bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Autoimmune Alleles at the Major Histocompatibility Locus Modify Melanoma Susceptibility

Authors: James Talwar[1,2], David Laub[1], Meghana Pagadala[3], Andrea Castro[1,2], McKenna Lewis[4], Georg E. Luebeck[5], Gerald P. Morris[6], Rany M. Salem[7], Wesley K. Thompson[8], Kit Curtius[2,9,10], Maurizio Zanetti[9,11,12], Hannah Carter[1,2,9]

**Affiliations:**

1 Department of Medicine, Division of Medical Genetics, University of California San Diego, La Jolla, CA 92093, USA
2 Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA 92093, USA
3 Biomedical Science Program, University of California San Diego, La Jolla, CA 92093, USA
4 Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA
5 Public Health Sciences Division, Herbold Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
6 Department of Pathology, University of California San Diego, La Jolla, CA 92093, USA
7 Division of Epidemiology, Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA 92093, USA
8 Division of Biostatistics and Department of Radiology, Population Neuroscience and Genetics Lab, University of California San Diego, La Jolla, CA 92093, USA
9 Moores Cancer Center, University of California San Diego, La Jolla, CA 92093, USA
10 Division of Biomedical Informatics, Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA
11 The Laboratory of Immunology, University of California San Diego, La Jolla, CA 92093, USA
12 Department of Medicine, Division of Hematology and Oncology, University of California San Diego, La Jolla, CA 92093, USA

**Keywords**: Melanoma, autoimmunity, major histocompatibility complex, polygenic risk scores

**Correspondence to:**
hkcarter@health.ucsd.edu

## Abstract:

Autoimmunity and cancer represent two different aspects of immune dysfunction. Autoimmunity is characterized by breakdowns in immune self-tolerance, while impaired immune surveillance can allow for tumorigenesis. The class I major histocompatibility complex (MHC-I), which displays derivatives of the cellular peptidome for immune surveillance by CD8+ T cells, serves as a common genetic link between these conditions. As melanoma-specific CD8+ T-cells have been shown to target melanocyte-specific peptide antigens more often than melanoma-specific

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

antigens, we investigated whether vitiligo and psoriasis predisposing MHC-I alleles conferred a melanoma protective effect. In individuals with cutaneous melanoma from both The Cancer Genome Atlas (N = 451) and an independent validation cohort (N = 586), MHC-I autoimmune allele carriers are significantly associated with a later age of melanoma diagnosis. MHC-I autoimmune allele contributions to melanoma risk are not captured by current polygenic risk scores (PRS) and incorporation of autoimmune MHC allele presence can improve relative risk stratification. Mechanisms of autoimmune protection were neither associated with improved melanoma-driver mutation association nor improved gene-level conserved antigen presentation relative to common alleles (population frequency ≥ 1%). However, autoimmune alleles showed a marked affinity relative to common alleles for particular windows of melanocyte conserved antigens suggesting a potential relationship between antigen processing, binding, and cell-surface presentation. Overall, this study presents evidence that MHC-I autoimmune risk alleles modulate melanoma risk currently unaccounted for by current PRS.

## **Introduction**:

The incidence of cutaneous melanoma, the most common form, has seen an increase globally, particularly in Western countries.[1,2] Early detection is a major determinant of overall disease prognosis with the 5-year survival rate dropping precipitously from 99% to 27.3% for local versus distant disease, respectively.[3] Models developed for early disease detection are often built around well-known environmental and host risk factors including ultraviolet radiation exposure,[4–6] pigmentary phenotypes,[7–9] melanocytic nevi count,[10,11] sex,[12,13] age,[12,13] telomere length,[14,15] immunosuppression,[16,17] and family history.[18,19] However, though cutaneous melanoma ranks among the most heritable forms of cancer with an estimated heritability of 58%,[20] the majority of genetic susceptibility remains unaccounted for.

Cutaneous melanoma is also considered among the most immunogenic forms of cancer. Melanoma exhibits one of the highest mutation burdens across cancers, which is driven primarily by the mutagenic influence of ultraviolet radiation exposure.[21,22] This increases the number of neoepitopes presented to the immune system and plays an essential role in immune surveillance. Immunosuppression, though, impairs the immune system's cytotoxic potential and is a documented risk factor for increased melanoma incidence.[16,17] Lymphocyte infiltration and melanoma-specific antibodies have been shown to be powerfully prognostic factors as well.[23,24] Immune traits themselves show considerable heritability,[25–27] and early investigations suggest that heritable immune alleles also contribute to melanoma risk.[28–30] In contrast to cancer, where poor immune function is a risk factor,[31,32] increased sensitivity of the immune system can lead to autoimmune disorders.[33] This dichotomy has led to speculation that induction of autoimmunity in cancer patients could lead to tumor regression and better immunotherapy efficacy,[34–39] though studies investigating the relationship between autoimmunity and cancer risk have returned mixed findings.[40–42] If so, incorporating information about immune variation related to autoimmune disorders into genetic risk scores may further improve PRS for projecting melanoma risk.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The class I major histocompatibility complex (MHC-I) represents a fundamental component of the antigen-directed immune response common to cancer and autoimmunity. MHC-I binds and displays peptide antigens derived primarily from intracellular proteins on the cell surface for immune surveillance by CD8+ T-cells.[43] In cancer, neopeptides containing somatic mutations unique to the tumor genome, when displayed by MHC-I, can be recognized as foreign by CD8+ T cells, triggering the release of cytotoxic granules.[44,45] The peptide-binding specificity of MHC-I is determined by three highly polymorphic genes, *HLA-A*, *HLA-B* and *HLA-C*, encoded at the Human Leukocyte Antigen (HLA) locus on chromosome 6. The specific set of HLA alleles carried by an individual has been found to impose selective constraints on the developing tumor genome [46–48] and modify response to immunotherapy.[49–51] MHC-I loss, often due to deletion or mutation of HLA genes, is one mechanism of immune evasion during tumor development.[52,53]

MHC-I also plays a role in several skin-specific autoimmune disorders. In vitiligo, destruction of melanocytes and consequent loss of skin pigmentation is mediated by CD8+ T-cell responses to self-antigens displayed by MHC-I.[54,55] Another skin autoimmune disorder involving CD8+ T cell responses is psoriasis,[56,57] which is characterized by dermal leukocyte infiltration and hyperproliferation of keratinocytes.[57] CD8+ T-cells in psoriasis have been shown to target melanocytes in patients carrying particular MHC-I alleles.[58,59] Both of these conditions also share a risk variant rs9468925 in HLA-B/HLA-C,[60] supporting a shared MHC-I driven etiology, which is consistent with disease co-occurrence findings in clinical investigations.[61,62]

Intriguingly, multiple vitiligo risk alleles implicated by genome-wide association studies have been shown to exhibit protection from cutaneous melanoma,[28,29] and emergence of vitiligo during immunotherapy treatment of melanoma patients has been associated with better responses.[63–66] Though psoriasis has also been documented as an immune-related adverse effect of immunotherapy, the association with melanoma prognosis is far less characterized.[67,68]

Altogether these findings support the potential for class I autoimmune HLA alleles to modify melanoma risk and inform risk prediction. To further investigate this possibility, we evaluated autoimmune HLA carrier status in cutaneous melanoma samples from The Cancer Genome Atlas (TCGA). Skin autoimmune alleles were associated with a significantly later age at diagnosis among melanoma cases, which was recapitulated in a validation cohort of 586 individuals assembled from the UKBiobank and 4 other published melanoma genome sequencing studies. We further investigated the peptide-specificity of autoimmune alleles for a set of 215 melanoma-specific driver mutations spanning 172 genes, and for conserved and cancer antigens previously implicated in melanoma-directed immunity. These analyses highlight a protective role for MHC-I in the context of melanoma development.

## **Results**:

### *MHC-I Autoimmune Alleles Associate with a Later Age of Diagnosis in Melanoma*

To investigate the relationship between autoimmunity and melanoma development, we sought evidence that MHC-I autoimmune alleles could provide protection from melanoma. We defined a

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

set of 8 MHC-I alleles based on documented links to skin autoimmune conditions. This set included 3 alleles linked to psoriasis (HLA-B*27:05, HLA-B*57:01, HLA-C*12:03),[69–73] 2 alleles linked to vitiligo (HLA-A*02:01, HLA-B*13:02),[74–79] and 1 allele linked to both conditions (HLA-C*06:02).[78–81] We also included 2 alleles (HLA-B*39:06, HLA-B*51:01) with postulated psoriasis associations,[69,82] but strong associations with other autoimmune conditions, specifically type 1 diabetes [83–85] and Behcet's disease,[86] respectively.

Using individuals with skin cutaneous melanoma tumors (SKCM) from the TCGA as our discovery cohort (**Supplementary Table 1**), we called MHC-I genotypes using two exome based methods: POLYSOLVER and HLA-HD (*Methods: Datasets*).[87,88] Autoimmune allele frequencies in the discovery cohort were present at the population distributions as reported by the National Marrow Donor Program (NMDP)[89] (**Supplementary Table 2**). Individuals under 20 years of age were also excluded from further analysis given their increased likelihood of harboring rare germline predisposing risk variants.

To evaluate the effect of carrying an autoimmune (AI) allele, we partitioned the discovery cohort into two groups: those with at least one AI allele and those lacking any of these alleles. As HLA-A*02:01 had high frequency relative to other alleles, and thus could potentially dominate the analysis, we considered AI carrier status both excluding and including the HLA-A*02:01 allele. We first assessed the potential for AI allele carrier status to be confounded by sex or UV exposure, two factors that influence melanoma incidence. Men have a well-documented higher risk of developing melanoma,[90,91] however, we found no significant sex specific age differences across AI allele status (**Supplementary Fig. 1A**, $p = 0.247$). UV-associated mutational signatures correlated strongly with overall tumor mutation burden in the discovery cohort (Pearson $R = 0.984$; **Supplementary Fig. 1D**). There were also no significant differences in mutation burden or UV-associated mutational signatures between AI allele carriers and non-carriers (**Supplementary Fig. 1B-D**; A*02:01 included: $p_{mutation} = 0.365$, $p_{UV} = 0.186$; A*02:01 excluded: $p_{mutation} = 0.352$, $p_{UV} = 0.415$).

We hypothesized that in an all-cancer cohort, a protective effect would manifest as delayed disease onset relative to individuals without these alleles. Across the discovery cohort, AI carrier status including HLA-A*02:01 failed to significantly associate with age (**Supplementary Fig. 1E**, $p = 0.316$). However, when HLA-A*02:01 was excluded, AI allele carrier status was significantly associated with a median later age of melanoma diagnosis of 5 years (**Fig. 1A**, $p = 0.002$). For subsequent analyses we therefore defined AI allele carrier status based on the other 7 AI alleles; individuals carrying only HLA-A*02:01 were considered non-carriers.

For the revised definition of carrier status, later age of diagnosis was conserved across sex with a median age difference of 6 years in males and 4 years in females (**Fig. 1B**, $p_{male} = 0.011$, $p_{female} = 0.033$). We used an ordinary least squares (OLS) regression to model the effect of carrying an AI allele on age at diagnosis with sex and mutation burden as covariates and these findings remained significant with a predicted 4.07 delayed years to melanoma diagnosis (**Fig. 1C**, $p = 0.005$). To ensure no single AI allele drove these findings we conducted a leave-one-out analysis, evaluating AI age effects across all 7 single allele exceptions. Regardless of which

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

allele was held out, we consistently observed a significant relationship between AI status and age of diagnosis (**Supplementary Fig. 2**), with a 5 year later median age of melanoma diagnosis (median age without AI = 57; median age with AI = 62). We did not observe significant differences at the level of HLA supertype, suggesting that the effects are specific to individual autoimmune alleles (**Supplementary Fig. 3**), though the B44 supertype, to which none of the AI alleles belong, trended towards an earlier age of diagnosis (p = 0.176, median earlier age of diagnosis difference = 4 years). We further evaluated whether carrying multiple AI alleles had an additive effect. Using a linear model to predict age of diagnosis as a function of an individual's total number of autoimmune alleles, the discovery cohort showed a significant 2.6 delayed years to melanoma diagnosis per autoimmune allele (**Fig. 1D**, p = 0.002; CI = [0.935, 4.262]).

To confirm the generality of these findings, we investigated the age of melanoma diagnosis and autoimmune allele presence in an independent validation cohort of 586 individuals diagnosed with cutaneous melanoma, compiled from 4 published dbGaP studies [92–96] and the UKBioBank. Despite our best efforts to construct a validation cohort similar to our discovery cohort, we observed notable differences in the distribution of sex and age (**Supplementary Table 1**), especially for the UKBioBank. In contrast to TCGA, females in the validation cohort were significantly associated with an earlier age of melanoma diagnosis independent of AI status (**Supplementary Fig. 4A**; p = 0.0034, median earlier age of diagnosis difference = 3 years). Given not only the existence, but also the direction of observed sex-specific age effects (i.e, females exhibiting an earlier age of diagnosis, rather than males), this potentially represents an intrinsic selection bias within the studies from which our validation cohort was compiled. Despite this, we again observed that having at least one autoimmune-linked MHC-I allele was significantly associated with a later age of diagnosis (**Fig. 1E**, p = 0.0435; median age separation = 1 year). While the direction of this effect was conserved across males with a median age difference of 1 year, we did not observe any significant differences in females (**Supplementary Fig. 4B**; $p_{male}$ = 0.055, $p_{female}$= 0.236).

As most validation samples lacked tumor sequencing data, it was not possible to estimate UV exposure. Thus, our regression analysis was limited to presence of an AI allele and sex. Validation individuals with at least one AI allele had a predicted 2.09 delayed years to melanoma diagnosis relative to those without any of these alleles (p = 0.083; CI = [-0.273, 4.460]). By comparison, the discovery cohort showed a predicted 4.0 delayed years to melanoma diagnosis when only AI status and sex are considered (p = 0.006; CI = [1.154, 6.850]). In the 559 validation cohort individuals with fully-resolved HLA types, we observed a similar trend for total AI allele burden to associate with later age at diagnosis, with a predicted 0.727 year delay to melanoma diagnosis per autoimmune allele, though these results did not reach statistical significance (**Fig. 1F**, p = 0.328; CI = [-0.731, 2.184]).

We also attempted to identify other HLA alleles associated with age at diagnosis by comparing individuals with a particular allele to all individuals without that allele. While none of the associations were significant after multiple hypothesis testing correction, we did note that HLA-B*27:02, showed a marked effect across cohorts with an earlier median age of diagnosis of 10 years in the SKCM-TCGA (15 individuals; p = 0.0496; $p_{adj}$ = 0.483) and 11 years in the

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

validation cohort (3 individuals; $p = 0.0332$; $p_{adj} = 0.388$) respectively (**Supplementary Fig. 5**). However, given the limited cohort sizes with sequencing data, coupled with the highly polymorphic nature of the HLA, this analysis is likely underpowered and may warrant further investigation in larger cohorts.

Finally, we evaluated whether these 7 alleles associated with skin autoimmune disorders also correlated with age at diagnosis in other cancer types. Across the TCGA, four other cancers (ACC, ESCA, PCPG, SARC) exhibited a similar later age of diagnosis association with AI allele presence, though these effects did not reach significance after multiple testing correction (**Supplementary Fig. 6A**; Delayed median age of diagnosis in those with AI allele vs. those without: ACC = 11 years, ESCA = 4 years, PCPG = 7.5 years, SARC = 4.5 years; $p = 0.166$). However, there was a noticeable size discrepancy between these four cancer types, which ranged between 89 and 255 individuals, and our discovery cohort (N = 451), potentially limiting the statistical power of these findings. We also repeated this analysis for HLA-B*27:02 and observed two other cancers (KIRC and READ) that exhibited the same early age of diagnosis effect as SKCM before multiple testing correction (**Supplementary Fig. 6B**). These trends merit further investigation with larger sample sizes, and analysis could be extended to include autoimmune alleles associated with other tissues besides skin.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

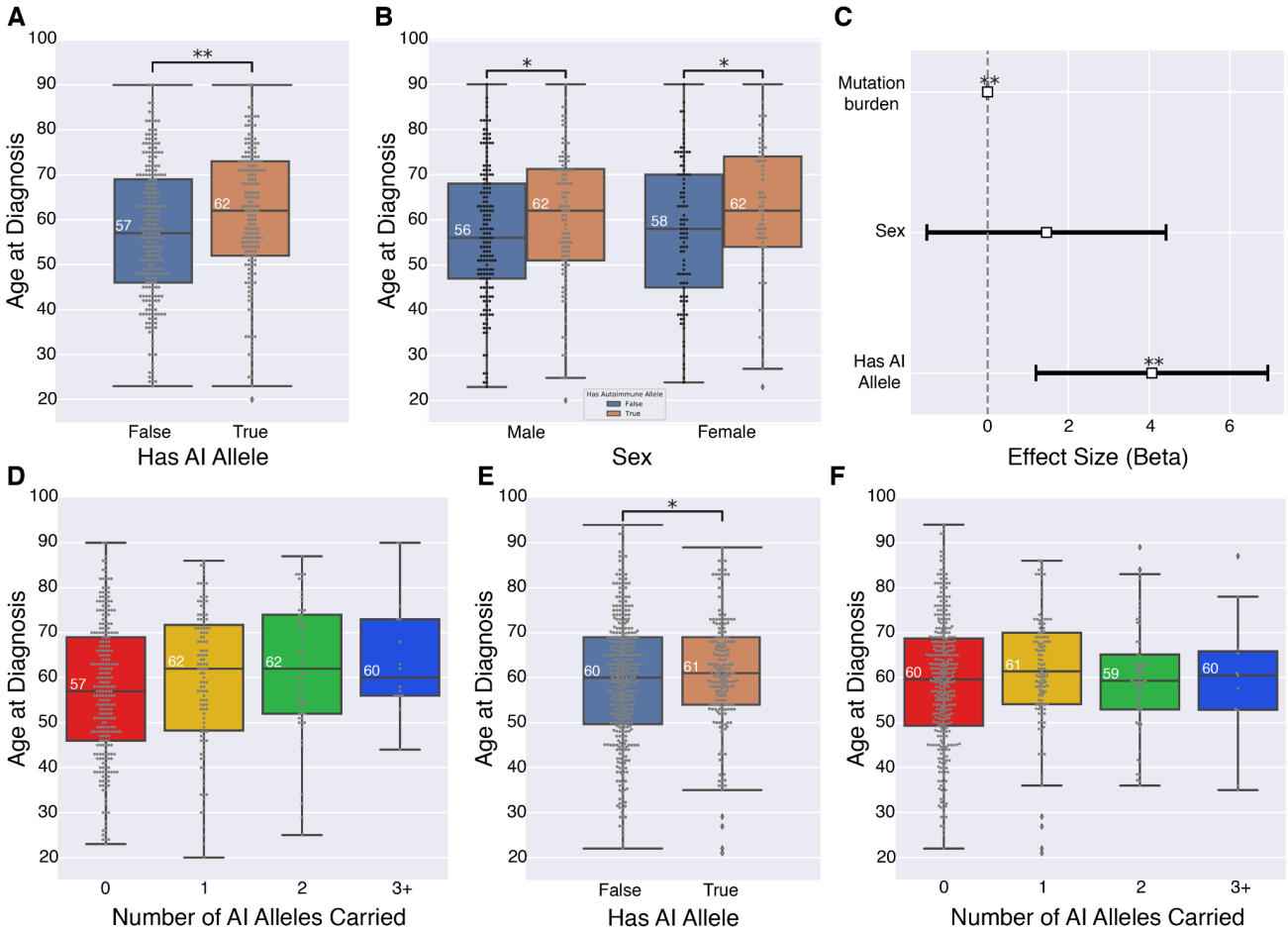


**Figure 1**: Effect of MHC-I autoimmune alleles on age at diagnosis in melanoma. A) *TCGA*: Having at least one MHC-I linked autoimmune allele is associated with a significant median later age of melanoma diagnosis of 5 years (p = 0.002). B) *TCGA*: MHC-I linked autoimmune allele age of melanoma diagnosis effect is conserved across sex ($p_{male}$ = 0.011, $p_{female}$= 0.033). C) *TCGA:* Having at least one MHC-I linked autoimmune allele is significantly associated with 4.07 delayed years to melanoma diagnosis after controlling for sex and mutation burden ($p_{autoimmune}$ = 0.005). Mutation burden also remained significant, but had a minimal contribution to age of diagnosis with 0.004 delayed years until melanoma diagnosis ($p_{mutation}$ = 0.002) D) *TCGA:* Age of diagnosis significantly increases with an individual's total number of autoimmune MHC-I autoimmune alleles (p = 0.002). E) *Validation*: Having at least one MHC-I links autoimmune allele is associated with a significant median later age of melanoma diagnosis of 1.0 years (p = 0.0435). F) *Validation:* Age of diagnosis increases with an individual's total number of autoimmune MHC-I autoimmune alleles (p = 0.328).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

*MHC-I Autoimmune Alleles Modify Melanoma Risk*

Polygenic risk scores (PRS) use information about genetic risk factors to predict individual disease risk. To evaluate the utility of incorporating MHC-I AI allele carrier status into a risk scoring framework, we evaluated it in the context of the PRS developed by Gu *et al.* [97] (Methods: *PRS Implementation)*. This PRS comprises 204 SNPs, of which 16 are found on chromosome 6, though none fall within the HLA class-I region (**Figure 2A**). Although the HLA class-I genes are not among the genes associated with each SNP as reported by Gu *et al.*,[97] we assessed whether HLA-proximal PRS SNPs (i.e., those SNPS within 3MB of the HLA-coding region) associated with MHC-I AI allele carrier status across our discovery cohort, but observed no significant relationship (p = 0.24).

As the PRS score was developed to stratify cases from controls, we ensured that it could be generalized to capture age-specific effects in a case-only cohort through regression analysis. We observed that higher PRS associated significantly with earlier age of diagnosis in the TCGA (*Discovery*: p = 0.002, Beta = -7.499 , CI = [-12.141, -2.857]), but not in the validation cohort (*Validation*: p = 0.710, Beta = 0.7551, CI = [-3.237, 4.747]). Given the disparity in PRS generalization, we evaluated the PRS age stratification in melanoma cases from the Melanostrum Consortium (N = 3001), the original PRS validation set.[97] Here we observed that higher risk scores were associated with an earlier age of diagnosis (p = 0.055, Beta = -1.780, CI = [-3.598, 0.038]). We also compared the minor allele frequency (MAF) of risk SNPs across these three datasets and observed strong correlations in all three pairwise dataset comparisons. However, we did observe certain SNPs with a MAF difference $\geq$ 0.1 across datasets. Four SNPs exhibited this difference between our discovery set and Melanostrum (rs1464510, rs187989493, rs7041168, rs7164220). Between our validation set and Melanostrum, seven SNPs exhibited this MAF gap (rs187989493, rs1393350, rs7164220, rs13338146, rs12919293, rs75570604, rs2092180) including the maximum PRS effect SNP (rs75570604). Finally, between our discovery and validation sets one SNP exhibited a MAF difference $\geq$ 0.1 (rs1464510) (**Supplementary Fig. 7A-C**). These discrepancies may explain the varying performance of PRS in age stratification across cohorts.

We next evaluated the relationship between PRS and age at diagnosis in AI allele carriers versus non-carriers. As expected, PRS score distributions did not differ between those with and without MHC-I AI alleles in either cohort (**Supplementary Fig. 8A-D**; $p_{disc}$ = 0.999, $p_{val}$= 0.901). Interestingly we observed that increases in PRS showed a greater negative effect on age at diagnosis in those with an autoimmune allele relative to those without one in the discovery cohort (**Figure 2B**). Interaction effects from a linear model between PRS and AI MHC-I allele genotype trended negatively (p = 0.055, Beta = -9.124, CI = [-18.432, 0.183]). This could suggest that interactions between known risk SNPs and MHC-I genotypes play a role in melanoma predisposition, or that known high risk variants can overwhelm the protective effect of carrying an MHC-I AI allele. We did not observe any significant interaction effects between PRS and AI MHC-I allele genotype in the validation cohort (**Figure 2C**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

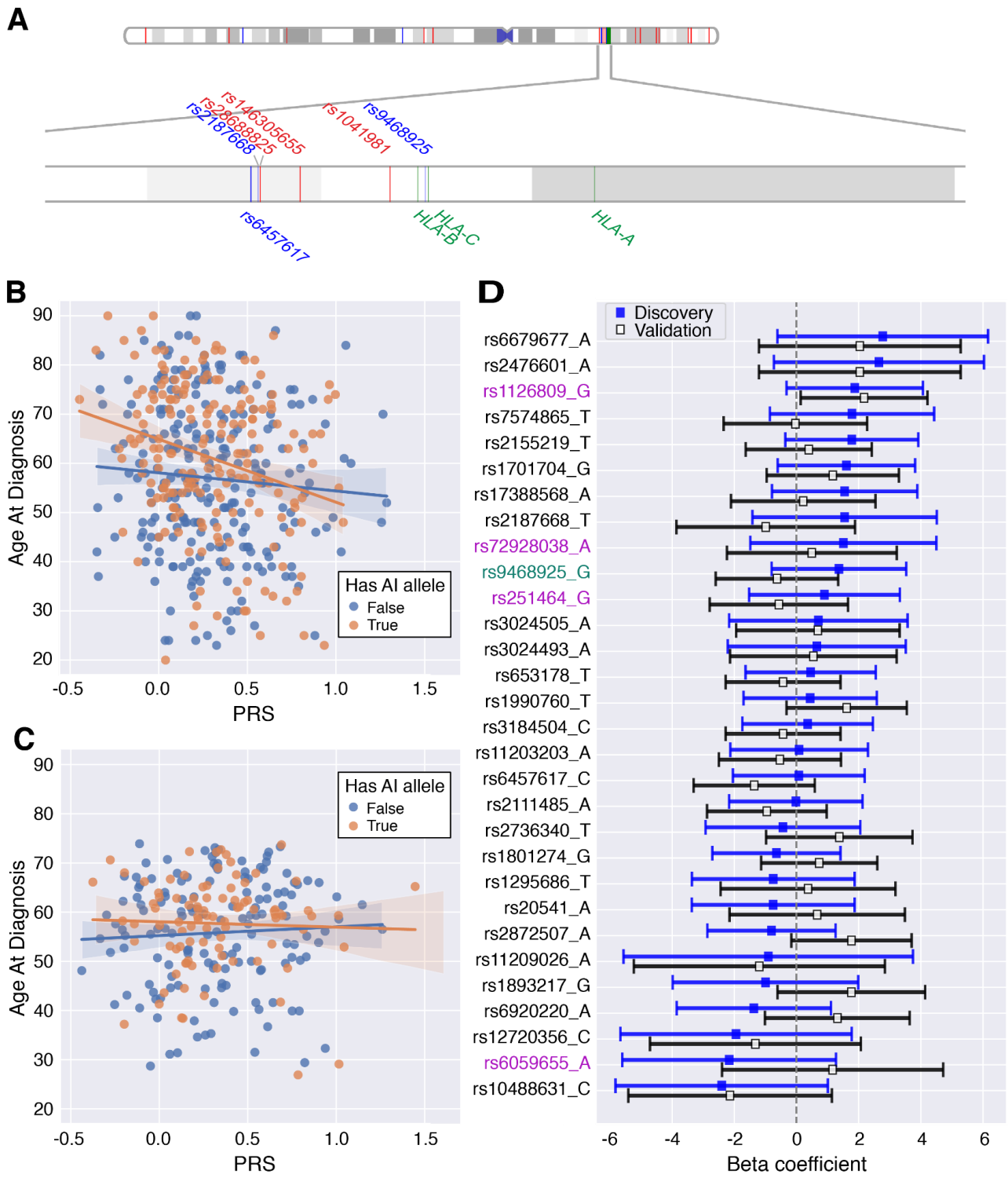


**Figure 2**: A) A PhenoGram [98] plot of chromosome 6 shows melanoma PRS SNPs (red) and autoimmune-associated SNPs (blue) fall outside of the HLA coding region (green). The distance from the closest PRS SNP (rs1041981) to the class-I coding region (HLA-B) is 215,819 bp. The closest AI SNP (rs9468925) falls in between HLA-C/HLA-B. B) *TCGA:* Age of diagnosis as a function of PRS and

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

autoimmune allele presence. Individuals with autoimmune MHC-I alleles show a steeper decrease in age of diagnosis as PRS increases (p = 0.055, Beta = -9.124, CI = [-18.432, 0.183]). C) *Validation:* Age of diagnosis as a function of PRS and autoimmune allele presence. Individuals with autoimmune MHC-I alleles again show a steeper decrease in age of diagnosis as PRS increases, but interaction effects were insignificant (p = 0.499, Beta = -2.878, CI = [-11.254, 5.498]). D) *Autoimmune SNP effect on age of diagnosis*: Autoimmune SNPs showed varying associations with age of diagnosis ranging from strongly protective (high Beta) to strongly predisposing (low Beta). Overall though these effects were highly variable and did not reach statistical significance after multiple-hypothesis correction. Joint vitiligo-melanoma associated SNPs are marked in red, the lone HLA-C/HLA-B psoriasis and vitiligo SNP is marked in green, and the remaining 25 broad autoimmune SNPs are marked in black. Error bars correspond to +/- 2 standard-deviations.

We next evaluated whether the association of AI alleles with melanoma risk extended to AI risk SNPs outside of the HLA. In total we examined 30 AI SNPs, including four with established vitiligo-melanoma associations either as the joint lead risk SNP for both conditions (rs1126809, rs6059655) or in strong linkage disequilibrium with known cutaneous melanoma risk SNPs (rs72928038, rs251464),[30] and one (rs9468925) that is associated with both psoriasis and vitiligo and falls in between HLA-C/HLA-B.[60] The remaining 25 AI SNPs are broadly associated with autoimmunity (i.e., associated with at least three autoimmune conditions, and at least one of which surpassed a GWAS significance of $p = 10^{-7}$) and were previously investigated in the context of immune-checkpoint inhibitor success in melanoma by Chat *et al*.[38] While coefficients for the relationship between AI SNP genotype and age of melanoma diagnosis ranged from strongly protective (e.g., rs6679677; $Beta_{Disc} = 2.775$; $Beta_{Val} = 2.036$) to strongly predisposing (e.g., rs10488631; $Beta_{Disc} = -2.407$; $Beta_{Val} = -2.137$) across cohorts, overall these effects exhibited large variability and were not significant after multiple-hypothesis testing (**Figure 2D**). In contrast to MHC-I AI alleles, including non MHC-I AI SNPs as covariates with PRS did not improve prediction of age at diagnosis.

While cancer cohorts document age at diagnosis, the ideal value to consider for a risk analysis is age at onset, as this would suggest optimal screening times for early detection. To estimate the potential “window of opportunity” for earlier melanoma detection, we estimated the expected time between the initial transformed malignant cell (in a surviving malignant clone that escapes extinction) and clinical detection using a multistage carcinogenesis model for melanoma (Methods: *Multistage Carcinogenesis Model for Melanoma*). Briefly, we developed a cell-based stochastic branching process model for the development of independent premalignant clones (such as nevi) that can arise and clonally expand in normal skin epithelium. Each cell in these clones has the propensity to transform to a malignant cell with a certain probability, the malignant clone population can expand in size or go extinct through a stochastic birth-death process, and clinical detection may occur with a size-based detection probability. Mathematically, the expectation of the lag-time variable, or the time between the founder cell of a persistent malignant clone and clinical detection, can be interpreted as the average “age” or sojourn time of the detected tumor.[99]

We analyzed Surveillance, Epidemiology, and End Results (SEER9) melanoma age- and cohort-specific incidence data from 1975-2018.[100] The hazard function from a “two-stage” model

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

corresponded to the best fit to SEER incidence for both men and women (**Supplementary Fig. 9**). Adding additional stages to the model (i.e., more than 2 rate-limiting events or "hits" such as driver mutations required before malignant transformation) did not improve the fits. With estimated model parameters, we found that the expected tumor sojourn time in males was 8.35 years (Markov chain Monte Carlo [MCMC] 95% CI: 6.61 - 9.73) and similarly in females was 9.64 years (95% CI: 8.48 - 10.66). Previous studies have estimated melanoma doubling times that can be used to then calculate the corresponding tumor sojourn time. With a mean doubling time of 144 days,[101] the mean growth rate in an exponentially growing tumor is approximately 1.76 per year. Assuming malignant tumors are detected on average at $10^8$ or $10^9$ cells in size, this implies a melanoma sojourn time of 10.5 years and 11.8 years, respectively. This is in line with our above estimates, along with those found in a previous modeling study of melanoma doubling times (mean = 3.78 months).[102] Although estimates may vary based on patient-specific factors, our findings suggest that it takes approximately a decade on average for a melanoma to be detected after it is first initiated in an individual.

Subtracting the ~10 year sojourn time from age of diagnosis, we further partitioned our discovery cohort into PRS quintiles and stratified by AI-carrier status. AI carrier status exhibited significant later predicted ages of onset for the lowest ($p = 0.005$) and second-highest risk quintiles ($p = 0.034$) (**Supplementary Fig. 8E**). Across quintiles we observed the median predicted onset age ranged from 44-55, within the proposed melanoma screening range. In the future customizing melanoma onset estimates from tumor genetic and epigenetic marks has the potential to markedly improve screening approaches based on germline genetic risk factors.

*Investigating Autoimmune Alleles in Melanoma ICPI Response*

Immune-checkpoint inhibitors (ICPI) induce immune activation and stimulate anti-cancer responses through blockade of the inhibitory proteins CTLA-4 (cytotoxic T lymphocyte-associated protein-4), PD-1 (programmed cell death protein-1), and PD-L1 (programmed death-ligand 1). While ICPIs have revolutionized cancer therapeutics, individual responses are highly variable. When successful, response can be marked including induction of total remission. Practically, however, many individuals either fail to respond or suffer from severe immune-related adverse events (irAEs). Given the variability in responses, stratifying likely responders before ICPI administration is an essential problem. Melanoma, though, is known to be one of the best ICPI responders across cancer types with estimated response rates of 20% for anti-CTLA4 monotherapy,[103,104] 30-40% for anti-PD-1 or PD-L1 monotherapy,[105–107] and 60% for combination therapy.[107,108] Thus far, tumor mutation burden and PD-L1 positivity have been found to be associated with response. However, models based on these markers still have significant false positive and negative rates. Given the importance of neoepitope presentation in mounting an immune response, it is likely that MHC-genotype plays an essential role as well.[49–51]

Interestingly, vitiligo as an irAE to ICPI administration is associated with improved prognosis and tumor regression.[63–66] If ICPIs can induce remission through broad melanocytic destruction

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mediated by CD8+ T-cells, then the presence of MHC-I AI predisposing alleles may have a role in pre-treatment stratification. Therefore, we investigated whether any relationship existed between our set of MHC-I AI alleles and clinical ICPI response in melanoma using a subset of our validation cohort ($N_{anti\text{-}CTLA4}$ = 103, $N_{anti\text{-}PD\text{-}1}$ = 35). Response was characterized in accordance with the methodology of the original studies. Specifically, those with (ir)RECIST criteria of either complete response, partial response, or stable disease with an overall survival exceeding one year were labelled as responders ($N_{responders}$ = 45). Overall, we did not observe any significant associations between MHC-I AI allele carrier status with ICPI response (OR = 0.69, p = 0.36) or with HLA-A*02:01 carrier status either in isolation (OR = 1.06, p = 1.0) or in the context of broad MHC-I AI allele carrier status (OR = 1.15, p = 0.85). Granular ICPI separation to a treatment-specific level of anti-CTLA4 (OR = 0.59, p = 0.36) or anti-PD1 (OR = 0.97, p = 1.0) similarly failed to show any significant associations.

### *Identifying Mechanisms of MHC-I Autoimmune Allele Protection*

Sequence variation across HLA alleles results in differences in the amino acid composition of the MHC-I peptide-binding groove. This leads to allele-specific binding affinity for peptides, such that HLA alleles effectively constrain the subspace of antigens that can be presented to CD8+ T-cells during immune surveillance. In psoriasis it has been observed that melanocyte antigens such as ADAMTS-like protein 5 presented by HLA-C*06:02 (one of the seven AI alleles) can induce a targeted CD8+ T-cell response against melanocytes.[59] Similarly, in vitiligo, CD8+ T-cells target antigens from melanosomal proteins such as PMEL, MART1/MLANA, TYR, TRP-1, and TRP-2.[76,109,110] Interestingly melanoma-specific CD8+ T-cells appear to recognize peptides derived from these conserved melanocytic antigens more often than melanoma-specific antigens.[76,111–113] Given this, a protective effect in cancer could indicate that AI alleles mediate more effective immune surveillance against conserved cancer antigens (i.e., self-antigens overexpressed in tumors) or even against somatic mutations that promote tumor development. To evaluate this possibility, we analyzed the binding potential of AI alleles for melanoma-specific conserved and neoantigens relative to other alleles.

We first established a list of candidate peptides that could serve as conserved or driver neoantigens in melanoma. For conserved antigens, we included peptides derived from antigens associated with melanocytes [76,109] and melanoma, including the melanoma antigen gene (MAGE) family.[114–120] We also expanded our list of conserved antigens to include genes constitutively expressed in melanocytes (Methods: *Identification and Differential Expression of Conserved Antigens*). In total, we considered 91 genes as sources of conserved antigens (**Supplementary Table 3**). For putative neoantigens, we hypothesized that a protective effect would require specificity against a mutation capable of promoting melanoma development. We focused on single nucleotide variants which dominate the landscape of melanoma and account for the majority of driver events.[121–123] To identify mutations that might serve as early drivers in melanomagenesis we used a combination of joint high DNA and RNA variant allelic fraction (VAF) and recurrence. We also included a set of high DNA VAF mutations that were predicted to be drivers by the CHASM algorithm (Methods: *Identifying Driver Mutations*).[124,125] In total, we considered 215 mutations, including well-known driver mutations in BRAF, NRAS and CDKN2A

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(**Fig. 3 A-B**). BRAF V600E was the most recurrent mutation across the discovery cohort with 164 unique occurrences (**Fig. 3A**).

We computed MHC-I binding affinity percentile rank across 2,915 alleles for peptides derived from conserved and neo-cancer antigens using NetMHCpan-4.1 (Methods: *Predicting Binding Affinities*). Following convention, a rank threshold <0.5 corresponds to strong binding and weak binding occurs down to a rank threshold <2.[126] To comprehensively evaluate antigen binding differences across the relevant range, we compared allele-specific differences continuously across binding ranks from 0 to 2. For putative neoantigens we first looked at whether AI alleles exhibited any binding differences relative to one another. For this purpose, we defined coverage as the fraction of considered mutations where at least one peptide is expected to bind to MHC-I at a given affinity cutoff. Across AI alleles, HLA-C*06:02 and HLA-C*12:03 had the greatest neopeptide coverage at higher rank scores with coverage fractions of 0.474 and 0.460 at a rank of 2. However, at the lower rank scores (i.e., stronger peptide affinity) up to 0.41, we observe HLA-B27:05 to be the top neoantigen binding allele (**Fig. 3C**). Given the noticeable coverage differences as a function of peptide affinity, this potentially points to innate binding preferences between HLA-B and HLA-C.

We next compared HLA-B and HLA-C AI alleles to population representations for matched common non-AI alleles (Methods: *HLA Population Allele Representations*; **Fig. 3D-E**). These population representations encapsulate a non-AI coverage span, such that any AI alleles falling outside of this range suggest either enhanced or reduced coverage relative to population. For HLA-B, 3 of the 5 AI alleles fell in this population allele range, while HLA-B*13:02 exhibited reduced coverage between rank scores of 0.27 and 0.84 and HLA-B*27:05 showed reduced coverage at higher rank scores, specifically $\geq$ 1.66 (**Fig. 3D**). For HLA-C, HLA-C*12:03 tracked within the population neopeptide coverage span with a small range of reduced neopeptide presentation between 0.29 and 0.33. HLA-C*06:02 on the other hand showed two extended ranges of enhanced presentation relative to the maximum population allele representation, specifically in the rank score ranges of 0.51-0.73 and 1-1.24 (**Fig. 3E**). In general, maximum and minimum population B-alleles exhibited greater coverage than their corresponding C-alleles at lower rank scores, but C-alleles outperformed B-alleles as the rank score exceeded 0.5 (**Supplementary Fig. 10A**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

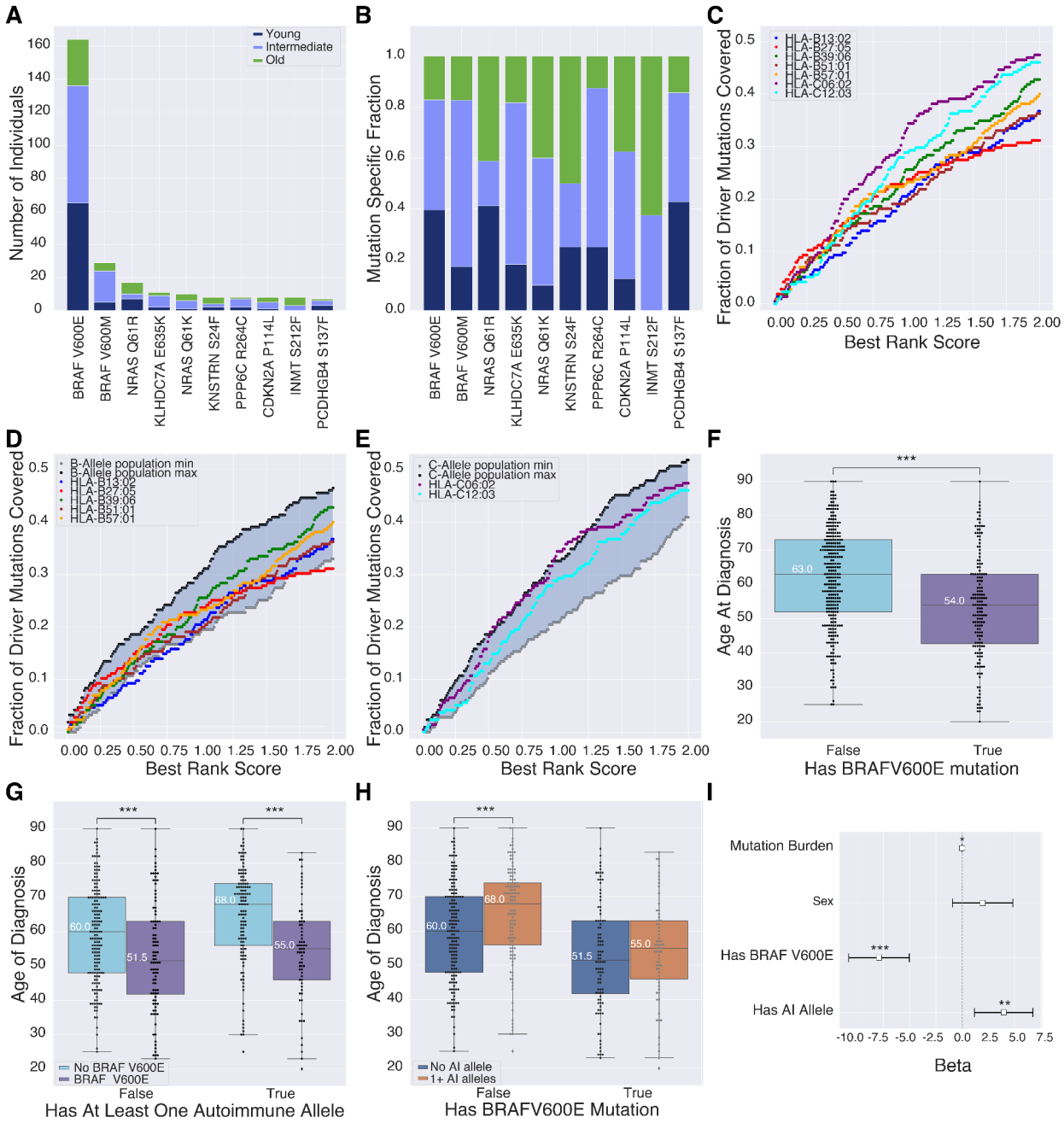


**Figure 3**: A) Frequency of the 10 most recurrent mutations across the TCGA by age group. Young (< 50) and old (≥ 69) age groups correspond to the bottom and top 30% of individuals by age respectively. The intermediate (50 ≤ x < 69) age group corresponding to the remaining 40% of individuals. B) Relative age group distribution of the 10 most recurrent mutations across the TCGA. C) Fraction of driver mutations presented by autoimmune alleles as a function of BR score. D) Fraction of driver mutations presented by HLA-B autoimmune alleles relative to common (≥ 1% population frequency; 19 alleles) HLA-B alleles. Maximum and minimum population allele coverage corresponds to the maximum and minimum fraction of driver mutations capable of being presented across common alleles at each best rank score respectively. E) Fraction of driver mutations presented by HLA-C autoimmune alleles relative to common (≥ 1% population frequency; 13 alleles) HLA-C alleles. F) Individuals with a BRAF V600E mutation show a significant 9 year earlier age of melanoma diagnosis relative to those without this mutation ($p = 7.58 \times 10^{-6}$).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

G) Discovery cohort individuals with an AI allele show a significant median later age of diagnosis of 8 years in the absence of a BRAF V600E mutation ($p = 2.98x10^{-4}$). However BRAF V600E mutation presence appears to counter the AI allele protective effect with a loss of significance between those with and without AI alleles ($p = 0.232$) H) BRAF V600E significantly reduces melanoma age of diagnosis across discovery cohort individuals independent of autoimmune allele presence with median earlier ages of diagnosis of 8.5 years in those without an AI allele ($p = 7.39x10^{-4}$) and 13 years in those with an AI allele ($p = 8.12x10^{-7}$). I) BRAF V600E mutation presence is significantly associated with 7.91 earlier years to melanoma diagnosis ($p_{BRAFV600E} = 7.12x10^{-8}$), while having at least one MHC-I linked autoimmune allele is significantly associated with 3.97 delayed years to melanoma diagnosis ($p_{autoimmune} = 4.59x10^{-3}$) after controlling for sex and mutation burden. Mutation burden also remained significant, but had a minimal contribution to age of diagnosis with 0.003 delayed years until melanoma diagnosis ($p_{mutation} = 0.014$)

We next examined whether specific mutations significantly influenced age of diagnosis across the discovery cohort. If so, this could suggest mutation-specific coverage as a mechanism for AI melanoma protection. One mutation, BRAF V600E, showed significant age differences with mutation carriers being diagnosed with melanoma on average 9 years earlier than those without (**Fig. 3F**; $p = 7.58x10^{-6}$). However, rather than observing a correlation between lack of AI allele presence and having a BRAF V600E mutation, BRAF V600E status served as an indiscriminate melanoma catalyst, significantly shifting age of diagnosis earlier regardless of AI allele status (**Fig. 3G**). Moreover, it appeared to counter the AI allele protective effect with a reduced age gap between those with and without AI alleles in BRAF V600E tumors (**Fig. 3H**). Regression analysis showed similar results with V600E mutation presence having a larger effect size than AI carrier status by almost 4 years. (**Fig. 3I**).

Across AI alleles, only three were predicted to present BRAF V600E based on rank score. HLA-B*27:05 was the only allele with a predicted affinity below the strong binding cutoff (best rank score = 0.22). HLA-B*27:05-restricted cytotoxic T-cell responses have been observed against V600E.[127] HLA-B*39:06 and HLA-B*57:01 had scores of 1.78 and 0.61 respectively, showing potential for weak V600E binding. We found no association with occurrence (OR = 1.04, $p = 0.84$) or expression of BRAF V600E ($p=0.34$) in individuals carrying AI alleles in general, or those carrying one or more of these three AI alleles specifically. As association might have been expected if the mutant allele was subject to strong counter selection by immune surveillance. Notably, BRAF V600E has been suggested to avoid immune surveillance by accelerating internalization of cell surface MHC-I.[128]

As we observed no obvious differences in AI-allele specific presentation of putative driver neoantigens, we next evaluated conserved antigens. We collected a set of known conserved cancer antigens in melanoma (**Supplementary Table 3**),[76,109,114–120] and further expanded this set to include other genes that could potentially serve as conserved antigens using the criteria that they be *(i)* stably expressed in melanocytes, *(ii)* specific to melanocytes and *(iii)* expressed in melanomas. We first identified genes that exhibited uniquely stable expression across melanocytes by comparing stably expressed genes (SEGs) across a cohort of melanocytes and 53 tissues from the Genotype-Tissue Expression (GTEx) project. We developed a score to capture stable expression and added to our set 52 genes (**Supplementary Table 3**) that scored as well as the majority of canonical melanocyte genes and were exclusive to melanocytes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

These SEGs included four well-known melanocyte genes, PMEL, MLANA, TYRP1, and TYR, and the rest were enriched for the folate metabolism pathway which is important for DNA repair in melanocytes [129] (Methods: *Identification and Differential Expression of Conserved Antigens*). We next evaluated whether these genes were differentially expressed (DE) between melanomas in our discovery cohort and normal melanocytes. Consistent with published findings, we observed that two MAGE genes, MAGEA10 and MAGEE1, were significantly upregulated and that several canonical melanocyte and stably expressed genes were downregulated (**Fig. 4A**). Specifically, MAGE genes have been shown to be specific to reproductive tissues [130] and tumors [114–120] and canonical melanocyte genes such as TYRP1 have been shown to be minimally expressed, if not undetectable, in melanoma.[112,131,132]

Conserved antigens encompass a broader space of presentable peptides than neoantigens as any peptide across a protein could serve as a MHC-I binding target. To evaluate broad allele-specific differences in conserved antigen-derived peptide repertoires, we compared the fraction of 8-11mers from each conserved antigen predicted to bind at a given percentile rank (Methods: *Predicting Binding Affinities*). Across AI alleles we observed minimal variation with HLA-B*51:01 being the most promiscuous AI allele and HLA-B*13:02 being the second most promiscuous across the binding range (**Fig. 4B**). However, in general, AI alleles exhibited a narrower conserved antigen repertoire relative to common alleles, with HLA-B*39:06, HLA-B*57:01, and HLA-B*27:05 exhibiting the narrowest repertoires (**Fig. 4C**). Similarly, HLA-C AI alleles exhibited relatively small repertoires within the range exhibited by common alleles (**Fig. 4D**). HLA-B though generally presented a broader range of peptides across conserved antigens than HLA-C (**Supplementary Fig. 10B**). This discrepancy again draws attention to potential HLA type specific binding preferences and further suggests innate binding preferences between HLA-B and HLA-C.

Considering individual conserved genes, AI alleles were generally predicted to bind a smaller fraction of derived peptides than common alleles. Comparison of the distributions of the top 10 differences by mean at a gene-specific level between AI and common alleles shows that almost all AI alleles present conserved antigens no better than common alleles (**Supplementary Fig. 11**). There were a few exceptions, however, as HLA-B*27:05 presented SRSF8 and TOMM6 remarkably better than common alleles (**Supplementary Fig. 11A-B**).

Recent analyses of eluted peptides bound to MHC-I and MHC-II found that presented peptides are not uniformly sampled across the entire parent protein sequence.[133] We therefore revisited our analysis with a focus on regional differences in binding affinity for the set of conserved antigens. We considered position-wise best percentile ranks from AI and common alleles by computing the best percentile rank across all 8-11mers overlapping a position in each gene (**Fig. 4E**). We then looked for regions where common alleles were non-binders (eluted at > 2% rank) and autoimmune alleles were strong binders (eluted at < 0.5% rank). We observed that HLA-B*27:05, an especially strong binder to the conserved antigen SRSF8 (**Supplementary Fig. 11A-B**), also exhibited the greatest position-wise advantage to SRSF8 over common alleles (**Fig 4F**). This is in spite of having the smallest conserved antigen repertoire overall (**Fig. 4B-C**). We additionally observed that amongst AI alleles, there was a substantial degree of variability in

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

binding affinity at the position-wise level and it was rare for more than one allele to present the same position well. This is counter-intuitive given the minimal variability seen more broadly (**Fig. 4B**). Thus, even though AI alleles have similar overall repertoire size, they appear to present different regions of proteins. Furthermore, the regions for which AI alleles have better presentation are sparse and short rather than long or contiguous (**Fig. 4F**, median length = 2 amino acids, s.d. = 1.7 amino acids). We also found that several canonical melanocyte and MAGE genes had positions with marked differences between HLA-B AI and common alleles, specifically the canonical melanocyte genes *PMEL, GPNMB*, *TYR*, and *MITF* and the MAGE genes *MAGEA10 and MAGEE1* (**Fig. 4F, Supplementary Fig. 12**). However, the autoimmune alleles with the greatest advantages over common alleles, HLA-B*27:05 and HLA-B*57:01, did not exhibit better binding to the upregulated MAGE genes (**Fig. 4F, Supplementary Fig. 12**). Unlike with HLA-B alleles, HLA-C AI alleles exhibited only one marked advantage in regional binding affinity over common alleles. MBD5 showed modestly better affinity for HLA-C*06:02 with a percentile rank of 0.5 compared to 2.11 for common alleles for 8-11mers overlapping amino acid positions 1490-1491. Overall, despite the fact that most 8-11mers from conserved antigens are better presented by common alleles, there were regional differences that could potentially facilitate AI allele-specific immunogenic responses.

Finally, we attempted to validate the positions that had better affinity for AI alleles than common alleles using the HLA Ligand Atlas. The HLA Ligand Atlas contains the immunopeptidomes of 21 individuals measured via mass spectrometry and identified their corresponding MHC-I molecules with strong and weak binding via allele-specific binding predictions.[133] However we were unable to find overlap of ligands detected in the HLA Ligand Atlas with peptides derived from conserved antigens. However, the HLA Ligand Atlas did not specifically measure the human melanocyte immunopeptidome. At the time of writing, it contains samples from 21 individuals of which only 8 contributed to information about MHC class I in skin (of which melanocytes are only a small fraction) and only 4 of the AI alleles are represented. Given these limitations, functional validation of these findings is a future direction.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

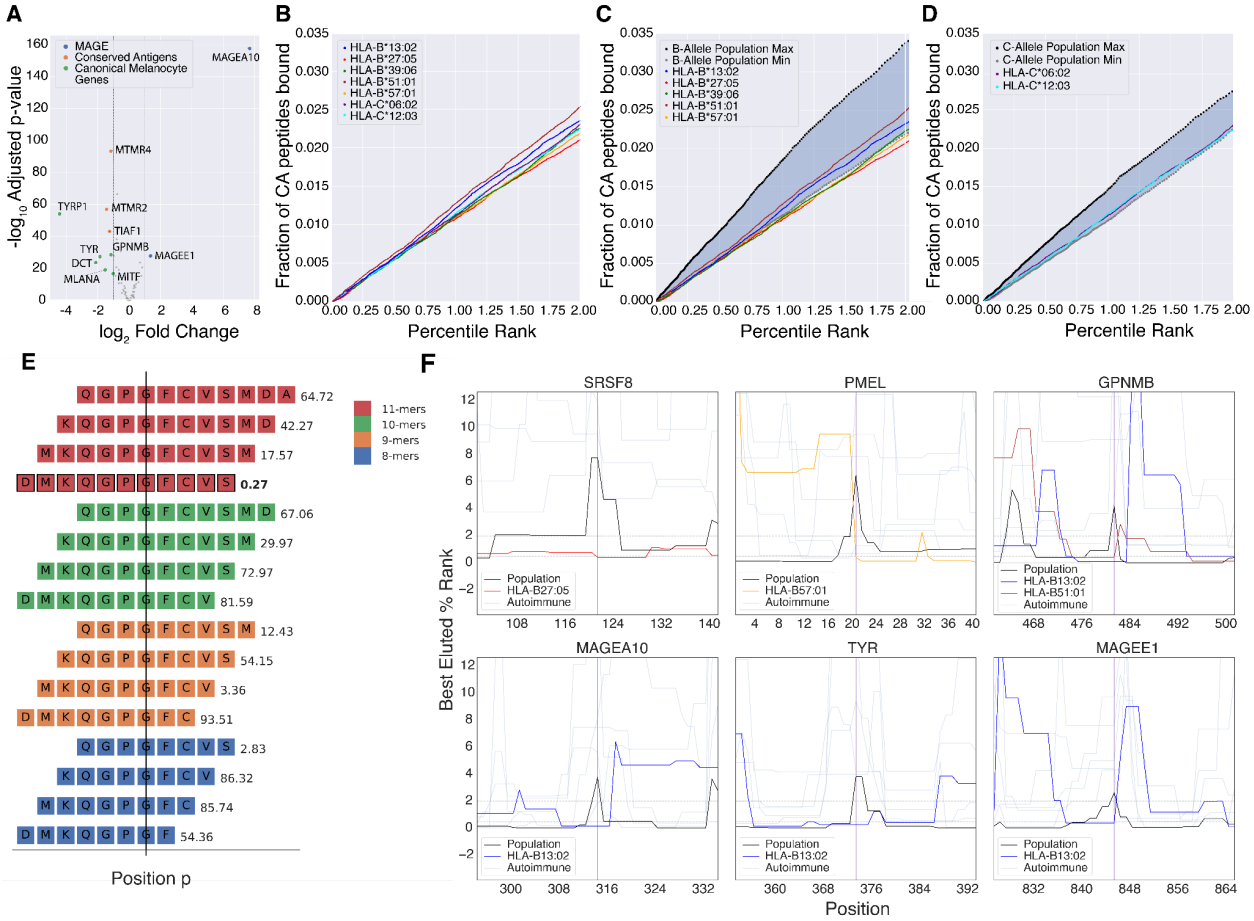


**Figure 4**: A) Differential expression of conserved antigens between normal melanocytes and melanoma. Labeled genes have an adjusted p-value > 0.5 and an absolute fold change > 2. B) Fraction of peptides from conserved antigens (CA) presented by each autoimmune allele. C) Fraction of peptides from conserved antigens (CA) presented by HLA-B autoimmune alleles relative to common HLA-B alleles (≥ 1% population frequency; 19 alleles). Maximum and minimum population allele representations correspond to the maximum and minimum fraction of conserved antigen peptides bound across common alleles at each rank score respectively. D) Fraction of peptides from conserved antigens (CA) presented by HLA-C autoimmune alleles relative to common HLA-C alleles (≥ 1% population frequency; 13 alleles). E) Position-wise best percentile rank scores were computed from predictions for 8-11mers by taking the best percentile rank of all overlapping peptides at any given position. F) Greatest position-wise differences in best percentile rank between HLA-B autoimmune and common alleles amongst the top 2 differences (*SRSF8* and the canonical melanocyte gene *PMEL*) and several DE conserved antigens (*GPNMB, MAGEA10, TYR, MAGEE1*). Common alleles are shown in orange and autoimmune alleles are shown in transparent blue unless they are predicted to elute at better percentile ranks than common alleles. Vertical purple lines demarcate positions where autoimmune alleles are predicted to elute at better percentile ranks than common alleles; plots show up to +/- 20 amino acids of the demarcated positions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Discussion:

Immunosurveillance has been implicated in melanomagenesis prevention,[16,17,31] yet HLA contributions to melanoma risk have largely remained uncharacterized. Here we investigated whether predisposing MHC-I alleles for CD8+ T-cell driven skin-associated autoimmune disorders (vitiligo and psoriasis) could protect against melanoma. Our findings support this hypothesis, with AI-allele carriers exhibiting a significant later age of melanoma diagnosis in both the TCGA and an independent validation cohort. Moreover, AI-allele specific protection appears not only to be uncaptured by current melanoma PRS,[97] but also can augment PRS performance in relative risk stratification. While at least 6 vitiligo risk SNPs are protective from melanoma,[28,29] our results show that autoimmune risk effects can be extended to MHC-I as well, and suggest a broader space of joint autoimmune predisposition and melanoma protection.

We further investigated potential mechanisms linking MHC-I AI alleles to delayed age at diagnosis. Immune activity against melanoma has been linked to high mutation burden due to UV exposure,[21,22] suggesting neoantigens could provide a potential substrate. Focusing on mutations that drive melanomagenesis, we did not see obvious differences in the affinity for MHC-I AI alleles for neopeptides relative to other alleles. Interestingly, lack of AI allele association with improved response to immunotherapy would seem to support that neoantigens are not the mechanism by which AI alleles boost immune surveillance in melanoma. While AI alleles did not seem to interact with specific mutations, we did note that BRAF V600E was associated with an earlier age at diagnosis independent of AI allele carrier status. Potential of autoimmune alleles to present BRAFV600E did not appear to impact the incidence of the mutation, which is consistent with reports that this mutation associates with impaired immune surveillance.[128]

As CD8+ T-cells target healthy melanocytes through conserved antigens in both psoriasis and vitiligo, responses against conserved antigen could also provide an explanation. We observed positions with greater presentability by AI alleles in genes spanning all subcategories of conserved antigens including canonical melanocyte genes, melanoma antigen genes, and genes stably expressed in melanocytes. In particular, AI alleles had greater presentability of amino acids in PMEL and TYR (**Fig. 4F**), which are melanocyte-specific antigens recognized by CD8+ T-cells in melanoma-associated vitiligo.[76,109] However, we did not observe position-wise advantages for AI alleles in other CD8+ T-cell targets reported in melanoma-associated vitiligo, such as MLANA or TYRP2. MHC-I AI alleles also uniquely target positions in the melanoma antigen genes MAGEA10 and MAGEE1 which are upregulated in melanoma relative to melanocytes (**Fig. 4A**). Finally, we observed positions uniquely targeted by AI alleles within melanocyte stably expressed genes. Notably, SRSF8 exhibited the position most favored by AI alleles and is overall substantially better targeted by the AI allele, HLA-B*27:05 (**Supplementary Fig. 11**). This suggests that melanocyte-specific conserved antigens other than those already identified in melanoma-associated vitiligo may also contribute to the protective effect of AI alleles. Additionally, while canonical melanocyte genes have been known to be recognized by CD8+ T cells, they are typically downregulated in melanoma (**Fig. 4A**) or otherwise inconsistently expressed.[134,135] In contrast, the novel conserved antigens described in this study exhibit stable expression across melanocytes and melanoma and might serve as

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

more consistent immune targets.

Altogether, these findings support that AI alleles' unique immunopeptidomes could contribute to their protective effect against melanoma. However, further investigation is needed. In particular, some studies have suggested that stability of the MHC-I-peptide complex distinguishes AI alleles from other MHC-I alleles,[136–138] so it is possible that the mechanism is not fully dependent on antigen specificity. We noted that protective effects were confined to 7 specific AI-associated HLA alleles, and were not shared by broader HLA supertype groupings supporting that allele-specific characteristics, whether related to unique antigen specificities, stability or some other characteristic, are likely to account for any protective effect.

In general, AI alleles presented a narrower peptide repertoire compared to common alleles (**Fig. 4C-D**). Notably, HLA-B*27:05 and HLA-B*57:01, both of which covered a smaller fraction of potential conserved peptides than the minimum HLA-B population allele representation, are known for their fastidiousness, or narrow peptide-binding repertoire, as studied in the context of progression from HIV infection to AIDS.[139] Košmrlj et al. also observed that fastidious alleles present peptides not found in common alleles' repertoires, which we similarly observed with the most fastidious AI alleles, HLA-B*27:05 and HLA-B*57:01, exhibiting the greatest position-wise advantages over common alleles. Additionally, fastidious class I alleles are expressed on the cell-surface at much higher levels than their promiscuous class I counterparts,[140] and therefore likely offer more opportunities for both neoepitope and self-antigen presentation from their binding repertoires. Interestingly, we note that outside of autoimmune associations four of our seven AI-alleles (HLA-B*27:05, HLA-B*51:01, HLA-C*06:02 and HLA-B*57:01) are among the strongest HIV-protective alleles.[70,141,142]

Peptide-MHC (pMHC) affinity is driven largely by MHC-I sequence variation where specific polymorphisms shape the binding pockets in the peptide-binding groove. Unsurprisingly, given the skin-specific autoimmune associations for which these alleles were selected, certain AI-alleles have similar binding pocket characteristics. For example, HLA-C*06:02 and HLA-C*12:03 both share a strongly negative E-pocket.[143] HLA-C*06:02 shares an electronegative B-pocket with HLA-B*27:05 as well.[143,144] An allele's affinity-based presentable peptide repertoire though is an idealistic representation that fails to account for antigen processing pathway contributions. Ultimately the space of bound peptides presented on the cell-surface is far narrower. Endoplasmic reticulum aminopeptidases (ERAP) 1 and 2 play an essential role in MHC-I antigen processing. They function to clip peptides to the appropriate size for MHC-I binding,[145,146] but can also destroy potential ligands through overtrimming.[147–150] ERAP1 and ERAP2 are prominent risk factors in MHC-I linked autoimmune conditions and both are associated with psoriasis risk.[151–153] Epistatic effects between ERAP1 and AI-risk alleles have been observed, particularly with HLA-C*06:02 in psoriasis,[151] and suggests this risk interaction is tied to an increased likelihood of specific autoantigens making it to the cell surface. Taken together, this suggests skin-specific autoimmune predisposing ERAP genes may also confer melanoma protection and is an interesting area for further pursuit.

Vitiligo has been reported as a favorable irAE correlating with melanoma immunotherapy response.[63–66] However, we did not observe any associations of AI carrier status with ICPI

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

response. Work by Chowell *et al.*[49] and Cummings *et al.*[154] showed that in melanoma the B44 supertype associates with extended survival after treatment with ICPIs. While none of our AI alleles fall within this supertype, we surprisingly observed that the B44 supertype trended towards an earlier age of diagnosis in our discovery cohort (**Supplementary Fig. 3**). The B44 supertype has an electropositive B-pocket, with an affinity for negatively charged residues at P2 such as glutamate (E).[155] Given this strong glutamate affinity, it may be that in the context of ICPI, the B44 supertype is capable of inducing an immune response through the binding and presentation of the highly recurrent BRAF V600E mutation. In fact certain AI alleles, such as HLA-B*27:05 and HLA-C*06:02, have electronegative B-pockets with strong affinities for the positively charged arginine at position 2,[143,144] and effectively serve as B44 antonyms. This suggests noticeable peptide repertoire differences between the B44 supertype and the AI allele set, and further suggests the potential for a dichotomy between MHC-I associated melanoma protection and ICPI response, which is associated with somatic mutation presentation and high tumor mutation burden. We note that sample sizes for assessing the role of AI alleles in ICPI responses are small, and revisiting this analysis in larger studies in the future may provide further insight.

In conclusion, our study supports that skin-specific auto-immune MHC alleles have a protective effect in melanoma. We note the potential for MHC-I mediated autoimmunity to interact with cancer development more broadly; we observed four other cancers (ACC, ESCA, PCPG, SARC) for which skin AI-allele carrier status showed similar later age of diagnosis effects (**Supplementary Fig. 6A**). Some MHC-I alleles are associated with multiple autoimmune conditions, such as HLA-B*27:05 in both psoriasis and ankylosing spondylitis.[69,70,72,73,156,157] Ankylosing spondylitis is a targeted form of spinal arthritis primarily affecting the entheses and leads to bone erosion and broad vertebral fusion.[158] This may explain the observation of delayed diagnosis with Sarcoma (SARC), which is primarily a bone and connective tissue cancer. Our study did not address MHC-II alleles which are generally expressed more specifically by antigen presenting cells, but have nonetheless been implicated in both skin autoimmune disorders and immunosurveillance. Taken together, tissue-specific MHC AI carrier status may broaden the scope of the autoimmune-cancer risk interplay and remains an interesting area for further exploration.

**Acknowledgements:**

This work was supported by a grant from the Harry J. Lloyd Charitable Trust (20191857) to H.C., an Emerging Leader Award from The Mark Foundation for Cancer Research (18-022-ELA) to H.C., a RO1 CA220009 grant to M.Z. and H.C., a R01 MH122688-02 grant to W.K.T., and an NIH (National Institutes of Health) National Library of Medicine training grant (T15LM011271) to A.C. The results shown here are in large part based upon data generated by the TCGA Research Network (https://www.cancer.gov/tcga), the UKBB (Project #37671), and the following studies: phs000452.v2.p1.c2, phs000933.v2.p1.c1, phs001550.v2.p1.c1, phs001500.v1.p1.c1, phs000424.v7.p2, GSE78220.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Disclosure of Interests:**
The authors declare no competing interests.

**Author Contributions:**
Original Concept, J.T., H.C.; Project Supervision, H.C.; Project Planning and Design, J.T., H.C.; Data Acquisition, Processing, and Analysis, J.T., D.L., M.P., A.C.; Melanoma Sojourn Time Estimation, M.L., G.E.L., K.C.; Statistical Advising, W.K.T.; UKBB Data Acquisition, R.M.S.; Immunological Interpretation Advising, G.P.M. and M.Z.; Preparation of the Manuscript, J.T., D.L., and H.C.

**Data Availability:**
All data for this project were obtained from public sources. *Discovery Cohort:* Data were obtained from The Cancer Genome Atlas (TCGA) Research Network (http://cancergenome.nih.gov/). Normal exome sequences and clinical data were downloaded from the GDC on June 23-26th, 2018 and April 25th, respectively, using the gdc-client v1.3.0. Somatic mutations were accessed from the NCI Genomic Data Commons (https://portal.gdc.cancer.gov/) on May 14th, 2017. Genotype calls for TCGA were accessed from GDC on April 26, 2019. *Validation and Stably Expressed Gene Cohorts*: dbGaP data from cohorts phs000452.v2.p1.c2, phs001550.v2.p1.c1, phs000933.v2.p1.c1 and phs001500.v1.p1.c1 were obtained using AsperaConnect v3.9.5.172984. SRA toolkit v2.9.2 was used to obtain WXS/WGS data from the Sequence Read Archive (SRA) (including from the following studies: SRP067938, SRP090294). UKBB data was retrieved under project ID 37671. *Melanostrum Cohort:* Genotype data from Gu *et al.* [97] were obtained by direct communication with the authors.

**Code Availability:**
Code for all analyses can be found at https://github.com/cartercompbio/MelMHC/

## Methods:

*Datasets:*

TCGA skin cutaneous melanoma tumors (SKCM) were used as the discovery set (N = 470). Cases were retained if they had appropriate clinical information for downstream analysis (i.e., age of diagnosis; 11 did not have this information), were microsatellite-stable (3 MSI tumors were removed), and were at least 20 years of age at time of diagnosis (5 were < 20). Individuals below 20 years of age were excluded due to their increased likelihood of harboring rare predisposing risk variants. After filtering there were 451 tumor samples. MHC-I genotypes were called using the exome based methods POLYSOLVER and HLA-HD.[87,88]

An independent validation cohort of melanoma cases with germline WXS/WGS data was built from 5 separate melanoma studies. Two of these studies (*Hugo et al.* [93]: SRP067938, SRP090294, *Van Allen et al.* [92]: phs000452.v2.p1.c2) focused on melanoma response to immune-checkpoint inhibitors (ICPI) and were used to evaluate autoimmune associations in the

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

context of ICPI response. Two of these studies (*Melanoma Exome Sequencing:* phs000933.v2.p1.c1 [95,96], *The genetic and transcriptomic evolution of melanoma:* phs001550.v2.p1.c1 [94]), were broader melanoma studies focusing on the genetic basis of sun-exposed melanoma and melanoma evolution. The final study consisted of individuals from the UKBB with WXS/WGS with ICD10 codes: C433, C434, C436, and C437. Validation cohort individuals were also filtered by age to exclude individuals under 20 years old. MHC-I genotypes for the validation cohort were called using HLA-HD.[88] Discovery and validation cohorts are described in **Supplementary Table 1**.

Cases from the Melanostrum Consortium (N = 3001) were used to evaluate PRS generalizability from absolute risk (i.e., cases vs. controls) to relative risk (i.e., age-specific effects). Genotype was available at 204 risk SNPs used in the development of the PRS by Gu *et al*.[97] Cases from this cohort were filtered to include individuals $\geq$ 20 years in age.

To identify putative conserved antigens we leveraged RNA-seq from a cohort of healthy melanocytes (dbGaP: phs001500.v1.p1) derived from newborn foreskins (N=106) and version 7 of the Genotype-Tissue Expression (GTEx) project (dbGaP: phs000424.v7.p2).

*Identifying Driver Mutations:*

We downloaded WES-based mutation calls from the TCGA GDC portal from four different mutation callers: VarScan,[159] MuSE,[160] MuTect,[161] and SomaticSniper.[162] RNA Variant Allelic Fraction (VAF) was obtained with bam-readcount. We considered a mutation to be a potential driver if it: 1) altered protein sequence, 2) was found in both the DNA and RNA in at least one individual, and 3) had a median DNA and RNA variant allele fraction (VAF) percentile less than or equal to 40%. DNA mutations were only considered at a patient-specific level if they were called by at least two of the mutation callers mentioned above. In total 51,062 mutations satisfied these criteria.

We further filtered the list of putative drivers based on recurrence. Specifically, if a specific mutation was detected in 4 or more different tumors we categorized it as a likely driver. In total 109 mutations satisfied this criterion (0.215% of the 51,062 candidate mutations). For those mutations that failed to reach this recurrence, we calculated mutation-specific contributions to melanoma pathogenicity using scores from a melanoma-specific CHASM classifier.[124,125] Mutations with a CHASM score greater than or equal to 0.9 were deemed to be likely melanoma drivers; 106 mutations satisfied this criterion (0.206% of the 51,062 candidate mutations). Combining these recurrent and predicted driver singleton mutations yielded a final set of 215 melanoma drivers.

*Identification and Differential Expression of Conserved Antigens:*

GTEx V7 contains 11,688 RNA-seq samples from 714 donors across 53 tissue types and was aligned with STAR v2.4.2a to GENCODE v19 and quantified with RNA-SeQC v1.1.8. RNA-seq reads from healthy melanocytes were aligned with STAR v2.5.0b to GENCODE v19 and quantified with RSEM v1.2.31. After quantification, both cohorts were filtered such that genes

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with ≤ 0.5 RSEM or with counts < 6 in > 93% of samples were removed. Additionally, ribosomal RNA, Y chromosomal, and histone genes were removed. Ribosomal and histone mRNA are not polyadenylated. Notably, the melanocyte cohort is exclusively male, while GTEx is not, which would potentially lead to Y chromosomal genes being falsely identified as stably expressed downstream.

Across healthy melanocytes and each tissue type in GTEX, genes were scored as stably expressed genes (SEGs) using the output from the scoring method described in scMerge.[163] Briefly, the method first fits the expression of each gene from each tissue sample to a Gamma-Gaussian mixture. For the expression of a gene $x_i$, the Gamma component corresponds to samples with low expression and the Gaussian component corresponds to samples with high expression. This mixture has the joint density function

$$f(x_i; \alpha_i, \beta_i, \mu_i, \sigma_i, \lambda_i) = \lambda_i \cdot Gamma(\alpha_i, \beta_i) + (1 - \lambda_i) \cdot N(\mu_i, \sigma_i)$$

where $\alpha_i$ and $\beta_i$ are the shape and rate parameters of the gamma component, $\mu_i$ and $\sigma_i$ are the mean and standard deviation of the Gaussian component, and the mixing proportion $\lambda_i$ is bounded by [0,1]. The SEG scoring method also takes into account the proportion of zeros in each gene $\omega_i$ for each tissue. Each gene is then scored by the percentile ranks of its mixing proportion $\lambda_i$, coefficient of variation (CV) $\sigma_i/\mu_i$, and proportion of zeros $\omega_i$ such that the average percentile rank across all three metrics is minimal. The highest scoring genes have lower mixing proportions, CVs, and proportion of zeros.

Using the aforementioned method, each gene from every tissue was fit to a Gamma-Gaussian distribution using scMerge v1.6.0 and given a score from 0 to 1 for stable expression. Then, the set of genes across all GTEx tissues that had scores > 0.69 was removed from the set of genes in melanocytes that had scores > 0.69. The threshold of 0.69 was chosen based on the observation that the scores of canonical melanocyte genes were, with the exception of DCT/TYRP2, all ~0.7 and by visual inspection of the distributions of scores across tissues (**Supplementary Fig. 13**). This process yielded 4 canonical melanocyte genes (PMEL, MLANA, TYRP1, TYR) and 52 additional protein coding genes for downstream analysis. The 52 non-canonical melanocyte genes were also passed to PANTHER [164] for Reactome pathway [165] overrepresentation analysis using Fisher's Exact test, which identified several genes as members of the folate metabolism pathway (FDR 0.0288).

To determine differential expression of conserved antigens in melanoma relative to melanocytes, we applied the same expression quantification pipeline and gene filtering steps to both healthy melanocytes and SKCM. Specifically, we quantified HLA-allele specific expression using HLApers v1.0 [166] and the Kallisto v0.44.0 [167] pipeline for HLApers. Reads were aligned to GENCODE v30 [168] and IMGT HLA v3.41.0.[169] After quantification, both cohorts were filtered in the same way as the conserved antigen identification pipeline described above. Additionally, 11 samples with missing age of diagnosis were removed from the SKCM cohort. We then

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

performed a differential expression analysis with DESeq2 v1.30.1 [170] conditioned on ancestry. However, several potentially relevant covariates were also incompatible across the two cohorts. Namely, age, sex, tumor type (primary or metastatic), tumor purity,[171] melanocytic plasticity score,[172] and TIDE score.[173] Due to these discrepant covariates between the cohorts, we also checked that no covariates were associated with significant DE for any conserved antigens. The only gene subject to DE was MAGEA10 with a -1.3 +/- 0.4 log fold change (LFC) in primary vs. metastatic melanoma. This is substantially less than the +8 LFC of MAGEA10 observed in melanocytes vs. melanoma (**Fig. 4A**).

Finally, we used the HLA Ligand Atlas [133] to validate the peptides from conserved antigens that were observed to have preferential binding to autoimmune alleles. At time of writing, the HLA Ligand Atlas has identified 90,428 HLA-I ligands from the tissues of 21 individuals. We queried the HLA Ligand Atlas for ligands from our conserved antigens that were predicted to be strong or weak binders to autoimmune alleles.

*Predicting Binding Affinities:*

MHC-I allele binding affinities were computed across the available 2,915 unique MHC-I alleles for both driver mutations and conserved antigens. Since driver mutations altered protein sequence, we evaluated MHC-I alleles' ability to present neoepitopes by generating all unique 8-11mers found in a mutation relative to the wild-type (corresponding to the set of novel peptides a MHC-I allele can present to the immune system). To circumvent cross-allele and cross-peptide variabilities that are inherent in predicted IC50 comparisons, we used percentile ranks relative to a random set of peptides provided by NetMHCpan-4.1[174] to approximate binding affinity for every MHC-I allele peptide pair. These percentile rank scores correspond to how strongly an allele binds a particular peptide relative to a set of random natural peptides. From peptide level rank scores, MHC-I mutation specific binding affinities were assigned according to the best rank score, the minimum allele-specific rank score across all unique 8-11mers for a mutation.

For conserved antigens, we partitioned proteins into their entire set of 8-11mers across the full length of the protein. MHC-I allele peptide pair percentile rank scores were again generated using NetMHCpan-4.1. Several metrics were derived from the percentile rank scores for downstream analyses. For broad and gene-level comparisons, we defined an allele-specific conserved antigen repertoire as the set of 8-11mers presented at or below a given percentile rank (**Fig. 4B-D**, **Supplementary Fig. 10B, 11**). For position-wise presentability (**Fig. 4E-F**) we used the best percentile rank of all overlapping peptides at each position along a protein.

*HLA Population Allele Representations:*

To compare HLA-B and HLA-C autoimmune (AI) alleles to common non-AI alleles, defined as those alleles with a population frequency $\geq$ 1% as given by the NMDP (19 common HLA-B alleles, 13 common HLA-C alleles), we established maximum and minimum population allele representations. For driver neoantigens, a maximum population allele representation was assigned to have coverage at each rank equating to the coverage of the best presenting

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

common allele at that rank. Similarly a minimum population allele representation was assigned to have coverage at each rank equating to the coverage of the worst presenting common allele at the rank. For maximum and minimum population allele representations in conserved antigens (CA), the metric of assignment was the fraction of CA peptides bound as opposed to coverage, as was consistent with CA analyses.

*PRS Implementation*:

We implemented the melanoma PRS developed by Gu *et al.* [97] comprising 204 SNPs. For the discovery cohort we were able to extract 190 of the 204 risk SNP genotypes using Plink. For the validation cohort, datasets lacking sufficient SNP data for PRS construction were excluded leaving 239 individuals for which 201 of the 204 risk variants were extracted using Plink2. For the Melanostrum Consortium (N=3001), all 204 risk SNPs were extracted. Risk SNP MAFs were compared across cohorts to ensure no significant differences (**Supplementary Fig. 7**). The final PRS for each cohort was generated as a weighted sum across extracted risk SNPs in each cohort in accordance with the optimal melanoma risk model by Gu *et al*.[97]

*Multistage Carcinogenesis Model for Melanoma:*

Dating back to the 1950's Armitage-Doll model [175] of cancer incidence and those created soon after by Knudsen and Moolgavkar [176] and others, multistage models of cancer are among the most developed mathematical methods for defining carcinogenesis and determining timescales of tumor formation in human populations.[177–181] These models assume evolutionary stages from normal cells to development of clinically detected symptomatic cancers. These stages typically include intermediate premalignant and preclinical malignant stages that represent field cancerization dynamics of stochastically growing and shrinking clonal populations in a tissue. These models can be described mathematically as stochastic multi-type branching processes with probabilities of events occurring with certain rates (**Supplementary Fig. 9A**). By calculating an age-dependent hazard function for cancer incidence using solutions to equations from the probability generating functions starting from birth, we can calibrate these models to fit hazard rates derived from cancer incidence registry data such as SEER in the US.[100] Importantly, this modeling framework provides a link between cell-level dynamics and population-level incidence data so that we can estimate parameters governing clonal growth, dwell times, and mutational "hits" in at-risk individuals.

In previous work we found that the "two-stage" model (2 "hits" for development of a first malignant cell) shown in **Supplementary Fig. 9** is closely approximated by a model that includes an effective malignant transformation rate and a characteristic lag-time or "sojourn" time between malignant transformation and clinical detection (see Luebeck *et al.* 2013 for mathematical details [99]). Here we created a two-stage model for melanoma incidence that adjusts for birth cohort trends, similar to methods used previously in esophageal squamous cell carcinoma (ESCC).[182] In this way, our models capture trends for both age and birth cohort (and thus calendar period) to enable robust estimation via Markov Chain Monte Carlo simulation of cell-level parameters for tumor evolution by sex and race/ethnicity (see **Supplementary Fig. 9B** for examples of model fits). We obtained estimates and 95% confidence intervals in the main

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

text for tumor sojourn times in men and women via MCMC posterior estimates for the lag-time parameter. Chains were run for 100,000 cycles with a 4000 cycle burn-in and checked for convergence. All code for hazard function calculation and parameter estimation was written in Fortran. The ICD-O-3 codes used for extraction of SEER data melanoma, all races combined, from SEER*Stat include: 8720/3, 8721/3, ,8722/3, 8723/3, 8726/3, 8727/3, 8728/3, 8730/3, 8740/3, 8741/3, 8742/3, 8743/3, 8744/3, 8745/3, 8746/3, 8761/3, 8770/3, 8771/3, 8772/3, 8773/3, 8774/3, 8780/3, 8790/3.

*Statistical Analyses:*

All box plot statistical tests comparing age of diagnosis effects between groups were assessed using the default Mann-Whitney U statistical test. Leave-one-out analysis was conducted by narrowing the AI allele set into all 7 unique sets of 6 AI alleles and stratifying individuals accordingly. Performing leave-one-out analysis by dropping all carriers of each allele yielded similar results, with AI status significantly associated with a later age of diagnosis in each holdout set. T-tests were used to compare PRS distributions across AI-allele status in both discovery and validation cohorts. Fisher's exact tests were used to evaluate associations between: 1) MHC-I AI allele carrier status and ICPI response status 2) Major and minor AI SNP alleles and ICPI response status 3) HLA-proximal PRS SNPs and MHC-I AI allele carrier status. These statistical tests were all implemented via the default scipy.stats Python package. Regression analyses were modeled using ordinary least squares linear models through the statsmodels.formula.api Python package.[183] All multiple hypothesis testing correction utilized the Benjamini-Hochberg procedure, and was implemented by means of the statsmodels.stats.multitest package in Python.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## **References**:


1. Ali, Z., Yousaf, N. & Larkin, J. Melanoma epidemiology, biology and prognosis. *EJC Suppl* **11**, 81–91 (2013).
2. Matthews, N. H., Li, W.-Q., Qureshi, A. A., Weinstock, M. A. & Cho, E. Epidemiology of Melanoma. in *Cutaneous Melanoma: Etiology and Therapy* (eds. Ward, W. H. & Farma, J. M.) (Codon Publications, 2018).
3. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2017. *National Cancer Institute*.
4. Gilchrest, B. A., Eller, M. S., Geller, A. C. & Yaar, M. The pathogenesis of melanoma induced by ultraviolet radiation. *N. Engl. J. Med.* **340**, 1341–1348 (1999).
5. Lea, C. S. *et al.* Ambient UVB and melanoma risk in the United States: a case-control analysis. *Ann. Epidemiol.* **17**, 447–453 (2007).
6. Garibyan, L. & Fisher, D. E. How sunlight causes melanoma. *Curr. Oncol. Rep.* **12**, 319–326 (2010).
7. Lin, J. Y. & Fisher, D. E. Melanocyte biology and skin pigmentation. *Nature* **445**, 843–850 (2007).
8. Rees, J. L. Genetics of hair and skin color. *Annu. Rev. Genet.* **37**, 67–90 (2003).
9. Olsen, C. M., Carroll, H. J. & Whiteman, D. C. Estimating the attributable fraction for melanoma: a meta-analysis of pigmentary characteristics and freckling. *Int. J. Cancer* **127**, 2430–2445 (2010).
10. Chang, Y.-M. *et al.* A pooled analysis of melanocytic nevus phenotype and the risk of cutaneous melanoma at different latitudes. *Int. J. Cancer* **124**, 420–428 (2009).
11. Olsen, C. M., Carroll, H. J. & Whiteman, D. C. Estimating the attributable fraction for

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cancer: A meta-analysis of nevi and melanoma. *Cancer Prev. Res.* **3**, 233–245 (2010).

12. Schuchter, L. *et al.* A prognostic model for predicting 10-year survival in patients with primary melanoma. The Pigmented Lesion Group. *Ann. Intern. Med.* **125**, 369–375 (1996).
13. Williams, L. H., Shors, A. R., Barlow, W. E., Solomon, C. & White, E. Identifying Persons at Highest Risk of Melanoma Using Self-Assessed Risk Factors. *J. Clin. Exp. Dermatol. Res.* **2**, (2011).
14. Han, J. *et al.* A prospective study of telomere length and the risk of skin cancer. *J. Invest. Dermatol.* **129**, 415–421 (2009).
15. Bataille, V. *et al.* Nevus Size and Number Are Associated with Telomere Length and Represent Potential Markers of a Decreased Senescence In vivo. *Cancer Epidemiol. Biomarkers Prev.* **16**, 1499–1502 (2007).
16. Green, A. C. & Olsen, C. M. Increased risk of melanoma in organ transplant recipients: systematic review and meta-analysis of cohort studies. *Acta Derm. Venereol.* **95**, 923–927 (2015).
17. Greene, M. H., Young, T. I. & Clark, W. H., Jr. Malignant melanoma in renal-transplant recipients. *Lancet* **1**, 1196–1199 (1981).
18. Olsen, C. M., Carroll, H. J. & Whiteman, D. C. Familial melanoma: a meta-analysis and estimates of attributable fraction. *Cancer Epidemiol. Biomarkers Prev.* **19**, 65–73 (2010).
19. Ford, D. *et al.* Risk of cutaneous melanoma associated with a family history of the disease. The International Melanoma Analysis Group (IMAGE). *Int. J. Cancer* **62**, 377–381 (1995).
20. Mucci, L. A. *et al.* Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA* **315**, 68–76 (2016).
21. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
22. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

23. Clemente, C. G. *et al.* Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma. *Cancer* **77**, 1303–1310 (1996).

24. Morton, D., Eilber, F. R., Malmgren, R. A. & Wood, W. C. Immunological factors which influence response to immunotherapy in malignant melanoma. *Surgery* **68**, 158–63; discussion 163–4 (1970).

25. Orrù, V. *et al.* Genetic variants regulating immune cell levels in health and disease. *Cell* **155**, 242–256 (2013).

26. Mangino, M., Roederer, M., Beddall, M. H., Nestle, F. O. & Spector, T. D. Innate and adaptive immune traits are differentially affected by genetic and environmental factors. *Nat. Commun.* **8**, 1–7 (2017).

27. Marderstein, A. R. *et al.* Demographic and genetic factors influence the abundance of infiltrating immune cells in human tissues. *Nat. Commun.* **11**, 1–14 (2020).

28. Jin, Y. *et al.* Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat. Genet.* **44**, 676–680 (2012).

29. Jin, Y. *et al.* Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat. Genet.* **48**, 1418–1424 (2016).

30. Landi, M. T. *et al.* Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. *Nat. Genet.* **52**, 494–504 (2020).

31. Corthay, A. Does the immune system naturally protect against cancer? *Front. Immunol.* **5**, 197 (2014).

32. Teng, M. W. L., Swann, J. B., Koebel, C. M., Schreiber, R. D. & Smyth, M. J. Immune-mediated dormancy: an equilibrium with cancer. *J. Leukoc. Biol.* **84**, 988–993 (2008).

33. Gregersen, P. K. & Behrens, T. W. Genetics of autoimmune diseases--disorders of immune homeostasis. *Nat. Rev. Genet.* **7**, 917–928 (2006).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

34. Nanda, N. K. & Sercarz, E. E. Induction of anti-self-immunity to cure cancer. *Cell* **82**, 13–17 (1995).

35. Pardoll, D. M. Inducing autoimmune disease to treat cancer. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 5340–5342 (1999).

36. Spritz, R. A. The genetics of generalized vitiligo: autoimmune pathways and an inverse relationship with malignant melanoma. *Genome Med.* **2**, 78 (2010).

37. Young, A., Quandt, Z. & Bluestone, J. A. The Balancing Act between Cancer Immunity and Autoimmunity in Response to Immunotherapy. *Cancer Immunol Res* **6**, 1445–1452 (2018).

38. Chat, V. *et al.* Autoimmune genetic risk variants as germline biomarkers of response to melanoma immune-checkpoint inhibition. *Cancer Immunol. Immunother.* **68**, 897–905 (2019).

39. Khan, Z. *et al.* Polygenic risk for skin autoimmunity impacts immune checkpoint blockade in bladder cancer. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 12288–12294 (2020).

40. Askling, J. *et al.* Risks of solid cancers in patients with rheumatoid arthritis and after treatment with tumour necrosis factor antagonists. *Ann. Rheum. Dis.* **64**, 1421–1426 (2005).

41. Song, L. *et al.* The risks of cancer development in systemic lupus erythematosus (SLE) patients: a systematic review and meta-analysis. *Arthritis Res. Ther.* **20**, 270 (2018).

42. Teulings, H. E. *et al.* Decreased risk of melanoma and nonmelanoma skin cancer in patients with vitiligo: a survey among 1307 patients and their partners. *Br. J. Dermatol.* **168**, 162–171 (2013).

43. Rock, K. L., Reits, E. & Neefjes, J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol.* **37**, 724–737 (2016).

44. Durgeau, A., Virk, Y., Corgnac, S. & Mami-Chouaib, F. Recent Advances in Targeting CD8 T-Cell Immunity for More Effective Cancer Immunotherapy. *Front. Immunol.* **9**, 14 (2018).

45. Woodsworth, D. J., Dunsing, V. & Coombs, D. Design Parameters for Granzyme-Mediated

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cytotoxic Lymphocyte Target-Cell Killing and Specificity. *Biophys. J.* **109**, 477–488 (2015).

46. Marty, R. *et al.* MHC-I Genotype Restricts the Oncogenic Mutational Landscape. *Cell* **171**, 1272–1283.e15 (2017).
47. Marty Pyke, R. *et al.* Evolutionary Pressure against MHC Class II Binding Cancer Mutations. *Cell* **175**, 416–428.e13 (2018).
48. Yang, F. *et al.* Quantifying immune-based counterselection of somatic mutations. *PLoS Genet.* **15**, e1008227 (2019).
49. Chowell, D. *et al.* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* **359**, 582–587 (2018).
50. Chowell, D. *et al.* Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nat. Med.* **25**, 1715–1720 (2019).
51. Goodman, A. M. *et al.* MHC-I genotype and tumor mutational burden predict response to immunotherapy. *Genome Med.* **12**, 45 (2020).
52. McGranahan, N. *et al.* Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* **171**, 1259–1271.e11 (2017).
53. Castro, A. *et al.* Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes. *BMC Med. Genomics* **12**, 107 (2019).
54. van den Boorn, J. G. *et al.* Autoimmune destruction of skin melanocytes by perilesional T cells from vitiligo patients. *J. Invest. Dermatol.* **129**, 2220–2232 (2009).
55. Boehncke, W.-H. & Brembilla, N. C. Autoreactive T-Lymphocytes in Inflammatory Skin Diseases. *Front. Immunol.* **10**, 1198 (2019).
56. Conrad, C. *et al.* α 1 β 1 integrin is crucial for accumulation of epidermal T cells and the development of psoriasis. *Nat. Med.* **13**, 836–842 (2007).
57. Cai, Y., Fleming, C. & Yan, J. New insights of T cells in the pathogenesis of psoriasis. *Cell. Mol. Immunol.* **9**, 302–309 (2012).
58. Prinz, J. C. Human Leukocyte Antigen-Class I Alleles and the Autoreactive T Cell

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Response in Psoriasis Pathogenesis. *Front. Immunol.* **9**, 954 (2018).

59. Arakawa, A. *et al.* Melanocyte antigen triggers autoimmunity in human psoriasis. *J. Exp. Med.* **212**, 2203–2212 (2015).
60. Zhu, K.-J. *et al.* Psoriasis regression analysis of MHC loci identifies shared genetic variants with vitiligo. *PLoS One* **6**, e23089 (2011).
61. Sharquie, K. E., Salman, H. A. & Yaseen, A. K. Psoriasis and vitiligo are close relatives. *Clin. Cosmet. Investig. Dermatol.* **10**, 341–345 (2017).
62. Vashist, S. *et al.* Association of Psoriasis with Autoimmune Disorders: Results of a Pilot Study. *Indian Dermatol. Online J.* **11**, 753–759 (2020).
63. Rosenberg, S. A. & White, D. E. Vitiligo in patients with melanoma: normal tissue antigens can be targets for cancer immunotherapy. *J. Immunother. Emphasis Tumor Immunol.* **19**, 81–84 (1996).
64. Hua, C. *et al.* Association of Vitiligo With Tumor Response in Patients With Metastatic Melanoma Treated With Pembrolizumab. *JAMA Dermatol.* **152**, 45–51 (2016).
65. Nakamura, Y. *et al.* Correlation between vitiligo occurrence and clinical benefit in advanced melanoma patients treated with nivolumab: A multi-institutional retrospective study. *J. Dermatol.* **44**, 117–122 (2017).
66. Indini, A. *et al.* Immune-related adverse events correlate with improved survival in patients undergoing anti-PD1 immunotherapy for metastatic melanoma. *J. Cancer Res. Clin. Oncol.* **145**, 511–521 (2019).
67. Hofmann, L. *et al.* Cutaneous, gastrointestinal, hepatic, endocrine, and renal side-effects of anti-PD-1 therapy. *Eur. J. Cancer* **60**, 190–209 (2016).
68. Bonigen, J. *et al.* Anti-PD1-induced psoriasis: a study of 21 patients. *J. Eur. Acad. Dermatol. Venereol.* **31**, e254–e257 (2017).
69. Okada, Y. *et al.* Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. *Am. J. Hum. Genet.* **95**, 162–172 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

70. Chen, H. *et al.* Psoriasis patients are enriched for genetic variants that protect against HIV-1 disease. *PLoS Genet.* **8**, e1002514 (2012).

71. Helms, C. *et al.* Localization of PSORS1 to a haplotype block harboring HLA-C and distinct from corneodesmosin and HCR. *Hum. Genet.* **118**, 466–476 (2005).

72. Winchester, R. *et al.* HLA associations reveal genetic heterogeneity in psoriatic arthritis and in the psoriasis phenotype. *Arthritis Rheum.* **64**, 1134–1144 (2012).

73. FitzGerald, O., Haroon, M., Giles, J. T. & Winchester, R. Concepts of pathogenesis in psoriatic arthritis: genotype determines clinical phenotype. *Arthritis Res. Ther.* **17**, 115 (2015).

74. Ogg, G. S., Rod Dunbar, P., Romero, P., Chen, J. L. & Cerundolo, V. High frequency of skin-homing melanocyte-specific cytotoxic T lymphocytes in autoimmune vitiligo. *J. Exp. Med.* **188**, 1203–1208 (1998).

75. Lang, K. S. *et al.* HLA-A2 restricted, melanocyte-specific CD8(+) T lymphocytes detected in vitiligo patients are related to disease activity and are predominantly directed against MelanA/MART1. *J. Invest. Dermatol.* **116**, 891–897 (2001).

76. Palermo, B. *et al.* Specific cytotoxic T lymphocyte responses against Melan-A/MART1, tyrosinase and gp100 in vitiligo by the use of major histocompatibility complex/peptide tetramers: the role of cellular immunity in the etiopathogenesis of vitiligo. *J. Invest. Dermatol.* **117**, 326–332 (2001).

77. Jin, Y. *et al.* Next-generation DNA re-sequencing identifies common variants of TYR and HLA-A that modulate the risk of generalized vitiligo via antigen presentation. *J. Invest. Dermatol.* **132**, 1730–1733 (2012).

78. Zhang, X.-J. *et al.* Association of HLA class I alleles with vitiligo in Chinese Hans. *J. Dermatol. Sci.* **35**, 165–168 (2004).

79. Quan, C. *et al.* Genome-wide association study for vitiligo identifies susceptibility loci at 6q27 and the MHC. *Nat. Genet.* **42**, 614–618 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

80. Nair, R. P. *et al.* Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. *Am. J. Hum. Genet.* **78**, 827–851 (2006).

81. Gudjonsson, J. E. *et al.* Psoriasis patients who are homozygous for the HLA-Cw*0602 allele have a 2.5-fold increased risk of developing psoriasis compared with Cw6 heterozygotes. *Br. J. Dermatol.* **148**, 233–235 (2003).

82. Yamamoto, T., Yokozeki, H. & Nishioka, K. Psoriasis arthropathy and HLA-B51: report of 5 cases. *J. Dermatol.* **32**, 606–610 (2005).

83. Nejentsev, S. *et al.* Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* **450**, 887–892 (2007).

84. Noble, J. A. *et al.* HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. *Diabetes* **59**, 2972–2979 (2010).

85. Howson, J. M. M., Walker, N. M., Clayton, D., Todd, J. A. & Type 1 Diabetes Genetics Consortium. Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. *Diabetes Obes. Metab.* **11 Suppl 1**, 31–45 (2009).

86. Ohno, S. *et al.* Close Association of HLA-Bw51 With Behcet's Disease. *Archives of Ophthalmology* vol. 100 1455–1458 (1982).

87. Shukla, S. A. *et al.* Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152–1158 (2015).

88. Kawaguchi, S., Higasa, K., Shimizu, M., Yamada, R. & Matsuda, F. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. *Hum. Mutat.* **38**, 788–797 (2017).

89. Gonzalez-Galarza, F. F. *et al.* Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* **48**, D783–D788 (2019).

90. Magnus, K. Prognosis in malignant melanoma of the skin. Significance of stage of disease,

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

anatomical site, sex, age and period of diagnosis. *Cancer* **40**, 389–397 (1977).

91. Natale, C. A. *et al.* Activation of G protein-coupled estrogen receptor signaling inhibits melanoma and improves response to immune checkpoint blockade. *Elife* **7**, (2018).
92. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
93. Hugo, W. *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* vol. 165 35–44 (2016).
94. Shain, A. H. *et al.* The Genetic Evolution of Melanoma from Precursor Lesions. *N. Engl. J. Med.* **373**, 1926–1936 (2015).
95. Krauthammer, M. *et al.* Exome sequencing identifies recurrent mutations in NF1 and RASopathy genes in sun-exposed melanomas. *Nat. Genet.* **47**, 996–1002 (2015).
96. Lazova, R. *et al.* Spitz nevi and Spitzoid melanomas: exome sequencing and comparison with conventional melanocytic nevi and melanomas. *Mod. Pathol.* **30**, 640–649 (2017).
97. Gu, F. *et al.* Combining common genetic variants and non-genetic risk factors to predict risk of cutaneous melanoma. *Human Molecular Genetics* (2018) doi:10.1093/hmg/ddy282.
98. Wolfe, D., Dudek, S., Ritchie, M. D. & Pendergrass, S. A. Visualizing genomic information across chromosomes with PhenoGram. *BioData Min.* **6**, 18 (2013).
99. Luebeck, E. G., Curtius, K., Jeon, J. & Hazelton, W. D. Impact of tumor progression on cancer incidence curves. *Cancer Res.* **73**, 1086–1096 (2013).
100. SEER Research Data 1975-2018 when Using SEER*Stat: Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Data, 9 Registries, Nov 2020 Sub (1975-2018) - Linked To County Attributes - Time Dependent (1990-2018) Income/Rurality, 1969-2019 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021, based on the November 2020 submission.
101. Carlson, J. A. Tumor doubling time of cutaneous melanoma and its metastasis. *Am. J.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

*Dermatopathol.* **25**, 291–299 (2003).

102. Kay, K., Dolcy, K., Bies, R. & Shah, D. K. Estimation of Solid Tumor Doubling Times from Progression-Free Survival Plots Using a Novel Statistical Approach. *AAPS J.* **21**, 27 (2019).

103. Schadendorf, D. *et al.* Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. *J. Clin. Oncol.* **33**, 1889–1894 (2015).

104. Maio, M. *et al.* Five-year survival rates for treatment-naive patients with advanced melanoma who received ipilimumab plus dacarbazine in a phase III trial. *J. Clin. Oncol.* **33**, 1191–1196 (2015).

105. Hamid, O. *et al.* Safety, Clinical Activity, and Biological Correlates of Response in Patients with Metastatic Melanoma: Results from a Phase I Trial of Atezolizumab. *Clin. Cancer Res.* **25**, 6061–6072 (2019).

106. Weber, J. S. *et al.* Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol.* **16**, 375–384 (2015).

107. Wolchok, J. D. *et al.* Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* **377**, 1345–1356 (2017).

108. Larkin, J. *et al.* Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N. Engl. J. Med.* **373**, 23–34 (2015).

109. Teulings, H.-E. *et al.* The antibody response against MART-1 differs in patients with melanoma-associated leucoderma and vitiligo. *Pigment Cell Melanoma Res.* **27**, 1086–1096 (2014).

110. Rodrigues, M. *et al.* New discoveries in the pathogenesis and classification of vitiligo. *J. Am. Acad. Dermatol.* **77**, 1–13 (2017).

111. Kawakami, Y. *et al.* Identification of a human melanoma antigen recognized by tumor-infiltrating lymphocytes associated with in vivo tumor rejection. *Proc. Natl. Acad. Sci.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

*U. S. A.* **91**, 6458–6462 (1994).

112. Chen, Y. T., Stockert, E., Tsang, S., Coplan, K. A. & Old, L. J. Immunophenotyping of melanomas for tyrosinase: implications for vaccine development. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 8125–8129 (1995).

113. Wang, R. F., Appella, E., Kawakami, Y., Kang, X. & Rosenberg, S. A. Identification of TRP-2 as a human tumor antigen recognized by cytotoxic T lymphocytes. *J. Exp. Med.* **184**, 2207–2216 (1996).

114. Brasseur, F. *et al.* Expression of MAGE genes in primary and metastatic cutaneous melanoma. *International Journal of Cancer* **63**, 375–380 (1995).

115. Mori, M. *et al.* Expression of MAGE genes in human colorectal carcinoma. *Ann. Surg.* **224**, 183–188 (1996).

116. Scarcella, D. L. *et al.* Expression of MAGE and GAGE in high-grade brain tumors: a potential target for specific immunotherapy and diagnostic markers. *Clin. Cancer Res.* **5**, 335–341 (1999).

117. Tajima, K. *et al.* Expression of cancer/testis (CT) antigens in lung cancer. *Lung Cancer* **42**, 23–33 (2003).

118. Gure, A. O. *et al.* Cancer-testis genes are coordinately expressed and are markers of poor outcome in non-small cell lung cancer. *Clin. Cancer Res.* **11**, 8055–8062 (2005).

119. Barrow, C. *et al.* Tumor antigen expression in melanoma varies according to antigen and stage. *Clin. Cancer Res.* **12**, 764–771 (2006).

120. Kim, Y.-D. *et al.* Pattern of cancer/testis antigen expression in lung cancer patients. *Int. J. Mol. Med.* **29**, 656–662 (2012).

121. Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251–263 (2012).

122. Cancer Genome Atlas Network. Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681–1696 (2015).

123. Conway, J. R. *et al.* Integrated molecular drivers coordinate biological and clinical states in

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

melanoma. *Nat. Genet.* **52**, 1373–1383 (2020).

124. Tokheim, C. & Karchin, R. CHASMplus Reveals the Scope of Somatic Missense Mutations Driving Human Cancers. *Cell Syst* **9**, 9–23.e8 (2019).

125. Carter, H. *et al.* Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res.* **69**, 6660–6667 (2009).

126. Nielsen, M. & Andreatta, M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med.* **8**, 33 (2016).

127. Andersen, M. H. *et al.* Immunogenicity of constitutively active V599EBRaf. *Cancer Res.* **64**, 5456–5460 (2004).

128. Bradley, S. D. *et al.* BRAFV600E Co-opts a Conserved MHC Class I Internalization Pathway to Diminish Antigen Presentation and CD8+ T-cell Recognition of Melanoma. *Cancer Immunol Res* **3**, 602–609 (2015).

129. Williams, J. D., Jacobson, E. L., Kim, H., Kim, M. & Jacobson, M. K. Folate in Skin Cancer Prevention. *Subcell. Biochem.* **56**, 181–197 (2012).

130. Simpson, A. J. G., Caballero, O. L., Jungbluth, A., Chen, Y.-T. & Old, L. J. Cancer/testis antigens, gametogenesis and cancer. *Nat. Rev. Cancer* **5**, 615–625 (2005).

131. Fang, D. *et al.* Expression of Microtubule-Associated Protein 2 in Benign and Malignant Melanocytes. *Am. J. Pathol.* **158**, 2107–2115 (2001).

132. Vachtenheim, J., Novotna, H. & Ghanem, G. Transcriptional Repression of the Microphthalmia Gene in Melanoma Cells Correlates with the Unresponsiveness of Target Genes to Ectopic Microphthalmia-Associated Transcription Factor. *J. Invest. Dermatol.* **117**, 1505–1511 (2001).

133. Marcu, A. *et al.* HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. *J Immunother Cancer* **9**, (2021).

134. Aris, M. *et al.* MART-1- and gp100-Expressing and -Non-Expressing Melanoma Cells Are

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Equally Proliferative in Tumors and Clonogenic In Vitro. *J. Invest. Dermatol.* **132**, 365–374 (2012).

135. Journe, F. *et al.* TYRP1 mRNA expression in melanoma metastases correlates with clinical outcome. *Br. J. Cancer* **105**, 1726–1732 (2011).

136. Wieczorek, M. *et al.* Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Front. Immunol.* **8**, 292 (2017).

137. Bodis, G., Toth, V. & Schwarting, A. Role of Human Leukocyte Antigens (HLA) in Autoimmune Diseases. *Rheumatol Ther* **5**, 5–20 (2018).

138. Miyadera, H., Ohashi, J., Lernmark, Å., Kitamura, T. & Tokunaga, K. Cell-surface MHC density profiling reveals instability of autoimmunity-associated HLA. *J. Clin. Invest.* **125**, 275–291 (2015).

139. Košmrlj, A. *et al.* Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature* **465**, 350–354 (2010).

140. Chappell, P. E. *et al.* Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. *Elife* **4**, e05345 (2015).

141. International HIV Controllers Study *et al.* The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* **330**, 1551–1557 (2010).

142. Zhang, Y. *et al.* Multilayered defense in HLA-B51-associated HIV viral control. *J. Immunol.* **187**, 684–691 (2011).

143. Mobbs, J. I. *et al.* The molecular basis for peptide repertoire selection in the human leucocyte antigen (HLA) C*06:02 molecule. *J. Biol. Chem.* **292**, 17203–17215 (2017).

144. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature* **353**, 321–325 (1991).

145. Saveanu, L. *et al.* Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat. Immunol.* **6**, 689–697 (2005).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

146. Saric, T. *et al.* An IFN-γ–induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I–presented peptides. *Nat. Immunol.* **3**, 1169–1176 (2002).

147. York, I. A. *et al.* The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. *Nat. Immunol.* **3**, 1177–1184 (2002).

148. Hammer, G. E., Gonzalez, F., James, E., Nolla, H. & Shastri, N. In the absence of aminopeptidase ERAAP, MHC class I molecules present many unstable and highly immunogenic peptides. *Nat. Immunol.* **8**, 101–108 (2006).

149. Reeves, E., Edwards, C. J., Elliott, T. & James, E. Naturally occurring ERAP1 haplotypes encode functionally distinct alleles with fine substrate specificity. *J. Immunol.* **191**, 35–43 (2013).

150. Martín-Esteban, A., Gómez-Molina, P., Sanz-Bravo, A. & López de Castro, J. A. Combined effects of ankylosing spondylitis-associated ERAP1 polymorphisms outside the catalytic and peptide-binding sites on the processing of natural HLA-B27 ligands. *J. Biol. Chem.* **289**, 3978–3990 (2014).

151. Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2 *et al.* A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat. Genet.* **42**, 985–990 (2010).

152. Tsoi, L. C. *et al.* Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat. Genet.* **44**, 1341–1348 (2012).

153. Yin, X. *et al.* Genome-wide meta-analysis identifies multiple novel associations and ethnic heterogeneity of psoriasis susceptibility. *Nat. Commun.* **6**, 1–11 (2015).

154. Cummings, A. L. *et al.* Mutational landscape influences immunotherapy outcomes among patients with non-small-cell lung cancer with human leukocyte antigen supertype B44. *Nature Cancer* **1**, 1167–1175 (2020).

155. DiBrino, M. *et al.* Identification of the peptide binding motif for HLA-B44, one of the most common HLA-B alleles in the Caucasian population. *Biochemistry* **34**, 10130–10138 (1995).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

156. Brewerton, D. A. *et al.* Ankylosing spondylitis and HL-A 27. *Lancet* **1**, 904–907 (1973).

157. Brewerton, D. A., Caffrey, M., Nicholls, A., Walters, D. & James, D. C. HL-A 27 and arthropathies associated with ulcerative colitis and psoriasis. *Lancet* **1**, 956–958 (1974).

158. Tam, L.-S., Gu, J. & Yu, D. Pathogenesis of ankylosing spondylitis. *Nat. Rev. Rheumatol.* **6**, 399–405 (2010).

159. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).

160. Fan, Y. *et al.* MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol.* **17**, 178 (2016).

161. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).

162. Larson, D. E. *et al.* SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* **28**, 311–317 (2012).

163. Lin, Y. *et al.* scMerge leverages factor analysis, stable expression, and pseudoreplication to merge multiple single-cell RNA-seq datasets. *Proceedings of the National Academy of Sciences* **116**, 9775–9784 (2019).

164. Mi, H. *et al.* PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, D394–D403 (2021).

165. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).

166. Aguiar, V. R. C., Masotti, C., Camargo, A. A. & Meyer, D. HLApers: HLA Typing and Quantification of Expression with Personalized Index. in *Bioinformatics for Cancer Immunotherapy: Methods and Protocols* (ed. Boegel, S.) 101–112 (Springer US, 2020).

167. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

168. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).

169. Robinson, J. *et al.* IPD-IMGT/HLA Database. *Nucleic Acids Res.* **48**, D948–D955 (2020).

170. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

171. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).

172. Pérez-Guijarro, E. *et al.* Multimodel preclinical platform predicts clinical response of melanoma to immunotherapy. *Nat. Med.* **26**, 781–791 (2020).

173. Jiang, P. *et al.* Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558 (2018).

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

175. Armitage, P. & Doll, R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* **8**, 1–12 (1954).

176. Moolgavkar, S. H. & Knudson, A. G., Jr. Mutation and cancer: a model for human carcinogenesis. *J. Natl. Cancer Inst.* **66**, 1037–1052 (1981).

177. Moolgavkar, S. H., Dewanji, A. & Venzon, D. J. A stochastic two-stage model for cancer risk assessment. I. The hazard function and the probability of tumor. *Risk Anal.* **8**, 383–392 (1988).

178. Moolgavkar, S. H. & Luebeck, G. Two-event model for carcinogenesis: biological, mathematical, and statistical considerations. *Risk Anal.* **10**, 323–341 (1990).

179. Brouwer, A. F., Meza, R. & Eisenberg, M. C. Parameter estimation for multistage clonal expansion models from cancer incidence data: A practical identifiability analysis. *PLoS*

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.12.456166; this version posted August 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

*Comput. Biol.* **13**, e1005431 (2017).

180. Meza, R., Jeon, J., Moolgavkar, S. H. & Georg Luebeck, E. Age-specific incidence of cancer: Phases, transitions, and biological implications. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 16284–16289 (2008).

181. Tan, W.-Y. *Stochastic models of carcinogenesis*. (CRC Press, 2020).

182. Luebeck, G. E., Vaughan, T. L., Curtius, K. & Hazelton, W. D. Modeling historic incidence trends implies early field cancerization in esophageal squamous cell carcinoma. *PLoS Comput. Biol.* **17**, e1008961 (2021).

183. Seabold, S. & Perktold, J. Statsmodels: Econometric and statistical modeling with python. in *Proceedings of the 9th Python in Science Conference* vol. 57 61 (Austin, TX, 2010).