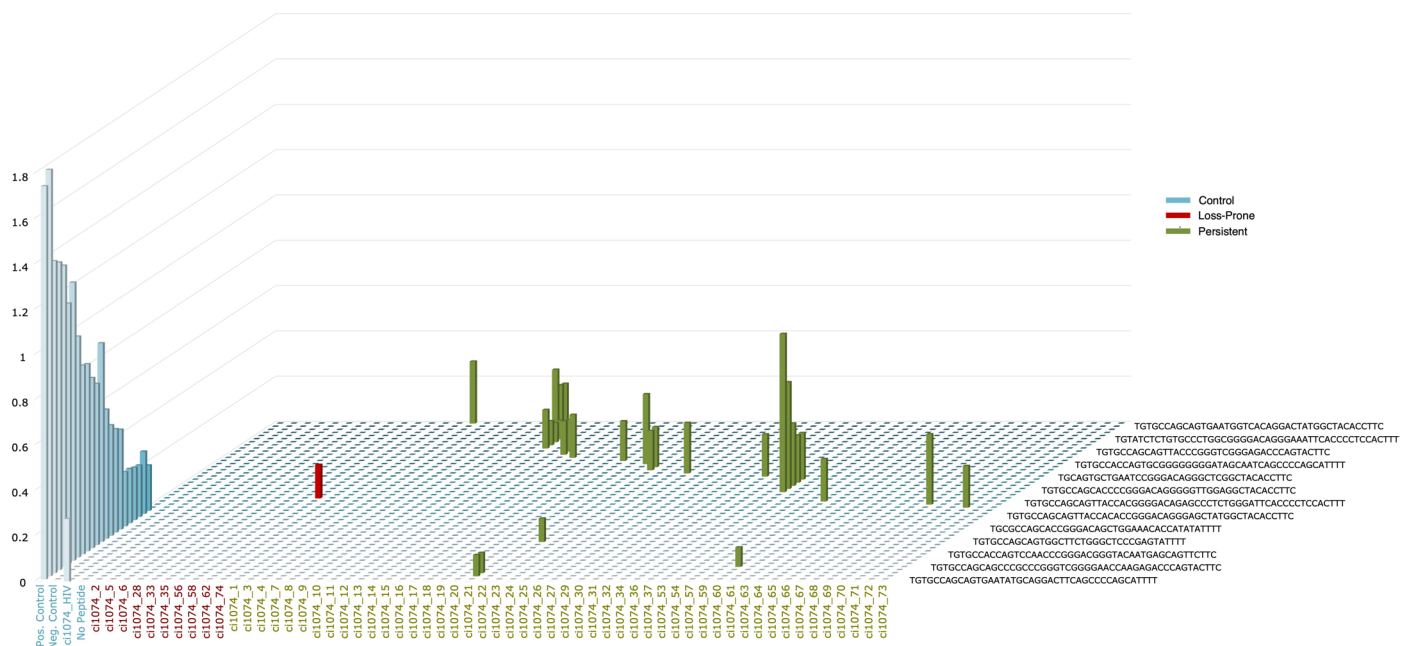
Extended Data Fig. 8 | Evaluation of the association between pTMB computed by gene panel NGS and response to immune checkpoint blockade. TMB and pTMB estimates were derived by restricting analysis to the targeted intervals of a commonly used commercial 309 gene panel (FoundationOne-CDx) through an *in silico* simulation and compared between response groups. pTMB better distinguished between responding and non-responding tumors in melanoma (**a**; pTMB MW $p = 1.3e-07$ vs TMB MW $p = 1.2e-05$, NR $n = 115$, R $n = 87$; two-sided test) and the NSCLC-Shim cohort (**c**; pTMB MW $p = 6.7e-04$ vs TMB MW $p = 0.014$,

NR $n = 120$, R $n = 49$; two-sided test). Comparison of TMB and pTMB between response groups using original estimates from whole exome sequence (WES) are included in each row for reference. pTMB and TMB had similar performance in the NSCLC-Anagnostou cohort (**b**), which was likely driven by the smaller size of this cohort (NR $n = 41$, R $n = 33$). Box plots depict the median value and hinges correspond to the first and third quartiles. The whiskers extend from the corresponding hinge to the furthest value within $1.5 \times$ the interquartile range from the hinge.



Extended Data Fig. 9 | Transcriptomic profiling of on-immunotherapy melanomas reveals an upregulation of inflammatory pathways in tumors harboring a high pTMB. Gene set enrichment analyses of on-treatment pTMB-high (n = 9) vs. pTMB-low (n = 22) melanomas revealed a marked enrichment of interferon- γ and inflammatory responses in the tumor microenvironment of pTMB-high vs pTMB-low tumors after ICB. In contrast, this profile was not observed in the tumor microenvironment of TMB-high (n = 8) compared to TMB-low (n = 23) melanomas. Pathways with a minimum adjusted p-value of 1e-05 in pTMB high vs low comparison are shown. The T Cell Inflamed GEP gene set was

derived from Cristescu et al., *Science*, 2018 and the Inflammatory Response gene set was derived by Ayers et al., *J Clin Invest*, 2017, while the remainder of gene sets were included in the Molecular Signatures Database (Methods). Nominal two-sided p-values are calculated using permutation testing and FDR adjusted p-values shown for gene set differential expression are provided for comparison of pTMB/TMB-high vs low groups. Abbreviations; BC: Biocarta, HM: Hallmark, KG: KEGG, RT: Reactome, EMT: Epithelial-mesenchymal transition, Cyt: Cytokine, Rec: Receptor, Med: Mediated.



Extended Data Fig. 10 | Neoantigen-specific T-cell expansion of persistent mutations. The MANAFEST assay was used to test pMANA-specific T-cell recognition. A collection of 58 neopeptides arising from 41 somatic persistent and loss-prone mutations, selected blindly with respect to pMANA status, were synthesized and tested in autologous T cell cultures from peripheral blood for patient CGLU310. A higher frequency of antigen-specific T-cell expansion was

identified in neopeptides associated with persistent (green bars, 15 out of 48) vs loss-prone (red bar, 1 out of 10) mutations. T-cell clonotypes (as nucleotide sequence) and neopeptides are listed alongside the plane while the height of the bars indicates the abundance of TCR clones. A floor value of frequency of 0.08 informed by the values observed in the negative control condition was applied to TCR clonotypic abundances.

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

Data analysis

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](#)

Source data for the TCGA tumor samples were retrieved from <http://cancergenome.nih.gov>. WES-derived somatic mutation calls from the TCGA PanCancer Atlas MC3 project were retrieved from the NCI Genomic Data Commons (<https://gdc.cancer.gov/about-data/publications/mc3-2017>). Somatic copy number profiles (<https://gdc.cancer.gov/about-data/publications/pancanatlas>) and clinical data (<https://gdc.cancer.gov/about-data/publications/PanCan-Clinical-2018>) were accessed from Genomic Data commons. Previously published genomic data, re-analyzed here, were obtained from the material of the original publications and from dbGaP under accession code phs000452.v3.p17, and Sequence Read Archive (SRA) under accession codes SRP0958096, SRP0679388 and SRP0902948. WES sequence data for the HNSCC and NKI cohorts from patients who consented to data deposition can be retrieved from the European Genome-phenome Archive

(EGA accession number EGAS00001006660).

The de-identified clinical and genomic data used for the analyses in this study are available in the Supplementary Tables and all publicly available data elements have been referenced.

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	Tumor samples from the TCGA and 7 retrospective cohorts of patients treated with immune checkpoint blockade (ICB) were used in this study. The TCGA set comprised of 10,742 tumors, for which whole exome sequence data were analyzed with respect to their copy number profile and used to quantify the background rate of genomic loss. In 9,242 tumor samples, somatic mutation calls and copy number profiles were combined to characterize the prevalence of single-copy, multi-copy, and persistent mutations. The ICB cohorts comprised of a total of 524 patients, whose tumors were analyzed to determine the association between persistent tumor mutation burden and clinical outcome.
Data exclusions	In the TCGA cohort, tumor samples with missing copy number profile, somatic mutation calls, or clinical data were excluded. In the ICB cohorts, tumor samples where the somatic copy number profile could not be resolved were excluded. In the Liu et al. melanoma cohort, samples with prior anti-CTLA4 treatment were excluded.
Replication	The study reports the proof-of-concept evaluation of a new genomic feature (persistent tumor mutation burden) in 10 independent cohorts treated with ICB. Sex was not considered in study design. Sex-stratified analyses were not performed as these were not within the scope of work presented. For the newly sequenced samples, self-reported sex is included in supplementary tables.
Randomization	The study was a retrospective cohort study, no randomization was performed.
Blinding	All data was de-identified.

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

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input 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 type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

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

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Population characteristics including self-reported sex and self-reported race are explicitly described in the manuscript. Sex or race-stratified analyses were not performed.
Recruitment	This is not applicable.
Ethics oversight	Sidney Kimmel Comprehensive Cancer Center Institutional Review Board, NKI Institutional Review Board, University of Chicago Comprehensive Cancer Center IRB 8980

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