Cell

Title: Immunoediting restricts clonal neoantigens in primary tumors

Elizabeth S. Borden[1,2], David Castellano[3], Alysia Hughes[4], Tao Ma[4], Xing Li[4], Aaron R. Mangold[4], Ryan N. Gutenkunst[3], Bonnie J. LaFleur[5,6], Kenneth H. Buetow[7,8], Melissa A. Wilson[7,8], and Karen Taraszka Hastings[1,2,9,10]*

[1]Department of Dermatology, College of Medicine-Phoenix, University of Arizona, Phoenix, AZ, USA
[2]Phoenix Veterans Affairs Health Care System, Phoenix, AZ, USA
[3]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ, USA
[4]Department of Dermatology, Mayo Clinic Health System, Scottsdale, AZ, USA
[5]BIO5 Institute, University of Arizona, Tucson, AZ, USA
[6]R. Ken Coit College of Pharmacy, University of Arizona, Tucson, AZ 85724, USA
[7]School of Life Sciences, Arizona State University, Tempe, AZ, USA
[8]Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA
[9]University of Arizona Cancer Center, University of Arizona, Tucson, AZ, USA
[10]Lead contact

*Correspondence: khasting@arizona.edu; Tel.: 1-602-827-2106

Summary:
Evidence consistent with immunoediting in human cancer is associative. To provide direct evidence that differences in the neoantigen profile are immune-mediated, we compared neoantigen profiles of cutaneous squamous cell carcinoma (cSCC) from immunocompetent and immunosuppressed patients. Despite consistency in the overall mutational burden and signature, immunocompetent patients have a lower clonal mutational burden. Clonal mutations in cSCC from immunocompetent patients have a lower proportion of neoantigens predicted to bind MHC class I, supporting the immune system's role in shaping the neoantigen profile. We observe an increase in exhausted CD8+ T cells in tumors from immunocompetent patients, suggesting an explanation for the persistence of binding neoantigens subclonally. Finally, neoantigens with a higher predicted MHC class I binding relative to the unmutated peptide are enriched subclonally in immunocompetent patients. Overall, this work supports the immune system's role in sculpting the neoantigen profile of cSCC.

Keywords: Immune evasion, Tumor escape, Immunologic surveillance, T-cell exhaustion, Squamous cell carcinoma, Skin cancer, Neoplasm antigens, Immunotherapy, Tumor microenvironment

Introduction:
Immunoediting is the process by which the immune system targets and removes cancer cells visible to the immune system, allowing the populations least visible to the immune system to persist. Immunoediting was defined in mouse models, where a number of studies observed increased tumor formation and growth in immunocompromised mice compared to mice with intact immune systems.[1, 2, 3, 4, 5] Furthermore, chemically induced sarcomas that form in immunocompetent mice grow when transplanted into immunocompetent mice, whereas sarcomas that develop in immunocompromised RAG-/- mice fail to grow in immunocompetent mice,[6] demonstrating that tumors that form in immunocompromised mice fail to develop mechanisms to escape immune-mediated destruction. More recent work in autochthonous mouse models has demonstrated that the immune system primarily restricts clonal neoantigens in immunocompetent mice, leading to greater tumor heterogeneity in tumors that arise in immunocompetent compared to immunocompromised mice.[7] These mouse studies have provided fundamental advances in our understanding of the extent and mechanisms of immunoediting in cancer. In human tumors, findings consistent with immunoediting have been found in 1) invasive cutaneous squamous cell carcinoma (cSCC) vs. precursor actinic keratoses,[8] 2) late vs. early recurrent tumors,[9] 3) tumors after vs. before immune checkpoint inhibition,[10] 4) in tumor regions with high immune infiltration and intact MHC alleles,[11] 5) compared to a neutral mutational profile in a pan-cancer analysis,[12] and 6) comparing synonymous to non-synonymous mutations within the same tumor.[13] However, the findings in human tumors that are consistent with immunoediting are associative in nature and have not been shown to be due to the immune system.

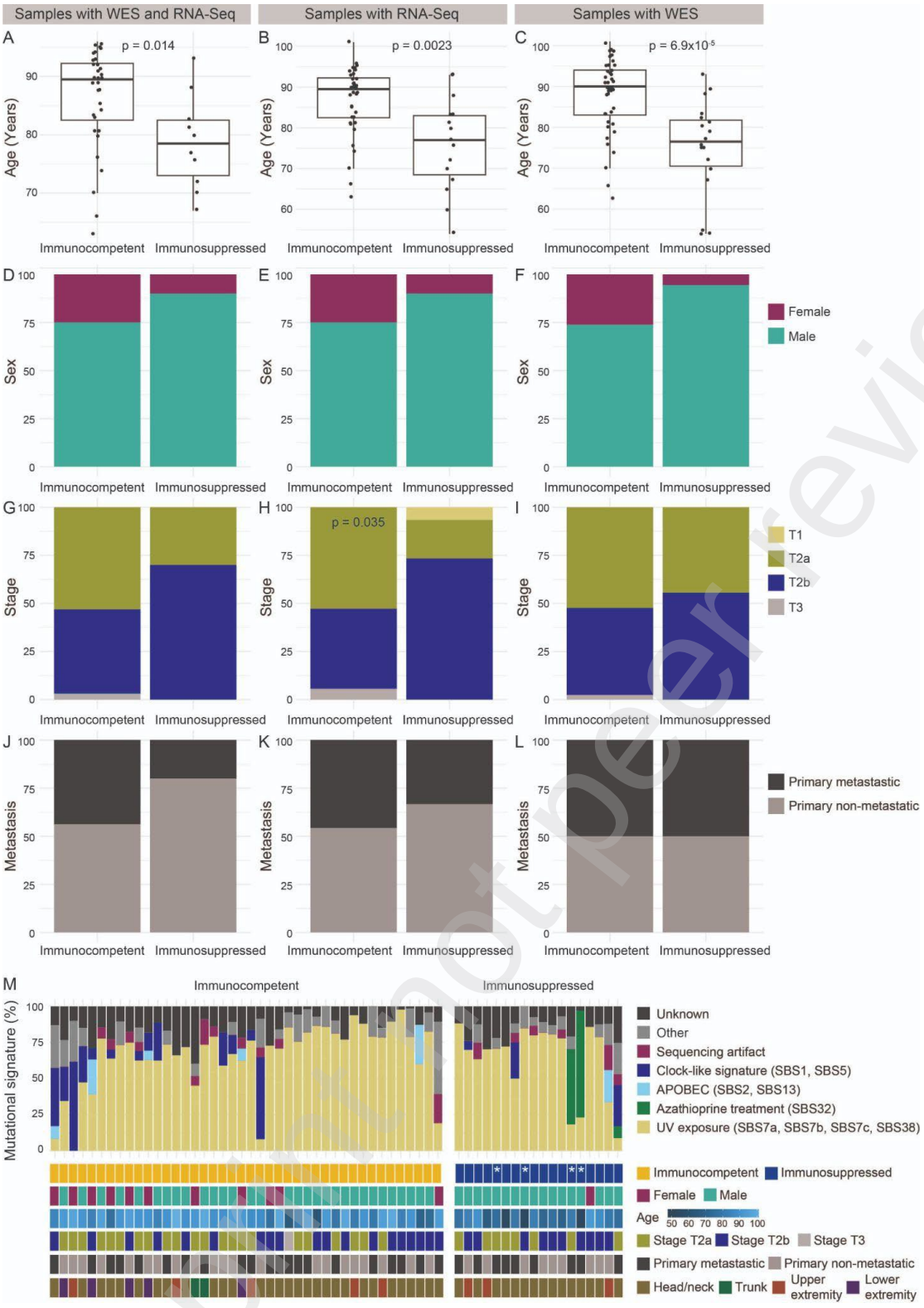
Further complicating the challenges in analyzing immunoediting in human cancer is an incomplete understanding of the characteristics of neoantigens that elicit an immune response. While many prediction models have been created to date,[9, 14, 15, 16, 17, 18, 19, 20] two central challenges remain. First, testing of neoantigens in human studies is performed after *in silico* prediction, leading to the potential to reaffirm starting biases.[19, 21, 22, 23, 24] Secondly, immunogenicity of neoantigens is tested in human studies through the identification of neoantigen-specific, IFNɣ-secreting, CD8+ T cells (e.g. by ELISPOT) but no confirmation of the ability of individual neoantigens to mediate tumor rejection is directly measured. Since neoantigens that elicit IFNɣ-secreting CD8+ T cells are imperfectly correlated with neoantigens that elicit T cell-mediated immune destruction in mouse models,[25, 26] this method of testing may introduce additional biases. Therefore, there is a need for further understanding of the characteristics of neoantigens that elicit immune responses in human cancer.

Here, we compared characteristics of the neoantigen and immune profile between cSCC in immunocompetent and immunosuppressed patients to illuminate the effects of immunoediting in primary tumors due to an intact immune response. Immunosuppressed solid organ transplant recipients have a 65-253-fold higher risk of developing cSCC,[27, 28, 29, 30, 31] suggesting that cSCC is highly constrained by the immune system in immunocompetent patients. Therefore, tumors from immunocompetent patients are expected to have developed mechanisms of immune evasion that are absent in the tumors from immunosuppressed patients. As a further control, we used *in silico* mutational profiles as a patient-specific comparison of the expected mutational profile in the absence of an immune response. These comparisons allow for the direct identification of changes in the neoantigen profile that are affected by the presence of an intact immune response. Finally, after demonstrating immunoediting using known characteristics of immunogenic neoantigens, we used the comparison of cSCC in immunocompetent and immunosuppressed patients as an unbiased approach to identify characteristics of neoantigens that are immunoedited in immunocompetent, but not immunosuppressed patients. Overall, this study demonstrates alterations in the neoantigen landscape in primary cSCC due to the immune system and uses the comparison of tumors in immunocompetent and immunosuppressed patients to determine characteristics that allow the immune system to recognize neoantigens.

Results:

Comparable mutational burden and profile in cSCC from immunocompetent and immunosuppressed patients

To isolate the role of immunoediting in cSCC, we first compared the demographics, overall mutational burden, and signature between cSCC in immunocompetent and immunosuppressed patients. The distribution of sex, stage, and metastatic status was equivalent for immunocompetent and immunosuppressed patients; however, immunocompetent patients were significantly older than immunosuppressed patients (Supplementary Figure 1A-L, Supplementary Table 1). We observed a wide range in mutational burden in the cSCC tumors, as previously reported by our group and others,[8, 32, 33, 34, 35] but there was no significant difference in the number of somatic or missense mutations in the cSCC from immunocompetent and immunosuppressed patients (Figure 1, Supplementary Table 2). We then calculated the proportion of mutations in each tumor attributable to established mutational signatures (Figure 1 and Supplementary Figure 1M). A majority of the tumors had a predominant signature of UV-induced mutations. The other predominant mutational signature, especially in low mutational burden tumors, was the `clock-like` signature, which consists of mutations known to accumulate in human cells at a steady rate over time and is associated with aging.[36] Of note, consistent with lower rates of sun exposure, all cSCC from the lower extremity and trunk were clustered in the lower mutational burden samples with higher rates of `clock-like` signatures. Two out of four patients treated with azathioprine had a strong azathioprine signature, consistent with the literature.[33, 34] No significant difference in the proportion of the signature attributable to UV exposure and clock-like signature was observed between cSCC from immunocompetent and immunosuppressed patients. Overall, no significant differences were observed in the mutational burden and signature in cSCC from immunocompetent and immunosuppressed patients, supporting that differences in the neoantigen profile are unlikely to be attributable to differences in the mutational burden and signature.



Supplementary Figure 1: Comparison of age, sex, stage, and metastasis in cSCC from immunocompetent and immunosuppressed patients. Summary of demographic characteristics for all cSCC samples, segregated by immune status. (A-C) Patient age for (A) samples with WES and RNA-Seq data (n=42), (B) samples with RNA-Seq data (n=50), and (C) samples with WES data (n=60). For all boxplots, the bold line indicates the median and the upper and lower limits of the boxes indicate the 75th and 25th percentiles, respectively. The lower and upper whiskers indicate the minimum and maximum after excluding outliers. Dots outside of the box and whiskers indicate outliers. Mean age differences hypothesis testing used a

Wilcoxon rank sum test statistic at the 0.05 level of significance. (D-E) Patient sex for (D) samples with WES and RNA-Seq data, (E) samples with RNA-Seq data, and (F) samples with WES data. (G-I) Tumor stage for (G) samples with WES and RNA-Seq data, (H) samples with RNA-Seq data, (I) samples with WES data. (J-L) Metastatic status of (J) samples with WES and RNA-Seq data, (K) samples with RNA-Seq data, and (L) samples with WES data. Fisher's exact test statistic, at 0.05 level of significance was used to test differences in proportions in panels D-L. P-values > 0.05 are not shown. (M) Mutational signatures deconvoluted with the decongstructSigs package from R for comparison to the results from sigminer shown in Figure 1. "Other" includes: SBS4, SBS10b, SBS11, SBS12, SBS16, SBS18, SBS19, SBS20, SBS23, SBS24, SBS28, SBS29, SBS30, SBS31, SBS36, SBS39, SBS40, SBS42, SBS44, SBS45, SBS46, SBS50, SBS51, SBS52, SBS53, SBS54, SBS58, SBS59. For skin site, head/neck includes ear, cheek, forehead, jaw, lip, temple, scalp, and periocular; trunk includes shoulder and back; upper extremity includes arm, forearm, and hand; and lower extremity includes leg, lower leg, and foot.

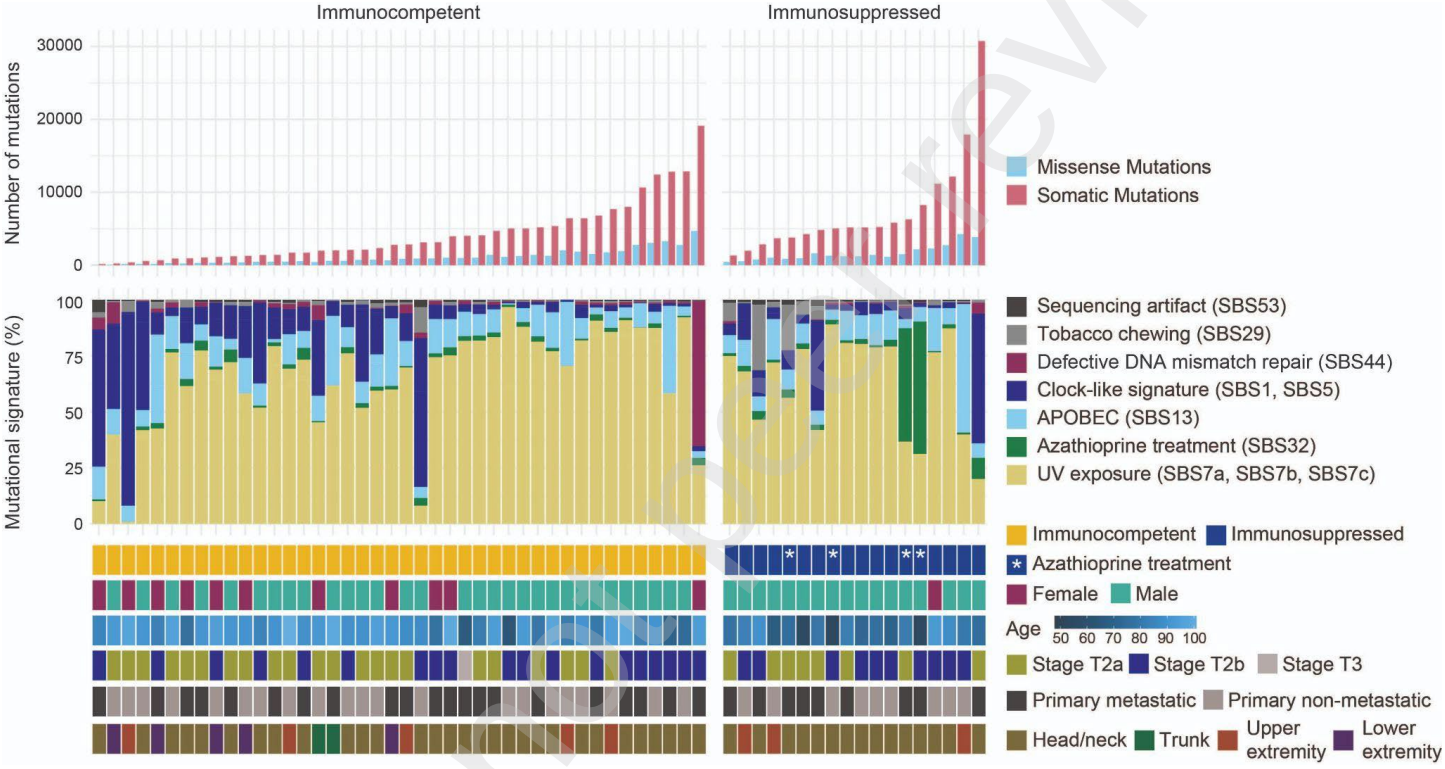


Figure 1: Comparable mutational burden and profile in cSCC from immunocompetent and immunosuppressed patients. Somatic mutations were identified, and missense mutations were annotated, as a subset, in 60 primary cSCC tumors from immunocompetent and immunosuppressed patients. Top: Bar graph showing the somatic and missense mutational burdens subset by immune status. Middle: Stacked bar graph showing the proportion of the mutations in each tumor attributable to known COSMIC mutational signatures. Bottom: Annotations for immune status, sex, age, stage, metastatic status, and skin site of each sample. Immunosuppressed patients are annotated for treatment with azathioprine given the known impact of azathioprine in the mutational profile of cSCC tumors. For skin site, head/neck includes ear, cheek, forehead, jaw, lip, temple, scalp, and periocular; trunk includes shoulder and back; upper extremity includes arm, forearm, and hand; and lower extremity includes leg, lower leg, and foot.

Lower clonal mutational burden in immunocompetent patients

Despite the consistency in the mutational burden and profile of cSCC from immunocompetent and immunosuppressed patients, cSCC from immunocompetent patients had a significantly lower mean variant allele frequency (VAF; proportion of sequencing reads containing the variant) compared to cSCC from immunosuppressed patients (Figure 2A-B). This difference persisted in the cancer cell fraction (CCF; VAF adjusted for purity and copy number, Supplementary Figure 2A-B). Furthermore, cSCC from

immunocompetent patients had a lower total number of clonal mutations (Figure 2C) and a lower proportion of clonal mutations relative to the overall mutational burden (Supplementary Figure 2C-D) than cSCC from immunosuppressed patients, tested by two different methods for assigning clonality. Notably, immunocompetent patients were significantly older than immunosuppressed patients (Supplementary Figure 1A); more time to accumulate UV-induced mutations should increase the absolute number of clonal mutations in the older immunocompetent patients, contrary to what was observed. When adjusted by age, tumors from immunocompetent patients accumulated fewer clonal mutations per year (Supplementary Figure 2E). Together these results demonstrate a substantial difference in the observed clonal and subclonal mutational profiles in tumors from immunocompetent and immunosuppressed patients.

To identify the cause of this difference, first, the purity of tumors was evaluated between immunocompetent and immunosuppressed patients. A decrease in purity would be expected to decrease the VAF distributions since fewer sequencing reads would be attributable to the tumor and able to contain the variant of interest. Despite the consistency in the preparation of the tumors for sequencing, tumors from immunocompetent patients had a decreased purity compared to tumors from immunosuppressed patients (Figure 2D). The difference in purity was inversely correlated with the immune cell infiltrate estimated from immune cell deconvolution (Figure 2E), suggesting that the differences in the purity were likely attributable to the differences in immune infiltration in cSCC (discussed in more detail later). As expected, the mean VAF was significantly associated with purity (Figure 2F); however, the differences in VAF distributions between immunosuppressed and immunocompetent patients could not be attributed to purity alone. If only clonal mutations were called/detectable, then the expected mean VAF would equal half the empirically measured purity (Figure 2F). Thus, a substantially higher number of subclonal mutations must have been present in immunocompetent patients to explain the significantly lower mean VAF observed in immunocompetent patients compared to immunosuppressed patients.

Two additional factors that may influence the distribution of clonal and subclonal mutations in the tumor are 1) the growth rate of the tumor and 2) immune selective pressure. Tumors with a high growth rate accumulate many subclonal mutations at very low frequency compared to slower-growing populations or populations with constant population size. Because of this, most variants may remain unique to a small proportion of cells and may be undetectable by standard variant-calling pipelines, resulting in a single peak of clonal mutations with a frequency equal to half the empirically measured purity (Supplementary Figure 2F). In contrast, for tumors that are slowly growing or have a constant population size, we expect a much higher number of detectable subclonal mutations (Supplementary Figure 2G), leading to a lower mean VAF. The second factor that may influence the distribution of clonal and subclonal mutations is immune selective pressure. Several studies have demonstrated that the immune system is more apt to recognize and target neoantigens at higher frequencies in the tumor.[7, 37, 38, 39] If the immune system selectively targets neoantigens, this would drive down the VAF/CCF distributions. To account for the potential effects of both growth rate differences and targeting of neoantigens by the immune system on individual patient VAF distributions, we used an Approximate Bayesian Computational (ABC) approach for parameter estimation.[40] VAF distributions were simulated across a range of growth rates and fractions of clonal mutations, using patient-specific purities (Figure 2D) and coverage distributions (Supplementary Figure 2H), to find the values of both parameters that best explain the observed VAF (Figure 2G, Supplementary Figure 3). The most likely parameter fits for the observed VAF distributions had a wide range in growth rates (Figure 2H) and a significantly lower fraction of clonal mutations in immunocompetent patients (Figure 2I). Qualitative results were consistent across assumptions of the number of cells, the use of relative vs. absolute mutation counts, and the ranges of parameters simulated (Methods). These results suggest that, in the presence of immune selective pressure, the tumors had significantly fewer clonal neoantigens and lower growth rates than those in immunosuppressed patients.

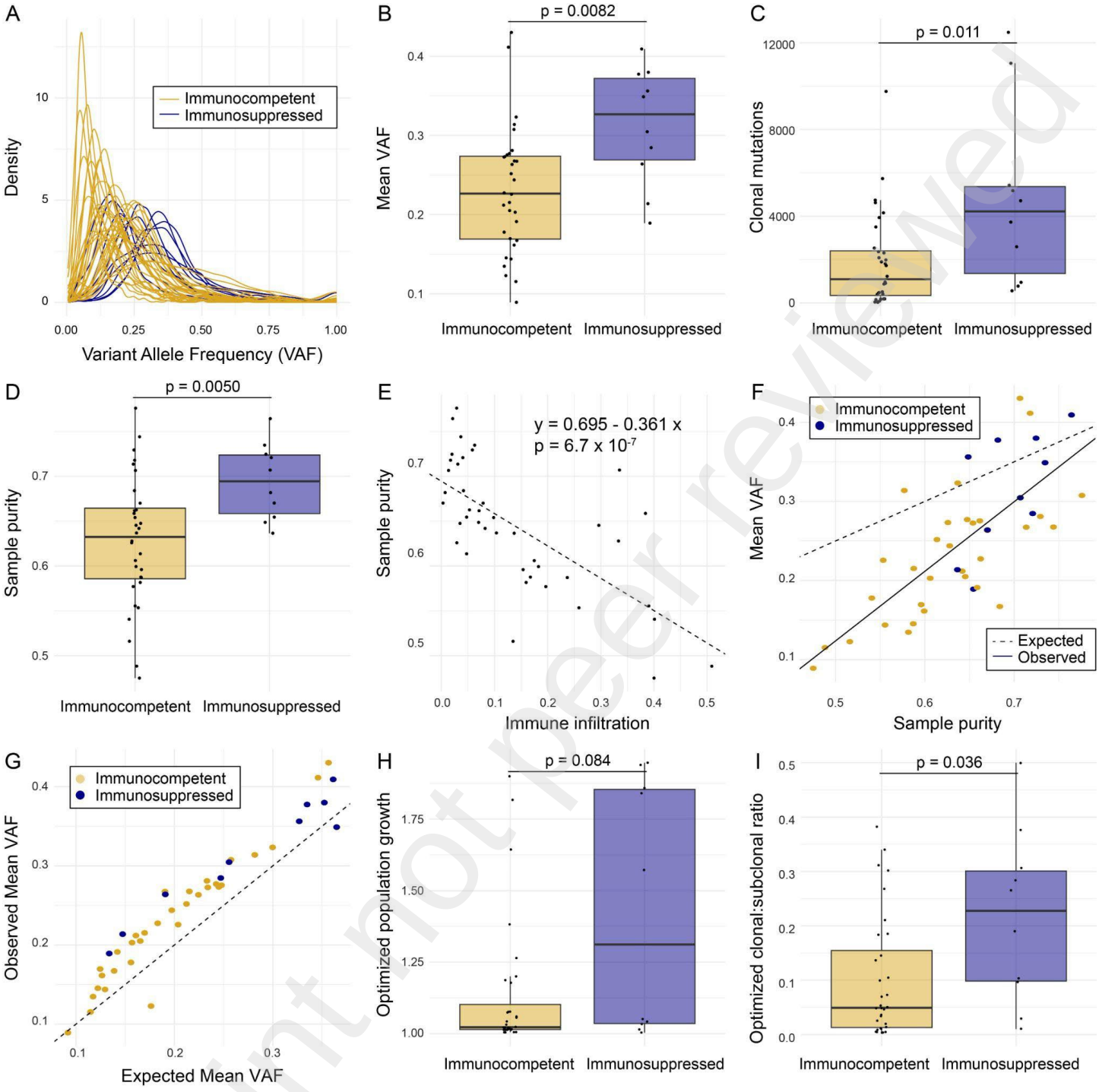
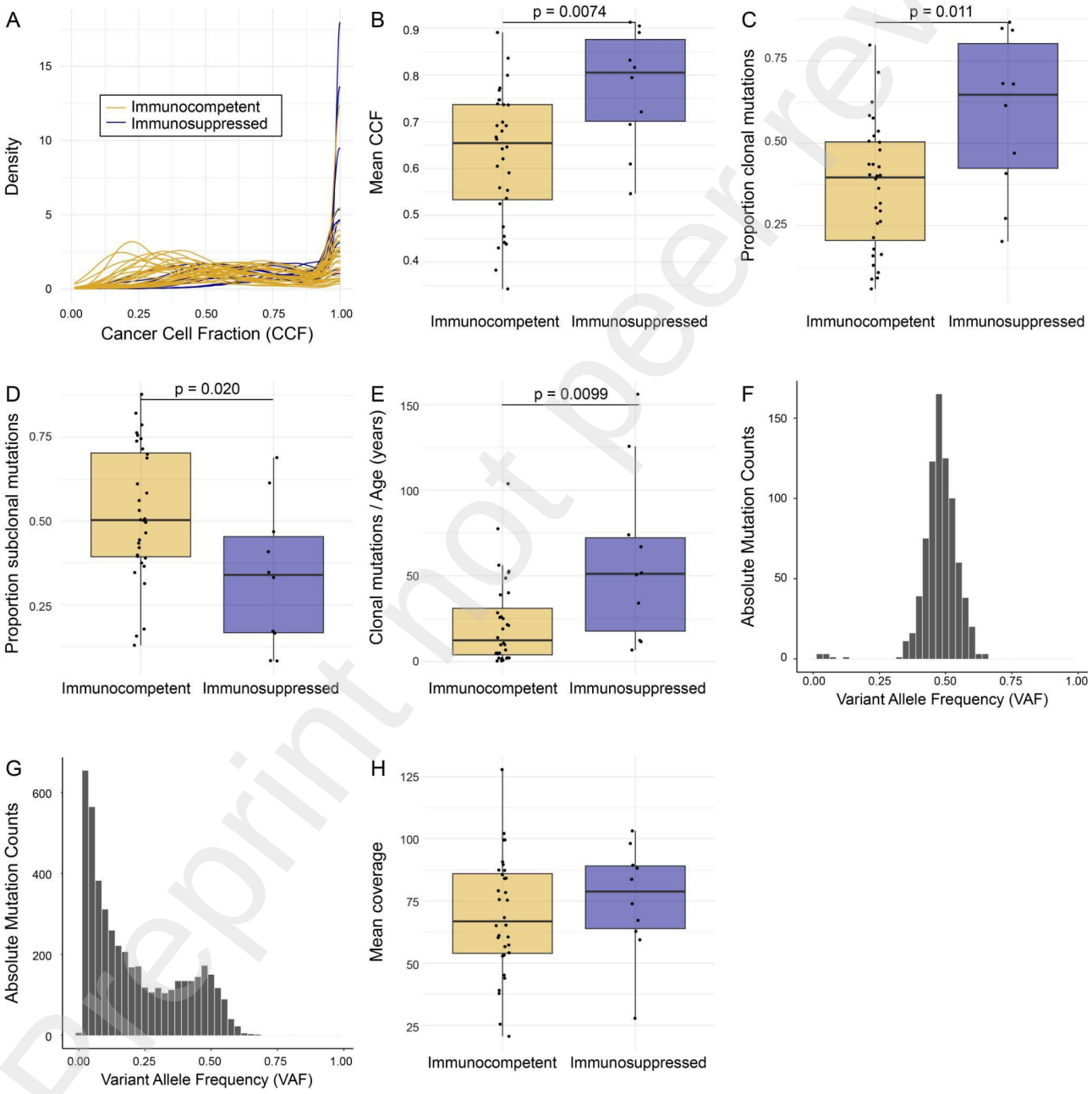
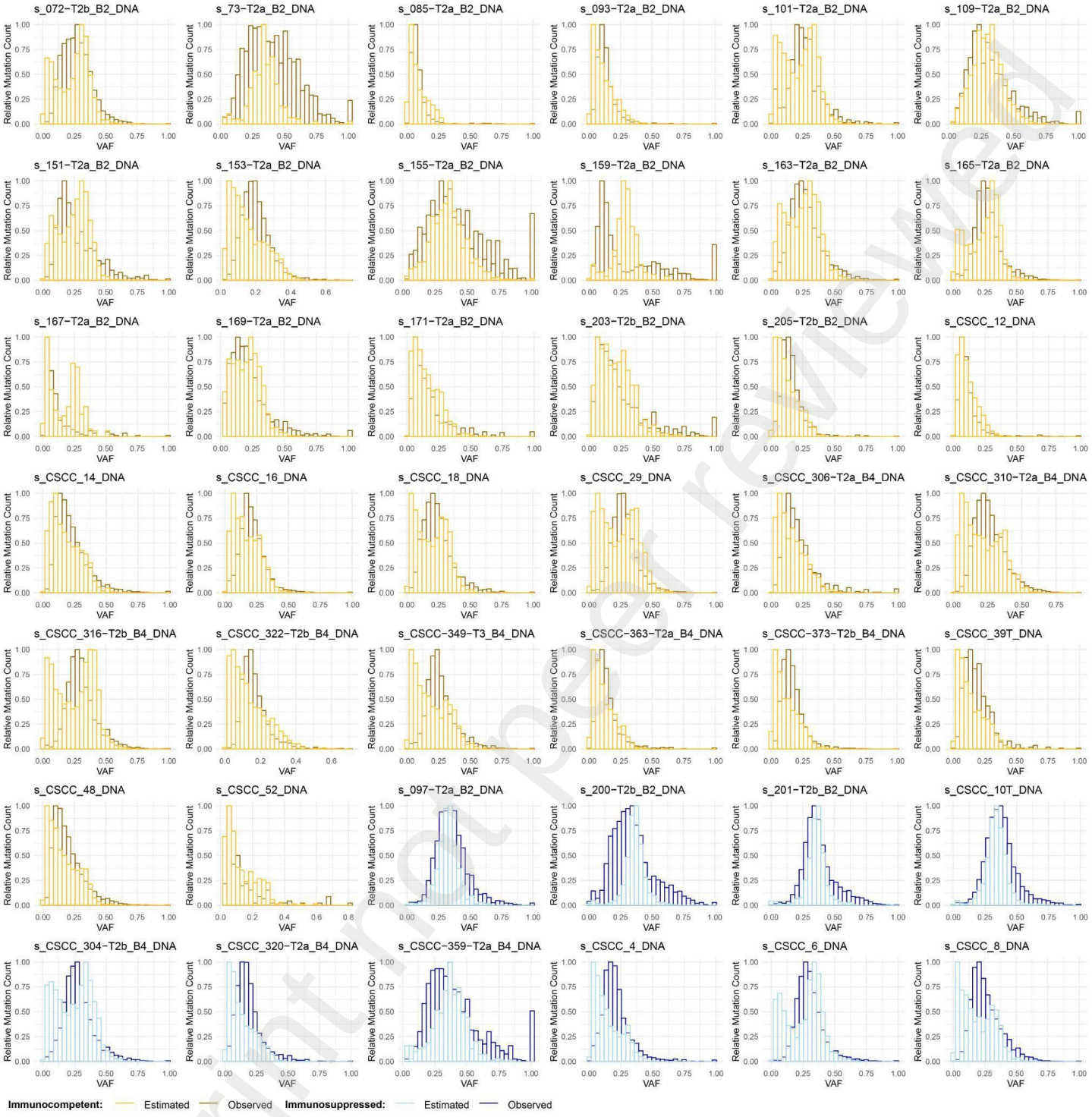


Figure 2: Lower clonal mutational burden in immunocompetent patients. (A) Variant allele frequency (VAF) distributions for all mutations from each tumor, colored by immune status (n = 32 immunocompetent, n=10 immunosuppressed, tumors with WES and RNA-Seq). Each line represents a tumor with the peak height indicating mutation density at specific VAFs. (B) Mean VAF for each tumor, segregated by immune status. (C) Number of clonal mutations in each tumor, segregated by immune status. Clonal mutations are defined as mutations with a cancer cell fraction (CCF; VAF adjusted for purity and copy number) > 0.75. (D) Tumor purity estimates from RNA-Seq gene signatures, segregated by immune status. (E) Scatter plot showing the relationship between sample purity and immune infiltration (estimated from xCell), with a fitted linear regression (dashed line). (F) Scatter plot of mean VAF versus sample purity. The black dashed line indicates the expected relationship between purity and mean VAF, presuming only clonal mutations, whereas the solid black line indicates the observed relationship between purity and mean variant allele frequency. Points are colored by

immune status. (G-I) VAF distributions were simulated across a range of growth rates and fractions of clonal mutations to find the values of both parameters that best explain the observed VAF. Approximate Bayesian Computation (ABC) was implemented to select the normalized simulated VAFs that had the best chi-square goodness-of-fit statistic between the observed and simulated distributions. (G) Comparison of the observed mean VAF to the estimated mean VAF from the models selected using ABC methods. The dotted black line indicates the 1:1 correspondence between observed and estimated mean VAFs. (H) Optimized growth rates from the estimated VAF distribution results. (I) Optimized estimate of the underlying clonal:subclonal ratio from the estimated VAF distribution results (includes both observed subclonal mutations and subclonal mutations below the limits of detection for variant calling). For all boxplots, the bold line indicates the median and the upper and lower limits of the boxes indicate the 75th and 25th percentiles, respectively. The lower and upper whiskers indicate the minimum and maximum after excluding outliers. Dots outside of the box and whiskers indicate outliers. Differences in scale parameters for panels B, C, D, H, and I were all tested using a Wilcoxon rank sum test at the 0.05 level of significance.



Supplementary Figure 2: Additional support for lower clonal mutational burden in immunocompetent patients. (A) Cancer cell fraction (CCF; variant allele frequency adjusted for purity and copy number) distributions for all mutations from each tumor, colored by immune status (n = 32 immunocompetent, n=10 immunosuppressed, tumors with WES and RNA-Seq). Each line represents a tumor with the peak height indicating mutation density at specific CCFs. (B) Mean CCF for each tumor, segregated by immune status. (C) Proportion of clonal mutations (CCF > 0.75) out of all observed mutations, segregated by immune status. (D) Proportion of subclonal mutations out of all observed mutations, segregated by immune status. As a confirmatory approach, subclonal mutations here are defined as mutations that were significantly lower than a purity-adjusted threshold for clonality using a one-sided binomial test. Elsewhere in the manuscript `clonal` and `subclonal` are defined by a CCF threshold of 0.75. (E) Comparison of the number of clonal mutations (CCF>0.75) divided by the patient's age, segregated by immune status. (F-G) Simulated variant allele frequencies (VAFs) with varying growth rates in (F) an exponentially growing population and in (G) a constant population size. The remaining simulation parameters are constant in the two scenarios: the number of sequenced cells (150,000), the fraction of clonal mutations in the sequenced sample (5%), the number of mutations per cell (10), tumor purity (100%), the number of required alternative reads to call a mutation (2), and the mean coverage (80) and variance (15) of the sequencing reads. It should be noted that the total number of mutations is the same in the two scenarios, but in (F) most variants remain undetectable at very low frequency while in (G) most detected mutations are subclonal. (H) Mean sequencing coverage across all variants, segregated by immune status. For all boxplots, the bold line indicates the median and the upper and lower limits of the boxes indicate the 75th and 25th percentiles, respectively. The lower and upper whiskers indicate the minimum and maximum after excluding outliers. Dots outside of the box and whiskers indicate outliers. A Wilcoxon rank sum test at the 0.05 level of significance was used to examine scale differences in panels B, C, D, and E.



Supplementary Figure 3. Comparison of estimated and observed VAF histograms. Overlaid histograms of the variant allele frequency (VAF) distributions estimated by the posterior distributions from implementation of Approximate Bayesian Computational (ABC) compared to the observed VAF distributions. Tumors from immunocompetent patients are shown in yellow with tumors from immunosuppressed patients in blue. Dark yellow and blue indicate estimated distributions with light yellow and blue indicating observed distributions.

Immunoediting at the DNA level selectively occurs in immunocompetent patients

To control for non-neoantigen-based immune escape mechanisms, we evaluated the tumors for loss of heterozygosity of HLA alleles and mutations in *B2M* (a subunit of the MHC class I complex, loss of which has

been previously implicated in immune evasion).[41] HLA alleles were identified in both the tumor and normal samples to assess for loss of heterozygosity of HLA alleles within the tumor as previously reported.[11] 6/32 immunocompetent and 3/10 immunosuppressed patients had homozygosity of at least one HLA allele unique to the tumor sample, suggesting loss of heterozygosity. There was no significant difference in the homozygosity of HLA alleles unique to the tumor between immunocompetent and immunosuppressed patients ($p = 0.66$, Fisher's exact test, Supplementary Table 3). Additionally, no missense, splice-site, or nonsense mutations were identified in *B2M*. Three tumors had mutations in *B2M*, one in the 3' untranslated region, one in an intronic region, and one on the 5' flank; however, expression of *B2M* at an mRNA level was maintained in these tumors. Therefore, all patients were considered subject to neoantigen-based immunoediting, but prediction of binding neoantigens was restricted to HLA alleles maintained in the tumor since neoantigens specific to lost HLA alleles should not be subject to the same immune-mediated selective pressure.

To explore whether the distribution of clonal and subclonal mutations was influenced by pressure from the immune system, we compared the predicted neoantigen:MHC class I binding since MHC binding has consistently been identified as a key characteristic of neoantigen immunogenicity.[9, 15, 17, 19, 20, 42] One possibility is that, in an immunocompetent setting, MHC class I binding neoantigens would be recognized and eliminated from the population, leading to a relatively lower proportion of binding neoantigens in immunocompetent patients. Since each patient has a different set of HLA alleles and some individuals have fewer alleles due to the homozygosity of an HLA allele, cross-patient comparisons may be influenced by many factors. Therefore, to set a baseline for the proportion of neoantigens expected to bind the set of patient-specific HLA alleles in the absence of immune-mediated selection, 100 *in silico* mutational profiles were generated for each patient. *In silico* mutational profiles were generated using the trinucleotide-aware mutation spectrum and the inferred mutability of each gene based on the density of synonymous mutations (Supplementary Figure 4). The proportion of binding neoantigens was calculated for each set of *in silico* mutations and averaged across 100 iterations. Subsequently, the observed proportion of binding neoantigens was adjusted by the expected proportion of binding neoantigens from the *in silico* mutations. There were no obvious differences in the proportion of binding neoantigens observed between the overall *in silico* and true mutations for either immunocompetent or immunosuppressed patients, as demonstrated by a ratio of one across both groups (Figure 3A-B). Therefore, we concluded that the proportion of binding neoantigens, including both clonal and subclonal neoantigens, was not different than expected given the mutational landscape.

Alternatively, the strength of the immune response may be a function of the frequency of the neoantigen in the population. To examine this research hypothesis, we compared the proportion of binding neoantigens between clonal and subclonal neoantigens. We found that clonal neoantigens in cSCC from immunocompetent patients had a significantly lower proportion of neoantigens predicted to bind MHC class I compared to subclonal neoantigens from the same patient (Figure 3A, Supplementary Table 4). No difference in the fraction of binding neoantigens was observed in the clonal and subclonal neoantigens in immunosuppressed patients (Figure 3B), supporting that the observed differences are attributable to immune pressure. As an additional negative control, we calculated the binding for all patient neoantigens to an equal number of random, non-patient HLA alleles. To account for the overlap in peptides that bind multiple alleles, only neoantigens that bind uniquely to non-patient HLA alleles were considered. There was no apparent difference in the proportion of the clonal and subclonal neoantigen populations predicted to bind to only non-patient HLA alleles in either immunocompetent or immunosuppressed patients (Figure 3C). Together, these data demonstrate a relative decrease in binding neoantigens among clonal neoantigens compared to subclonal neoantigens that only occur in immunocompetent patients, supporting the role of the immune system in shaping the neoantigen profile in the tumor.

One potential mechanism for the decrease in clonal binding neoantigens would be the loss of clonal neoantigens through copy number alterations. To test whether this was a likely mechanism, we tested the proportion of binding neoantigens in regions with and without copy number alterations and found no significant difference in either immunocompetent or immunosuppressed patients (Figure 3D). Thus, immunogenic neoantigens are prevented from becoming clonal during tumorigenesis or eliminated prior to tumorigenesis, rather than first becoming clonal and later being removed by loss of heterozygosity or downregulated by duplication of the non-mutated alleles.

We then sought to determine whether the observed changes in the clonal binding ratio are likely to impact the overall tumor recognition. Therefore, we compared the maximum immunogenicity of clonal and subclonal neoantigens in the cSCC from immunocompetent and immunosuppressed patients. We calculated the immunogenicity of each neoantigen using the NeoScore, which combines characteristics of the neoantigen:MHC class I binding interaction with mRNA expression into an overall immunogenicity score.[15] We then compared the NeoScore of the most immunogenic neoantigen (maximum NeoScore) between clonal and subclonal neoantigens in immunocompetent and immunosuppressed patients, because a high maximum NeoScore has been previously associated with improved response to immune checkpoint blockade.[15] Subclonal neoantigens had a higher maximum NeoScore compared to clonal neoantigens in immunocompetent patients, whereas the clonal neoantigens had a higher maximum NeoScore compared to subclonal neoantigens in immunosuppressed patients (Figure 3E). Differences in the maximum NeoScore may be due to the prevalence of clonal and subclonal neoantigens in immunocompetent vs. immunosuppressed patients, as the larger subclonal population in immunocompetent patients increases the chances of finding the most immunogenic neoantigen in the subclonal population. Despite this caveat, these data demonstrate that the decrease in binding neoantigens in the clonal population decreases the immunogenicity of the clonal neoantigens in cSCC from immunocompetent patients whereas in cSCC from immunosuppressed patients the immunogenicity is retained in the clonal neoantigens.

In addition to changes in the clonal profile of the tumor, previous work has suggested that neoantigens can accumulate in low-expressing regions of the genome as a mechanism of immune escape, either through downregulation of genes with neoantigens or preferential elimination of cells with high expression of neoantigens.[11, 12, 43, 44] Therefore, we modified a prior approach[12] to test whether neoantigens were enriched at low expression compared to what would be expected from the *in silico* mutations. Expression values were numerically ranked from low to high and enrichment scores were calculated as the difference in the average sum of the ranks for binding and non-binding neoantigens individually, where a higher value indicates more enrichment of binding neoantigens at low expression. Consistent with previous work,[12] positive enrichment scores were observed across the majority of the tumors (Figure 3F). However, contrary to what might be expected if the enrichment of binding neoantigens at low expression was immune-mediated, the observed enrichment was less than expected given the mutational landscape alone (Figure 3F). This observation is consistent with a prior report that the trinucleotide mutational context derived from UV exposure biases towards more binding mutations in lowly expressed genes.[45, 46] Additionally, there was no significant difference in the enrichment scores for immunocompetent and immunosuppressed patients, providing further evidence that the enrichment of binding neoantigens at low expression is not attributable to immune-mediated selection. Overall, we did not detect a signature of RNA-level immunoediting in cSCC from immunocompetent patients.

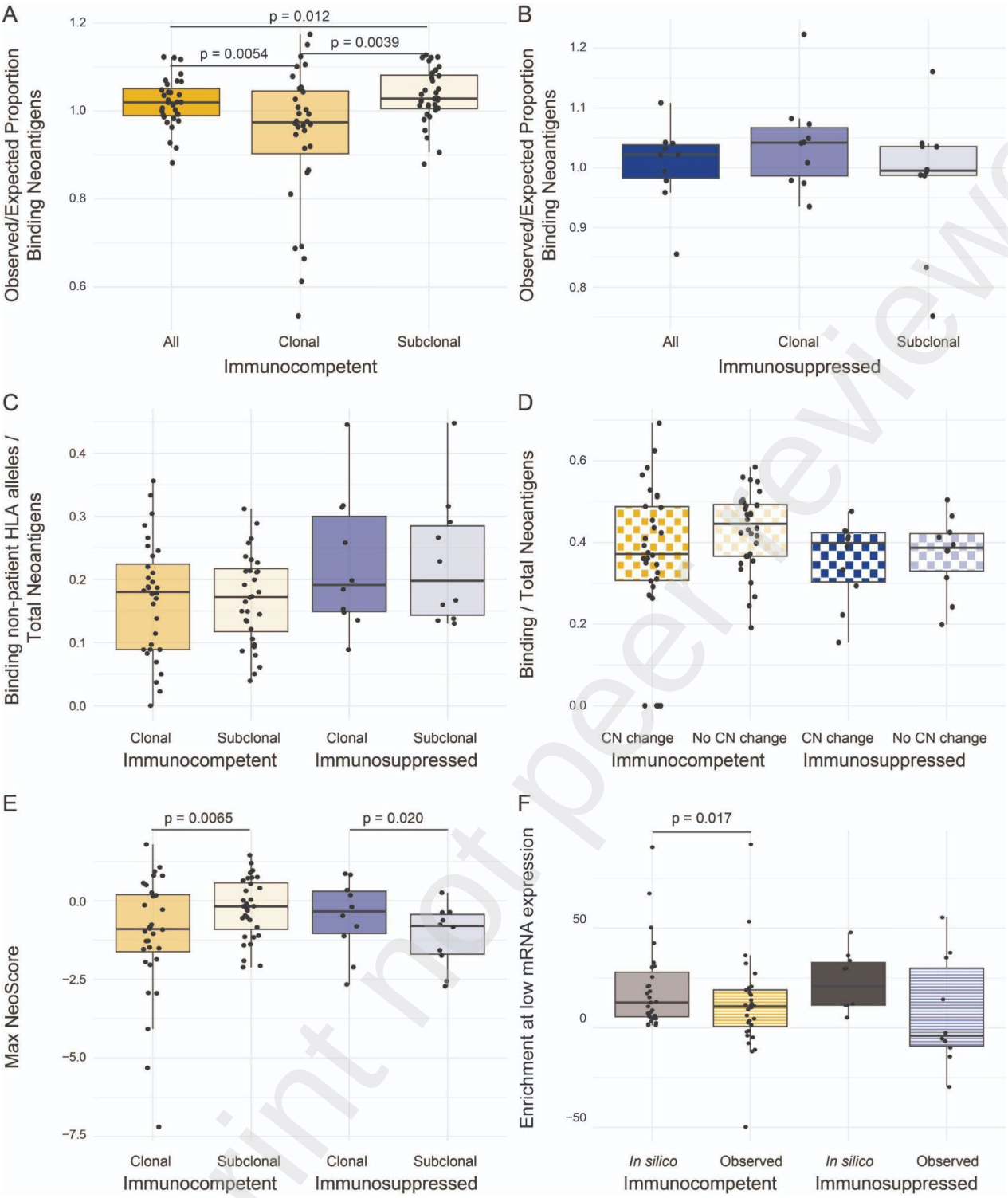
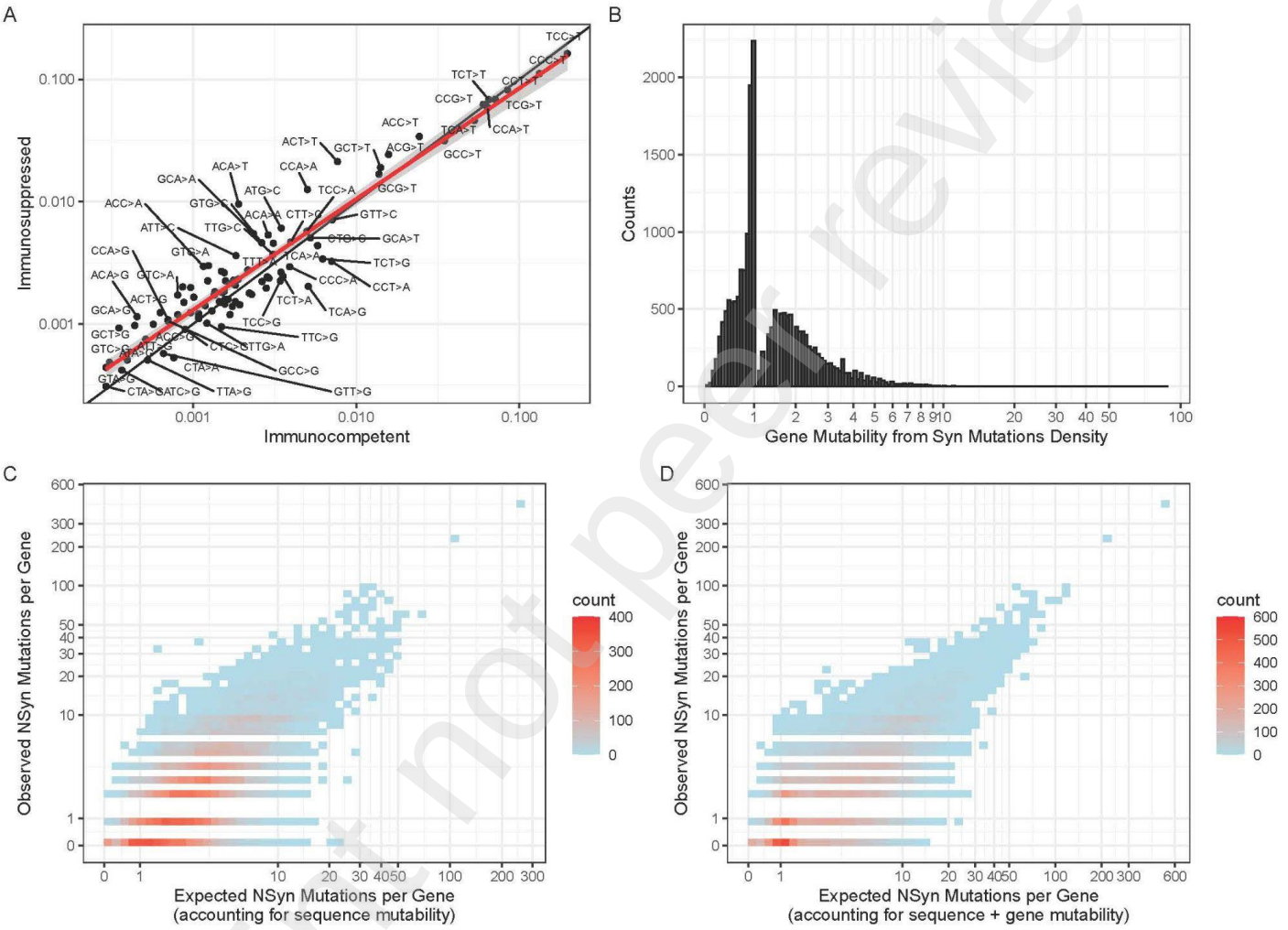


Figure 3: Immunoediting at the DNA level in cSCC selectively occurs in immunocompetent patients. (A-B) Ratio of the observed proportion of binding neoantigens (dissociation constant < 500 nM) over the expected proportion of binding neoantigens, calculated as the average proportion of binding neoantigens from 100 iterations of *in silico* mutations. The ratio is shown across all mutations, clonal mutations (Cancer Cell Fraction, CCF > 0.75), and subclonal mutations (CCF < 0.75). (A) Tumors from immunocompetent patients (n=32, tumors with WES and RNA-Seq). (B) Tumors from immunosuppressed patients (n=10, tumors with WES and RNA-Seq). (C) Comparison of the ratio of binding to total neoantigens with predicted binding to random, non-patient HLA alleles, stratified by immune status. (D) The proportion of binding neoantigens in regions with copy number alterations compared to regions without copy number alterations, stratified by immune status. (E)

Comparison of the maximum NeoScore (linear combination of neoantigen:MHC dissociation constant, neoantigen:MHC binding stability, and mRNA expression) in clonal and subclonal neoantigens, segregated by immune status. (F) Comparison of the enrichment of neoantigens at low mRNA expression for neoantigens from *in silico* mutations compared to the enrichment of neoantigens at low mRNA expression for neoantigens from true patient mutations. Higher enrichment scores indicate that binding neoantigens are enriched at low expression. For all boxplots, the bold line indicates the median and the upper and lower limits of the boxes indicate the 75th and 25th percentiles, respectively. The lower and upper whiskers indicate the minimum and maximum after excluding outliers. Dots outside of the box and whiskers indicate outliers. The paired signed rank sum test statistic was used for all comparisons in panels A-F. No significant differences were observed for panels B, C, and F.



Supplementary Figure 4: Generation of *in silico* mutations. (A) Relationship between the normalized mutation spectrum in immunosuppressed and immunocompetent patients, (B) histogram with the inferred gene mutability per gene inferred from synonymous (Syn) mutations, (C) relationship between the observed and expected number of nonsynonymous (NSyn) mutations considering only sequence mutability, and (D) relationship between the observed and expected number of nonsynonymous mutations considering both sequence mutability and gene mutability. In C and D, to avoid overplotting, individual points were binned into rectangles and the density of the points was mapped to the fill color of the rectangles.

Increased CD8+ T cell and T regulatory cell infiltration in cSCC from immunocompetent patients. Next, we compared the immune cell populations within the tumor microenvironment of cSCC from immunocompetent and immunosuppressed patients. Unsupervised hierarchical clustering of deconvolution

data from quanTIseq identified high and low infiltrate groups that were visually separated by CD8+ T cells and T regulatory (Treg) cells (Figure 4A). The high infiltrate group was predominated by immunocompetent patients, which was reflected by an overall increase in CD8+ T cell and Treg cell infiltration in cSCC from immunocompetent compared to immunosuppressed patients (Figure 4B-C). As confirmation, we performed immune cell deconvolution with two additional, independent software (Supplementary Figure 5A-B). The high infiltration group from quanTIseq was used as an annotation across all three heatmaps, demonstrating a high degree of concordance in the samples identified as having high immune infiltrate. An increase in CD8+ T cells and Treg cells was observed in the additional deconvolution approaches (Supplementary Figure 5C-E, note MCP counter does not include a prediction of Treg cells). Consistent with higher T cell infiltration, the cytotoxicity score calculated by MCP counter was significantly increased in immunocompetent compared to immunosuppressed patients (Supplementary Figure 5F), reflecting increased CD8+ T cell activity. To further characterize the immune microenvironment of cSCC from immunocompetent and immunosuppressed patients, we compared the expression of immune checkpoint molecules *CD274* (encodes PD-L1)*, PDCD1, CTLA4,* and *LAG3*, and markers of cytolytic activity *GZMB* and *PRF1* in cSCC from immunocompetent and immunosuppressed patients (Figure 4D-I, Supplementary Table 5). We found that consistent with increased CD8+ T cell and Treg cell infiltration, cSCC from immunocompetent patients had increased expression of markers of cytotoxicity and exhaustion. Overall, the increase in T cell infiltration and markers of cytotoxicity and exhaustion support that the cSCC from immunocompetent patients are under increased surveillance.

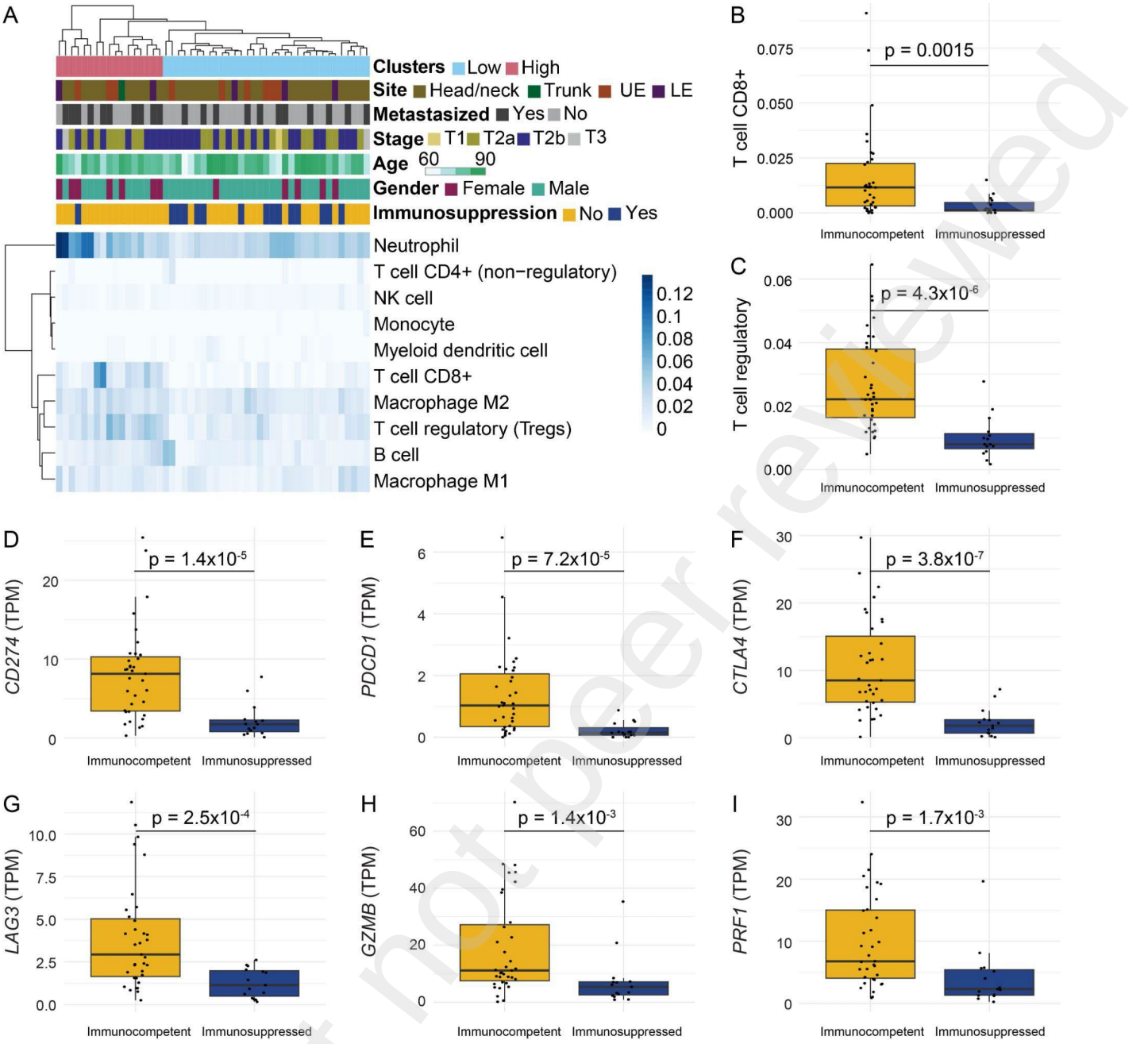
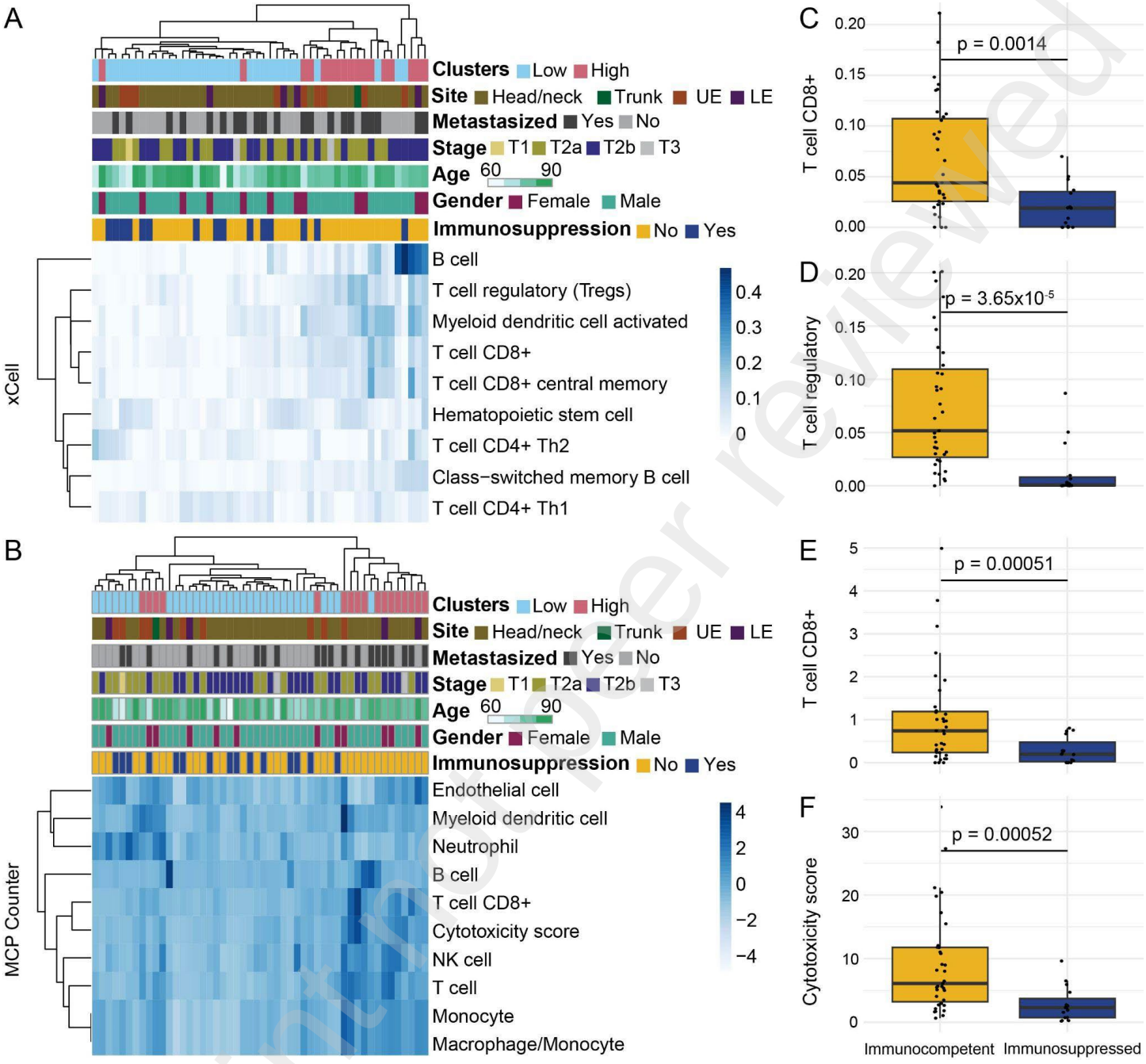


Figure 4: Increased CD8+ T cell and T regulatory cell infiltration in cSCC from immunocompetent patients. (A-C) Immune cell deconvolution of cSCC from 35 immunocompetent and 15 immunosuppressed patients using quanTiseq. (A) Heatmap illustrating the frequencies of immune cell types designated as high vs. low immune cell infiltration by hierarchical clustering. Heatmap is annotated with immune status, sex, age, stage, metastatic status, and skin site of each sample. For skin site, head/neck includes ear, cheek, forehead, jaw, lip, temple, scalp, and periocular; trunk includes back; upper extremity (UE) includes arm, forearm, and hand; and lower extremity (LE) includes leg, lower leg, and foot. (B-C) Boxplots comparing the expression of (B) CD8+ T cells and (C) T regulatory (Treg) cells between immunocompetent and immunosuppressed patients. from quanTIseq. (D-I) Boxplots comparing the expression of immune checkpoint molecules and markers of cytolytic activity in cSCC from immunocompetent and immunosuppressed patients. (D) *CD274* (encodes PD-L1), (E) *PDCD1*, (F) *CTLA4*, (G) *LAG3*, (H) *GZMB*, (I) *PRF1*. For all boxplots, the bold line indicates the median and the upper and lower limits of the boxes indicate the 75th and 25th percentiles, respectively. The lower and upper whiskers indicate the minimum and maximum after excluding outliers. Dots outside of the box and whiskers indicate outliers. A Wilcoxon rank sum test statistic was used to examine scale differences in panels B-I.



Supplementary Figure 5: Additional support for increased CD8+ T cell and T regulatory cell infiltration in cSCC from immunocompetent patients. Immune cell deconvolution of cSCC from 35 immunocompetent and 15 immunosuppressed patients. Heatmaps illustrating the frequencies of immune cell types from (A) xCell (cell types were included if they had a score of at least 0.05 in at least 10 samples) and (B) MCP Counter. Clusters are designated as high vs. low immune cell infiltration using the clusters derived from quanTIseq deconvolution in Figure 4. Heatmaps are annotated with immune status, sex, age, stage, metastatic status, and skin site of each sample. For skin site, head/neck includes ear, cheek, forehead, jaw, lip, temple, scalp, and periocular; trunk includes back; upper extremity (UE) includes arm, forearm, and hand; and lower extremity (LE) includes leg, lower leg, and foot. (C-F) Boxplots comparing the proportion of specific cell populations between immunocompetent and immunosuppressed patients. (C) CD8+ T cells from xCell, (D) Treg cells from xCell, (E) CD8+ T cells from MCP Counter, (F) Cytotoxicity scores from MCP Counter. For all boxplots, the bold line indicates the median and the upper and lower limits of the boxes indicate the 75th and

25th percentiles, respectively. The lower and upper whiskers indicate the minimum and maximum after excluding outliers. Dots outside of the box and whiskers indicate outliers. A Wilcoxon rank sum test statistic was used to examine scale differences in panels C-F.

Predominant increase in exhausted CD8+ T cells in cSCC from immunocompetent patients

To further examine differences in immune infiltrate in immunocompetent and immunosuppressed patients, a publicly available single-cell RNA sequencing (scRNA-Seq) and T cell receptor sequencing (TCRseq) dataset[47] was obtained for tumor-infiltrating CD8+ T cells from cSCC from immunocompetent and immunosuppressed patients. Confirming the findings from bulk RNA sequencing, cSCC from immunocompetent patients had greater CD8+ T cell infiltration (Figure 5A) and expansion of CD8+ T cell clones (Figure 5B) compared to cSCC from immunosuppressed patients. We performed unsupervised clustering followed by annotation of the clusters by comparing them to an atlas of tumor-infiltrating T cells[48] (Figure 5C-F). As further confirmation of the annotation of the CD8+ T cell clusters, we performed pseudotime analysis and demonstrated a progression from naïve T cells, to central and effector memory T cells, and then to exhausted phenotypes, consistent with the known trajectory of CD8+ T cells (Figure 5G-H). A very low number of mucosal-associated invariant CD8+ T cells (MAIT) cells were identified, and they were annotated at a similar pseudotime as effector memory cells, despite not being a true part of the progression of CD8+ T cells. Past studies have noted that MAIT cells frequently cluster with effector memory cells due to similarities in their transcriptional profile.[49] When comparing the phenotypes of the CD8+ T cells between immunocompetent and immunosuppressed patients, there was no difference in the overall proportion of CD8+ T cells in each phenotype (Figure 5C-D). However, when we restricted CD8+ T cells from expanded clones, which are anticipated to be enriched for tumor-specific T cells, we demonstrated a significant increase in the exhausted CD8+ T cell phenotype in immunocompetent compared to immunosuppressed patients (Figure 5E-F). The predominance of tumor-infiltrating CD8+ T cells with an exhausted phenotype in immunocompetent patients is consistent with the persistence of binding neoantigens in the subclonal population in the tumor.

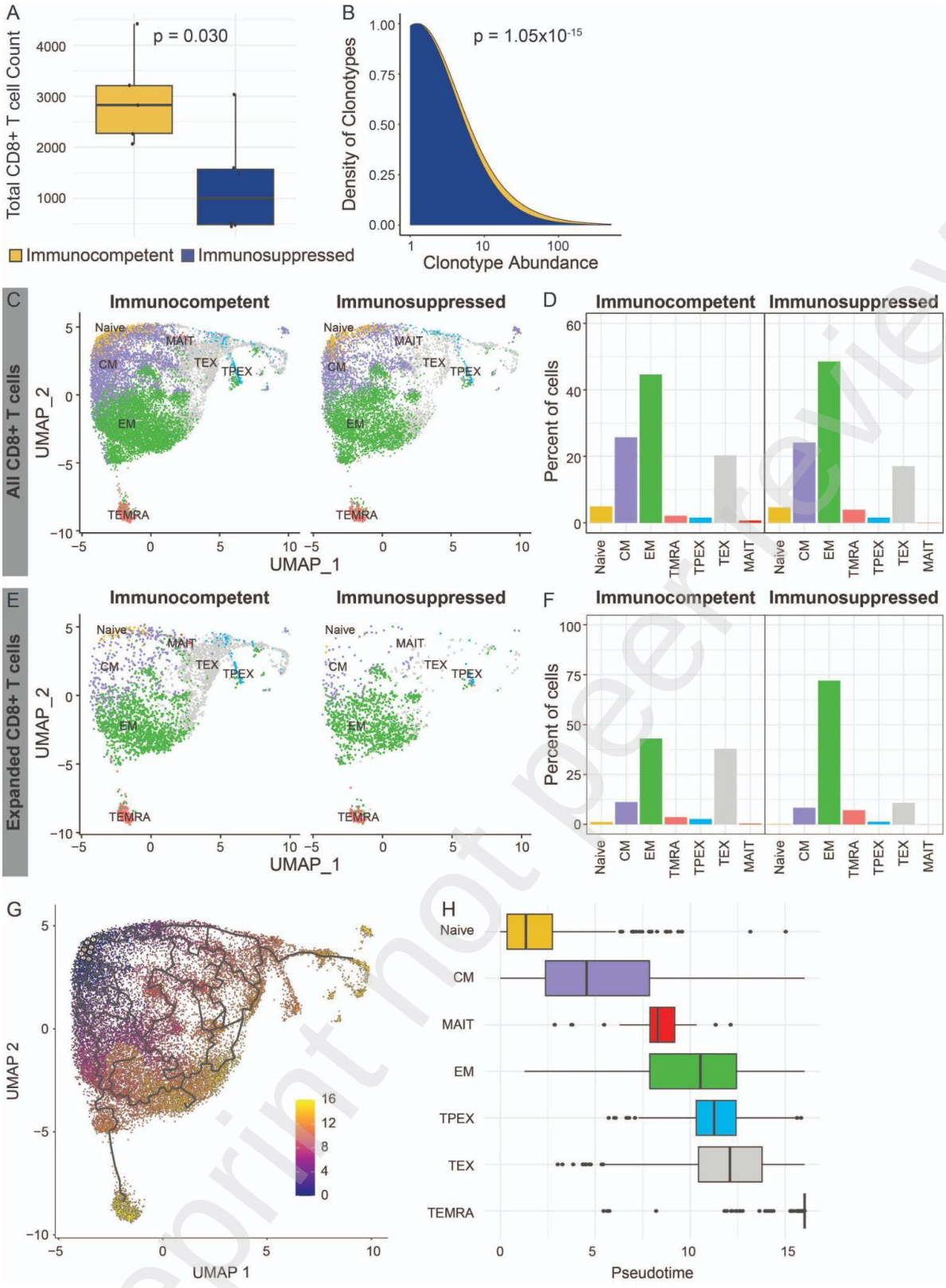


Figure 5: Predominant increase in exhausted CD8+ T cells in cSCC from immunocompetent patients. scRNA-Seq and TCRseq data from 5 cSCC tumors from 5 immunocompetent patients, and 6 tumors from 5 immunocompetent patients. (A) Quantification of the total CD8+ T cell infiltrates in immunocompetent and immunosuppressed patients. (B) The density of clonotypes at each clonotype abundance was compared using a Wilcoxon rank sum test. (C) UMAP projection of all CD8+ T cells in immunocompetent and immunosuppressed patients. Clusters are annotated based on a standard atlas of tumor-infiltrating

lymphocytes. Naïve; naïve CD8+ T cells, MAIT; mucosal-associated invariant CD8+ T cells, CM; central memory CD8+ T cells, EM; effector memory CD8+ T cells, TEMRA; terminally differentiated effector memory CD8+ T cells, TEX; exhausted CD8+ T cells; TPEX; progenitor exhausted CD8+ T cells. (D) Barplot quantifying the percentage of CD8+ T cells belonging to each phenotype. (E) UMAP projection restricted to CD8+ T cells from expanded clones (>1%). (F) Bar graph quantifying the percentage of CD8+ T cells from expanded clones belonging to each phenotype. (G-H) Pseudotime analysis of CD8+ T cell clusters. (A) UMAP projection colored by pseudotime. (B) Boxplot of clusters, ordered by pseudotime. Naïve; naïve CD8+ T cells, CM; central memory CD8+ T cells, MAIT; mucosal-associated invariant CD8+ T cells, EM; effector memory CD8+ T cells, TPEX; progenitor exhausted CD8+ T cells, TEX; exhausted CD8+ T cells, TEMRA; terminally differentiated effector memory CD8+ T cells. For all boxplots, the bold line indicates the median and the upper and lower limits of the boxes indicate the 75th and 25th percentiles, respectively. The lower and upper whiskers indicate the minimum and maximum after excluding outliers. Dots outside of the box and whiskers indicate outliers.

Neoantigens with high differential agretopic index enriched in subclonal neoantigens

Given the evidence of immunoediting in cSCC from immunocompetent patients, the comparison of clonal and subclonal neoantigens in immunocompetent and immunosuppressed patients was applied as an unbiased approach to identifying neoantigen characteristics associated with an immune response. First, the neoantigens from all patients were grouped and annotated with a set of characteristics involved with antigen processing and presentation and foreignness from self (Supplementary Table 4). For processing and presentation, the mRNA level expression of the gene encoding the peptide, the proteasomal cleavage potential of the peptide, and the TAP transport potential of the peptide were considered.[50] For binding, the dissociation constant,[42, 51] and stability[52] of the neoantigen:MHC class I complex were considered since those have previously been demonstrated to provide separate information for prioritization of immunogenic neoantigens.[15] Finally, for foreignness from self, peptides can differ either in the binding to the MHC class I molecule or in the interaction with the T cell receptor (TCR). The difference in the binding to the MHC class I molecule was calculated as the differential agretopic index (DAI; ratio of the dissociation constant of the wild-type to the mutant peptide).[25, 26] For the difference in the TCR recognition, the cross-reactivity distance was calculated which quantifies the likelihood that
a T cell would be able to discriminate the mutation from the wild-type peptide based on the position of the amino acid change and the change in the size and hydrophobicity of the new amino acid relative to the original amino acid.[9]

Next, we used k-means clustering to identify discrete clusters of neoantigens with similar sets of characteristics (Figure 6A). The optimal number of clusters was determined as the number that maximized the gap statistic, which compares the intra-cluster variation (sums-of-squares) under several choices of the number of clusters to the sums-of-squares under the null reference distribution (e.g., uniform distribution); the final number of clusters is the number that maximizes the difference from the null reference (Supplementary Figure 6).[53] The fraction of neoantigens attributable to each cluster was compared between the clonal and subclonal neoantigens in immunocompetent and immunosuppressed patients. Cluster 2, which is uniquely defined by a high DAI, was enriched in subclonal neoantigens in cSCC from immunocompetent, but not immunosuppressed, patients (Figure 6B). Of note, low dissociation constants and high stability have previously been applied to prioritize immunogenic neoantigens;[15, 19, 42] however, cluster 7, which was defined by lower dissociation constants and greater predicted binding stability than cluster 2, showed no enrichment in subclonal neoantigens from immunocompetent patients (Figure 6C). Both cluster 2 and 7 were characterized by an average dissociation constant well below the standard 500 nM threshold for binding (127.78 nM C2, 45.74 nM C7), but cluster 2 was characterized by neoantigens with consistently high DAI scores (Figure 6D-E). These findings suggest that, out of the characteristics evaluated, DAI was the predominant characteristic of neoantigens underrepresented in the clonal neoantigen population, suggesting that DAI is associated with an effective T cell response to the neoantigen.

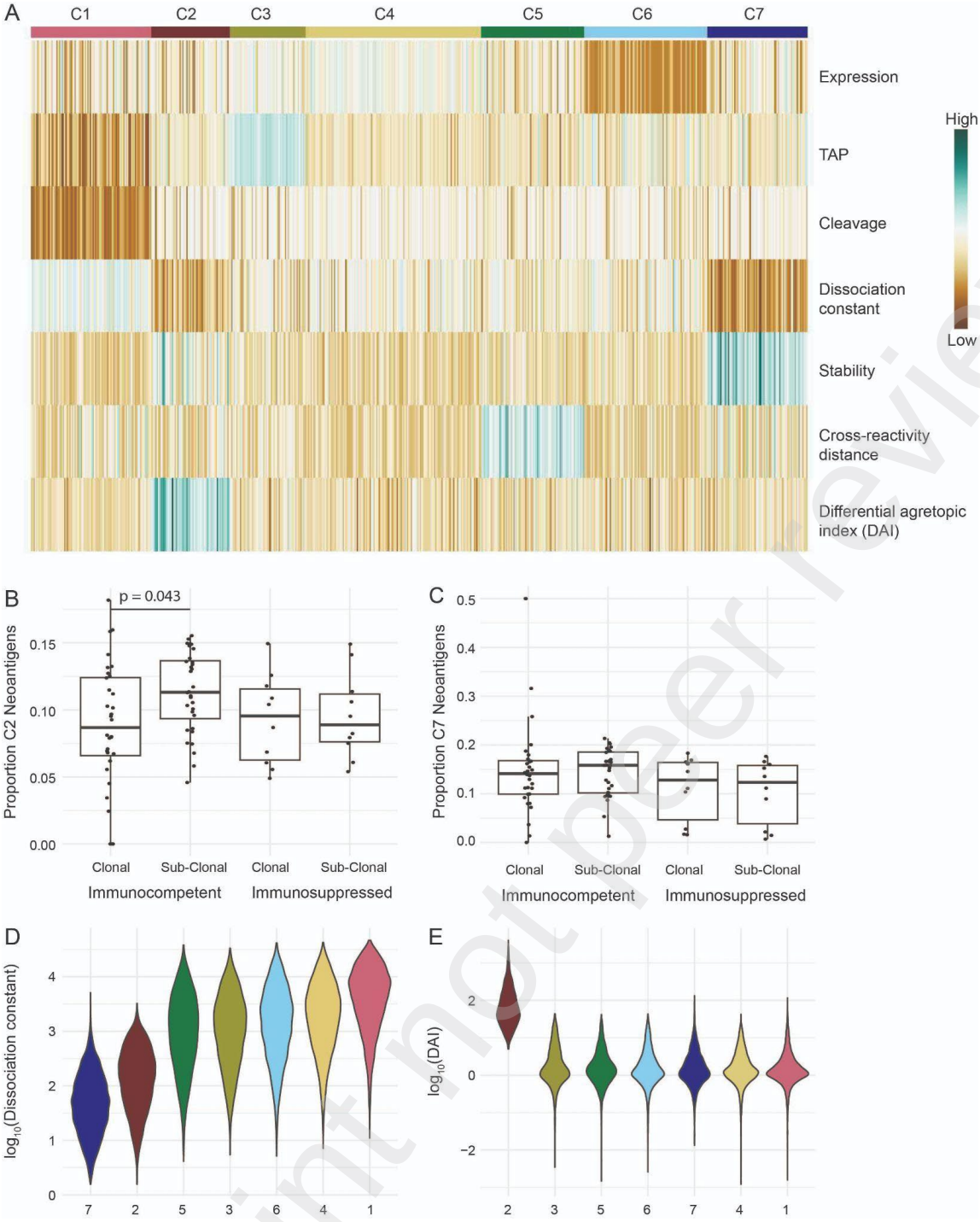
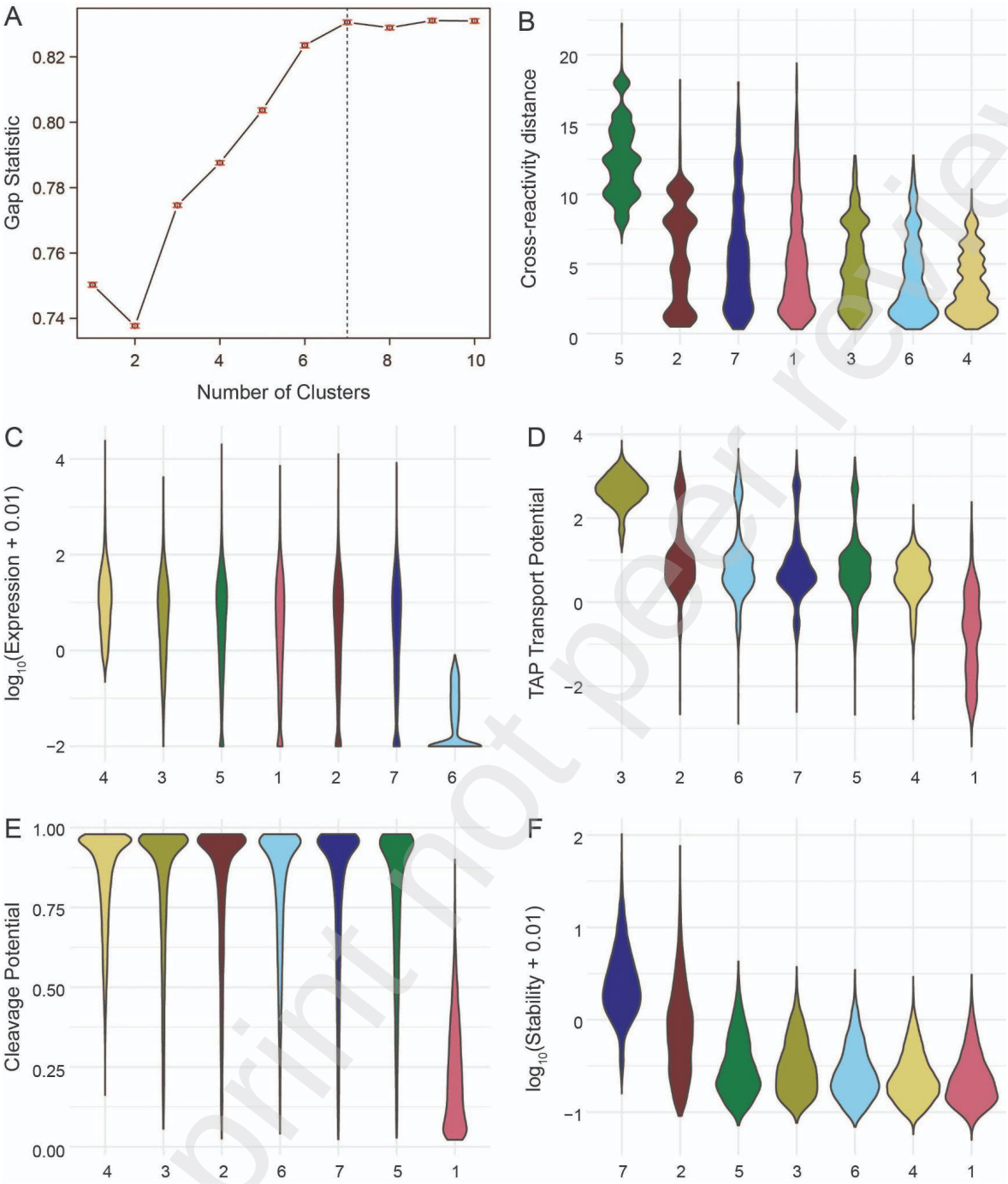


Figure 6: Differential agretopic index identified as a central characteristic in defining immunogenic neoantigens in unbiased assessment of all neoantigens. (A) Clustering analysis across all predicted neoantigen characteristics for all neoantigens from immunocompetent and immunosuppressed patients (n = 32 immunocompetent, n=10 immunosuppressed, tumors with WES and RNA-Seq). (B-C) Boxplots of comparisons of the proportion of selected clusters of neoantigens in the clonal and subclonal populations from immunocompetent patients. The bold line indicates the median and the upper and lower limits of the boxes indicate the 75th and 25th percentiles, respectively. The lower and upper whiskers indicate the minimum and maximum after excluding outliers. Dots outside of the box and whiskers indicate outliers. Differences in scale were tested using the signed rank test statistic for within-patient comparisons. (B) The proportion of cluster 2 neoantigens in the clonal and subclonal populations from immunocompetent and immunosuppressed patients. (C) Proportion of cluster 7 neoantigens in the clonal and subclonal populations from immunocompetent and

immunosuppressed patients. (D-E) Violin plots of select characteristics included in the overall clustering analysis, separated by cluster. The width of each curve demonstrates the density of the data points at that value. (D) Neoantigen:MHC class I dissociation constant, log base 10 transformed. (E) Differential agretopic index calculated as the ratio of the dissociation constant for the normal peptide compared to the neoantigen, log base 10 transformed.



Supplementary Figure 6: Determination of optimal number of clusters with gap statistic. (A) K-means clustering was performed with 1-10 groups and the gap statistic was calculated for each. The optimal number of clusters was selected as the one that maximized the gap statistic. (B-F) Violin plot of each characteristic included in the overall clustering analysis, separated by cluster. The width of each curve demonstrates the density of the data points at that value. (B) Cross-reactivity distance between the mutant and wild-type peptides, (C) mRNA-level expression, log base 10 transformed, (D) TAP transport potential from netCTLpan,

(E) proteasomal cleavage potential from netCTLpan, (F) neoantigen:MHC stability calculated with netMHCstab, log base 10 transformed.

Discussion:

While evidence consistent with immunoediting has been demonstrated in multiple associative studies, a central challenge in the field has been determining to what extent changes in the neoantigen profile are attributable to immune-mediated selection. This is the first immunoediting study, to our knowledge, to use the comparison of tumors from immunocompetent and immunosuppressed patients to evaluate the extent to which changes in the neoantigen profile are attributable to immune activity. This work demonstrated 1) DNA-level immunoediting in cSCC from immunocompetent patients that was not detectable in immunosuppressed patients, 2) exhaustion of expanded CD8+ T cells in tumors from immunocompetent patients, and 3) the importance of DAI in determining which neoantigens are immunoedited in immunocompetent patients.

DNA-level immunoediting was demonstrated first through a decrease in the VAF distribution and a lower number of clonal mutations in immunocompetent compared to immunosuppressed patients. The impact of the immune system on this shift was further supported by the lower proportion of binding neoantigens in the clonal population in immunocompetent patients. These findings are consistent with multiple reports that the clonal mutational burden is more strongly associated with response to immune checkpoint inhibition than the total mutational burden, suggesting that the immune system is more apt to recognize and respond to neoantigens above a certain frequency.[37, 38, 39, 54] Additionally, previous work showed a lower VAF distribution in autochthonous tumors from immunocompetent mice compared to immunocompromised mice.[7] Consistent with a previous report,[46] we demonstrate no overall decrease in the proportion of binding neoantigens compared to what is expected from the *in silico* mutational profiles. However, we demonstrate the importance of considering neoantigens from clonal and subclonal populations. Our data provides strong evidence of the role of the immune system in shaping the clonal mutational profile in primary tumors in immunocompetent patients.

One question arising from these findings is the mechanism by which the immune system shapes the clonal neoantigen burden in immunocompetent patients. Three possible mechanisms are 1) survivorship bias, 2) immunoediting of new subclonal mutations after tumorigenesis, and 3) partial elimination of clonal mutations. The first hypothesis is that there is immunoediting before tumorigenesis, leading to a survivorship bias. Under this hypothesis, if a tumor driver mutation occurs in a cell with many binding neoantigens, all its descendants will be eliminated by the immune system. Therefore, the only tumors that become clinically detectable are those with a lower clonal mutational burden that is less recognizable to the immune system. This hypothesis is consistent with the lower number of clonal mutations per year and the mutational signature of the tumors which is predominantly attributable to UV exposure (and therefore likely to accumulate over time). However, it does not offer a rationale for the increased proportion of observed binding neoantigens in the subclonal population relative to the expected proportion using *in silico* mutations. Secondly, immunoediting could occur on new mutations after tumorigenesis, preventing them from reaching a high frequency in the tumor. In this case, a binding neoantigen that arose after tumorigenesis would be prevented from becoming high-frequency through an active immune response, restricting clonal binding neoantigens. To explain the lower number of clonal mutations per year under this hypothesis, a significant number of clonal (or apparently clonal) mutations in our tumor samples must have occurred after tumorigenesis. Importantly, these mechanisms are not mutually exclusive, therefore, we consider that it is most likely that the immunoediting observed is attributable to a combination of survivorship bias and restriction of new mutations that encode binding neoantigens from becoming clonal. Finally, it is possible that previously clonal mutations might be targeted by the immune system and partially removed, leaving a remnant in the subclonal population. While this would explain the decrease in the proportion of binding neoantigens in the clonal population and the lower number of clonal mutations per year in immunocompetent patients, the change from a clonal mutation to a subclonal mutation would require either back-mutation or changes of copy number of an allele, both of which are unlikely to occur with enough frequency to drive the difference in clonal mutations. Additionally, there was not an increase in the proportion of binding neoantigens in regions with copy number alterations as would be expected if this were a dominant mechanism. Therefore, while it is feasible that this mechanism could contribute, it is unlikely to be a dominant cause of the immunoediting detected through this study.

With regards to immune infiltration, there was a predominant increase in the proportion of expanded CD8+ T cell clones with markers of T cell exhaustion in immunocompetent patients. We hypothesize that T cell

exhaustion could explain why the observed immunoediting is in the accumulation of binding neoantigens in the subclonal population rather than the complete elimination of binding neoantigens. In the setting of chronic neoantigen exposure, the CD8+ T cell response can become exhausted, leading to decreased immune surveillance over time.[55] This is also consistent with the work from Rosenthal et al. demonstrating that the signal of immunoediting was decreased in subclonal neoantigens, particularly in tumors that no longer had high immune infiltration, suggesting that the immune response may have decreased over time.[11] Of note, a few mouse and human studies have demonstrated complete genetic loss of neoantigens, which contrasts with the results presented here.[10, 56] Several possible explanations exist for the discrepancy in the results. In the case of the mouse models, differences have previously been observed in the immune response to transplantable and autochthonous tumors,[57] and the transplantable mouse model may not fully replicate the immune reaction to an autochthonous tumor. In human disease, the loss of neoantigens was established following immunotherapy, which could decrease CD8+ T cell exhaustion and lead to a more complete immune response to a given neoantigen. Additionally, both the mouse and human studies referenced here track the fate of a small number of neoantigens. Elimination of a few neoantigens may not impact the profile of the whole tumor sufficiently to be detected in this study design. Further studies should explore the extent to which complete elimination of immunogenic neoantigens compared to restriction of immunogenic neoantigens from the clonal population is observed in repeated sampling of a single tumor.

Finally, this work applied the comparison of clonal and subclonal neoantigens in immunocompetent and immunosuppressed patients to identify characteristics of neoantigens associated with immunoediting. Through this approach, we find that, of the characteristics tested, the difference in the MHC class I binding affinity between the mutated and unmutated peptides (measured by the statistic DAI) was the most important for identifying neoantigens enriched in the subclonal population in immunocompetent patients. Of note, work to date has been inconsistent on the role of DAI in predicting immunogenic neoantigens in human disease.[15, 17, 19, 58] This inconsistency may be attributable to the training of models on sets of neoantigens that were prioritized *in silico* on the binding affinity of the neoantigen:MHC class I before testing.[14, 15, 16, 17, 18, 19, 20] Prioritization of the neoantigens with the highest binding affinity increases the chances of prioritizing neoantigens where the mutated peptide binds better, which may bias against the identification of DAI as a characteristic that distinguishes immunogenic from non-immunogenic neoantigens. When unbiased testing was performed in a mouse model, DAI was also identified as a key characteristic of immunogenic neoantigens.[25]

Our work supports that immunoediting in human primary tumors occurs through the suppression of clonal mutations at the DNA level and the importance of DAI as a characteristic of the immunoedited neoantigens. These findings suggest that clonality and DAI may be critical characteristics in improving the selection of immunogenic neoantigens for inclusion in personalized neoantigen vaccines.

Acknowledgements:
This work was supported by an American Cancer Society – Simone Charitable Foundation - Discovery Boost Grant, AZ-DBG-23-1151311-01-IBCD, https://doi.org/10.53354/ACS.AZ-DBG-23-1151311-01-IBCD.pc.gr.175465 (K.T.H) and National Institutes of Health grants F30CA281056 (E.S.B.) and R35GM149235 (R.N.G.). This work benefited from support for a related project by the Merit Review Award I01-BX005336 from the United States Department of Veterans Affairs (VA), Biomedical Laboratory Research and Development Service (K.T.H.). The contents do not represent the views of the VA or the United States Government. This material is based upon High-Performance Computing (HPC) resources supported by the University of Arizona TRIF, UITS, and Research, Innovation, and Impact (RII) and maintained by the UArizona Research Technologies department. We thank Sara Willis and Soham Pal for their assistance with the creation of singularity environments on the HPC, effectively parallelizing jobs, and optimizing our use of the available computational time, which was made possible through the University of Arizona Research Technologies Collaborative Support program. We thank Dr. Andrew Ji for his feedback and advice on the analysis of the scRNA-Seq data. We thank Dr. Diego Mallo for his technical advice on cancer cell fraction and heterogeneity. We thank Dr. Seema Plaiser for her technical advice on estimating tumor purity. We thank Anne Macy for her assistance with manuscript editing.

Author contributions:
Conceptualization E.S.B., K.H.B., M.A.W., K.T.H. Methodology E.S.B., D.C., R.N.G., B.J.L., K.H.B., M.A.W., K.T.H. Formal Analysis E.S.B., D.C., T.M., X.L. Investigation E.S.B., D.C. Data Curation A.H. Resources A.R.M., K.T.H. Writing -- Original Draft E.S.B., D.C., K.T.H. Writing -- Reviewing and Editing E.S.B., D.C., R.N.G., B.J.L., M.A.W., K.T.H. Visualization E.S.B., D.C. Supervision K.T.H. Funding acquisition E.S.B., K.T.H.

Declaration of interests:
A.R.M. consulted for Phelecs BV, Kyowa, Eli Lilly, Momenta, and UCB (greater than 24 months ago); Incyte, Soligenix, Clarivate, Argenyx, and Bristol Myers Squibb (less than 12 months ago); and currently consults for Nuvig, Tourmaline Bio, Janssen, Boehringer Ingelheim (without payments to the institution; and Regeneron and Pfizer (with payments to the institution). He had grant support from Sun Pharma, Elorac, Novartis, and Janssen (greater than 24 months ago) and from Kyowa, Miragen, Regeneron, Corbus, Pfizer, Incyte, Eli Lilly, Argenx, Palvella, Abbvie, Priovant, Merck (less than 24 months ago). He has received royalties from Adelphi Values and Clarivate. His current patents include Methods and Materials for assessing and treating cutaneous squamous cell carcinoma (provisional PCT/US2023/078902); Use of Oral Jaki in Lichen Planus (PCT/US2024/020149); Topical Ruxolitinib in Lichen Planus (PCT/US2021/053149, 2023-520085, & 21805700.8); and Methods and Materials for Treating Lichen Planopilaris (Registration Number: 53,103).

Supplemental information
Document S1. Figures S1-S6 (included in main text for primary submission)
Supplementary Table 1. Excel file with all demographic data, related to Figure 1, Supplementary Figure 1, Figure 4, and Supplementary Figure 5.
Supplementary Table 2. Variants identified from all WES data, related to Figure 1, Supplementary Figure 1, Figure 2, and Supplementary Figure 2.
Supplementary Table 3. HLA types related to statements in the text.
Supplementary Table 4. Mutations annotated with neoantigen characteristics, related to Figure 3, Figure 6, and Supplementary Figure 6.
Supplementary Table 5. Expression data for all samples, related to Figure 4 and Supplementary Figure 5.

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Karen Taraszka Hastings (khasting@arizona.edu)

Materials available

This study did not generate new unique reagents.

Data and code availability

This paper analyzes existing data from our group and others. The accession numbers for the datasets are listed in the key resources table. All original code has been deposited in GitHub and is publicly available as of the date of publication. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

cSCC samples

This study analyzed WES and/or RNA-Seq data previously obtained by our collaborative team for a total of 71 patients with cSCC (46 immunocompetent and 25 immunosuppressed). Primary tumors and normal tissue controls were manually macrodissected from serial sections, guided by the annotations of a board-certified dermatologist. Specimens were prepared for sequencing as described.[59] WES was performed with an average coverage of approximately 80x and RNA-Seq with approximately 60 million reads per sample. Immunosuppression was due to either treatment following solid organ transplantation (renal, lung, or heart, n=19), for myasthenia gravis (n=3), or for severe asthma (n=1, WES only). Immunosuppressed patients were treated with azathioprine, tacrolimus, prednisone, sirolimus, mycophenolate mofetil, cyclosporin, or a combination of the aforementioned medications. Two patients considered immunosuppressed in the prior study[59] were excluded because the treatment was local (i.e. ophthalmic cyclosporine for CLL) or the treatment inhibited only the Th2 response (dupilumab for Crohns disease). One patient with only RNA-Seq data was excluded due to an unknown metastatic status. The resulting sample sizes and full demographic data are provided in Supplementary Table 1.

Mutational burden and signature

Somatic variants were identified using the methods described in the Genome_GPA v5.0.3 algorithm (formerly called TREAT).[60] Briefly, read mapping was performed to the GRCh38 genome using bwa-mem (v0.7.10),[61] realignment was performed with GATK (v.3.4-46),[62] and somatic variants were identified with a combination of Mutect2 and Strelka2.[62, 63] Consistent with previous work, only variants identified by both GATK Mutect2 and Strelka2 were kept, to ensure high fidelity calls.[15] Variants were filtered using the standard filters from GATK Mutect2. Mutations were annotated using the Variant Effects Predictor (VEP) from Ensmebl with cache version 109.[64] The *Downstream* plugin was applied for Vep to account for any downstream effects of frameshift variants on the protein sequence. VEP-annotated variant call files were converted to MAF files with MAF tools for efficient downstream mutational analyses.[65] Mutational signatures were deconvoluted with two different approaches, using the decongstructSigs (v.1.9.0)[66] and the sigminer (v.2.3.1)[67] packages from R.

Read count quantification

Processing of the RNA-Seq paired-end reads was performed with the Mayo RNA-Seq bioinformatics pipeline, MAP-RSeq (v.3.1.4).[68] RNA-Seq data was aligned to the GRCh38 reference genome using STAR.[69] Gene-level read count quantification was performed with Subread *geneCounts* function to obtain raw reads (Supplementary Table 5).[70, 71] Read counts were transformed to transcripts per million (TPM) for downstream analyses.

Clonality and heterogeneity

VAFs were obtained from GATK Mutect2 during the identification of somatic variants. To better compare the heterogeneity of the tumors, CCFs were calculated from the VAF using the following equation:

$$CCF = \frac{VAF}{P}\ x\ (CN_v * P + CN_n * (1 - P))$$

Where P is the purity, calculated from the RNA-Seq data with Puree,[72] the CN_v is the copy number of the variant allele calculated with CNVkit, and CN_n is the copy number of the normal allele calculated with CNVkit.[73] Somatic variants with a CCF greater than 1 were set to 1 to more accurately reflect the percent of the tumor containing a given mutation. The number of clonal neoantigens was quantified as the number of neoantigens with a CCF > 0.75, consistent with a prior study.[7] Given the uncertainty involved in defining neoantigens as subclonal or clonal based on a CCF cutoff, we performed additional testing to support that the subclonal and clonal neoantigens are correctly assigned. A one-sided binomial test was performed to assign neoantigens as subclonal only if they had a VAF significantly lower than the purity-adjusted VAF threshold for clonality.

Inference of growth rate and clonal ratio

We lack analytical expectations for the shape of the VAF under different models of tumor growth. Fortunately, we have analytical expectations for a closely related summary statistic well-known in population genetics, the site frequency spectrum (SFS). The SFS is a histogram showing the distribution of allele frequencies among loci within a sample. It summarizes the frequencies of derived alleles across the sequenced haplotypes.[74] Under the standard neutral model (SNM), that is, a model of constant population size with no selection,[75] the SFS decays as 1/x^1, where x is the number of haplotype copies in a sample of size of *n* haplotypes (x ranges from 1, mutations present only in one haplotype, to *n*, mutations present in all haplotypes copies in the sample). In contrast, under exponential growth the SFS decays as 1/x^2.[76, 77] This results in a faster decay of the SFS under population expansion and an excess of low-frequency variants relative to the SNM.

Unlike the SFS, the sample size or number of sequenced haplotypes (or cells/individuals) per locus in a given VAF is not visually readily apparent. To generate a VAF, a large pool or sample of diploid cells (at least 1,667 cells in this study) is used, rather than sequencing a predefined number of individuals/cells as in the construction of an SFS. However, the average sequencing coverage (~80X in this study) is substantially smaller than the pool or sample size. This is equivalent to subsampling with replacement from large SFS to a more modest sample size. The coverage also varies across different sites, adding more uncertainty to the recovered allele frequency distribution. The VAF is therefore a noisy version of the SFS, and it is particularly underpowered for the detection of low-frequency mutations. Moreover, in tumor VAFs, mutations tend to occur only once in a chromosome copy, so the highest achievable allele frequency in the sample is not n but n/2. To circumvent these peculiarities, we simulate the entire sequencing-calling-filtering process used to generate our tumor VAFs under the SNM and a model of an exponentially growing population. To account for intermediate "growth rates", we allow the exponent to vary between 1 and 2 (see below).

Three simulation parameters were held constant across patients:

- Number of sequenced cells present in the sample (*n* = 1,667). The sequencing protocol requires 10 nanograms to 1 microgram of DNA. Since a diploid human cell contains about 6 picograms (pg) of DNA, this is the lower limit of sequenced cells. This lower limit is more than an order of magnitude larger than the average coverage (~80X), making it unlikely, but not impossible, that the same piece of chromosome will be sequenced more than once.
- The number of mutated sites per cell (μ > 10). The exact value of μ is not critical, provided that the total number of mutations in the theoretically large sample of cells (n = 1,667) is sufficiently large to generously populate all SFS frequency bins. Note we use normalized VAF counts as summary statistics rather than absolute VAF counts. This normalization allows our inference to remain independent of the underlying mutation rate and reduces the optimization to just two parameters (see below).

- The number of required alternative reads to call a mutation (r_i = 2). This number is extracted from our mutation calling protocols. This filter removes sequencing errors, but also all true singletons found in SFS and many other low-frequency variants.

Finally, patient-tailored VAF distributions were simulated using patient-specific coverage distributions (with mean coverage across patients ~80X, Supplementary Figure 2H) and estimated RNA purities (Figure 2D). Then 5,151 simulations were performed per patient across a range of values for two parameters:

- The "growth rate" (simulated range [1, 2] in 0.01 steps). Analytical expectations were used to vary the exponent in the SFS decay, which we interpret as a proxy for the "growth rate".
- Fraction of clonal mutations, *C* (simulated range [0, 0.49] in 0.01 steps). The normalized SFS is a mixture of the (1-C) times the original normalized SFS plus C times an SFS with weight only at frequency n/2.

The results of each simulated VAF were stored as the normalized number of mutations observed in binned intervals (ranging from 0 to 1 with a bin size of 0.01), summing to one. This is our vector of summary statistics which is compared to the patient normalized VAF with the same binned intervals. Approximate Bayesian Computation (ABC) was used to approximate the posterior probabilities that estimate the growth rates and fraction of clonal mutations from the normalized VAF simulations. The rejection method used a tolerance of 0.001; all analyses were performed with the *abc* package[78] implemented in R. We subsequently evaluated the mean of the parameters from the accepted simulations.

In silico mutations

We developed a null mutation model that integrates mutation rate variation at two genomic scales: the trinucleotide context (using the 96-type pyrimidine-oriented mutation spectrum) and the gene level. Note this model is not strand-oriented. To capture DNA sequence context-dependent mutability, we aggregated all single nucleotide somatic coding mutations across patients into two, not independent, pan-cancer substitution rate vectors per trinucleotide context—one for synonymous (putatively neutral) mutations in all genes (drivers+passengers) and another for coding mutations in passenger genes. Hereafter, we refer to these as the synonymous and passenger mutation spectra, respectively. We confirmed that the mutation spectra for both immunosuppressed and immunocompetent patients are remarkably similar (Pearson correlation coefficient = 0.99, $p < 0.0005$, Supplementary Figure 4A).

To account for the variation in mutability across genes, we first estimated the expected number of synonymous mutations per gene. This expected number for a given gene *i* ($\xi_{s,i}$) is calculated by multiplying the number of synonymous mutation opportunities per gene ($L_{s,i}$), which is determined by the gene's 3-mer composition and length and by the synonymous mutation spectrum. Next, we calculated the expected synonymous substitution rate per site ($\lambda_{s,i}$) for each gene by dividing $\xi_{s,i}$ by $L_{s,i}$. It is important to note that $\lambda_{s,i}$ varies between genes due to differences in their 3-mer composition for synonymous mutation opportunities. For instance, a gene with more synonymous CpG>TpG opportunities than the average gene will have a higher $\lambda_{s,i}$ on average.

Subsequently, for each gene, we modeled the distribution of $\xi_{s,i}$ using a Poisson distribution (*dpois()* function in R) with the gene's $\lambda_{s,i}$ and $L_{s,i}$ values. We then assessed how the observed number of synonymous mutations in gene *i* ($O_{s,i}$) compares to this distribution. If $O_{s,i}$ matches the mean of the distribution of $\xi_{s,i}$, the gene's mutability ($u_{s,i}$) is one, indicating that the gene mutates as expected given its length and 3-mer composition. However, this is not typically the case, as some genes will mutate more or less than expected due to variations in regional mutation rate covariates, such as replication time or expression levels. To adjust for these factors, we define a corrected version of $\lambda_{s,i}$ which accounts for gene mutability ($\lambda_{s,i}' = \lambda_{s,i} \times u_{s,i}$). We again use a Poisson distribution and maximum likelihood estimation to find the value of $\lambda_{s,i}'$ that best explains $O_{s,i}$. Supplementary Figure 4B displays the distribution of inferred $u_{s,i}$ across genes. To validate the usefulness of including inferred gene mutabilities in our null mutation model, we compared the fit between the expected number of nonsynonymous mutations per gene (calculated using the passenger mutation spectrum and the gene's nonsynonymous mutation opportunities) and the observed number of nonsynonymous mutations per gene across all patients (Supplementary Figure 4C). Although the correlation is strong (R^2 = 65%, $p < 0.0005$), it becomes even stronger (R^2 = 74%, $p < 0.0005$) when we use the inferred gene mutability to calculate the

expected number of nonsynonymous mutations per gene, particularly for long genes (Supplementary Figure 4D). This result suggests that incorporating gene mutability using synonymous mutation density is a reasonable choice.

To generate *in silico* mutations, we begin by computing a large mutation probability vector that spans all coding sequences in human autosomes. This vector is constructed by concatenating substitution rate vectors for each site in a gene, which are determined by the passenger mutation spectrum and the gene's 3-mer composition (for both synonymous and nonsynonymous mutation opportunities). However, simply concatenating these substitution rate vectors does not account for the known significant variability in gene mutability (Supplementary Figure 4B). To address this, we adjust each gene's substitution rate vector by multiplying it by the corresponding gene's mutability vector. The resulting large vector is then normalized to ensure its sum equals one. Finally, we apply a hypergeometric distribution (*rhyper()* function in R), using the observed number of coding mutations in a patient's genome and the large normalized mutation probability vector, to generate 100 replicates per patient. Each replicate contains the same number of coding mutations as observed in the original patient. The null mutation model itself is not patient-specific. What is patient-specific, however, is the mutation burden within coding sequences and the individual's HLA alleles.

HLA alleles and *B2M* mutations

Patient-specific HLA alleles were identified in both tumor and normal samples using Polysolver (Supplementary Table 3).[79] Variant call files were queried for mutations in *B2M*. A comparison of the HLA alleles identified in tumor and normal samples was used to determine the frequency of loss of heterozygosity within the tumors. HLA alleles identified in the tumor samples were utilized for binding predictions to ensure that the predicted neoantigens could be presented within the tumor cells. To assign random HLA alleles, the total pool of HLA alleles was combined, and each patient was assigned 10 iterations of an equal number of randomly selected HLA alleles (e.g. if the patient had two unique HLA-A alleles, two unique HLA-B alleles, and two unique HLA-C alleles, the patient would be assigned to two random HLA-A alleles, two random HLA-B alleles, and two random HLA-C alleles).

Annotation of binding neoantigens

VEP-annotated mutations were processed into 21mer amino acid sequences using pVAC-Seq tools version 3.0.5.[80] The MHC class I:neoantigen dissociation constant (Kd) was then predicted for every 9mer sequence to each patient allele and non-patient allele with netMHCpan4.0.[42, 51] Dissociation constants were also predicted for every *in silico* mutation from each of the 100 iterations for each patient-specific HLA allele. Subsequently, we calculated the average proportion of binding neoantigens across the 100 iterations of *in silico* mutations and compared this expected proportion to the observed proportion in the true patient mutations. Then, we adjusted the observed proportion of neoantigens (across all mutations, clonal mutations, and subclonal mutations) by the expected proportion from the *in silico* mutations. The adjustment for the proportion of binding neoantigens expected from the *in silico* mutational profiles accounts for patient-specific differences in the number of HLA alleles and their different binding proclivities. As an additional negative control, the proportion of binding neoantigens was compared for neoantigens predicted to bind to only non-patient alleles between clonal and subclonal neoantigens. Next, to compare the immunogenicity of the subclonal and clonal neoantigen populations, MHC binding stability was calculated for every 9mer sequence to each patient allele using netMHCstab1.0.[52] The dissociation constant from the 9mer neoantigen with the best binding for each mutation was used as the score for each mutation. The stability score was assigned as the stability of the peptide bound to the allele with the minimum dissociation constant. The NeoScore was calculated as the linear combination of the dissociation constant, stability, and mRNA expression level, as described.[15]

Enrichment of neoantigens at low expression

Enrichment of predicted binding neoantigens at low expression was calculated with a method similar to previous methods.[12] All neoantigens were ranked in order from lowest to highest expression. Then each neoantigen was assigned a rank, with neoantigens at the same expression level assigned to the same rank.

The ranks for all binding neoantigens (to patient alleles and random, non-patient alleles) and non-binding neoantigens were summed and divided by the total number of neoantigens in that category. Then, the result for all binding neoantigens was subtracted from that for the non-binding neoantigens. Overall, the enrichment score was calculated as:

$$Enrichment = \frac{\sum_{i=1}^{i=n_{nb}} r_i}{n_{nb}} - \frac{\sum_{i=1}^{i=n_b} r_i}{n_b}$$

Where r_i is the rank of a given neoantigen, n_{nb} is the number of non-binding neoantigens, and n_b is the number of binding neoantigens. In this way, lower expression for the binding neoantigens leads to a lower adjusted rank score for the binding neoantigens, and a higher enrichment score when this value is subtracted from the adjusted rank score for the non-binding neoantigens.

Immune cell deconvolution

Immune cell deconvolution was performed across all samples using quanTIseq,[81] xCell,[82] and MCP counter,[83] applied using the immunedeconv (v.2.1.0) package in R.[84] For clear visualization, MCP counter was scaled for heatmap visualization, while the other two packages had results easily visualizable without adjustment. To compare across methods, the high-expressing groups were assigned a cluster in the quanTIseq results and the cluster annotation was applied to the other two methods.

Expression analysis

To quantify the expression of markers of cytotoxicity and exhaustion in the bulk RNA-Seq data, we compared the TPM expression level of a range of markers, *CD274*, *PDCD1*, *CTLA4*, *LAG3*, *GZMB*, and *PRF1* across all samples.

Single-cell RNA-Seq analysis

Publicly available scRNA-Seq and TCRseq data from tumor-infiltrating CD8+ T cells was obtained from Frazette *et al.*[47] for 5 tumors from 5 immunocompetent patients, and 6 tumors from 5 immunocompetent patients. Gene-cell matrices and assembled V(D)J data were processed in the original publication using standard methods from Cell Ranger (https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome). We performed standard quality control filters to remove cells with a unique feature count of less than 200 or with > 10% mitochondrial counts. TCR sequencing data was integrated using scRepertoire (v1.12.0) in R.[85] The total number of infiltrating CD8+ T cells was compared after filtering. The frequency of T cells from clonotypes at any given abundance was calculated and visualized using the *abundanceContig* function from scRepertoire. Next, normalization, feature selection, and integration were performed using the Seurat (v5.1.0) package in R.[86, 87] Scaling was performed to prepare the data for dimensional reduction and then the data was visualized with Uniform Manifold Approximation and Projection (UMAP) using the Seurat package. We then used ProjectTILs (v3.3.1)[48] to annotate the UMAP with known phenotypes of CD8+ tumor-infiltrating lymphocytes from the “human CD8+ TIL atlas” which was derived from 11,021 single-cell transcriptomes across 20 samples and 7 tumor types (https://github.com/ncborcherding/utility). The proportion of T cells attributable to each phenotype was plotted across all T cells and T cells with a clonal frequency greater than 1%. Finally, pseudotime analysis was performed with standard approaches from Monocle3.[88, 89]

Clustering analysis

A set of characteristics was calculated for each neoantigen from every patient (Supplementary Table 4). mRNA expression of each neoantigen was calculated at the gene level and quantified as TPM. TAP transport and proteasomal cleavage potential were predicted with NetCTLpan1.1.[50] Neoantigen:MHC class I dissociation constants were calculated with NetMHCpan4.0, and neoantigen:MHC class I binding stability was calculated

with NetMHCstab1.0. DAI was calculated as the ratio of the dissociation constant for the wild-type peptide compared to the neoantigen as described.[15, 17] Cross-reactivity was calculated using the methods and code previously described.[9] mRNA expression, dissociation constant, MHC class I:neoantigen binding stability, and DAI were all normalized on a log 10 scale, and all characteristics were scaled and centered on zero. Clustering was performed with kmeans clustering. The optimal number of clusters was determined by calculating the gap statistic at a range of clusters from 1 to 10 using the clusGap package from R (Supplementary Figure 6A).[53]

QUANTIFICATION AND STATISTICAL ANALYSES

Statistical analyses were performed in R. Statistical tests are indicated in all figure legends or the text. Non-parametric tests were used when parametric tests were questionable as well as when the primary comparison was around comparing density functions between groups or simulations/hypothesized distributions. The sample size for each comparison is noted in the text or in the figure legends, where relevant. The threshold for significance was set to $p < 0.05$.

References:

1. Dunn, G.P., Old, L.J. & Schreiber, R.D. The three Es of cancer immunoediting. *Annu Rev Immunol* **22**, 329-360 (2004).

2. Smyth, M.J. *et al.* Perforin-mediated cytotoxicity is critical for surveillance of spontaneous lymphoma. *J Exp Med* **192**, 755-760 (2000).

3. Street, S.E., Cretney, E. & Smyth, M.J. Perforin and interferon-gamma activities independently control tumor initiation, growth, and metastasis. *Blood* **97**, 192-197 (2001).

4. Street, S.E., Trapani, J.A., MacGregor, D. & Smyth, M.J. Suppression of lymphoma and epithelial malignancies effected by interferon gamma. *J Exp Med* **196**, 129-134 (2002).

5. van den Broek, M.E. *et al.* Decreased tumor surveillance in perforin-deficient mice. *J Exp Med* **184**, 1781-1790 (1996).

6. Shankaran, V. *et al.* IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410**, 1107-1111 (2001).

7. Westcott, P.M.K. *et al.* Mismatch repair deficiency is not sufficient to elicit tumor immunogenicity. *Nat Genet* **55**, 1686-1695 (2023).

8. Borden, E.S. *et al.* Neoantigen Fitness Model Predicts Lower Immune Recognition of Cutaneous Squamous Cell Carcinomas Than Actinic Keratoses. *Front Immunol* **10**, 2799 (2019).

9. Luksza, M. *et al.* Neoantigen quality predicts immunoediting in survivors of pancreatic cancer. *Nature* **606**, 389-395 (2022).

10. Anagnostou, V. *et al.* Evolution of Neoantigen Landscape during Immune Checkpoint Blockade in Non-Small Cell Lung Cancer. *Cancer Discov* **7**, 264-276 (2017).

11. Rosenthal, R. *et al.* Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479-485 (2019).

12. Wu, T. *et al.* Quantification of Neoantigen-Mediated Immunoediting in Cancer Evolution. *Cancer Res* **82**, 2226-2238 (2022).

13. Zapata, L. *et al.* Immune selection determines tumor antigenicity and influences response to checkpoint inhibitors. *Nat Genet* **55**, 451-460 (2023).

14. Bjerregaard, A.M., Nielsen, M., Hadrup, S.R., Szallasi, Z. & Eklund, A.C. MuPeXI: prediction of neo-epitopes from tumor sequencing data. *Cancer Immunol Immunother* **66**, 1123-1130 (2017).

15. Borden, E.S. *et al.* NeoScore Integrates Characteristics of the Neoantigen:MHC Class I Interaction and Expression to Accurately Prioritize Immunogenic Neoantigens. *J Immunol* **208**, 1813-1827 (2022).

16. Gartner, J.J. *et al.* A machine learning model for ranking candidate HLA class I neoantigens based on known neoepitopes from multiple human tumor types. *Nat Cancer* **2**, 563-574 (2021).

17. Luksza, M. *et al.* A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* **551**, 517-520 (2017).

18. Muller, M. *et al.* Machine learning methods and harmonized datasets improve immunogenic neoantigen prediction. *Immunity* **56**, 2650-2663 e2656 (2023).

19. Wells, D.K. *et al.* Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. *Cell* **183**, 818-834 e813 (2020).

20. Zhou, C. *et al.* pTuneos: prioritizing tumor neoantigens from next-generation sequencing data. *Genome Med* **11**, 67 (2019).

21. Carreno, B.M. *et al.* Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348**, 803-808 (2015).

22. Cohen, C.J. *et al.* Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. *J Clin Invest* **125**, 3981-3991 (2015).

23. Ott, P.A. *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217-221 (2017).

24. Stronen, E. *et al.* Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* **352**, 1337-1341 (2016).

25. Brennick, C.A. *et al.* An unbiased approach to defining bona fide cancer neoepitopes that elicit immune-mediated cancer rejection. *J Clin Invest* **131** (2021).

26. Duan, F. *et al.* Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J Exp Med* **211**, 2231-2248 (2014).

27. Gonzalez, J.L., Cunningham, K., Silverman, R., Madan, E. & Nguyen, B.M. Comparison of the American Joint Committee on Cancer Seventh Edition and Brigham and Women's Hospital Cutaneous Squamous Cell Carcinoma Tumor Staging in Immunosuppressed Patients. *Dermatol Surg* **43**, 784-791 (2017).

28. Hartevelt, M.M., Bavinck, J.N., Kootte, A.M., Vermeer, B.J. & Vandenbroucke, J.P. Incidence of skin cancer after renal transplantation in The Netherlands. *Transplantation* **49**, 506-509 (1990).

29. Jensen, P., Moller, B. & Hansen, S. Skin cancer in kidney and heart transplant recipients and different long-term immunosuppressive therapy regimens. *J Am Acad Dermatol* **42**, 307 (2000).

30. Lindelof, B., Sigurgeirsson, B., Gabel, H. & Stern, R.S. Incidence of skin cancer in 5356 patients following organ transplantation. *Br J Dermatol* **143**, 513-519 (2000).

31. Rabinovics, N. *et al.* Cancer of the head and neck region in solid organ transplant recipients. *Head Neck* **36**, 181-186 (2014).

32. Adams, A.C. *et al.* Distinct sets of molecular characteristics define tumor-rejecting neoantigens. *bioRxiv*, 2024.2002.2013.579546 (2024).

33. Chang, D. & Shain, A.H. The landscape of driver mutations in cutaneous squamous cell carcinoma. *NPJ Genom Med* **6**, 61 (2021).

34. Inman, G.J. *et al.* The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature. *Nat Commun* **9**, 3667 (2018).

35. Kim, Y.S., Bang, C.H. & Chung, Y.J. Mutational Landscape of Normal Human Skin: Clues to Understanding Early-Stage Carcinogenesis in Keratinocyte Neoplasia. *J Invest Dermatol* **143**, 1187-1196 e1189 (2023).

36. Alexandrov, L.B. *et al.* Clock-like mutational processes in human somatic cells. *Nat Genet* **47**, 1402-1407 (2015).

37. Lu, T. *et al.* Tumor neoantigenicity assessment with CSiN score incorporates clonality and immunogenicity to predict immunotherapy outcomes. *Sci Immunol* **5** (2020).

38. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463-1469 (2016).

39. Miao, D. *et al.* Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat Genet* **50**, 1271-1281 (2018).

40. Beaumont, M.A., Zhang, W. & Balding, D.J. Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025-2035 (2002).

41. Wang, H., Liu, B. & Wei, J. Beta2-microglobulin(B2M) in cancer immunotherapies: Biological function, resistance and remedy. *Cancer Lett* **517**, 96-104 (2021).

42. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* **48**, W449-W454 (2020).

43. DuPage, M., Mazumdar, C., Schmidt, L.M., Cheung, A.F. & Jacks, T. Expression of tumour-specific antigens underlies cancer immunoediting. *Nature* **482**, 405-409 (2012).

44. Westcott, P.M.K. *et al.* Low neoantigen expression and poor T-cell priming underlie early immune escape in colorectal cancer. *Nat Cancer* **2**, 1071-1085 (2021).

45. Claeys, A. & Van den Eynden, J. Quantification of Neoantigen-Mediated Immunoediting in Cancer Evolution-Letter. *Cancer Res* **83**, 971-972 (2023).

46. Van den Eynden, J., Jimenez-Sanchez, A., Miller, M.L. & Larsson, E. Lack of detectable neoantigen depletion signals in the untreated cancer genome. *Nat Genet* **51**, 1741-1748 (2019).

47. Frazzette, N. *et al.* Decreased cytotoxic T cells and TCR clonality in organ transplant recipients with squamous cell carcinoma. *NPJ Precis Oncol* **4**, 13 (2020).

48. Andreatta, M. *et al.* Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat Commun* **12**, 2965 (2021).

49. Garner, L.C. *et al.* Single-cell analysis of human MAIT cell transcriptional, functional and clonal diversity. *Nat Immunol* **24**, 1565-1578 (2023).

50. Stranzl, T., Larsen, M.V., Lundegaard, C. & Nielsen, M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics* **62**, 357-368 (2010).

51. Jurtz, V. *et al.* NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* **199**, 3360-3368 (2017).

52. Jorgensen, K.W., Rasmussen, M., Buus, S. & Nielsen, M. NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology* **141**, 18-26 (2014).

53. Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society* **63**, 411-423 (2001).

54. Chen, S. *et al.* Frequency-dependent selection of neoantigens fosters tumor immune escape and predicts immunotherapy response. *Commun Biol* **7**, 770 (2024).

55. Philip, M. & Schietinger, A. Heterogeneity and fate choice: T cell exhaustion in cancer and chronic infections. *Curr Opin Immunol* **58**, 98-103 (2019).

56. Matsushita, H. *et al.* Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* **482**, 400-404 (2012).

57. DuPage, M. *et al.* Endogenous T cell responses to antigens expressed in lung adenocarcinomas delay malignant tumor progression. *Cancer Cell* **19**, 72-85 (2011).

58. Capietto, A.H. *et al.* Mutation position is an important determinant for predicting cancer neoantigens. *J Exp Med* **217** (2020).

59. Nassir, S. *et al.* Whole Exome and Transcriptome Sequencing of Stage-Matched, Outcome-Differentiated Cutaneous Squamous Cell Carcinoma Identifies Gene Expression Patterns Associated with Metastasis and Poor Outcomes. *medRxiv*, 2024.2002.2005.24302298 (2024).

60. Asmann, Y.W. *et al.* TREAT: a bioinformatics tool for variant annotations and visualizations in targeted and exome sequencing data. *Bioinformatics* **28**, 277-278 (2012).

61. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).

62. Van der Auwera, G. & O'Connor, B. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition)*. O'Reilly Media, 2020.

63. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* **15**, 591-594 (2018).

64. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).

65. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C. & Koeffler, H.P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* **28**, 1747-1756 (2018).

66. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* **17**, 31 (2016).

67. Wang, S. *et al.* Copy number signature analysis tool and its application in prostate cancer reveals distinct mutational processes and clinical outcomes. *PLoS Genet* **17**, e1009557 (2021).

68. Kalari, K.R. *et al.* MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. *BMC Bioinformatics* **15**, 224 (2014).

69. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

70. Liao, Y., Smyth, G.K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41**, e108 (2013).

71. Liao, Y., Smyth, G.K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* **47**, e47 (2019).

72. Revkov, E., Kulshrestha, T., Sung, K.W. & Skanderup, A.J. PUREE: accurate pan-cancer tumor purity estimation from gene expression data. *Commun Biol* **6**, 394 (2023).

73. Talevich, E., Shain, A.H., Botton, T. & Bastian, B.C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* **12**, e1004873 (2016).

74. Braverman, J.M., Hudson, R.R., Kaplan, N.L., Langley, C.H. & Stephan, W. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**, 783-796 (1995).

75. Fu, Y.X. Statistical properties of segregating sites. *Theor Popul Biol* **48**, 172-197 (1995).

76. Durrett, R. Population Genetics of Neutral Mutations in Exponentially Growing Cancer Cell Populations. *Ann Appl Probab* **23**, 230-250 (2013).

77. Williams, M.J., Werner, B., Barnes, C.P., Graham, T.A. & Sottoriva, A. Identification of neutral tumor evolution across cancer types. *Nat Genet* **48**, 238-244 (2016).

78. Csilléry, K., François, O. & Blum, M. abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* **3**, 475-479 (2012).

79. Matey-Hernandez, M.L., Danish Pan Genome, C., Brunak, S. & Izarzugaza, J.M.G. Benchmarking the HLA typing performance of Polysolver and Optitype in 50 Danish parental trios. *BMC Bioinformatics* **19**, 239 (2018).

80. Hundal, J. *et al.* pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. *Genome Med* **8**, 11 (2016).

81. Plattner, C., Finotello, F. & Rieder, D. Deconvoluting tumor-infiltrating immune cells from RNA-seq data using quanTIseq. *Methods Enzymol* **636**, 261-285 (2020).

82. Aran, D., Hu, Z. & Butte, A.J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* **18**, 220 (2017).

83. Becht, E. *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* **17**, 218 (2016).

84. Sturm, G., Finotello, F. & List, M. Immunedeconv: An R Package for Unified Access to Computational Methods for Estimating Immune Cell Fractions from Bulk RNA-Sequencing Data. *Methods Mol Biol* **2120**, 223-232 (2020).

85. Borcherding, N., Bormann, N.L. & Kraus, G. scRepertoire: An R-based toolkit for single-cell immune receptor analysis. *F1000Res* **9**, 47 (2020).

86. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* **42**, 293-304 (2024).

87. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).

88. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979-982 (2017).

89. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386 (2014).