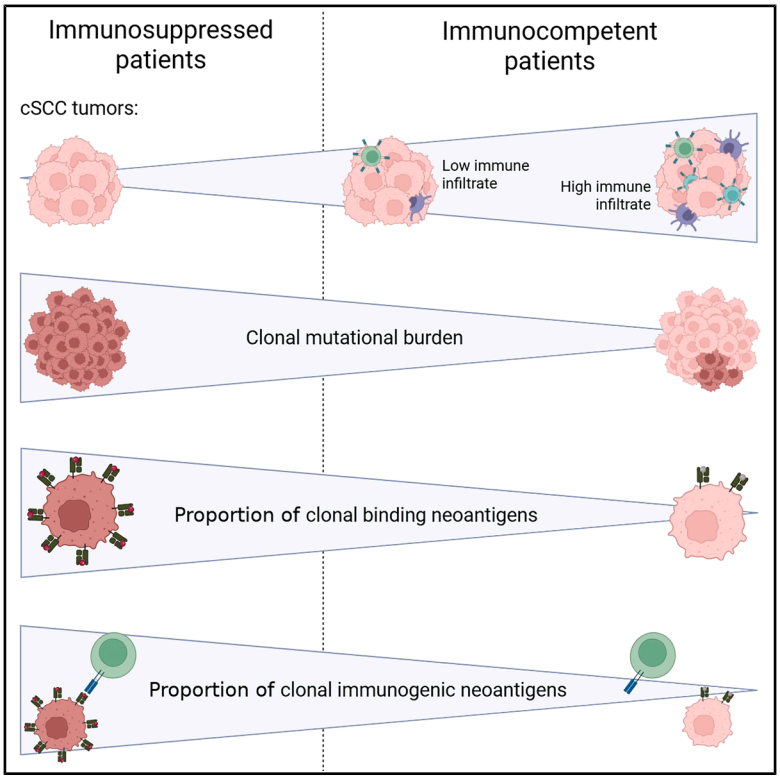
Article

Immunity

Immunoediting restricts clonal neoantigens in primary, treatment-naive human tumors

Graphical abstract



Authors

Elizabeth S. Borden, David Castellano, Emily A. Kantzos, ..., Kenneth H. Buetow, Melissa A. Wilson, Karen Taraszka Hastings

Correspondence

khasting@arizona.edu

In brief

Borden et al. examine primary, treatment-naive squamous cell carcinoma (SCC) tumors to determine the impact of immunoediting. Comparison of SCC tumors from immunocompetent and immunosuppressed patients demonstrates immune-mediated sculpting of the tumor neoantigen profile, where immunogenic neoantigens are restricted from becoming clonal within the tumor.

Highlights

- Comparison of SCC tumors from immunocompetent vs. immunosuppressed patients
- The primary mechanism of immunoediting is the restriction of clonal neoantigens
- Immunoedited neoantigens share features of validated immunogenic neoantigens

Borden et al., 2026, Immunity *59*, 1999–2014
July 14, 2026 Published by Elsevier Inc.
https://doi.org/10.1016/j.immuni.2026.05.005

CellPress

Immunity

CellPress

Article

Immunoediting restricts clonal neoantigens in primary, treatment-naive human tumors

Elizabeth S. Borden,[1,2] David Castellano,[3] Emily A. Kantzos,[1] Alysia Hughes,[4] Zachary Leibovit-Reiben,[4] Alyssa Stockard,[4] Anirudh Choudhary,[5] Tao Ma,[4] Xing Li,[4] Aaron R. Mangold,[4] Ryan N. Gutenkunst,[3] Bonnie J. LaFleur,[6,7] Kenneth H. Buetow,[8,9] Melissa A. Wilson,[8,9] and Karen Taraszka Hastings[1,2,10,11,*]

[1]Department of Dermatology, College of Medicine-Phoenix, University of Arizona, Phoenix, AZ 85004, USA
[2]Phoenix Veterans Affairs Health Care System, Phoenix, AZ 85012, USA
[3]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA
[4]Department of Dermatology, Mayo Clinic Health System, Scottsdale, AZ 85259, USA
[5]Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Champaign, IL 61801, USA
[6]BIO5 Institute, University of Arizona, Tucson, AZ 85719, USA
[7]R. Ken Coit College of Pharmacy, University of Arizona, Tucson, AZ 85724, USA
[8]School of Life Sciences, Arizona State University, Tempe, AZ 85281, USA
[9]Center for Evolution and Medicine, Arizona State University, Tempe, AZ 85281, USA
[10]University of Arizona Cancer Center, University of Arizona, Tucson, AZ 85719, USA
[11]Lead contact
*Correspondence: khasting@arizona.edu
https://doi.org/10.1016/j.immuni.2026.05.005

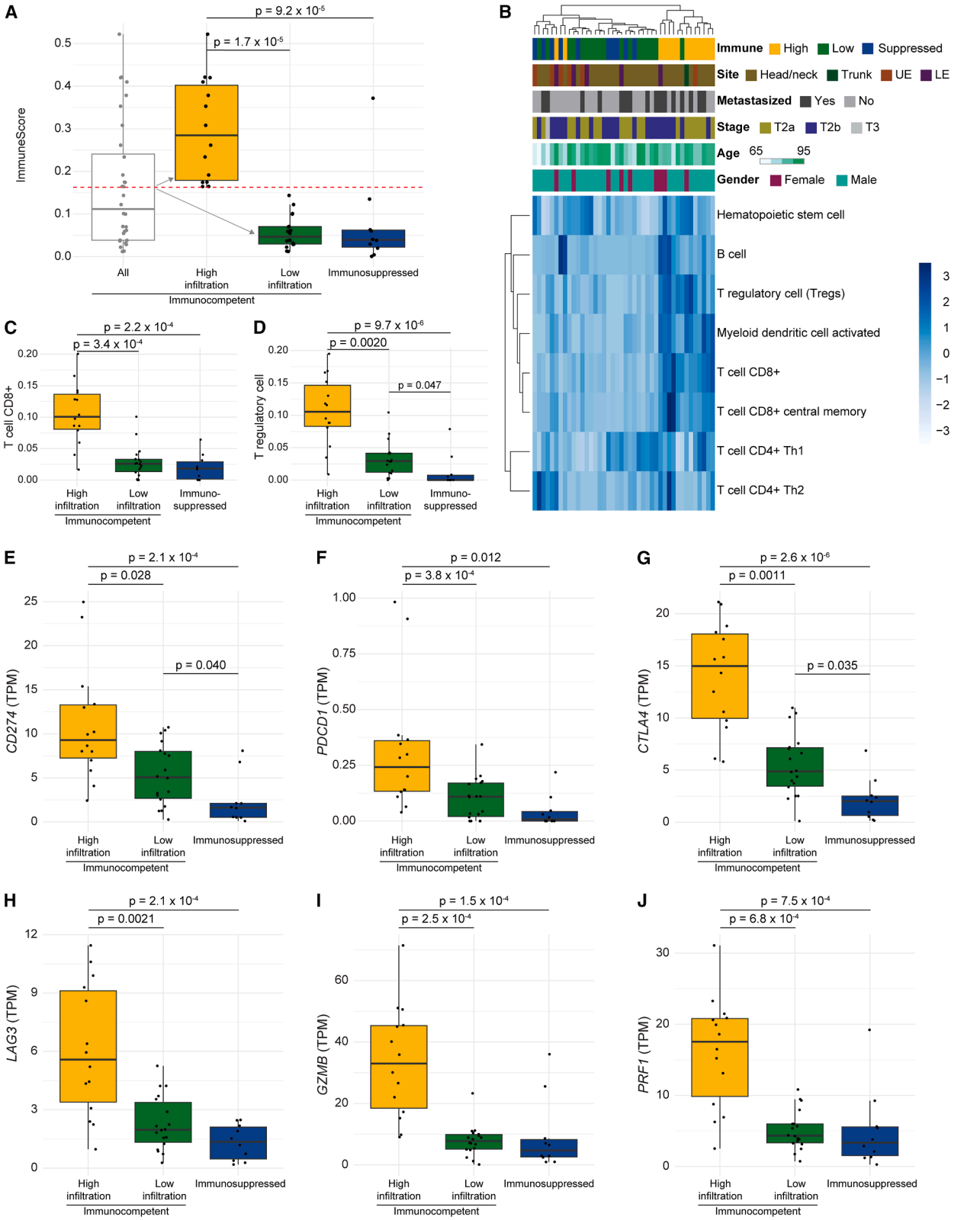
SUMMARY

T cell targeting of cancer cells alters the tumor antigen landscape in preclinical models. Here, we examined the impact of immunoediting on the antigenic landscape of primary, treatment-naive human tumors. Cutaneous squamous cell carcinoma tumors from immunocompetent and immunosuppressed patients revealed consistent tumor mutational signatures; however, high-immune-infiltrate tumors from immunocompetent patients had lower overall mutational burdens and lower clonal mutational burdens compared with low-infiltrate tumors from immunocompetent patients and tumors from immunosuppressed patients. The lower clonal mutational burden in high-immune-infiltrate tumors from immunocompetent patients persisted after accounting for tumor purity and growth rate. Predicted neoantigen: major histocompatibility complex (MHC) class I binding affinity decreased with increasing variant allele frequency, demonstrating restriction of mutations encoding MHC-binding neoantigens. Neoantigens with features shared with validated immunogenic neoantigens were decreased in clonal relative to subclonal cancer cell populations in high-immune-infiltrate tumors from immunocompetent patients. Thus, the immune system restricts cancer cells expressing immunogenic antigens from clonal populations in primary, treatment-naive human tumors.

INTRODUCTION

Immunoediting is the process by which the immune system targets and removes cancer cells visible to the immune system, allowing the least-visible populations to persist. Immunoediting was defined in mouse models, where a number of studies observed increased tumor formation and growth in immunodeficient compared with immunocompetent mice.[1–5] Furthermore, chemically induced sarcomas that form in immunocompetent mice grow when transplanted into immunocompetent mice, whereas sarcomas that develop in immunodeficient mice fail to grow in immunocompetent mice,[6] demonstrating that tumors that form in immunodeficient mice fail to develop mechanisms to escape immune-mediated destruction. These mouse studies have provided the fundamental basis for our understanding 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,[7] (2) late vs. early recurrent tumors,[8] (3) tumors after vs. before immune checkpoint inhibition,[9,10] (4) tumor regions with high immune infiltration and intact major histocompatibility complex (MHC) alleles vs. sparsely infiltrated regions,[11] (5) observed vs. *in silico* mutational profiles,[12] and (6) non-synonymous vs. synonymous mutations within the same tumor.[13–15] However, the findings in human tumors that are consistent with immunoediting are associative in nature and have not been conclusively attributed to the immune system.

Additionally, the predominant effects of immunoediting on the neoantigen profile remain unclear. Within mouse models, evidence has been identified for complete loss of mutations encoding neoantigens at the DNA level,[16] frequency-dependent elimination of neoantigens leading to a shift toward more



(legend on next page)

heterogeneous tumors,[17] RNA-level loss of neoantigen expression,[18] and loss of antigen-processing and presentation machinery.[18] A similar range of mechanisms has been explored in human cancer. Although studies have suggested that there is an overall decrease in the proportion of immunogenic neoantigens compared with what would be expected in the absence of an intact immune response,[13–15] another study demonstrated that there is no reduction in the expected proportion of immunogenic neoantigens when accounting for the mutational signature.[19] Similar debates have occurred around loss of expression of neoantigens, with some studies showing enrichment of neoantigens at low expression[12] and others questioning whether this result is an underlying feature of the mutational signature.[20] Therefore, the predominant effects of immunoediting on the neoantigen profile remain to be fully elucidated.

Here, we implemented three controls to isolate the effects of immunoediting in primary, treatment-naive human tumors. First, we compared the neoantigen profile in tumors from immunocompetent vs. immunosuppressed patients. We selected cSCC because immunosuppressed solid organ transplant recipients have a 65- to 253-fold higher risk of developing cSCC,[21–25] suggesting that cSCC is highly constrained by the immune system. Therefore, tumors from immunocompetent patients are expected to have developed mechanisms of immune evasion that are absent in tumors from immunosuppressed patients. Second, we subdivided the cSCC tumors into high- and low-immune-infiltrate groups to demonstrate the impact of immune infiltration in shaping the neoantigen profile. Finally, we employed *in silico* mutational profiles as a patient-specific comparison of the expected mutational profile, accounting for the mutational signature, in the absence of immune-mediated effects. Overall, this study demonstrates alterations in the neoantigen landscape of primary cSCC that are attributable to the immune system.

RESULTS

Only a portion of cSCC tumors from immunocompetent patients have higher immune infiltration than tumors from immunosuppressed patients

First, we characterized the immune microenvironment of tumors from immunocompetent and immunosuppressed patients. Although analysis of the immune infiltrates estimated by xCell deconvolution[26] demonstrated a difference in the degree of immune infiltration between tumors from immunocompetent and immunosuppressed patients, there was a wide range in the quantity of immune infiltrate in tumors from immunocompetent patients. Therefore, we split the immunocompetent group into high- and low-immune-infiltrate groups using the ImmuneScore from xCell (Figure 1A). The resulting low-infiltrate tumors had no significant difference in immune infiltrate (ImmuneScore) compared with tumors from immunosuppressed patients, and the high-infiltrate group had significantly more immune infiltration (Figure 1A). Unsupervised hierarchical clustering of deconvoluted immune cell types from xCell demonstrated separation of the high- and low-infiltrate groups (Figure 1B). In particular, expression attributable to $CD8^+$ T cells and T regulatory (Treg) cells distinguished high-infiltrate tumors from immunocompetent patients relative to tumors from immunosuppressed patients. Low-infiltrate tumors from immunocompetent patients had immune infiltrates more similar to tumors from immunosuppressed patients (Figures 1C and 1D).

As confirmation, we performed immune cell deconvolution with two independent approaches (Figures S1A and S1B).[27,28] The immune infiltration status identified by xCell was used as an annotation across all three heatmaps, demonstrating a high degree of concordance in the samples identified as having high immune infiltrates. An increase in $CD8^+$ T cells and Treg cells in the high-infiltrate group was observed in the additional deconvolution approaches, along with a higher level of Treg infiltration in the low-infiltrate tumors from immunocompetent patients compared with immunosuppressed patients (Figures S1C–S1E). The cytotoxicity score calculated by Microenvironment Cell Populations-counter (MCP-counter) was significantly increased in high-infiltrate tumors compared with tumors from immunosuppressed patients (Figure S1F), reflecting increased $CD8^+$ T cell activity.

Additionally, we compared the expression of immune checkpoint molecules *CD274* (encodes programmed death-ligand 1 [PD-L1]), *PDCD1*, *CTLA4*, and *LAG3* and markers of cytolytic activity *GZMB* and *PRF1* (Figures 1E–1J; Table S1). Consistent with increased $CD8^+$ T cell and Treg cell infiltration, high-infiltrate cSCC had increased expression of markers of cytotoxicity and exhaustion compared with low-infiltrate tumors from immunocompetent patients and tumors from immunosuppressed patients. Immune checkpoint molecules *CD274* and *CTLA4* were significantly higher in the low-infiltrate tumors from immunocompetent patients compared with tumors from immunosuppressed patients, whereas no difference was observed in markers of cytotoxicity (Figures 1E–1J). The increase in T cell infiltration and markers of cytotoxicity and exhaustion supports the conclusion that high-infiltrate cSCC is under increased surveillance, whereas immune surveillance in low-infiltrate cSCC from immunocompetent patients is expected to be more similar to that in cSCC from immunosuppressed patients.

Figure 1. Only a portion of cSCC tumors from immunocompetent patients have higher immune infiltration than tumors from immunosuppressed patients

(A–D) xCell immune cell deconvolution of all cSCC samples with available RNA-seq data (32 immunocompetent and 10 immunosuppressed). (A) Boxplots comparing xCell ImmuneScore in immunocompetent and immunosuppressed patients. Immunocompetent patients were divided into high- and low-infiltrate groups based on the mean ImmuneScore of the immunocompetent group (dashed red line). The xCell ImmuneScore is calculated as the sum of the estimated immune cell abundances (subtypes of B cells, T cells, dendritic cells, eosinophils, macrophages, monocytes, mast cells, neutrophils, and natural killer [NK] cells). (B) Heatmap illustrating the frequencies of immune cell types. Samples were arranged by hierarchical clustering. Cell types were included if they had a score of at least 0.05 in at least 10 samples. (C and D) Boxplots comparing infiltration of (C) $CD8^+$ T cells and (D) Treg cells across immunocompetent and immunosuppressed patients.

(E–J) Boxplots comparing the expression of (E) *CD274* (encodes programmed death-ligand 1 [PD-L1]), (F) *PDCD1*, (G) *CTLA4*, (H) *LAG3*, (I) *GZMB*, and (J) *PRF1* in cSCC tumors from immunocompetent and immunosuppressed patients. A Kruskal-Wallis test with Dunn's multiple comparisons test was used to examine differences in (A) and (C)–(J).

See also Figure S1.

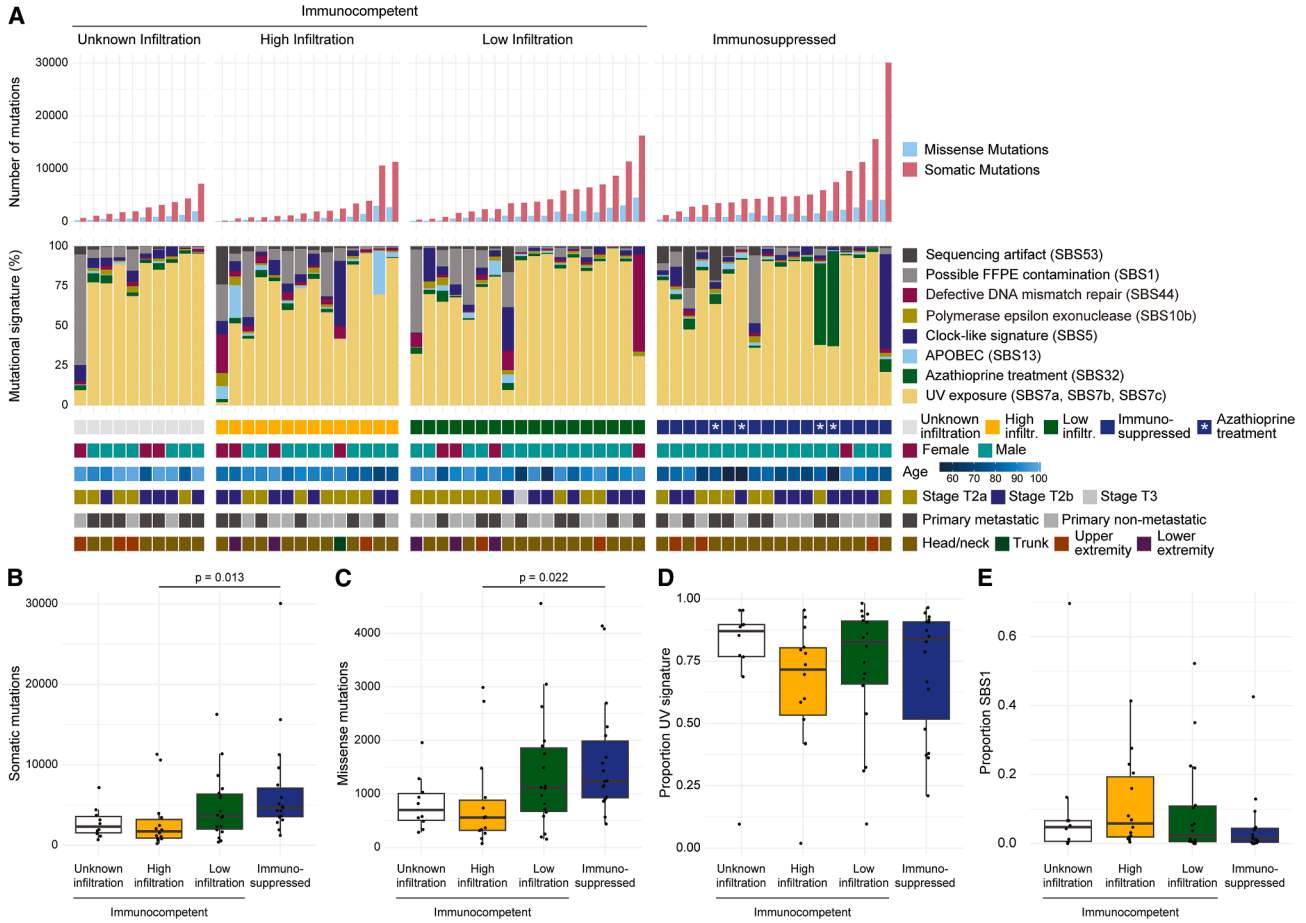


Figure 2. Lower mutational burden and comparable mutational profile in high-infiltrate cSCC compared with tumors from immunosuppressed patients

Somatic mutations were identified, and missense mutations were annotated, in cSCC tumors (42 immunocompetent and 18 immunosuppressed). Tumors without RNA-seq data are indicated as having unknown infiltration.

(A) Top: bar graph showing somatic and missense mutational burdens. Bottom: stacked bar graph showing the proportion of mutations in each tumor attributable to known COSMIC mutational signatures.

(B–E) Comparison of (B) somatic mutations, (C) missense mutations, (D) proportion of mutations attributable to UV signature (single-base substitutions [SBS] 7a, b, and c), and (E) proportion of mutations attributable to SBS1, a putative FFPE signature, across tumors segregated by immune status. Differences in scale parameters for (B)–(E) were tested using a Kruskal-Wallis test with Dunn's multiple comparisons test.

See also Figure S2.

To isolate the role of the immune system, we also compared the demographics across the high- and low-immune-infiltrate tumors from immunocompetent patients and the tumors from 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 (Figures S1G–S1J; Table S2).

Lower clonal mutational burden in high-infiltrate tumors from immunocompetent patients

Next, we compared the overall mutational burden and signature across high- and low-infiltrate cSCC tumors from immunocompetent patients and cSCC tumors from immunosuppressed patients. We observed a wide range in the mutational burden of cSCC tumors (Figure 2A; Table S3), as previously reported.[7,29–32] Tumors from immunosuppressed patients had a significantly larger number of overall mutations and missense mutations compared with high-infiltrate tumors from immunocompetent patients (Figures 2B and 2C), whereas low-infiltrate tumors from immunocompetent patients had an intermediate mutational burden. We then calculated the proportion of mutations in each tumor attributable to established mutational signatures (Figure 2A). Most tumors had a predominant signature of UV-induced mutations, consistent with the known etiology of skin cancer. The other predominant mutational signature, especially in low mutational burden tumors, was the "clock-like" signature (single-base substitutions (SBS) 1 and SBS5), which consists of mutations known to accumulate in human cells at a steady rate over time.[33] However, SBS1 is also the predominant signature of formalin-fixed, paraffin-embedded (FFPE)-induced mutations chemically corrected during the DNA extraction process.[34] Therefore, filtering criteria were selected to

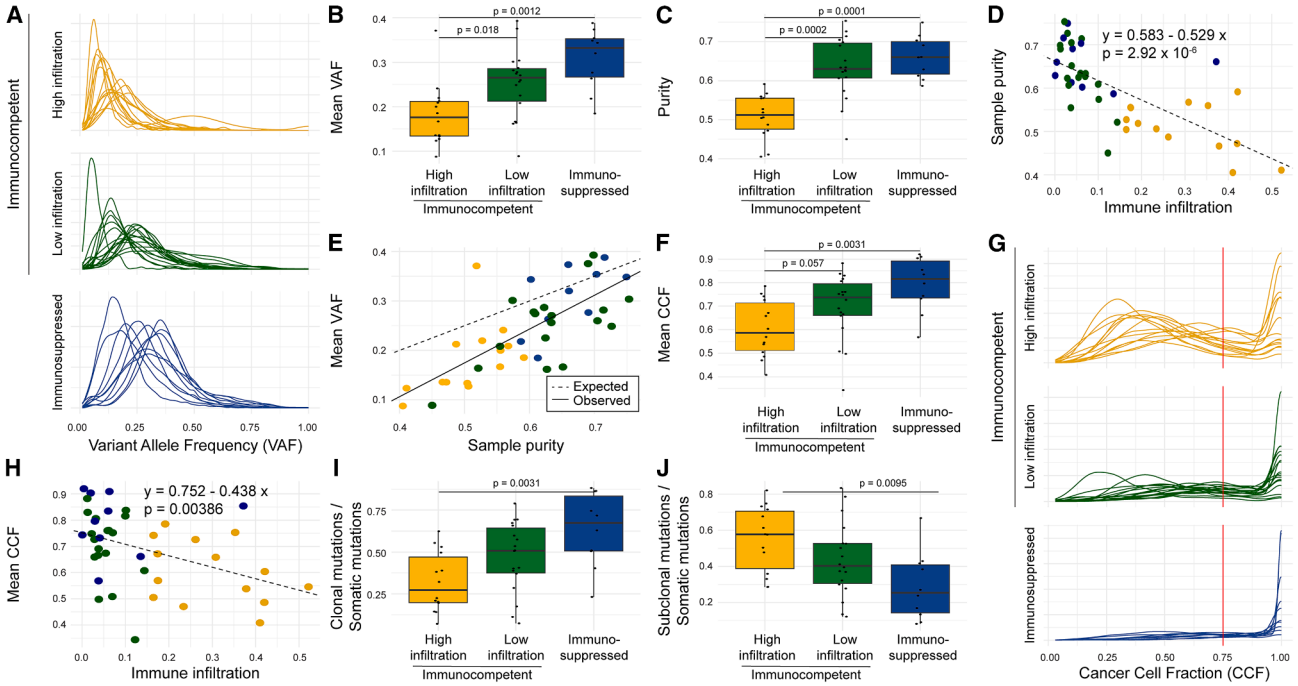


Figure 3. Lower clonal mutational burden in high-infiltrate tumors from immunocompetent patients

(A) VAF distributions for all mutations from each tumor with WES and RNA-seq (32 immunocompetent, 10 immunosuppressed). Each tumor is represented by a line indicating the mutation density at specific VAFs.

(B) Mean VAF for each tumor, segregated by immune status.

(C) Tumor purity estimates from RNA-seq gene signatures, segregated by immune status.

(D) Sample purity vs. immune infiltration (xCell ImmuneScore), with a fitted linear regression (dashed line).

(E) Mean VAF vs. sample purity. The dashed line indicates the expected relationship between purity and mean VAF, presuming only clonal mutations. The solid line indicates the observed relationship between purity and mean VAF.

(F) Mean CCF (VAF adjusted for purity and copy number) for each tumor, segregated by immune status.

(G) CCF distributions for all mutations from each tumor. Each tumor is represented by a line indicating the mutation density at specific CCFs. The red line indicates the cut point used to define clonal and subclonal mutations.

(H) Mean CCF vs. immune infiltration (xCell ImmuneScore), with a fitted linear regression (dashed line).

(I) Proportion of clonal mutations (CCF > 0.75) out of all observed somatic mutations, segregated by immune status.

(J) Proportion of subclonal mutations out of all observed mutations, segregated by immune status. 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. Differences in scale parameters for (B), (C), (F), (I), and (J) were tested using a Kruskal-Wallis test with Dunn's multiple comparisons test.

minimize excess SBS1 signature (Figures S2A–S2C). After filtering, no significant difference was observed in the proportion of the signature attributable to UV exposure or the putative FFPE signature between cSCC tumors from immunocompetent and immunosuppressed patients (Figures 2D and 2E). Two out of four patients treated with azathioprine had a strong azathioprine signature, as expected.[29,30] Overall, no significant differences were observed in the mutational signature in high- or low-infiltrate cSCC from immunocompetent patients and cSCC from immunosuppressed patients, supporting the conclusion that differences in neoantigen profile are unlikely to be attributable to underlying differences in mutational signature. The lower mutational burden in high-infiltrate tumors from immunocompetent patients in the context of a comparable mutational signatures suggests the possibility of immunoediting in tumors from immunocompetent patients.

In addition to a lower mutational burden, high-infiltrate cSCC tumors from immunocompetent patients had a significantly lower mean variant allele frequency (VAF; proportion of sequencing reads containing the variant) compared with low-infiltrate cSCC tumors from immunocompetent patients and cSCC tumors from immunosuppressed patients (Figures 3A and 3B). To identify the cause of this difference, first tumor purity was evaluated. A decrease in purity would be expected to decrease the VAF distributions because fewer sequencing reads would be attributable to the tumor and able to contain the variant of interest. Although all tumors were reviewed by a dermatopathologist and annotated by a dermatologist to guide macrodissection of the tumor, high-infiltrate tumors from immunocompetent patients had decreased purity compared with low-infiltrate tumors from immunocompetent patients and tumors from immunosuppressed patients (Figure 3C). The difference in purity was inversely correlated with the immune cell infiltrate (Figure 3D), suggesting that the differences in purity were attributable to differences in immune infiltration in cSCC. As expected, the mean VAF was significantly associated with purity (Figure 3E); however, the differences in VAF distributions across immunosuppressed and immunocompetent

patients could not be attributed to purity alone. If only clonal mutations were detectable, then the expected mean VAF would equal half the empirically measured purity (Figure 3E). Thus, a higher number of subclonal mutations must have been present in high-infiltrate tumors from immunocompetent patients to explain the significantly lower mean VAF compared with tumors from immunosuppressed patients.

To adjust for tumor purity, we calculated the cancer cell fraction (CCF; VAF adjusted for purity and copy number) and found a significant decrease in the mean CCF of high-infiltrate tumors from immunocompetent patients compared with tumors from immunosuppressed patients, with low-infiltrate tumors from immunocompetent patients showing an intermediate mean CCF (Figures 3F and 3G). There was also an inverse association of immune infiltration with the mean CCF, where the mean CCF increased with lower immune infiltrate (Figure 3H). Furthermore, high-infiltrate cSCC had a lower proportion of clonal mutations relative to the overall mutational burden compared with cSCC from immunosuppressed patients; low-infiltrate tumors from immunocompetent patients had an intermediate proportion of clonal mutations (Figure 3I). This finding was confirmed by an increase in the proportion of subclonal mutations in high-infiltrate tumors from immunocompetent patients, calculated by a second method of assigning clonality (Figure 3J). Together, these results demonstrate a substantial difference in the observed clonal and subclonal mutational profiles in tumors from immunocompetent and immunosuppressed patients.

Two additional factors that may influence the distribution of clonal mutations in the tumor are (1) the growth rate of the tumor and (2) immune selective pressure. Assuming neutral evolution, rapidly growing tumors acquire more subclonal mutations that remain at a very low allele frequency compared with slower-growing tumors or tumors with constant population size. Consequently, in simulated VAF distributions, most variants remain unique to a small proportion of cells and are predicted to 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 (Figure 4A). In contrast, we expect a higher number of detectable subclonal mutations for tumors that are slowly growing or have a constant population size, leading to a lower mean VAF (Figure 4B). The second factor that may influence the distribution of clonal and subclonal mutations is immune selective pressure. Studies have demonstrated that the immune system is more apt to recognize and target neoantigens at higher frequencies in the tumor.[17,35–37] If the immune system selectively targets neoantigens, this would drive down the VAF and CCF distributions. To account for the potential effects of both growth rate and targeting of neoantigens by the immune system on individual patient VAF distributions, we used an approximate Bayesian computation (ABC) approach for parameter estimation.[38] VAF distributions were simulated across a range of growth rates and ratios of clonal:subclonal mutations, using patient-specific purities (Figure 3C) and coverage distributions (Figure 4C), to find the values of both parameters that best explain the observed VAFs (agreement between the estimated and observed distributions within each patient is visualized in Figures 4D, 4E, and S3). The most likely parameter fits for the observed VAF distributions showed no significant difference in growth rates (Figure 4F) and a significantly lower ratio of clonal:subclonal mutations in high-infiltrate tumors from immunocompetent patients compared with tumors from immunosuppressed patients, with low-infiltrate tumors from immunocompetent patients occupying an intermediate level (Figure 4G). 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. These results suggest that immune selective pressure restricts clonal mutations in tumors with high immune infiltrate relative to tumors from immunosuppressed patients.

Lower clonal mutational burden in tumors from immunocompetent patients confirmed in an independent dataset

To increase the sample size, we sequenced 17 additional tumors, 14 from immunosuppressed patients and 3 from immunocompetent patients (Table S2). Sampling was preferentially skewed toward tumors from immunosuppressed patients to adjust for the previous imbalance in the data. Because differences in the sequencing technology prevented comparisons across the datasets, we independently compared the characteristics of the mutational burden and VAFs in the new data. Consistent with the original dataset, tumors from immunocompetent patients had lower mutational burdens (Figures 5A and 5B), though the difference did not reach statistical significance. Tumors from immunocompetent patients had significantly lower VAF distributions (Figures 5C and 5D) and significantly lower purity (Figure 5E) compared with tumors from immunosuppressed patients. After adjustment for purity, a significant difference in the CCF persisted (Figure 5F), and a significantly lower proportion of clonal mutations was observed in tumors from immunocompetent patients (Figure 5G). These results provide independent evidence corroborating the restriction of clonal neoantigens in tumors from immunocompetent patients.

Given that purity estimates are central to determining the accuracy of the CCF distributions, we employed an independent approach to corroborate the purity estimates. A random set of samples across the original and new cohorts were analyzed using machine learning on dermatologist-annotated whole-slide images. The percentage of neoplastic cells was calculated and compared with the RNA-based purity estimates. The independent estimates showed a significant association, and both estimates ranged between 40% and 80% (Figure S4A). Critically, the RNA estimates were lower for most samples than the machine learning estimates, suggesting that our interpretation of the CCF shifts may be conservative. Examples of cell annotations in a high-, medium-, and low-purity sample are included (Figures S4B–S4D). Overall, these independent, pathology-based estimates corroborate the accuracy of the RNA-based estimates and further demonstrate that purity alone does not explain the prominent shift in the VAF and CCF distributions observed between high-infiltrate tumors and tumors from immunosuppressed patients.

Immunoediting at the DNA level selectively occurs in high-infiltrate tumors from immunocompetent patients

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

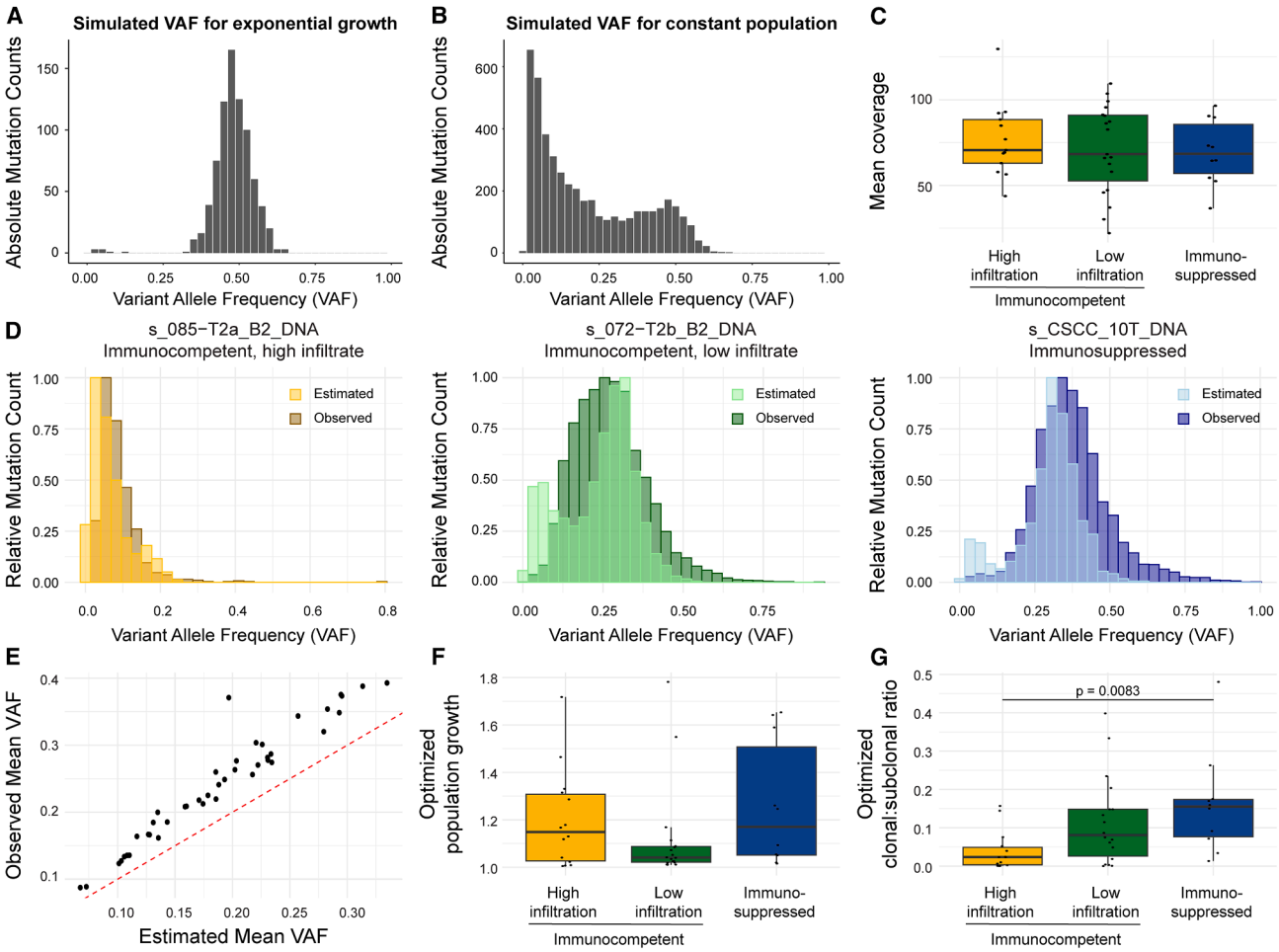


Figure 4. ABC approach to estimate growth rate and clonal:subclonal ratio reveals lower ratio of clonal:subclonal mutations in high-infiltrate cSCC tumors from immunocompetent patients

(A and B) Simulated VAFs with varying growth rates in (A) an exponentially growing population and (B) a constant population size. The remaining simulation parameters are constant in the two scenarios. The total number of mutations is the same in the two scenarios, but in (A), most variants remain undetectable at very low frequency, whereas in (B), most detected mutations are subclonal.

(C) Mean sequencing coverage across all variants, segregated by immune status.

(D–G) ABC was implemented to select the simulated VAFs that had the best chi-square goodness-of-fit statistic between the observed and estimated distributions. VAF distributions were simulated across a range of growth rates and ratios of clonal:subclonal mutations to find the values of both parameters that best explain the observed VAFs.

(D) Representative examples of histograms comparing the VAF distributions estimated by the posterior distributions from the implementation of ABC to the observed VAF distributions in specific patient tumors.

(E) Comparison of the observed mean VAF to the estimated mean VAF from the models selected using ABC methods. The dotted line indicates the 1:1 correspondence between observed and estimated mean VAFs.

(F) Optimized growth rates from the estimated VAF distribution results.

(G) 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). Differences in scale parameters for (C), (F), and (G) were tested using a Kruskal-Wallis test with Dunn's multiple comparisons test.

See also Figure S3.

immune evasion).[39] For this and all subsequent analyses, the old and new datasets were combined because the comparisons were either neutral to the sequencing platform or normalized to within-patient metrics. HLA alleles were identified in both the tumor and normal samples to assess for loss of heterozygosity of HLA alleles within the tumor, as reported.[11] Homozygosity of at least one HLA allele unique to the tumor sample, suggesting loss of heterozygosity, was observed in 5/17 high-infiltrate tumors from immunocompetent patients, 4/18 low-infiltrate tumors from immunocompetent patients, and 7/24 tumors from immunosuppressed patients. There was no significant difference in the proportion of tumors with loss of heterozygosity across the groups (p = 0.94, Fisher's exact test; Table S4). No mutations in *B2M* were identified that passed the coverage and allele count filters. Therefore, all patients were considered subject to neoantigen-based immunoediting, but prediction of

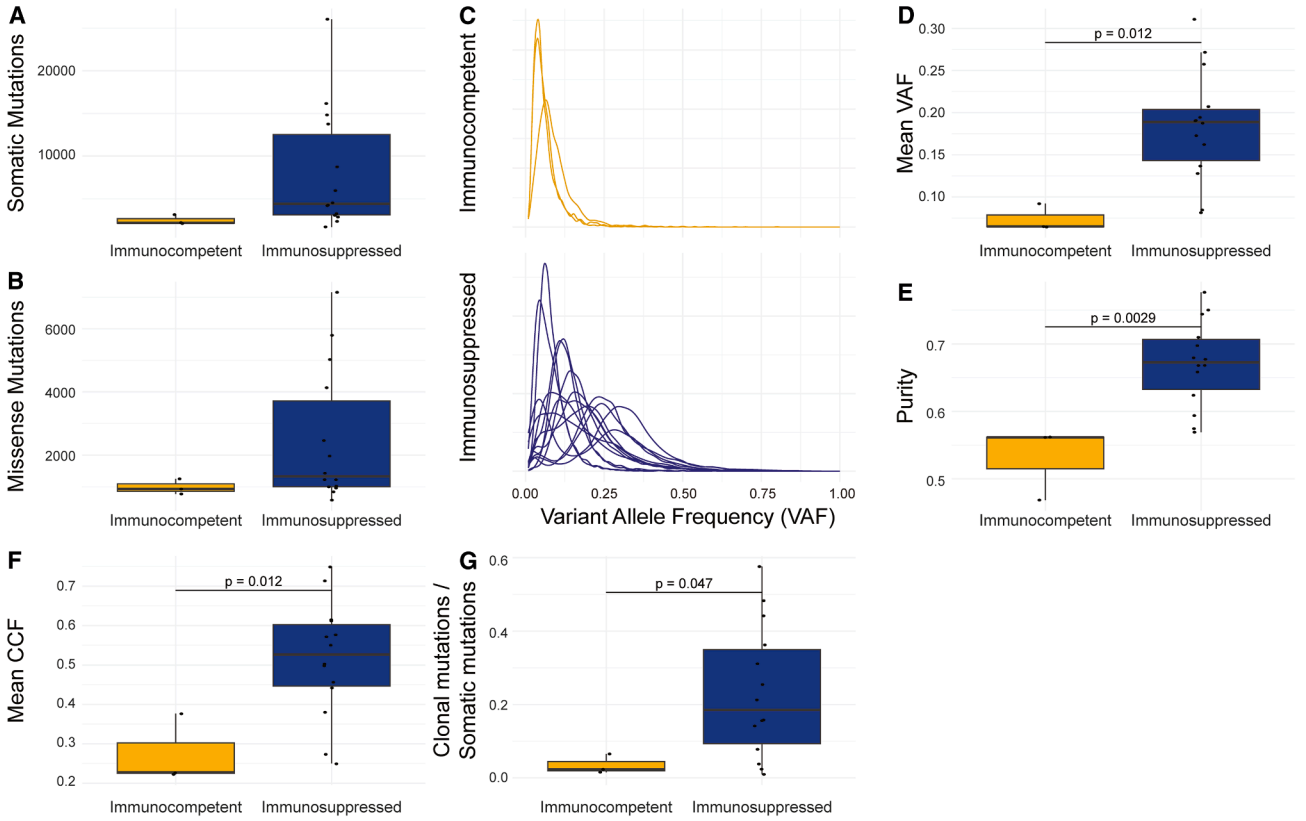


Figure 5. Lower clonal mutational burden in tumors from immunocompetent patients confirmed in an independent dataset

Characteristics of the mutational burden of tumors (3 immunocompetent and 14 immunosuppressed).

(A and B) Comparison of (A) somatic mutations and (B) missense mutations in tumors, segregated by immune status.

(C) VAF distributions for all mutations from each tumor. Each line represents a tumor, with the peak height indicating mutation density at specific VAFs.

(D) Mean VAF for each tumor, segregated by immune status.

(E) Tumor purity estimates from RNA-seq gene signatures, segregated by immune status.

(F) Mean CCF (VAF adjusted for purity and copy number) for each tumor, segregated by immune status.

(G) The proportion of clonal mutations (CCF > 0.75) out of all observed mutations, segregated by immune status. Differences in scale parameters for (A), (B), and (D)–(G) were tested using a Wilcoxon rank-sum test.

See also Figure S4.

binding neoantigens was restricted to HLA alleles maintained in the tumor because neoantigens specific to lost HLA alleles should not be subject to the same immune-mediated selective pressure.

To further characterize the role of the immune system in shaping the mutational profile, we compared the predicted neoantigen:MHC class I binding affinity, as MHC binding has consistently been identified as a key characteristic of neoantigen immunogenicity.[8,40–44] For simplicity, all mutated peptides examined are referred to as "neoantigens," although only a portion will be recognized by the immune system. Two hypotheses for the impact of the immune system on the neoantigen profile are (1) that MHC class I-binding neoantigens are recognized and eliminated in the setting of an intact immune system, leading to a decrease in the proportion of binding neoantigens in immunocompetent patients, or (2) that the immune response restricts MHC class I-binding neoantigens to lower VAF within the tumor. As each patient has a different set of HLA alleles and some individuals have fewer alleles due to homozygosity, we compared 100 *in silico* mutational profiles for each patient to assess for overall loss of binding neoantigens. *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 (Figures S5A–S5D). No significant differences in the proportion of neoantigens predicted to bind MHC class I were identified between the *in silico* and observed mutations for either immunocompetent or immunosuppressed patients (Figure 6A), suggesting that binding neoantigens were not completely eliminated from the tumor.

There appears to be a slight increase in the proportion of binding neoantigens in high-infiltrate tumors from immunocompetent patients compared with low-infiltrate tumors in immunocompetent patients and tumors from immunosuppressed patients. Although this difference is non-significant, it is striking given that the direction of change is opposite to what would be expected if the proportions of binding neoantigens were influenced by immunoediting. Additionally, the difference is conserved across both the observed and expected proportions of binding neoantigens. Because the *in silico* mutational profiles were generated using a mutational signature that was held constant

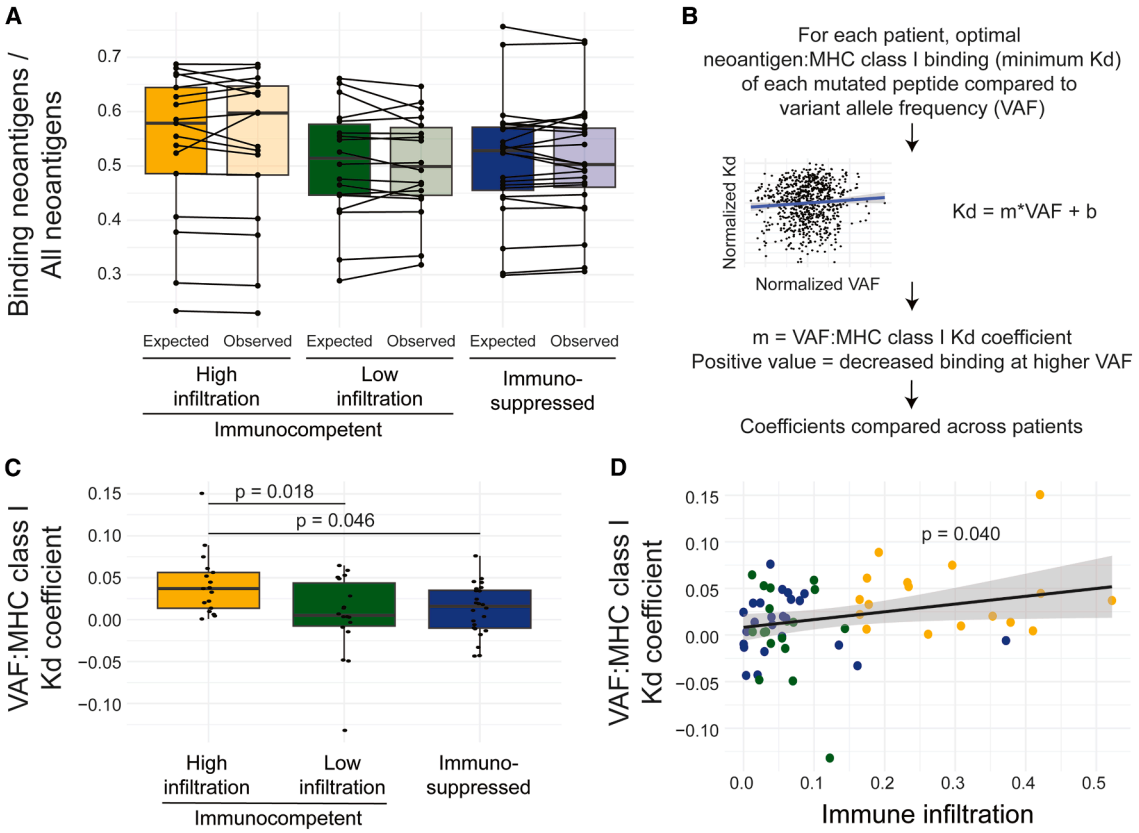


Figure 6. MHC class I-binding neoantigens are restricted to a lower VAF in an immune-dependent manner

(A) Comparison of the observed proportion of binding neoantigens (K_D < 500 nM) to the expected proportion of binding neoantigens, calculated as the average proportion of binding neoantigens from 100 iterations of *in silico* mutations.

(B) Schematic of the analysis performed in (C) and (D).

(C) Comparison of the VAF:MHC class I K_D coefficients, segregated by immune status.

(D) Linear regression of VAF:MHC class I K_D coefficients vs. the ImmuneScore xCell. A one-way ANOVA with a post hoc Tukey HSD test was used to examine differences in (C) due to the homogeneity of variances across the groups.

See also Figure S5.

across all three groups (high infiltrate, low infiltrate, and immunosuppressed), the only parameter that differed between the groups was the HLA alleles. Therefore, we further characterized the HLA profiles across the three groups. Although no individual HLA allele is significantly different between the groups, the overall HLA profiles differ substantially (Figures S5E–S5G). Given the differences in the overall HLA profiles, we restricted all subsequent analyses of neoantigen characteristics to intra-patient metrics.

To test for restriction of binding neoantigens to low VAF, we adapted previous methods[45] and performed linear regression analyses comparing the MHC class I dissociation constant (K_D) of the optimal neoantigen (minimum K_D) and the VAF of the variant within each patient and tracked the coefficients (Figure 6B). As lower K_D indicates higher binding affinity, a positive coefficient between VAF and K_D indicates worse binding of neoantigens at higher VAF. High-infiltrate tumors from immunocompetent patients had significantly higher VAF:MHC class I K_D coefficients compared with low-infiltrate tumors from immunocompetent patients and tumors from immunosuppressed patients (Figure 6C). All high-infiltrate tumors from immunocompetent patients showed positive associations between VAF and MHC class I K_D, suggesting that binding neoantigens are restricted to lower VAF in tumors with an intact immune response. When immune infiltration was used as a continuous variable, there was a significant association of higher VAF:MHC class I K_D coefficients with increased immune infiltration (Figure 6D). Overall, although there is no net decrease in binding neoantigens in high-infiltrate tumors from immunocompetent patients, the binding neoantigens present are at lower relative VAF, supporting the role of the immune system in shaping the neoantigen profile in the tumor.

One potential mechanism for the decrease in high-frequency mutations resulting in binding neoantigens would be the loss of clonal mutations or duplication of the non-mutated allele through copy-number alterations. Therefore, 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 S5H). These results suggest that a change in copy number is not a predominant mechanism for the loss of high-frequency binding neoantigens.

In addition to accumulation of binding neoantigens at low frequency, 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.[11,12,18,46] Therefore, we modified a prior approach[12] to test whether neoantigens were enriched at low expression compared with what would be expected from the *in silico* mutations. As expected,[12] positive enrichment scores were observed across the majority of the tumors, indicating enrichment of binding neoantigens at low expression (Figure S5I). However, the observed enrichment was not significantly different than the expected enrichment given the mutational landscape alone (Figure S5I). Additionally, there was no significant difference in the enrichment scores for immunocompetent and immunosuppressed patients. This observation is consistent with a prior report that the trinucleotide mutational context derived from UV exposure biases toward more binding mutations in lowly expressed genes.[19,20] Overall, we did not detect RNA-level immunoediting in cSCC tumors from immunocompetent patients.

Neoantigen characteristics shared with validated immunogenic neoantigens are decreased in clonal relative to subclonal populations in high-infiltrate tumors from immunocompetent patients

We next sought to identify characteristics of neoantigens that are preferentially targeted through immunoediting. Because our previous analyses demonstrated that immunoediting in cSCC primarily affected the clonality of neoantigens rather than eliminating them from the tumor, we compared neoantigen characteristics between the clonal and subclonal neoantigen populations. Intra-patient comparisons between the clonal and subclonal populations have the further advantage of controlling for the HLA allele profiles between the groups.

First, all neoantigens were annotated with characteristics involved in antigen processing, presentation, and T cell receptor (TCR) recognition (Table S5). For processing and presentation, the mRNA-level expression of the gene encoding the peptide, the proteasomal cleavage potential of the peptide, and the transporter associated with antigen processing (TAP) transport potential of the peptide were considered.[47] For binding, the K_D[42,48] and stability[49] of neoantigen:MHC class I were considered, as each has been shown to provide distinct and complementary information for prioritizing immunogenic neoantigens.[40] Finally, for TCR recognition, peptides must differ from the normal peptide in either their binding to MHC class I or their interaction with the TCR. The difference in MHC binding was calculated as the differential agretopic index (DAI; ratio of the K_D of the wild-type to the mutant peptide).[50,51] For differences in the interaction with the TCR, the cross-reactivity distance was calculated, which quantifies the likelihood that a T cell would be able to discriminate the mutated from the wild-type peptide based on the position of the amino acid change and the change in the size and hydrophobicity of the amino acid.[8]

Next, we used k-means clustering to identify clusters of neoantigens with similar characteristics (Figure 7A). The optimal number of clusters was determined as the number that resulted in a local maximum in the gap statistic (Figure S6A), which compares the intra-cluster variation (sum of squares) at each number of clusters to the sum of squares under the null reference distribution (e.g., uniform distribution).[52] The fraction of neoantigens attributable to each cluster was compared across the clonal and subclonal neoantigens in high- and low-infiltrate tumors from immunocompetent patients and tumors from immunosuppressed patients. Cluster 6, which is defined by low K_D (high binding affinity), high stability, and high expression (Figures 7A and S6B–S6D), represented a lower proportion of clonal neoantigens relative to subclonal neoantigens in high-infiltrate tumors from immunocompetent patients (Figure 7B). The lower proportion of cluster 6 neoantigens in the clonal compared with subclonal populations suggests restriction of cluster 6 neoantigens from the clonal population. Critically, no comparable change was observed between clonal and subclonal neoantigens in low-infiltrate tumors from immunocompetent patients or tumors from immunosuppressed patients (Figure 7B), supporting the conclusion that the lower proportion of cluster 6 neoantigens in the clonal population in high-infiltrate tumors is immune mediated.

By contrast, cluster 7 had a similar range of K_D and stability but low expression values (Figures 7A and S6B–S6D) and showed no difference in proportion between clonal and subclonal neoantigens in high-infiltrate cSCC from immunocompetent patients (Figure 7C), supporting that immune targeting of immunogenic neoantigens requires sufficient levels of expression. The majority of neoantigens had relatively high TAP and cleavage scores, with the exception of cluster 2 (Figures S6E and S6F). Cluster 6 was not defined by changes in the DAI or cross-reactivity distance (Figures S6G and S6H), and the clusters defined by DAI and cross-reactivity distance (clusters 1 and 5, respectively) did not have a difference in proportion between clonal and subclonal neoantigens (Figures S6I and S6J). The high DAI group (cluster 1) had lower binding affinity than cluster 6 ($p < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test), suggesting that many of the cluster 1 neoantigens may be insufficiently presented on MHC class I to elicit an immune response. Of the characteristics evaluated, a combination of high binding affinity and high stability of the peptide:MHC complex, along with high expression of the mutated gene, defined the neoantigens with a lower proportion in the clonal relative to the subclonal population in high-infiltrate tumors from immunocompetent patients.

To substantiate the finding that the decrease in cluster 6 neoantigens in the clonal population is attributable to an immune response, we tested whether this cluster was enriched in validated immunogenic neoantigens. We obtained validated immunogenic neoantigens from three available datasets.[43,53,54] We then annotated these neoantigens with the same set of characteristics and assigned them to the closest matched cluster from Figure 7A (Figures 7D and S6K). We found that the distributions of the cluster assignments differed significantly between immunogenic and non-immunogenic neoantigens (Figure 7E). These results remained largely stable when analyzed separately across the three datasets (Figure S6L). Strikingly, 52.3% of immunogenic neoantigens were assigned to cluster 6, whereas only 13.2% of non-immunogenic neoantigens were assigned to cluster 6 (Figure 7E). This reflected an increase from 3.8% of all neoantigens validated as immunogenic to 13.6% of cluster 6 neoantigens validated as immunogenic. The cluster with the next largest proportion of immunogenic neoantigens was cluster 1, defined by high DAI. Cluster 1 contained 19.9% of validated

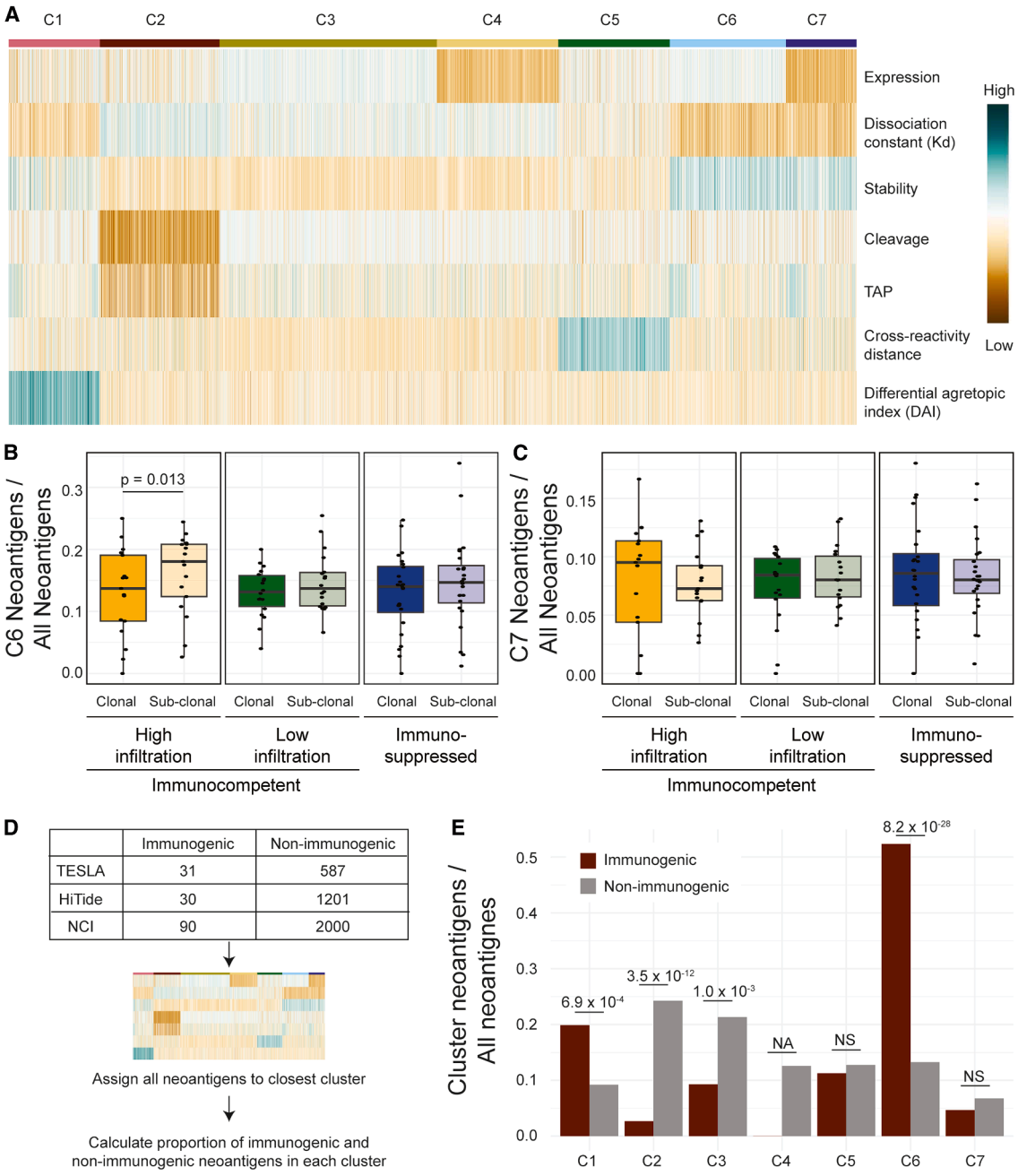


Figure 7. Neoantigen characteristics shared with validated immunogenic neoantigens are decreased in clonal relative to subclonal populations in high-infiltrate tumors from immunocompetent patients

(A) Clustering analysis across characteristics for all neoantigens from tumors with WES and RNA-seq (35 immunocompetent, 24 immunosuppressed).

(B and C) Proportion of (B) cluster 6 neoantigens and (C) cluster 7 neoantigens in the clonal and subclonal populations, segregated by immune status. Differences in scale were tested using the signed-rank test statistic for intra-patient comparisons.

(D) Schematic of the analysis applying the identified clusters to three sets of validated immunogenic neoantigens.

(E) Comparison of immunogenic neoantigens in each cluster out of all immunogenic neoantigens, and non-immunogenic neoantigens in each cluster out of all non-immunogenic neoantigens. Fisher's exact test statistic was used to compare the counts of immunogenic and non-immunogenic neoantigens in each cluster, with the reference being all other clusters combined; all *p* values were adjusted using a Bonferroni correction. NS, non-significant; NA, not applicable, included for the comparison with zero in one group.

See also Figure S6.

immunogenic neoantigens and 9.2% of non-immunogenic neoantigens, consistent with studies suggesting that DAI is a characteristic of at least a subset of neoantigens.[8,41,43,51] Overall, these findings demonstrate that neoantigens with a lower proportion in the clonal compared with subclonal populations in high-infiltrate tumors from immunocompetent patients are those most likely to be recognized by the immune system.

DISCUSSION

Although 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 study, to our knowledge, to use a comparison between tumors from immunocompetent and immunosuppressed patients to evaluate the role of the immune system in sculpting the neoantigen profile. Compared with low-infiltrate tumors from immunocompetent patients and tumors from immunosuppressed patients, high-infiltrate tumors from immunocompetent patients demonstrated (1) a lower proportion of clonal neoantigens that was not explained by purity or growth rate alone, (2) an association of increased binding to MHC class I with lower VAFs, and (3) lower proportions of neoantigens from the clonal population with characteristics defining immunogenic neoantigens. These findings provide robust evidence that the immune system sculpts the neoantigen profile.

In addition to providing evidence that immunoediting occurs, this study demonstrates the predominant effects of immunoediting on the neoantigen profile. Previous studies of immunoediting have examined whether the immune system decreases the proportion of MHC class I-binding neoantigens compared with what is expected by random chance.[11,13,19] Consistent with previous work,[19] we demonstrated that the overall proportion of binding neoantigens is equivalent to what is expected given the mutational signature of the tumor. However, there is a shift in the neoantigens most likely to be immunogenic toward lower frequency in the tumor. Restriction of clonal neoantigens is also supported by a shift toward lower CCF distributions in autochthonous tumors from immunocompetent compared with immunocompromised mice.[17] One study of the response to immune checkpoint inhibition also demonstrated a decrease in frequency of immunogenic neoantigens,[10] suggesting that restriction of clonal neoantigens is an escape mechanism that continues to predominate through tumor treatment. Supporting previous work,[20] we demonstrated that observed enrichment of immunogenic neoantigens at low expression is consistent with what is expected based on the underlying mutational signature. Together, these results suggest that the predominant impact of the immune system on the neoantigen profile of cSCC is the restriction of immunogenic neoantigens to a low frequency.

The difference in the proportion of cluster 6 neoantigens between the clonal and subclonal populations is significant but not large in magnitude, raising the question of whether the effect of immunoediting is relatively weak. However, the magnitude of the shift observed in cluster 6 neoantigens likely underestimates the impact of immunoediting, as a large proportion of that cluster is still expected to be non-immunogenic. Despite the observation that 52.3% of validated immunogenic neoantigens are assigned to cluster 6, only 13.6% of cluster 6 neoantigens are immunogenic due to the abundance of non-immunogenic neoantigens (comprising 96.2% of the starting population). The low overall frequency of immunogenic neoantigens within cluster 6 may reflect one of many things: (1) error in predicting each of the neoantigen characteristics, (2) characteristics of immunogenic neoantigens that have not yet been determined, (3) a need for smaller clusters to further refine the characteristics of immunogenic neoantigens (which would likely require a larger sample size), or (4) immunologic ignorance, where a neoantigen has the potential to elicit an immune response but has not for reasons not explored in this study (such as immunodominance hierarchies or insufficient priming and cross-presentation). Therefore, the ability to detect the influence of immunoediting despite our imperfect understanding of neoantigen immunogenicity demonstrates a central role for the immune system in sculpting the neoantigen profile.

One question arising from these findings is how the immune system restricts clonal neoantigens in immunocompetent patients. 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 in high-infiltrate tumors and the mutational signature of the tumors, which is predominantly attributable to UV exposure (and therefore likely to accumulate over time). Second, 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 under this hypothesis, a significant number of clonal (or apparently clonal) mutations must have occurred after tumorigenesis. Importantly, these mechanisms are not mutually exclusive; therefore, we consider it most likely that the immunoediting observed is attributable to a combination of survivorship bias and restriction of new mutations from becoming clonal. Finally, previously clonal mutations might be targeted by the immune system and partially removed, leaving a remnant in the subclonal population. Although this is consistent with a shift of binding neoantigens toward lower VAF, the change from a clonal mutation to a subclonal mutation would require either back-mutation or changes in copy number of an allele, both of which are unlikely to occur at high frequency. Additionally, there was no increase in the proportion of binding neoantigens in regions with copy-number alterations, as would be expected if this were a dominant mechanism. Although this mechanism could contribute, it is unlikely to be a dominant cause of the immunoediting observed.

Finally, a methodologic advance provided by this study is the conceptualization of clusters of neoantigens. We propose that the clustering analysis provides a unique approach to characterizing the overall neoantigen profile. Although previous approaches from our group and others have provided prediction tools for the immunogenicity of individual neoantigens,[8,40,41,43,44] it is consistently a challenge to summarize the

neoantigen landscape of the whole tumor. Although the exact clusters may be refined and expanded with additional data and characteristics, the ability to map new neoantigens onto the same set of clusters enables comparison of the neoantigen landscape across tumors. We envision that this may have significant utility in comparing the neoantigen landscape across cancer types as well as in detecting shifts in the neoantigen profile in other clinical scenarios, including in response to treatment with immune checkpoint inhibition or personalized neoantigen vaccines.

Overall, this work provides evidence that the immune system shapes the neoantigen profile in primary, treatment-naive human tumors by restricting clonal immunogenic neoantigens. A critical application of this work is the need to target multiple, clonal neoantigens when designing personalized vaccine strategies to prevent immune escape through increased heterogeneity. We further demonstrate that the immune infiltrate in the tumor significantly impacts the available neoantigen targets, suggesting that the immune microenvironment should be considered in determining the optimal treatment for an individual tumor. Previous studies have demonstrated an association of the clonal mutational burden with the response to immune checkpoint inhibition.[35–37,55] In combination with results showing that immunoediting restricts the clonal neoantigen burden, these findings suggest that low-immune-infiltrate tumors may have a more target-rich environment. However, effectively targeting the available clonal neoantigens in low-infiltrate tumors may require a combination approach with strategies aimed at modulating the underlying immune microenvironment, as recently reviewed.[56] These findings advance our understanding of the impact of the immune system on shaping tumor evolution and provide a foundation for further studies into the impact of these changes on individualized therapeutic approaches.

Limitations of the study

This study has several limitations that delineate opportunities for future investigation. To further elucidate the mechanisms of restriction of clonal neoantigens, future work should consider multi-regional sampling over time to allow for examination of temporal and spatial impacts on the clonal dynamics of primary human tumors. Additionally, although we demonstrate that the cluster with a lower proportion of clonal neoantigens is enriched in validated immunogenic neoantigens, we were not able to test the immunogenicity of the neoantigens identified in this study. Another open question is how the enrichment of immunogenic neoantigens in the subclonal population may influence the overall T cell response. Immunodominance hierarchies have various effects in determining T cell phenotypes.[57] Comparing the neoantigen profiles to T cell phenotypes may further our understanding of the impact of immunogenic neoantigens in the subclonal population. Finally, this study focused on MHC class I neoantigens, and additional studies are needed to examine the impact of the immune system on MHC class II neoantigens.

RESOURCE AVAILABILITY

Lead contact

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

Materials availability

This study did not generate new unique reagents.

Data and code availability

This paper analyzes a combination of existing and new data from our group and others. The newly generated RNA sequencing (RNA-seq) and whole-exome sequencing (WES) data are available on the Sequence Read Archive (SRA) under project number PRJNA1282586. The accession numbers for existing datasets are listed in the key resources table. All code has been deposited on GitHub and is available at the following link: https://github.com/HastingsLab/Immunoediting_cSCC. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by American Cancer Society - Simone Charitable Foundation grant AZ-DBG-23-1151311-01-IBCD (K.T.H.), the University of Arizona Cancer Center (E.S.B.), and National Institutes of Health grants F30CA281056 (E.S.B.) and R35GM149235 (R.N.G.). This work benefited from support from a US Department of Veterans Affairs (VA) Merit Review award I01-BX005336 (K.T.H.). The contents do not represent the views of the VA or US Government. This material is based upon high-performance computing (HPC) resources supported by the University of Arizona TRIF, UITS, and Research, Innovation, and Impact and maintained by the UArizona Research Technologies Department. We thank Sara Willis and Soham Pal for their assistance with HPC, Diego Mallo and Seema Plaiser for their technical advice, and Anne Macy for editorial assistance.

AUTHOR CONTRIBUTIONS

Conceptualization, E.S.B., K.H.B., M.A.W., and K.T.H.; methodology, E.S.B., D.C., R.N.G., B.J.L., K.H.B., M.A.W., and K.T.H.; formal analysis, E.S.B., D.C., A.C., T.M., and X.L.; investigation, E.S.B. and D.C.; data curation, E.S.B., A.H., Z.L.-R., and A.S.; resources, A.R.M. and K.T.H.; writing – original draft, E.S.B., D.C., and K.T.H.; writing – reviewing and editing, E.S.B., D.C., E.A.K., R.N.G., B.J.L., M.A.W., and K.T.H.; visualization, E.S.B., D.C., and E.A.K.; supervision, K.T.H.; funding acquisition, E.S.B. and K.T.H.

DECLARATION OF INTERESTS

A.R.M. consulted for Leo Pharma, PPD, Mallinckrodt, Nuvig, Tourmaline Bio, Janssen, Boehringer Ingelheim, and Biocryst; A.R.M. also consulted for Regeneron and Pfizer with payments to the institution. He has grant support from Kyowa, Miragen, Regeneron, Incyte, Eli Lilly, Argenx, Palvella, AbbVie, Priovant, Bristol Myers Squibb, Merck, Sci-Tech, Soligenix, and Insmed. He has received royalties from Priovant, Adelphi Values, and Clarivate. His current patents include Methods and Materials for Assessing and Treating cSCC (PCT/US2023/078902); Use of Oral JAKi in Lichen Planus (PCT/US2024/020149); Topical Ruxolitinib in Lichen Planus (PCT/US2021/053149, 2023-520085, and 21805700.8); Methods and Materials for Treating Lichen Planopilaris (registration number: 53,103); and Machine-Learning Models for Tumor Grading Using Rank-Aware Contextual Reasoning on Whole Slide Images (63/616,287).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - cSCC samples
  - Read count quantification
  - Immune cell deconvolution
  - Mutational burden and signature
  - Clonality and heterogeneity
  - Inference of growth rate and clonal ratio

- *In silico* mutations
- Machine learning estimates of tumor purities
- HLA alleles and *B2M* mutations
- Comparison of neoantigen characteristics with VAF distributions
- Enrichment of neoantigens at low expression
- Clustering analysis
- Validation of clustering analysis

- QUANTIFICATION AND STATISTICAL ANALYSES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.immuni.2026.05.005.

Received: July 2, 2025
Revised: December 20, 2025
Accepted: May 10, 2026
Published: June 11, 2026

REFERENCES

1. Dunn, G.P., Old, L.J., and Schreiber, R.D. (2004). The three Es of cancer immunoediting. Annu. Rev. Immunol. *22*, 329–360. https://doi.org/10.1146/annurev.immunol.22.012703.104803.
2. Smyth, M.J., Thia, K.Y., Street, S.E., MacGregor, D., Godfrey, D.I., and Trapani, J.A. (2000). Perforin-mediated cytotoxicity is critical for surveillance of spontaneous lymphoma. J. Exp. Med. *192*, 755–760. https://doi.org/10.1084/jem.192.5.755.
3. Street, S.E., Cretney, E., and Smyth, M.J. (2001). Perforin and interferon-gamma activities independently control tumor initiation, growth, and metastasis. Blood *97*, 192–197. https://doi.org/10.1182/blood.v97.1.192.
4. Street, S.E.A., Trapani, J.A., MacGregor, D., and Smyth, M.J. (2002). Suppression of lymphoma and epithelial malignancies effected by interferon gamma. J. Exp. Med. *196*, 129–134. https://doi.org/10.1084/jem.20020063.
5. van den Broek, M.E., Kägi, D., Ossendorp, F., Toes, R., Vamvakas, S., Lutz, W.K., Melief, C.J., Zinkernagel, R.M., and Hengartner, H. (1996). Decreased tumor surveillance in perforin-deficient mice. J. Exp. Med. *184*, 1781–1790. https://doi.org/10.1084/jem.184.5.1781.
6. Shankaran, V., Ikeda, H., Bruce, A.T., White, J.M., Swanson, P.E., Old, L.J., and Schreiber, R.D. (2001). IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature *410*, 1107–1111. https://doi.org/10.1038/35074122.
7. Borden, E.S., Kang, P., Natri, H.M., Phung, T.N., Wilson, M.A., Buetow, K.H., and Hastings, K.T. (2019). Neoantigen Fitness Model Predicts Lower Immune Recognition of Cutaneous Squamous Cell Carcinomas Than Actinic Keratoses. Front. Immunol. *10*, 2799. https://doi.org/10.3389/fimmu.2019.02799.
8. Łuksza, M., Sethna, Z.M., Rojas, L.A., Lihm, J., Bravi, B., Elhanati, Y., Soares, K., Amisaki, M., Dobrin, A., Hoyos, D., et al. (2022). Neoantigen quality predicts immunoediting in survivors of pancreatic cancer. Nature *606*, 389–395. https://doi.org/10.1038/s41586-022-04735-9.
9. Anagnostou, V., Smith, K.N., Forde, P.M., Niknafs, N., Bhattacharya, R., White, J., Zhang, T., Adleff, V., Phallen, J., Wali, N., et al. (2017). Evolution of Neoantigen Landscape during Immune Checkpoint Blockade in Non-Small Cell Lung Cancer. Cancer Discov. *7*, 264–276. https://doi.org/10.1158/2159-8290.CD-16-0828.
10. Alban, T.J., Riaz, N., Parthasarathy, P., Makarov, V., Kendall, S., Yoo, S.K., Shah, R., Weinhold, N., Srivastava, R., Ma, X., et al. (2024). Neoantigen immunogenicity landscapes and evolution of tumor ecosystems during immunotherapy with nivolumab. Nat. Med. *30*, 3209–3222. https://doi.org/10.1038/s41591-024-03240-y.
11. Rosenthal, R., Cadieux, E.L., Salgado, R., Bakir, M.A., Moore, D.A., Hiley, C.T., Lund, T., Tanić, M., Reading, J.L., Joshi, K., et al. (2019). Neoantigen-directed immune escape in lung cancer evolution. Nature *567*, 479–485. https://doi.org/10.1038/s41586-019-1032-7.
12. Wu, T., Wang, G., Wang, X., Wang, S., Zhao, X., Wu, C., Ning, W., Tao, Z., Chen, F., and Liu, X.S. (2022). Quantification of Neoantigen-Mediated Immunoediting in Cancer Evolution. Cancer Res. *82*, 2226–2238. https://doi.org/10.1158/0008-5472.CAN-21-3717.
13. Zapata, L., Caravagna, G., Williams, M.J., Lakatos, E., AbdulJabbar, K., Werner, B., Chowell, D., James, C., Gourmet, L., Milite, S., et al. (2023). Immune selection determines tumor antigenicity and influences response to checkpoint inhibitors. Nat. Genet. *55*, 451–460. https://doi.org/10.1038/s41588-023-01313-1.
14. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell *160*, 48–61. https://doi.org/10.1016/j.cell.2014.12.033.
15. Zapata, L., Pich, O., Serrano, L., Kondrashov, F.A., Ossowski, S., and Schaefer, M.H. (2018). Negative selection in tumor genome evolution acts on essential cellular functions and the immunopeptidome. Genome Biol. *19*, 67. https://doi.org/10.1186/s13059-018-1434-0.
16. Matsushita, H., Vesely, M.D., Koboldt, D.C., Rickert, C.G., Uppaluri, R., Magrini, V.J., Arthur, C.D., White, J.M., Chen, Y.S., Shea, L.K., et al. (2012). Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. Nature *482*, 400–404. https://doi.org/10.1038/nature10755.
17. Westcott, P.M.K., Muyas, F., Hauck, H., Smith, O.C., Sacks, N.J., Ely, Z.A., Jaeger, A.M., Rideout, W.M., 3rd, Zhang, D., Bhutkar, A., et al. (2023). Mismatch repair deficiency is not sufficient to elicit tumor immunogenicity. Nat. Genet. *55*, 1686–1695. https://doi.org/10.1038/s41588-023-01499-4.
18. DuPage, M., Mazumdar, C., Schmidt, L.M., Cheung, A.F., and Jacks, T. (2012). Expression of tumour-specific antigens underlies cancer immunoediting. Nature *482*, 405–409. https://doi.org/10.1038/nature10803.
19. Van den Eynden, J., Jiménez-Sánchez, A., Miller, M.L., and Larsson, E. (2019). Lack of detectable neoantigen depletion signals in the untreated cancer genome. Nat. Genet. *51*, 1741–1748. https://doi.org/10.1038/s41588-019-0532-6.
20. Claeys, A., and Van den Eynden, J. (2023). Quantification of Neoantigen-Mediated Immunoediting in Cancer Evolution-Letter. Cancer Res. *83*, 971–972. https://doi.org/10.1158/0008-5472.CAN-22-2497.
21. Gonzalez, J.L., Cunningham, K., Silverman, R., Madan, E., and Nguyen, B.M. (2017). 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. https://doi.org/10.1097/DSS.0000000000001038.
22. Hartevelt, M.M., Bavinck, J.N., Kootte, A.M., Vermeer, B.J., and Vandenbroucke, J.P. (1990). Incidence of skin cancer after renal transplantation in The Netherlands. Transplantation *49*, 506–509. https://doi.org/10.1097/00007890-199003000-00006.
23. Jensen, P., Møller, B., and Hansen, S. (2000). Skin cancer in kidney and heart transplant recipients and different long-term immunosuppressive therapy regimens. J. Am. Acad. Dermatol. *42*, 307. https://doi.org/10.1016/s0190-9622(00)90154-3.
24. Lindelöf, B., Sigurgeirsson, B., Gäbel, H., and Stern, R.S. (2000). Incidence of skin cancer in 5356 patients following organ transplantation. Br. J. Dermatol. *143*, 513–519. https://doi.org/10.1111/j.1365-2133.2000.03703.x.
25. Rabinovics, N., Mizrachi, A., Hadar, T., Ad-El, D., Feinmesser, R., Guttman, D., Shpitzer, T., and Bachar, G. (2014). Cancer of the head and neck region in solid organ transplant recipients. Head Neck *36*, 181–186. https://doi.org/10.1002/hed.23283.
26. Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220. https://doi.org/10.1186/s13059-017-1349-1.
27. Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., et al. (2016). Estimating the population abundance of tissue-infiltrating immune

and stromal cell populations using gene expression. Genome Biol. *17*, 218. https://doi.org/10.1186/s13059-016-1070-5.

28. Plattner, C., Finotello, F., and Rieder, D. (2020). Deconvoluting tumor-infiltrating immune cells from RNA-seq data using quanTIseq. Methods Enzymol. *636*, 261–285. https://doi.org/10.1016/bs.mie.2019.05.056.
29. Chang, D., and Shain, A.H. (2021). The landscape of driver mutations in cutaneous squamous cell carcinoma. NPJ Genom. Med. *6*, 61. https://doi.org/10.1038/s41525-021-00226-4.
30. Inman, G.J., Wang, J., Nagano, A., Alexandrov, L.B., Purdie, K.J., Taylor, R.G., Sherwood, V., Thomson, J., Hogan, S., Spender, L.C., et al. (2018). The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature. Nat. Commun. *9*, 3667. https://doi.org/10.1038/s41467-018-06027-1.
31. Kim, Y.S., Bang, C.H., and Chung, Y.J. (2023). Mutational Landscape of Normal Human Skin: Clues to Understanding Early-Stage Carcinogenesis in Keratinocyte Neoplasia. J. Investig. Dermatol. *143*, 1187–1196.e9. https://doi.org/10.1016/j.jid.2023.01.006.
32. Adams, A.C., Macy, A.M., Borden, E.S., Herrmann, L.M., Brambley, C.A., Sonar, S.A., Ma, T., Li, X., Hughes, A., Roe, D.J., et al. (2025). Structural changes from wild-type define tumor-rejecting neoantigens. J. Immunother. Cancer *13*, e013148. https://doi.org/10.1136/jitc-2025-013148.
33. Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. Nat. Genet. *47*, 1402–1407. https://doi.org/10.1038/ng.3441.
34. Guo, Q., Lakatos, E., Bakir, I.A., Curtius, K., Graham, T.A., and Mustonen, V. (2022). The mutational signatures of formalin fixation on the human genome. Nat. Commun. *13*, 4487. https://doi.org/10.1038/s41467-022-32041-5.
35. Lu, T., Wang, S., Xu, L., Zhou, Q., Singla, N., Gao, J., Manna, S., Pop, L., Xie, Z., Chen, M., et al. (2020). Tumor neoantigenicity assessment with CSiN score incorporates clonality and immunogenicity to predict immunotherapy outcomes. Sci. Immunol. *5*, eaaz3199. https://doi.org/10.1126/sciimmunol.aaz3199.
36. McGranahan, N., Furness, A.J.S., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S.K., Jamal-Hanjani, M., Wilson, G.A., Birkbak, N.J., Hiley, C.T., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science *351*, 1463–1469. https://doi.org/10.1126/science.aaf1490.
37. Miao, D., Margolis, C.A., Vokes, N.I., Liu, D., Taylor-Weiner, A., Wankowicz, S.M., Adeegbe, D., Keliher, D., Schilling, B., Tracy, A., et al. (2018). Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. Nat. Genet. *50*, 1271–1281. https://doi.org/10.1038/s41588-018-0200-2.
38. Beaumont, M.A., Zhang, W., and Balding, D.J. (2002). Approximate Bayesian computation in population genetics. Genetics *162*, 2025–2035. https://doi.org/10.1093/genetics/162.4.2025.
39. Wang, H., Liu, B., and Wei, J. (2021). Beta2-microglobulin(B2M) in cancer immunotherapies: Biological function, resistance and remedy. Cancer Lett. *517*, 96–104. https://doi.org/10.1016/j.canlet.2021.06.008.
40. Borden, E.S., Ghafoor, S., Buetow, K.H., LaFleur, B.J., Wilson, M.A., and Hastings, K.T. (2022). NeoScore Integrates Characteristics of the Neoantigen:MHC Class I Interaction and Expression to Accurately Prioritize Immunogenic Neoantigens. J. Immunol. *208*, 1813–1827. https://doi.org/10.4049/jimmunol.2100700.
41. Łuksza, M., Riaz, N., Makarov, V., Balachandran, V.P., Hellmann, M.D., Solovyov, A., Rizvi, N.A., Merghoub, T., Levine, A.J., Chan, T.A., et al. (2017). A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. Nature *551*, 517–520. https://doi.org/10.1038/nature24473.
42. Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). 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. https://doi.org/10.1093/nar/gkaa379.
43. Wells, D.K., van Buuren, M.M., Dang, K.K., Hubbard-Lucey, V.M., Sheehan, K.C.F., Campbell, K.M., Lamb, A., Ward, J.P., Sidney, J., Blazquez, A.B., et al. (2020). Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. Cell *183*, 818–834.e13. https://doi.org/10.1016/j.cell.2020.09.015.
44. Zhou, C., Wei, Z., Zhang, Z., Zhang, B., Zhu, C., Chen, K., Chuai, G., Qu, S., Xie, L., Gao, Y., et al. (2019). pTuneos: prioritizing tumor neoantigens from next-generation sequencing data. Genome Med. *11*, 67. https://doi.org/10.1186/s13073-019-0679-x.
45. Kim, J.Y., Cha, H., Kim, K., Sung, C., An, J., Bang, H., Kim, H., Yang, J.O., Chang, S., Shin, I., et al. (2023). MHC II immunogenicity shapes the neoepitope landscape in human tumors. Nat. Genet. *55*, 221–231. https://doi.org/10.1038/s41588-022-01273-y.
46. Westcott, P.M.K., Sacks, N.J., Schenkel, J.M., Ely, Z.A., Smith, O., Hauck, H., Jaeger, A.M., Zhang, D., Backlund, C.M., Beytagh, M.C., et al. (2021). Low neoantigen expression and poor T-cell priming underlie early immune escape in colorectal cancer. Nat. Cancer *2*, 1071–1085. https://doi.org/10.1038/s43018-021-00247-z.
47. Stranzl, T., Larsen, M.V., Lundegaard, C., and Nielsen, M. (2010). NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics *62*, 357–368. https://doi.org/10.1007/s00251-010-0441-4.
48. Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. J. Immunol. *199*, 3360–3368. https://doi.org/10.4049/jimmunol.1700893.
49. Jørgensen, K.W., Rasmussen, M., Buus, S., and Nielsen, M. (2014). NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. Immunology *141*, 18–26. https://doi.org/10.1111/imm.12160.
50. Brennick, C.A., George, M.M., Moussa, M.M., Hagymasi, A.T., Seesi, S.A., Shcheglova, T.V., Englander, R.P., Keller, G.L., Balsbaugh, J.L., Baker, B.M., et al. (2021). An unbiased approach to defining bona fide cancer neoepitopes that elicit immune-mediated cancer rejection. J. Clin. Investig. *131*, e142823. https://doi.org/10.1172/JCI142823.
51. Duan, F., Duitama, J., Al Seesi, S., Ayres, C.M., Corcelli, S.A., Pawashe, A.P., Blanchard, T., McMahon, D., Sidney, J., Sette, A., et al. (2014). Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. J. Exp. Med. *211*, 2231–2248. https://doi.org/10.1084/jem.20141308.
52. Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. *63*, 411–423. https://doi.org/10.1111/1467-9868.00293.
53. Gartner, J.J., Parkhurst, M.R., Gros, A., Tran, E., Jafferji, M.S., Copeland, A., Hanada, K.I., Zacharakis, N., Lalani, A., Krishna, S., et al. (2021). 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. https://doi.org/10.1038/s43018-021-00197-6.
54. Müller, M., Huber, F., Arnaud, M., Kraemer, A.I., Altimiras, E.R., Michaux, J., Taillandier-Coindard, M., Chiffelle, J., Murgues, B., Gehret, T., et al. (2023). Machine learning methods and harmonized datasets improve immunogenic neoantigen prediction. Immunity *56*, 2650–2663.e6. https://doi.org/10.1016/j.immuni.2023.09.002.
55. Chen, S., Xie, D., Li, Z., Wang, J., Hu, Z., and Zhou, D. (2024). Frequency-dependent selection of neoantigens fosters tumor immune escape and predicts immunotherapy response. Commun. Biol. *7*, 770. https://doi.org/10.1038/s42003-024-06460-7.
56. Baharom, F., Hermans, D., Delamarre, L., and Seder, R.A. (2025). Vax-Innate: improving therapeutic cancer vaccines by modulating T cells and the tumour microenvironment. Nat. Rev. Immunol. *25*, 195–211. https://doi.org/10.1038/s41577-024-01091-9.

57. Burger, M.L., Cruz, A.M., Crossland, G.E., Gaglia, G., Ritch, C.C., Blatt, S.E., Bhutkar, A., Canner, D., Kienka, T., Tavana, S.Z., et al. (2021). Antigen dominance hierarchies shape TCF1+ progenitor CD8 T cell phenotypes in tumors. Cell *184*, 4996–5014.e26. https://doi.org/10.1016/j.cell.2021.08.020.

58. Nassir, S., Yousif, M., Li, X., Severson, K.J., Hughes, A., Kechter, J., Hwang, A., Boudreaux, B., Bhullar, P., Zhang, N., et al. (2025). Multiomic Sequencing of Intermediate- to High-Risk Cutaneous Squamous Cell Carcinoma Identifies Critical Genes and Expression Patterns Associated with Disease and Poor Outcomes. J. Investig. Dermatol. *145*, 2060–2070.e5. https://doi.org/10.1016/j.jid.2025.01.015.

59. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

60. Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108. https://doi.org/10.1093/nar/gkt214.

61. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.

62. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

63. Auwera, G.A., and O'Connor, B.D. (2020). Genomics in the cloud: using Docker, GATK, and WDL in Terra (B. D O'Reilly Media, Inc.).

64. Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., et al. (2018). Strelka2: fast and accurate calling of germline and somatic variants. Nat. Methods *15*, 591–594. https://doi.org/10.1038/s41592-018-0051-x.

65. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.

66. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. *28*, 1747–1756. https://doi.org/10.1101/gr.239244.118.

67. Wang, S., Li, H., Song, M., Tao, Z., Wu, T., He, Z., Zhao, X., Wu, K., and Liu, X.S. (2021). Copy number signature analysis tool and its application in prostate cancer reveals distinct mutational processes and clinical outcomes. PLoS Genet. *17*, e1009557. https://doi.org/10.1371/journal.pgen.1009557.

68. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput. Biol. *12*, e1004873. https://doi.org/10.1371/journal.pcbi.1004873.

69. Revkov, E., Kulshrestha, T., Sung, K.W.K., and Skanderup, A.J. (2023). PUREE: accurate pan-cancer tumor purity estimation from gene expression data. Commun. Biol. *6*, 394. https://doi.org/10.1038/s42003-023-04764-8.

70. Hörst, F., Rempe, M., Heine, L., Seibold, C., Keyl, J., Baldini, G., Ugurel, S., Siveke, J., Grünwald, B., Egger, J., et al. (2024). CellViT: Vision Transformers for precise cell segmentation and classification. Med. Image Anal. *94*, 103143. https://doi.org/10.1016/j.media.2024.103143.

71. Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et al. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. *33*, 1152–1158. https://doi.org/10.1038/nbt.3344.

72. Sturm, G., Finotello, F., and List, M. (2020). 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. https://doi.org/10.1007/978-1-0716-0327-7_16.

73. Hundal, J., Carreno, B.M., Petti, A.A., Linette, G.P., Griffith, O.L., Mardis, E.R., and Griffith, M. (2016). pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. Genome Med. *8*, 11. https://doi.org/10.1186/s13073-016-0264-5.

74. Kalari, K.R., Nair, A.A., Bhavsar, J.D., O'Brien, D.R., Davila, J.I., Bockol, M.A., Nie, J., Tang, X., Baheti, S., Doughty, J.B., et al. (2014). MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. BMC Bioinform. *15*, 224. https://doi.org/10.1186/1471-2105-15-224.

75. Liao, Y., Smyth, G.K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. *47*, e47. https://doi.org/10.1093/nar/gkz114.

76. Fairley, S., Lowy-Gallego, E., Perry, E., and Flicek, P. (2020). The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. Nucleic Acids Res. *48*, D941–D947. https://doi.org/10.1093/nar/gkz836.

77. Saunders, C.T., Wong, W.S.W., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. Bioinformatics *28*, 1811–1817. https://doi.org/10.1093/bioinformatics/bts271.

78. Braverman, J.M., Hudson, R.R., Kaplan, N.L., Langley, C.H., and Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics *140*, 783–796. https://doi.org/10.1093/genetics/140.2.783.

79. Fu, Y.X. (1995). Statistical properties of segregating sites. Theor. Popul. Biol. *48*, 172–197. https://doi.org/10.1006/tpbi.1995.1025.

80. Csilléry, K., François, O., and Blum, M.G.B. (2012). abc: an R package for approximate Bayesian computation (ABC). Methods Ecol. Evol. *3*, 475–479. https://doi.org/10.1111/j.2041-210X.2011.00179.x.

81. Durrett, R. (2013). Population Genetics of Neutral Mutations in Exponentially Growing Cancer Cell Populations. Ann. Appl. Probab. *23*, 230–250. https://doi.org/10.1214/11-aap824.

82. Matey-Hernandez, M.L., Danish; Pan; Genome Consortium, Brunak, S., and Izarzugaza, J.M.G. (2018). Benchmarking the HLA typing performance of Polysolver and Optitype in 50 Danish parental trios. BMC Bioinform. *19*, 239. https://doi.org/10.1186/s12859-018-2239-6.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
GeneRead FFPE kit	Qiagen	discontinued
QIAmp Advanced DNA FFPE kit	Qiagen	Cat #56604
RNeasy FFPE kit	Qiagen	Cat #73504
Deposited data		
Whole exome sequencing data (new cohort)	This study	SRA: PRJNA1282586
Transcriptome data (new cohort)	This study	SRA: PRJNA1282586
Whole exome sequencing data (original cohort)	Nassir et al.[58]	SRA: PRJNA1202062
Transcriptome data (original cohort)	Nassir et al.[58]	GEO: GSE284467
Validated immunogenic neoantigens	Muller et al.[54]	https://figshare.com/s/147e67dde683fb769908
Software and algorithms		
FastQC	N/A	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
STAR	Dobin et al.[59]	https://github.com/alexdobin/STAR
Subread	Liao et al.[60]	https://subread.sourceforge.net/
quanTIseq	Plattner et al.[28]	https://icbi.i-med.ac.at/software/quantiseq/doc/
xCell	Aran et al.[26]	https://comphealth.ucsf.edu/app/xcell
MCP-counter	Becht et al.[27]	https://github.com/ebecht/MCPcounter
bbduk	N/A	https://sourceforge.net/projects/bbmap/
bwa-mem v. 0.7.17	Li et al.[61]	https://bio-bwa.sourceforge.net/
Samtools	Li et al.[62]	https://www.htslib.org/
Picard	Auwera et al.[63]	https://broadinstitute.github.io/picard/
GATK Mutect2 v4.1.8.1	Auwera et al.[63]	https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2
Strelka2 v2.9.2	Kim et al.[64]	https://github.com/Illumina/strelka
VEP	McLaren et al.[65]	https://useast.ensembl.org/info/docs/tools/vep/index.html
MAF tools	Mayakonda et al.[66]	https://bioconductor.org/packages/devel/bioc/vignettes/maftools/inst/doc/maftools.html
Sigminer v.2.3.1	Wang et al.[39,67]	https://github.com/ShixiangWang/sigminer
CNVkit	Talevich et al.[68]	https://cnvkit.readthedocs.io/en/stable/
Puree	Revkov et al.[69]	https://github.com/skandlab/PUREE
CellViT	Horst et al.[70]	https://github.com/TIO-IKIM/CellViT
Polysolver	Shukla et al.[71]	https://github.com/researchapps/polysolver
netMHCpan4.0	Jurtz et al.[48]	https://services.healthtech.dtu.dk/services/NetMHCpan-4.0/
netCTLpan1.1	Stranzl et al.[47]	https://services.healthtech.dtu.dk/services/NetCTLpan-1.1/
netMHCstab1.0	Jorgensen et al.[49]	https://services.healthtech.dtu.dk/services/NetMHCstabpan-1.0/
clusGap	Tibshirani et al.[52]	https://www.rdocumentation.org/packages/cluster/versions/2.1.8.1/topics/clusGap

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
immunedeconv v. 2.1.0	Sturm et al.[72]	https://github.com/omnideconv/immunedeconv
pVAC-Seq tools v.3.0.5	Hundal et al.[73]	https://pvactools.readthedocs.io/en/latest/pvacseq.html

METHOD DETAILS

cSCC samples

This study analyzed WES and RNA-Seq data previously obtained by our team from 60 patients with cSCC (42 immunocompetent and 18 immunosuppressed).[58] Samples with only RNA-Seq data were excluded from the study. An additional set of WES and RNA-Seq data was generated from 17 patients with cSCC (3 immunocompetent and 14 immunosuppressed). Immunosuppressed patients were included if they were solid organ transplant recipients or had been treated with systemic azathioprine, tacrolimus, prednisone, methylprednisone, sirolimus, mycophenolate mofetil, methotrexate, cyclosporin, or a combination of the aforementioned medications. One patient considered immunosuppressed in the prior study[58] was excluded because the treatment inhibited only the Th2 response (dupilumab for Crohn's disease). The resulting sample sizes and full demographic data are provided in Table S2. For skin site annotated throughout, head/neck includes ear, cheek, forehead, jaw, lip, temple, and scalp; trunk includes back; upper extremity (UE) includes arm and hand; and lower extremity (LE) includes leg and foot.

Primary tumors and normal tissue controls were manually macrodissected from serial sections, guided by the annotations of a board-certified dermatologist. A GeneRead formalin-fixed paraffin-embedded (FFPE) kit was used to isolate DNA from the FFPE tissues and chemically correct FFPE-induced mutations following the manufacturer's recommendations (Qiagen). For the second set of samples, the QIAmp Advanced DNA FFPE kit was used due to the discontinuation of the GeneRead FFPE kit; the only difference noted in the protocols was the double-lyse protocol included in the QIAmp Advanced DNA FFPE kit. RNA was isolated using the RNAeasy FFPE kit (Qiagen). Specimens were prepared for sequencing as described.[58] All exome libraries were prepared using the Agilent SureSelectXT Low Input prep reagents and the SureSelectXT Human All Exon v5 + UTR capture reagents. WES was performed with an average coverage of approximately 80x and RNA-Seq with approximately 60 million reads per sample. The WES samples had a read length of 100 bp for the majority of the existing data, with a read length of 150 bp for a subset of the existing data and all of the new data. All of the existing data (60 cSCC samples) were sequenced on the Illumina HiSeq 4000, while the new data were sequenced on the Illumina NovaSeq X due to the discontinuation of support for the Illumina HiSeq 4000.

Read count quantification

RNA FASTQ files were visualized for quality with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). RNA-Seq data were aligned to the GRCh38 reference genome using STAR.[59,74] Gene-level read count quantification was performed with Subread *featureCounts* function to obtain raw reads (Table S1).[60,75] Read counts were transformed to transcripts per million (TPM) for downstream analyses.

Immune cell deconvolution

Immune cell deconvolution was performed across all samples using quanTIseq,[28] xCell,[26] and MCP-counter,[27] applied using the immunedeconv (v.2.1.0) package in R.[72] For clear visualization, results were scaled by row for heatmap visualization. High and low infiltration groups were split on the mean immune infiltration calculated by xCell. Infiltration group was used as an annotation across the results from quanTIseq, xCell, and MCP-counter to demonstrate the consistency of high infiltration in these tumors, estimated by multiple approaches. xCell estimates the prevalence of the following cell types: B cells (naive, plasma, memory, and class-switched memory), CD4+ T cells (non-regulatory, central memory, effector memory, memory, naive, Th1, Th2, Tregs), CD8+ T cells (central memory, effector memory, naive, gamma delta, and NK), common lymphoid progenitor cells, common myeloid progenitor cells, granulocyte-monocyte progenitor cells, hematopoietic stem cells, dendritic cells (myeloid, myeloid activated, and plasmacytoid), eosinophils, macrophages (M1 and M2), monocytes, mast cells, neutrophils, and NK cells, cancer associated fibroblasts, endothelial cells. The ImmuneScore is calculated as a sum of the estimates of immune cell abundances from xCell.

Mutational burden and signature

WES FASTQ files were visualized for quality with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). WES FASTQ files were trimmed using bbduk (sourceforge.net/projects/bbmap/) with a quality-trim to Q30 using the Phred algorithm, a minimum length of 75 base pairs, and a minimum average quality of at least 20 for the read. Subsequently, quality was revisualized after trimming and showed mean quality scores above 30 for the full length of the reads as anticipated. Read mapping to the GRCh38 genome from 1000 genomes[76] was performed with bwa-mem (v0.7.17).[61] Bam files were converted to sam files with SAMtools, and mate coordinates were corrected with *fixmates* from SAMtools.[62] Duplicates were marked with Picard *MarkDuplicates*.[63] Somatic variants were identified with a combination of Mutect2 (v4.1.8.1)[63] and Strelka2(v2.9.2).[64,77] GATK Mutect2 variants were analyzed

for read biases using *LearnReadOrientationModel*, and these biases were incorporated into the GATK Mutect2 filtering. Consistent with previous work,[40] only variants identified by both GATK Mutect2 and Strelka2 were kept, to ensure high-fidelity calls. Subsequently, variants were further filtered for those with a minimum read depth of at least 10 total reads and at least 5 reads supporting the alternative allele (Table S3). Mutations were annotated using the Variant Effects Predictor (VEP) from Ensmebl with cache version 109.[65] 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.[66] Mutational signatures were deconvoluted using the Sigminer (v.2.3.1) package from R.[67]

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:

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

Where P is the purity, calculated from the RNA-Seq data with Puree,[69] the CN_t is the copy number of the allele in the tumor calculated with CNVkit, and CN_n is the normal copy number.[68] Somatic variants with a CCF greater than 1 were set to 1 to more accurately reflect the percentage 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.[46] 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.[78] Under the standard neutral model (SNM), that is, a model of constant population size with no selection,[79] 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.[80,81] This results in a faster decay of the SFS under population expansion and an excess of low-frequency variants relative to the SNM. To illustrate the expected differences in the SFS based on the tumor growth rate, we performed two simulations holding all parameters equal aside from the growth rate (Figures 4A and 4B). The remaining simulation parameters are as follows: the number of sequenced cells (150,000), the ratio of clonal:subclonal mutations in the sequenced sample (0.05), 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.

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 a 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 that 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 = 5). 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, Figure 4C) and estimated RNA purities (Figure 3C). 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".
- Ratio of clonal:subclonal mutations, *C* (simulated range [0, 0.5] 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 ratios of clonal:subclonal mutations from the normalized VAF simulations. The rejection method used a tolerance of 0.001; all analyses were performed with the *abc* package implemented in R.[80] 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 that 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, $p < 0.0005$, Figure S5A).

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 $O_{s,i}$. Figure S5B 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 (Figure S5C). Although the correlation is strong ($R^2 = 65\%$, $p < 0.0005$), it becomes even stronger ($R2 = 74\%$, $p < 0.0005$) when we use the inferred gene mutability to calculate the expected number of nonsynonymous mutations per gene, particularly for long genes (Figure S5D). 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 (Figure S5B). 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. The proportion of neoantigens expected to bind MHC class I was calculated for each set of *in silico* mutations and averaged across 100 iterations.

Machine learning estimates of tumor purities

We estimate tumor purity of the cSCC whole slide images using a two-step pipeline. First, an experienced, board-certified dermatologist manually annotated all tumor regions in each whole slide image that corresponded to sections used for downstream genomic

sequencing. These annotations were used to generate binary masks delineating tumor-containing regions, which served as input for nuclei segmentation. Next, we employed CellViT,[70] a Vision Transformer-based deep learning model, to perform nuclei segmentation and classification within the identified tumor regions based on binary masks. CellViT was pretrained on over 104 million histopathology image patches spanning 19 cancer types. The model classifies individual nuclei into five phenotypic categories: neoplastic, inflammatory, epithelial, dead, and connective. Tumor purity was computed as the proportion of neoplastic nuclei relative to the total number of nuclei detected within the annotated tumor region. The machine learning estimates of purity were then compared to the RNA estimates of purity, and examples of the cell annotations are provided to demonstrate the accuracy of the nuclei assignments.

HLA alleles and *B2M* mutations

Patient-specific HLA alleles were identified in both tumor and normal samples using Polysolver (Table S4).[71,82] 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. For the comparison of the HLA allele distributions, we used a χ^2 test statistic for proportions for each allele.

Comparison of neoantigen characteristics with VAF distributions

VEP-annotated mutations were processed into 21mer amino acid sequences using pVAC-Seq tools version 3.0.5.[73] A set of characteristics was calculated for each neoantigen from every patient (Table S5). mRNA expression of each neoantigen was calculated at the gene level and quantified as TPM. The MHC class I:neoantigen dissociation constant (Kd) was then predicted for every 9 and 10mer sequence to each patient allele and non-patient allele with netMHCpan4.0.[42,48] TAP transport and proteasomal cleavage potential were predicted with NetCTLpan1.1.[47] Neoantigen:MHC class I binding stability was calculated with NetMHCstab1.0.[49] DAI was calculated as the ratio of the dissociation constant for the wild-type peptide compared to the neoantigen as described.[40,41] Cross-reactivity was calculated using the methods and code previously described.[8] Since the previous model was trained only on 9mer sequences, the position factor for position 10 was presumed to be equivalent to that calculated for position 9. 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. A linear regression was then fitted between the Kd and the VAF for each patient, and the coefficient of the model was tracked. The coefficients were subsequently compared across high and low infiltrate tumors from immunocompetent patients and tumors from immunosuppressed patients. Finally, we compared the proportion of binding neoantigens out of all neoantigens in regions with copy number alterations compared to the proportion of binding neoantigens out of all neoantigens in regions without copy number alterations.

Enrichment of neoantigens at low expression

Enrichment of predicted binding neoantigens at low expression was calculated with a method similar to what has been previously described.[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 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. The same process was repeated for all 100 iterations of *in silico* mutations, and the average expected enrichment score was calculated for each patient.

Clustering analysis

A set of characteristics was calculated for each neoantigen from every patient as described above. 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 K-means 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 (Figure S6A).[52]

Validation of clustering analysis

Validated immunogenic neoantigens were obtained from the data curated by Müller et al.[54] This included neoantigens validated by Müller et al. (HiTide)[54] as well as those from Wells et al. (TESLA)[43] and Gartner et al. (NCI).[53] We subset the available data with the following inclusion criteria: 1) peptides directly tested for an immune response, 2) peptides that were 9 or 10 amino acids in length, 3)

peptides mapped uniquely to a single optimal HLA allele, and 4) peptides with a maximum of one nucleotide difference between the mutant and wild-type peptides. Because of the exceptionally large number of non-immunogenic neoantigens in the dataset from Gartner et al., we subset a random 2,000 non-immunogenic neoantigens. The final neoantigen counts were 30 immunogenic and 1,201 non-immunogenic neoantigens from HiTide, 60 immunogenic and 2000 randomly-selected non-immunogenic neoantigens from NCI, and 31 immunogenic and 587 non-immunogenic neoantigens from TESLA. The mRNA expression (in units of TPM), cleavage scores, and TAP scores were provided by Müller et al. We calculated the MHC class I binding for mutant and wild-type, the DAI, and the cross-reactivity distance for each of the peptides as described above. All data was normalized and scaled as described above. We then calculated the distance of each new neoantigen to the kmean centers previously determined, and assigned the neoantigen to the cluster with the minimal distance from the kmean center. This analysis was performed using the *flexclust* package in R statistical software. Finally, we compared the proportion of immunogenic and non-immunogenic neoantigens assigned to each cluster.

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 assumptions for parametric tests were violated (e.g., homogeneity of variance), as well as when the primary comparison included 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 $p < 0.05$. 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.