

# **Defining the genetic determinants of CD8<sup>+</sup> T cell receptor repertoire in the context of immune checkpoint blockade**

Esther S. Ng<sup>1</sup>, Orion Tong<sup>2</sup>, Chelsea Taylor<sup>2</sup>, Robert Watson<sup>2,3,4</sup>, Bo Sun<sup>2,5</sup>, Gusztav Milotay<sup>2</sup>, Sophie MacKay<sup>2</sup>, James J Gilchrist<sup>2,7</sup>, Martin Little<sup>2,3,4</sup>, Benjamin P Fairfax\*<sup>§2,3,4,6</sup>, and Yang Luo\*<sup>1</sup>

<sup>1</sup>*Kennedy Institute of Rheumatology; Oxford, UK*

<sup>2</sup>*MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK*

<sup>3</sup>*Department of Oncology, University of Oxford, Oxford, UK*

<sup>4</sup>*Cancer and Haematology Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK*

<sup>5</sup>*Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford, UK*

<sup>6</sup>*NIHR Oxford Biomedical Research Centre*

<sup>7</sup>*Department of Paediatrics, University of Oxford, Oxford, UK*

\**Equal contributions*

<sup>1</sup>*Corresponding author: benjamin.fairfax@oncology.ox.ac.uk*

## Abstract

CD8<sup>+</sup> T cells play a central role in the cancer response to Immune checkpoint blockade (ICB) treatment, with activity predicated upon antigen recognition by the associated T cell receptor (TCR) repertoire. The contribution of genetic variation to this in cancer treatment is under-explored. We have conducted a genome-wide and human leukocyte antigen (HLA)-focused analysis of CD8<sup>+</sup> T cell TCR repertoire to identify genetic determinants of variable gene (V-gene) and CDR3 K-mer usage from samples taken prior to and after ICB treatment (n=250). We find 11 significant *cis* associations and 10 *trans* associations, primarily to the HLA, with V-gene usage meeting permuted p-value thresholds. Notably, TCR clones containing V-genes associated with HLA were less likely to be persistent across treatment. Conversely, clones containing V-genes corresponding to those associated with Class I HLA alleles demonstrated increased expression of a signature of tumour reactivity in individuals carrying matching alleles. Our work indicates a complex relationship between genotype and TCR repertoire in the context of treatment with ICB, which has novel implications for understanding determinants of treatment response and patient outcomes.

**One Sentence Summary:** TCR repertoire is strongly associated with specific HLA alleles in cancer patients, but immune checkpoint blockade influences this association.

## INTRODUCTION

Immune checkpoint proteins, of which the cell-surface receptors programmed cell death protein 1 (PD-1) and cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) are archetypal examples, play key roles in restraining the magnitude of T cell responses, limiting deleterious inflammation and autoimmunity. In the context of cancer, up-regulation and ligation of PD-1 and CTLA-4 are implicated in the development of T-cell exhaustion and suppression of anti-tumour immunity. Correspondingly, Immune Checkpoint Blockade (ICB) treatment leads to the re-invigoration of exhausted T cells enhancing cell division and cytotoxic activity, and broadening the T cell repertoire[1, 2, 3] and has markedly improved outcomes in metastatic melanoma (MM) and numerous other cancers. Notably combination ICB with monoclonal antibodies to CTLA-4 (ipilimumab) and PD-1 (nivolumab) for MM is associated with extending median survival from months to beyond five years [4, 5].

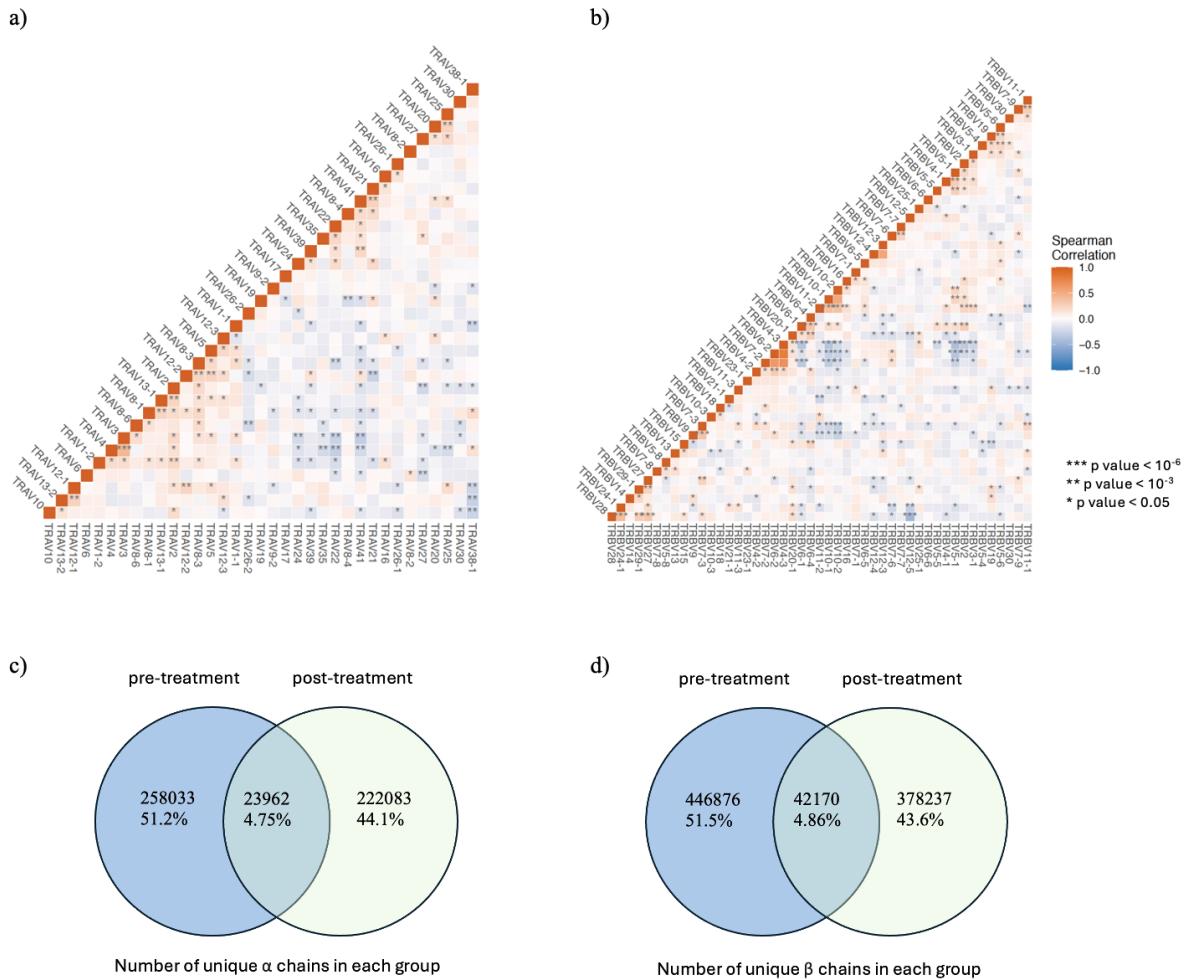
Effective anti-tumour responses requires antigen recognition by CD8<sup>+</sup> T cells, with Tumour Associated Antigens (TAA) presented by Class I Human Leukocyte Antigen (HLA) alleles in complex with  $\beta$ -2 microglobulin to reactive TCR. The TCR of HLA reactive CD8<sup>+</sup> T cells comprise a heterodimer of  $\alpha$  and  $\beta$  chains encoded on chromosome 14 and 9 respectively. Within the TCR  $\beta$  chain, there exists 3 complementarity-determining regions (CDR1, 2 and 3). CDR 1 and 2 are primarily interface with HLA  $\alpha$  helices whilst interactions with the hypervariable CDR3 region is the critical determinant of peptide recognition [6]. In both the  $\alpha$  and  $\beta$  chains, the CDR3 loops have the highest sequence diversity and are the principal determinants of receptor specificity [7].

The CDR1 and 2 loops are determined by the V-gene segment in the TCR, while the CDR3 is determined by the V, D and J gene segment. The TCR is formed through a process of V(D)J recombination in the early stage of T cell maturation in which these gene segments are recombined to encode a final full length transcript. This process results in a combinatorial possibility of  $10^{10}$  possible TCR sequences [8], vital for enabling recognition of a vast array self or pathogen derived peptides.

The TCR repertoire of an individual is determined by a complex interplay between antigen exposure and by germline genetics, impacting HLA and ability to present specific antigens. Early thymic selection is crucial in repertoire determination, with deletion from the repertoire of T cell clones expressing TCRs either unresponsive or overtly reactive to self-peptide MHC complexes[9][10]. The interaction between germline genetics at the MHC and TCR repertoire thus has significant implications on our understanding of the pathophysiology of directly immune-mediated diseases and those where immunity plays a key role in clinical responses to therapeutics including cancer. However, in the context of ICB therapy the relationship between germline genetics including HLA alleles and TCR usage are relatively unexplored and any impact on clinical response to treatment is unknown.

To address this we have investigated this relationship in CD8<sup>+</sup> T cells from 250 patients receiving ICB for cancer, looking at TCR usage and genetics in both the untreated and post-ICB states. By integrating germline genotyping and bulk TCR sequencing both prior to and post first cycle of ICB, we explored the germline determinants of TCR repertoire and impact of ICB treatment on this. We describe a subset of TCR that are strongly genetically determined in *trans* by HLA alleles. By integrating observations with scRNA-seq data we find associations are to specific CD8<sup>+</sup> T cell subsets, which behave distinctly on

treatment, and are associated with long-term survival in patients.



**Figure 1. Summary of TCR chains in our study.** Correlation between **a)** TCR  $\alpha$  V-gene usage **b)** TCR  $\beta$  V-gene usage in our cohort. In each plot, V-gene usage is calculated by normalising the number of unique clones per individual. **c)** Number of TCR  $\alpha$  chains and **d)** TCR  $\beta$  chains in three groups of cells - those seen only in individuals pre-treatment, those seen in individuals only post-treatment and those seen both pre- and post-treatment, where treatment represents first cycle of immune-checkpoint blockade.

## RESULTS

### Exploring CD8 $^{+}$ T cell receptor repertoire

We performed bulk RNA sequencing of CD8 $^{+}$  T cells isolated from whole blood prior to and after ICB treatment from 250 cancer patients (**Supplementary Table 1**). We explored the relationship between  $\alpha$  and  $\beta$  chain usage, identifying conserved blocks of correlating chains (**Figure 1a,b**). We performed principal component (PC) analyses of correlations between  $\alpha$  and  $\beta$  chain V-genes, and demonstrated conserved patterns of chain usage (**Supplementary Figure 1a,b**). Most notably, the first PC of the  $\alpha$  chain was dominated by TRAV1-2 which forms the key TCR for Mucosal-associated invariant T cells (MAIT) cells. Correspondingly, the PC of the  $\beta$  chain correlations reflected this pattern with the first PC dominated by TRBV6-4, the key partner of TRAV1-2.

For 179 patients we had TCR data from both prior to and after the first cycle of ICB. Across two timepoints, we identified 504,078 unique  $\alpha$  chains and 867,283 unique  $\beta$  chains. Analysis of chain conservation across ICB treatment demonstrated that only a small number of clones were resampled across both timepoints, likely reflecting the known increase in T cell mitosis induced by ICB (**Figure 1c,d**).

### GWAS of V-gene usage reveals *cis*-associations

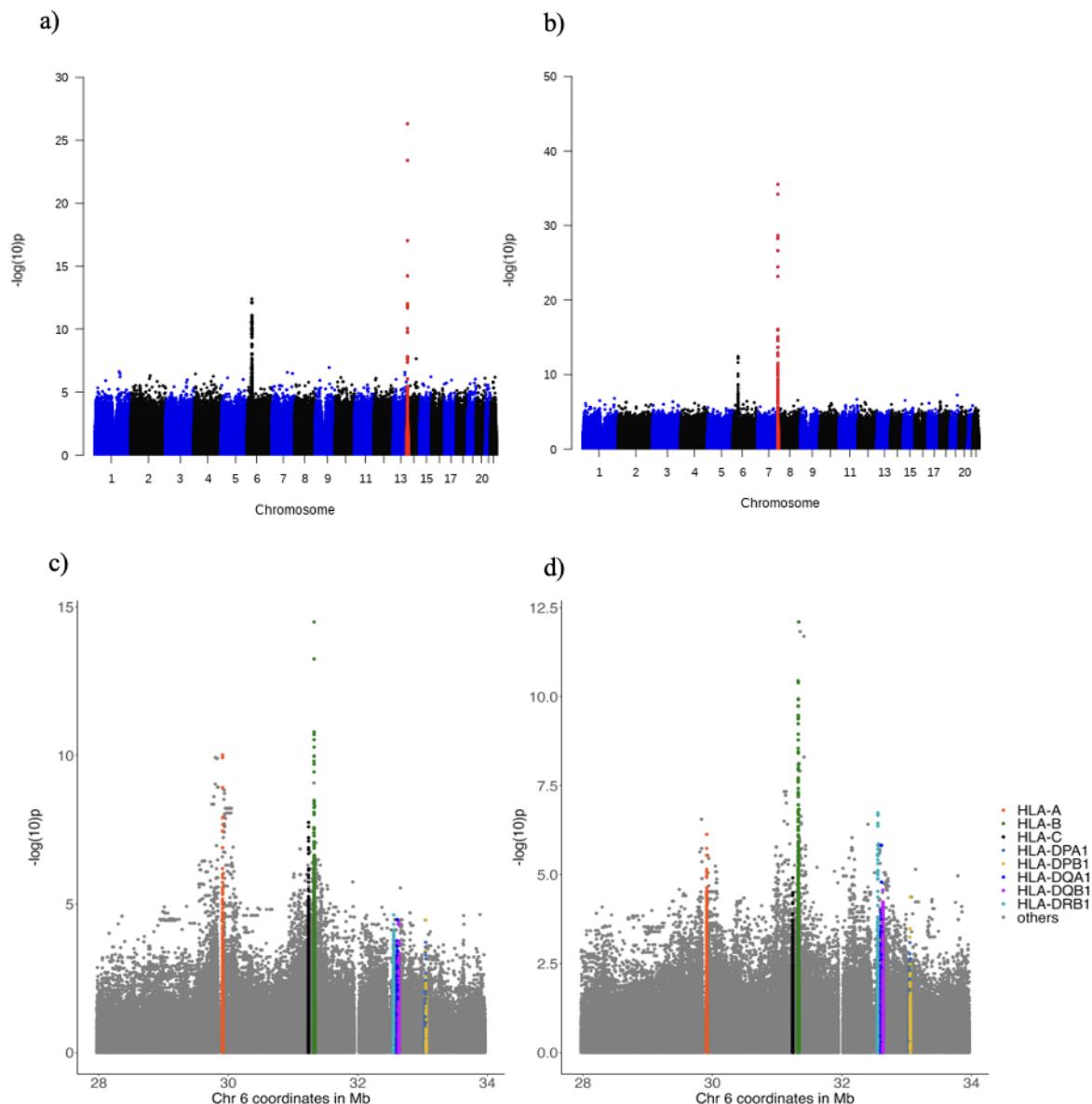
We proceeded to explore the relationship between genetics and TCR usage, focusing upon pre-treatment samples where our power was greatest to detect effect (n=250). Patients were genotyped genome-wide with 486,469 SNPs available for post-QC analysis. V-gene usage for  $\alpha$  and  $\beta$  chains was defined as normalised count of unique clones per V-gene and modelled as a quantitative trait. For each gene this was associated with genotype using an additive model correcting for two genetic PCs, two TCR PCs, age, gender and cancer type. In total, associations were performed for 47  $\beta$  chains and 42  $\alpha$  chains V-genes from pre-treatment samples. In total, we ran  $> 40$  million regression models, observing the previously described *cis* associations to the  $\alpha$  and  $\beta$  chain corresponding to their genomic loci  $\alpha$  (**Figure 2a**) and  $\beta$  chain V-genes (**Figure 2b**). Additionally we noted a strong secondary trans-acting signal at 6p23, corresponding to the MHC in both chains.

To account for multiple testing and also correlation between V-gene usage, we permuted the phenotype dataset 1,000 times, preserving the V-genes that each individual had but re-ordering the samples, resulting in a permutation p-value threshold of  $3.83 \times 10^{-9}$  for  $\beta$  chain V-gene (TRBV) and  $3.46 \times 10^{-9}$  for  $\alpha$  chain V-gene (TRAV) (**Supplementary Figure 2a,b**).

In total, we found nine TRBV-genes which were associated with SNPs at p-values below the permuted significance threshold (**Supplementary Table 2**). The most significant signal was an association between TRBV28 and rs4726571 ( $\beta=-0.895$ , p-value= $3.12 \times 10^{-36}$ ). TRBV28 is a significant prognostic marker in renal cancer, melanoma, endometrial, cervical, head and neck cancer[11]. In hepatocellular carcinoma, it has been found to be more highly expressed in tumour tissue compared to surrounding healthy tissue[12]. As for the  $\alpha$  chain, only TRAV26-2 and TRAV38-1 were significantly associated with SNPs in the TRAV-gene region. TRAV26-2 has been shown to be prognostic for survival in head and neck cancer[13] and bladder cancer[14].

We further performed conditional analysis on all the significant V-genes. The only secondary signal we found was in TRBV4-3 with top primary signal SNP being rs361489 ( $\beta=0.571$ , p-value= $9.83 \times 10^{-12}$ ) and top secondary signal SNP being rs17249 ( $\beta=-0.460$ , p-value= $2.66 \times 10^{-11}$ ), suggesting that there were multiple germline variants determining this V-gene.

There are several mechanisms through *cis* acting polymorphisms may influence V-gene usage. They could influence recombination by interrupting the recombination signal sequence [15] or they could be located within regulatory regions such as promoters and enhancers [16].



**Figure 2. Association between TCR V-gene usage and germline genetic variation.** a) Manhattan plot of association between TCR  $\alpha$  chain V-gene usage. The TRAV gene locus is highlighted in red. b) Manhattan plot of TCR  $\beta$  chain V-gene usage. The TRBV gene locus is highlighted in red. The strength of association is indicated by the  $-\log_{10}$  of the p-value of the linear model fitted between each V-gene usage and each variant on the y-axis. The x-axis shows the genomic position (in GRCh38). All V-genes are plotted together and TRAV and TRAV gene loci are highlighted in red. c) Locus plot of association between TRAV usage and variants in the MHC region d) Locus plot of association between TRBV usage and MHC variants. Classical HLA genes are annotated with different colours in the legend.

### V-gene usage is associated with classical HLA alleles

To further investigate the *trans* signal mapping to 6p23 which corresponds to the MHC region, we associated V-gene usage with imputed variation in the MHC region including classical HLA alleles and coding polymorphisms. This corresponds to 16,781 SNPs, 174 classical alleles, and 2,119 amino acids. To control for false positives, we performed a permutation analysis by randomly shuffling the phenotypes and fitting the regression model 1,000 times. We found that the empirical p-value thresholds for significance

were  $3.37 \times 10^{-7}$  for  $\beta$  and  $3.61 \times 10^{-7}$  for  $\alpha$  chain V genes respectively (**Supplementary Figure 2b,c**). In total, there were five  $\beta$  chain V-genes and five  $\alpha$  chain V-genes that had significant associations with variants within the MHC region (**Supplementary Table 3, Supplementary Figure 3, Figure 2**).

We found that the most significant association in the  $\alpha$  chain was HLA-A\*02 with TRAV12-2 ( $\beta=0.689$ , p-value= $3.04 \times 10^{-13}$ ) (**Figure 3a,b**). In a study on melanocyte antigen (Melan-A) specific TCR, there is a strong bias towards TRAV12-2 usage with predominant interaction between the TRAV12-2 chain and the HLA-A2/Melan-A peptide located in the CDR1loop (Gln31) [17]. These results suggest that germline determined V-gene usage may play a role in anti-melanoma immunity. There were no secondary signals in the conditional analysis of any of the TRAV-genes.

We found that the strongest association in  $\beta$  chain was between TRBV19 and rs2250287 ( $\beta=0.760$ , p-value= $7.93 \times 10^{-13}$ , **Figure 3c**). The top classical HLA allele associated with TRBV19 was HLA-B\*44 ( $\beta=0.734$ , p-value= $1.83 \times 10^{-10}$ ) (**Figure 3d**). rs2250287 and HLA-B\*44 have a linkage disequilibrium  $r^2$  of 0.882.

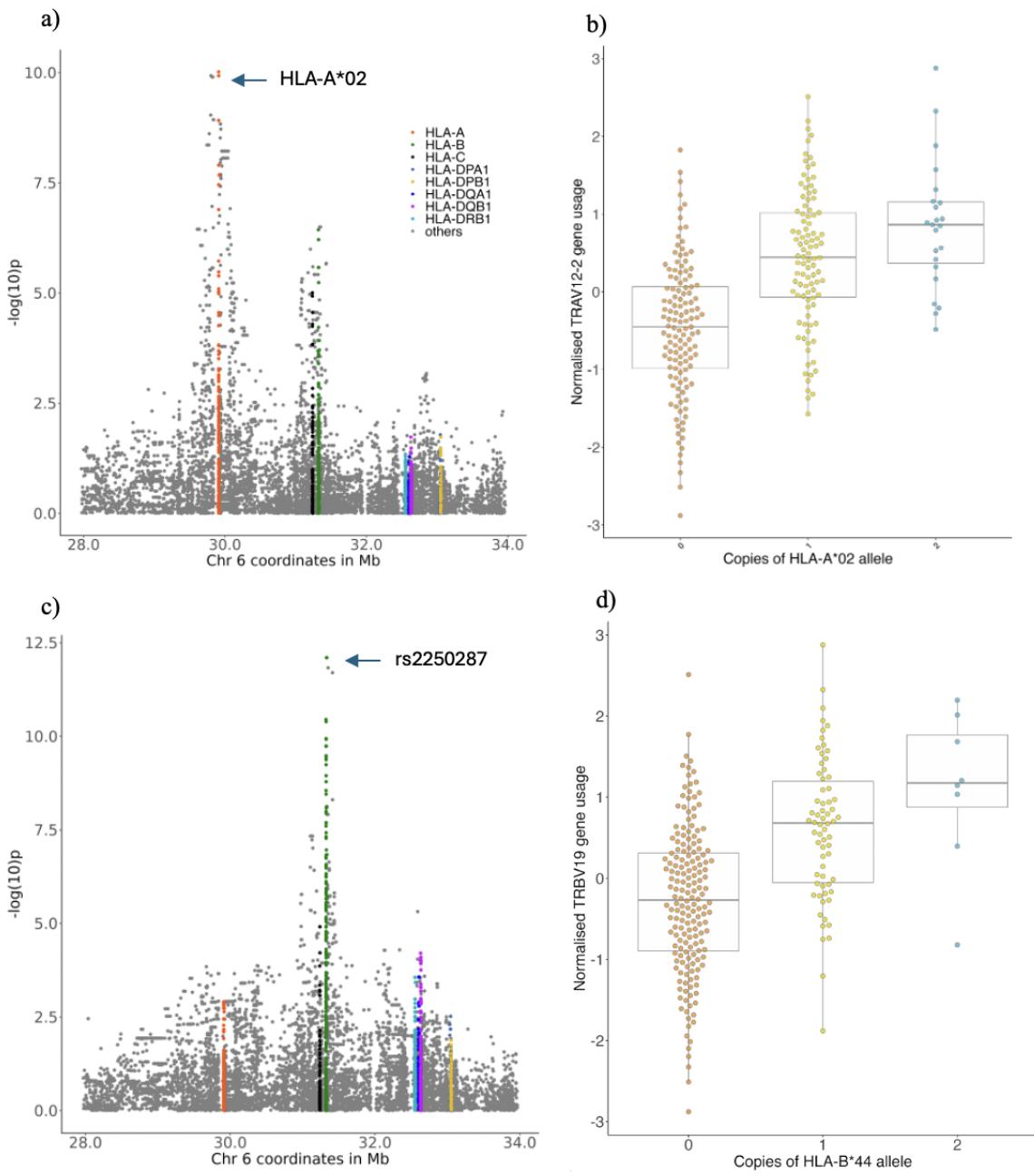
At a four-digit resolution, TRBV19 was associated with HLA-B\*44:02 ( $\beta=0.576$ , p-value= $2.78 \times 10^{-5}$ ) and HLA-B\*44:03 ( $\beta=0.725$ , p-value= $1.23 \times 10^{-4}$ ). HLA-B\*44:02 is associated with protection against multiple sclerosis in several studies. [18, 19, 20]. In the context of melanoma, carriage of the B-44 supertype has been associated with improved outcomes to ICB treatment[21], although this was not reproduced in a follow-up study[22]. Notably though, in a study involving Melan-A reactive T cells from patients with metastatic melanoma, TRBV19 was enriched for TCR reactive to this common antigen[17]. We further performed conditional analysis and observed only one secondary signal among the significantly associated V-genes - TRBV19 with the top secondary SNP being rs62395263 ( $\beta=0.705$ , p-value= $2.96 \times 10^{-7}$ ).

To determine whether specific amino acid positions were associated with V-gene usage, we tested each of the amino acid positions by grouping haplotypes carrying a specific residue at each position in an additive model (**Supplementary Table 4**). We found that the strongest the most significant association in  $\alpha$  chain was between TRAV12-2 and position 74 (exon2) in HLA-A ( $P_{omnibus}=2.93 \times 10^{-13}$ ), and the most significant association in  $\beta$  chain was between TRBV19 and position 45 (exon2) in HLA-B ( $P_{omnibus}=5.08 \times 10^{-11}$ ). These positions had stronger association signals than any single SNP or classical HLA allele, and fall within the peptide-binding groove of the respective HLA protein, indicating that variation in the amino acid content of the peptide-binding groove is the major genetic determinant of V-gene usage.

### **Refinement of genetic signal using single-cell sequencing**

To validate our associations we performed single-cell TCR sequencing data for 59 individuals, 55 of whom were also included in the bulk analysis. The advantage of single cell data is that we are able to separate out different cell types and look for cell specific association. We were able to replicate the top genome-wide associations in CD8<sup>+</sup> cells (**Supplementary Table 5**).

Our top  $\beta$  chain association, rs4726571 with TRBV28, was also seen in single-cell analysis for all CD8<sup>+</sup> cells together ( $\beta=-0.339$ , p-value=0.00412). It was also seen in all CD4<sup>+</sup> cells together albeit with a weaker association ( $\beta=-0.320$ , p-value=0.00551). The association was stronger in CD8 terminally differentiated



**Figure 3. Association between V-gene usage and MHC variants.** **a)** MHC locus plot of TRAV12-2 showing top signal in HLA-A. Each dot represents an MHC variant **b)** TRAV12-2 plotted against most significant HLA variant HLA-A\*02. Each dot represents an individual. **c)** MHC locus plot of TRBV19 showing top signal in HLA-B **d)** TRBV19 plotted against most significant classical HLA allele HLA-B\*44.

effector memory cells re-expressing CD45RA (TEMRA) cells ( $\beta=-0.489$ , p-value= $5.53 \times 10^{-5}$ ) and CD8 T Effector Memory (TEM) cells ( $\beta=-0.426$ , p-value= $1.35 \times 10^{-4}$ ) compared to CD8 Naive cells ( $\beta=-0.236$ , p-value=0.0700)

Our top V-gene MHC associations were also replicated in the CD8<sup>+</sup> single-cell analysis (Supplementary Table 6). The direction of effect of the TRBV19 association with HLA-B\*44 was the same across all CD8 T<sup>+</sup> cells ( $\beta=0.0823$ , p-value=0.391) but this association was not observed CD4 T cells ( $\beta$  negative). Again, looking at CD8<sup>+</sup> T cell subsets we noted a stronger signal in CD8 TEMRA cells ( $\beta=0.373$ ,

p-value=0.0662) compared to CD8 Naive cells ( $\beta=0.0119$ , p-value=0.950).

### CDR3 amino acid K-mers are associated with HLA alleles

Within the TCR  $\beta$  chain, the complementarity-determining region (CDR) 1 and 2 loops contact the MHC  $\alpha$  helices while the hypervariable CDR3 regions contact the peptide. In both  $\beta$  and  $\alpha$  chains, CDR3 loops have the greatest diversity in sequence and are the crucial determinants of receptor binding specificity[23]. We hypothesised that CDR3 would be associated with variants in the MHC region. To test this, we employed a K-mer based approach in bulk sequencing data, using a sliding window of 7 amino acids to quantify K-mer motif usage in the CDR3 region (6654  $\beta$  chain and 2306  $\alpha$  chain K-mers) across individuals. We then fitted a linear model to test the association between K-mers and HLA alleles using the same covariates included in the V-gene HLA association study.

There were two significantly associated K-mers, both of which were on the  $\beta$  chain (**Supplementary Figure 4**). There were no significant associations in the  $\alpha$  chain. The top  $\beta$  chain K-mer was TGDSNQP, associated with HLA-B\*35:01 ( $\beta=1.41$ , p-value= $1.57 \times 10^{-13}$ ). HLA-B\*35 correlates with a favourable outcome in melanoma following the administration of allogeneic melanoma vaccine [24]. The K-mer TSGDYNE was associated with rs3763288 ( $\beta=1.12$ , p-value= $2.66 \times 10^{-10}$ ). rs3763288 is an intron variant in the Major histocompatibility complex class I-like molecule A (*MICA*). Cancer cells may escape natural killer cell immune surveillance by releasing *MICA* from the cell surface into a soluble form (*sMICA*) which can be detected in blood [25]. rs3763288 is a quantitative trait locus for levels of soluble *MICA* in blood [26] and is also associated with Takayasu arteritis [27]. *MICA* is a major ligands for natural-killer group 2, member D (*NKG2D*) receptors [28]. *NKG2D* recognises stress-induced surface ligands and triggers a cytotoxic response [29]. In CD8+ T cells, simultaneous activation of the TCR is required for *NKG2D* to be functional. This could potentially help to explain how a SNP affecting *sMICA* levels would be associated with CDR3 K-mer usage.

To test whether the observed association between K-mer and HLA allele was influenced by V-gene usage, we performed a correlation analysis between CDR3 K-mers and V-gene usage. We found no V-genes associated with TGDSNQP indicating that the association is independent of V-genes. TSGDYNE was associated with TRBV19 ( $r=0.276$ , p. p-value= $9.61 \times 10^{-6}$ ). We repeated our analysis of TSGDYNE against rs3763288 this time conditioning on TRBV19 and the association remained significant ( $\beta=1.01$ , p-value= $1.06 \times 10^{-8}$ ), suggesting that the association between TSGDYNE and rs3763288 is independent of TRBV19.

### ICB alters HLA V-gene association

Our findings regarding the frequency of V-gene usage and HLA allele status were based on CD8<sup>+</sup> T cells from patients prior to treatment. We wanted to explore the effect of ICB treatment on potential HLA×V-gene selection. In the dataset, most patients had paired CD8<sup>+</sup> T cell samples before and after their first cycle of ICB treatment - a key strength of the dataset - allowing us to explore genotype×V-gene usage post-treatment.

Focusing on the 179 individuals for which paired samples were available, we sub-divided clones into three groups: those only observed pre-treatment (*Unstable*), those observed post-treatment only (*Novel*),

and those present across both samples (*Persistent*).

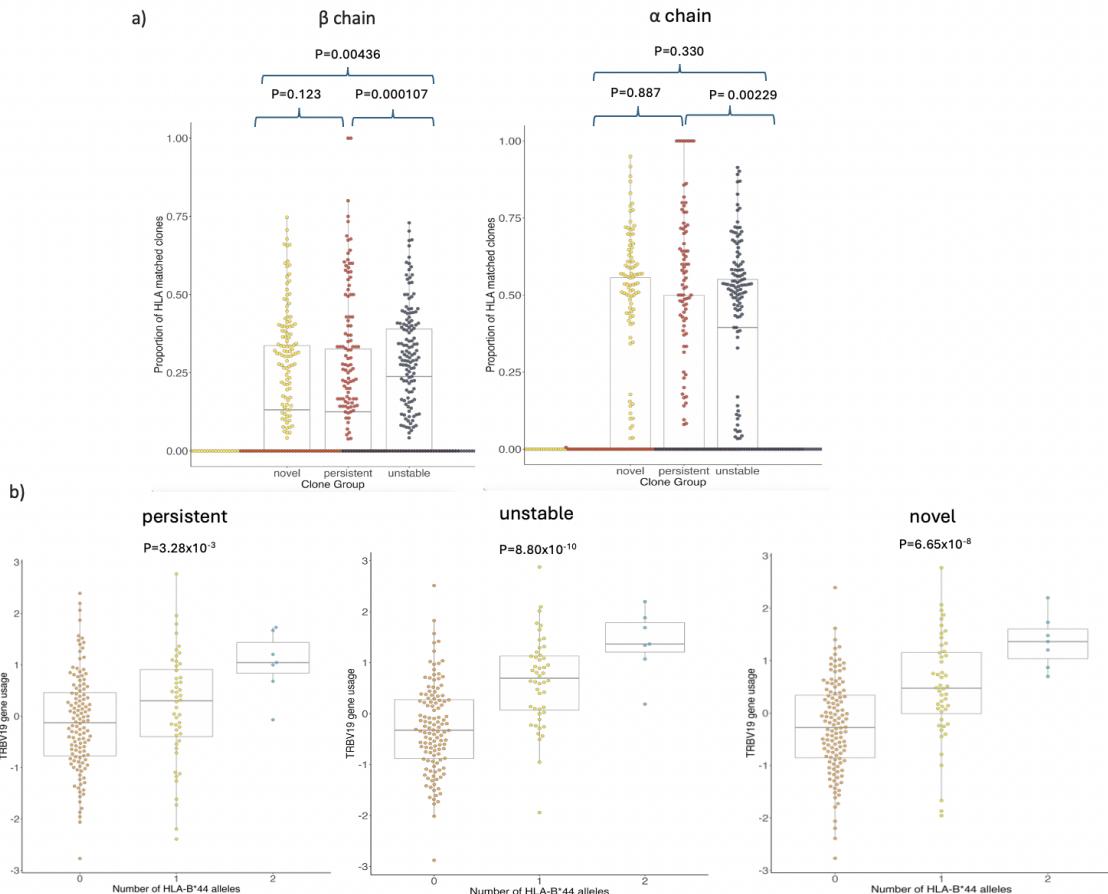
Using V-genes significantly associated with variants in the MHC region, we extracted the top classical HLA allele associated with each V-gene. HLA-matched clones were defined as clones which had this classical HLA-V-gene pairing. HLA-unmatched cells were defined as all other clones. We then calculated the proportion of HLA-matched clones for each individual within each clone group (Unstable, Persistent and Novel).

To investigate whether ICB alters the proportion of HLA-matched clones, we performed a paired Wilcoxon test on proportions of HLA-matched clones between unstable and novel groups. For the  $\beta$  chain, we found that the proportion of HLA matched clones in each individual in the unstable group was significantly higher (median = 0.238) than in the novel group (median = 0.132, p-value=0.00436). The proportion of HLA-matched clones in the unstable group was also significantly higher than the persistent group (median = 0.125, p-value=0.000107, **Figure 4a**). Although the median proportion of HLA-matched clones was higher in the novel compared to persistent group, this was not statistically significant.

For the  $\alpha$  chain, proportion of HLA-matched clones was higher in the unstable group than the persistent group (p-value = 0.00229). There was no significant difference in proportions between unstable and novel groups or between persistent and novel groups. (**Figure 4a**).

We also performed HLA V-gene association analysis within these three groups of clones (**Supplementary Table 7**). Our top  $\beta$  chain association (TRBV19 with HLA-B\*44) was strongest in the unstable group, followed by the novel group, and was weakest in the persistent group (**Figure 4b**).

In both  $\alpha$  and  $\beta$  chains, we observed that the median proportions of HLA-matched clones is highest pre-treatment only group, followed by the post-treatment only group and then the intersection of the two groups. This suggests that in both  $\alpha$  and  $\beta$  chains, naïve TCRs are more likely to be HLA-selected. Furthermore, HLA-selected TCRs are less long-lived than their non-HLA selected counterparts, suggesting that TCR persistence through time and ICB treatment could be driven by other selection processes such as cancer antigens.



**Figure 4. HLA-Vgene association across ICB treatment** a) This shows the proportion of HLA-matched clones in 3 groups of clones - those only observed pre-treatment (*Unstable*), those observed post-treatment only (*Novel*), and those present across both samples (*Persistent*). Each dot represents an individual b) TRBV19 plotted against HLA-B\*44 in each of the 3 groups of clones demonstrating that the weakest association is in the persistent clones

### Cells with HLA matched V-gene demonstrated increased tumour reactivity expression profiles

As our results show that naïve TCR repertoires are more likely to be HLA selected and also that HLA selected T cells are less likely to persist through ICB, we sought to determine whether HLA selected T cells were more or less likely to be tumour-reactive. To do this, we used a set of 20 genes identified to characterise tumour-reactive cells based on single-cell sequencing in melanoma patients [30] to generate a Tumour-Reactivity Score (TRS) (**Supplementary Table 8**). We calculated TRS for each cell in a combined single-cell RNA sequencing and TCR sequencing experiment across 59 individuals. We then separated the cells into two groups based on their HLA and V-genes using the same definition as the section above. Using V-genes significantly associated with variants in the MHC region, we extracted the top classical HLA allele associated with each V-gene. HLA-matched cells were defined as cells which had this classical HLA-V-gene pairing and HLA-unmatched cells were defined as all other cells. We calculated TRS for each cell and then fitted a linear model against HLA-matching status. We found HLA-matched pre-ICB CD8<sup>+</sup> cells displayed a significantly increased TRS ( $\beta=0.0854$ , p-value= $6.40 \times 10^{-4}$ ).

We subsequently explored whether the increased TRS seen in HLA-matched cells were also seen in pre-

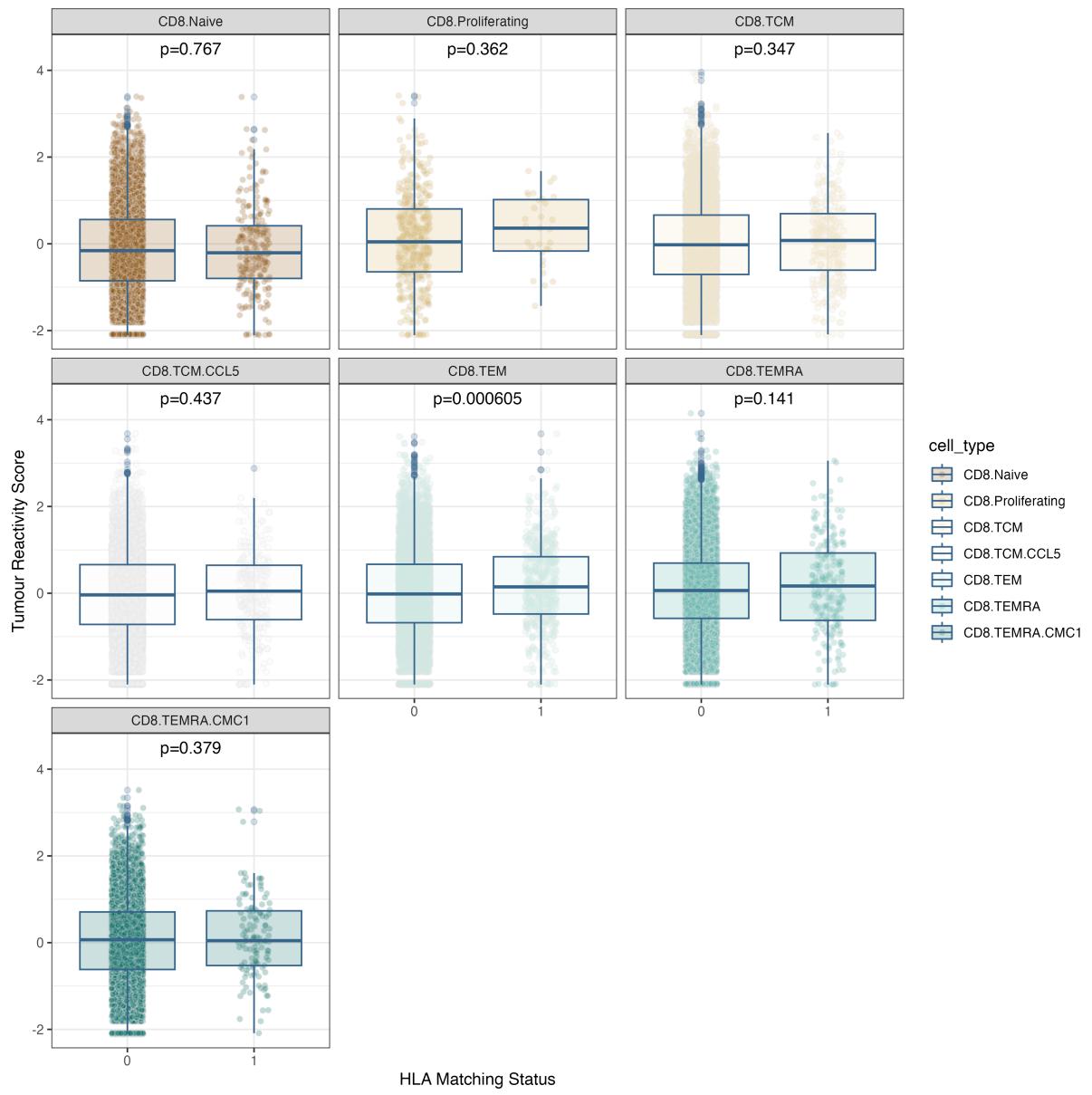
defined cell subtypes (**Figure 5**). Here, we found the most robust association was in T effector memory cells (CD8 TEM) ( $\beta=0.156$ , p-value= $6.00 \times 10^{-4}$ ). Further, we tested this association in both unique cells and cells belonging to a clone. We found that the association between TRS and HLA-matching status was much stronger when the cell belonged to a clone ( $\beta=0.141$ , p-value= $4.71 \times 10^{-5}$ ) compared to unique cells ( $\beta=0.0415$ , p-value=0.256). Furthermore, TRS was significantly correlated with clone size (p-value=1.62  $\times 10^{-7}$ , Spearman  $\rho$  test).

We then repeated this analysis for all cells from a subset of 56 individuals after first cycle of treatment. This time, the association between TRS and HLA matching among all CD8<sup>+</sup> cells was even stronger ( $\beta=0.132$ , p-value= $3.77 \times 10^{-8}$ ). Amongst the cell subsets, the strongest association was with CD8 TEMRA ( $\beta=0.225$ , p-value= $8.61 \times 10^{-7}$ ) but CD8 TEM cells also had a significant association ( $\beta=0.109$ , p-value=0.0142) as did naïve cells ( $\beta=0.260$ , p-value=0.000467). We also repeated the analysis separating the cells into whether they belonged to a clone or whether they were unique cells and this time unique cells had a stronger association ( $\beta=0.153$ , p-value= $2.16 \times 10^{-5}$ ) vs cells belonging to a clone ( $\beta=0.120$ , p-value=0.000188). TRS was also significantly correlated with clone size but less strongly than in pre-treatment samples (p-value= $8.58 \times 10^{-5}$ , Spearman  $\rho$  test). The fact that the association between tumour reactivity and HLA-Vgene is strongest in CD8 TEM cells before ICB treatment and strongest in CD8 TEMRA cells after treatment is noteworthy and could possibly be because TEMRA cells are terminally differentiated and carry higher levels of cytotoxic and exhausted genes compared to TEM cells [31]. The fact that naïve cells also had that association after ICB treatment suggests that these may be the clones that have been selected for by tumour antigen. This is further supported by the fact that the correlation between TRS and clone size is weaker in the post-treatment sample. Furthermore, the fact that both single cells as well as cells belonging to clones were significantly associated with HLA matching only after treatment suggests that these cells may have been selected for by the ICB treatment.

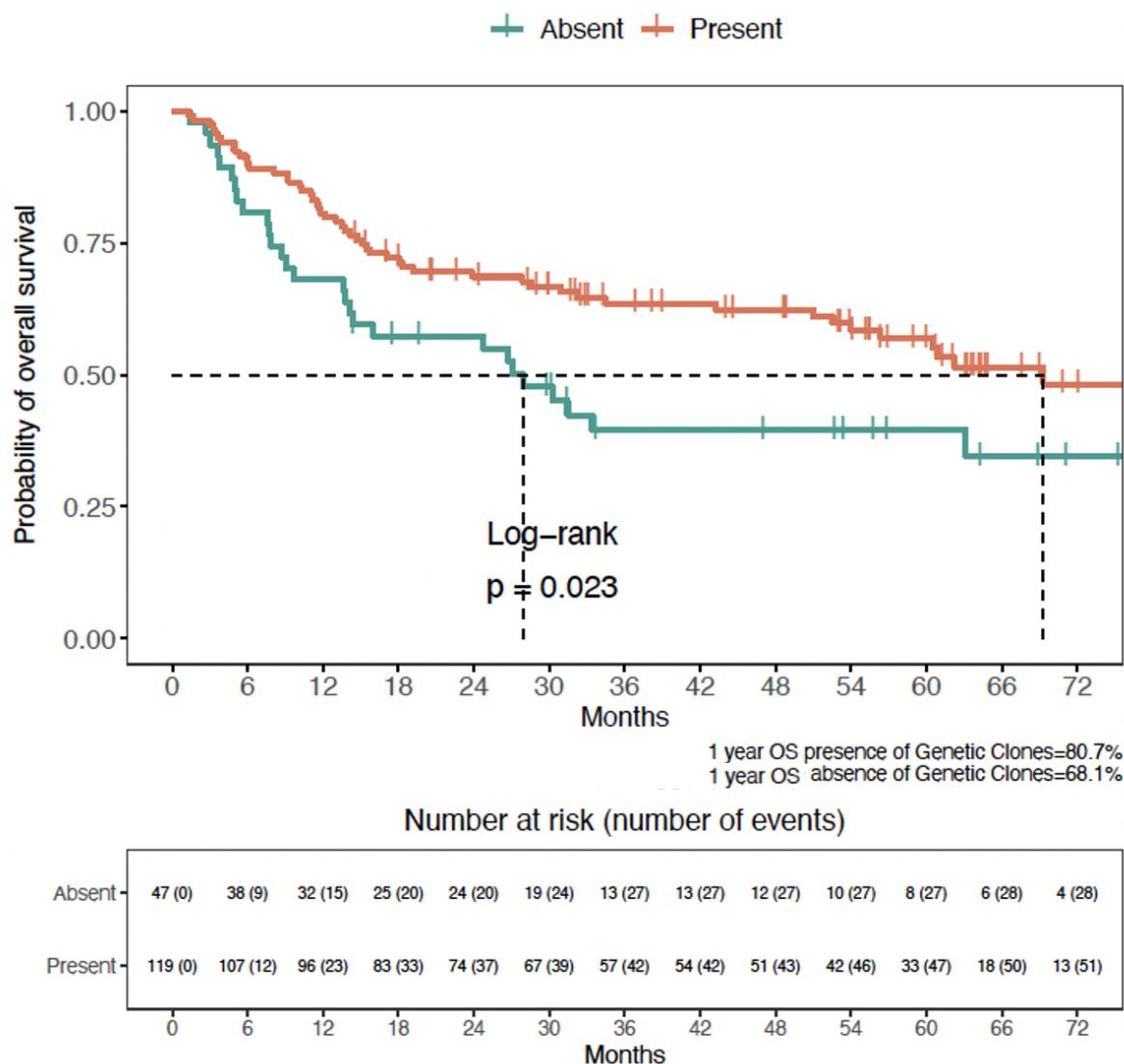
### **Carriage of HLA-matched clones is associated with improved overall survival**

Given that in the single-cell data HLA-matched clones were observed to demonstrate a higher TRS, we proceeded to examine the relationship between carriage of HLA-matched clones and long-term oncological outcomes. Survival analysis across all patients receiving ICB for metastatic disease (n=231) for which we had pre-treatment samples demonstrated that patients carrying HLA-matched clones prior to treatment had improved overall survival (p-value=0.0399, log-rank test, **Supplementary Figure 6a**).

We similarly examined samples for which we had both pre and post-treatment data, finding that 130/179 patients had HLA-matched clones on at least one of these two timepoints. In keeping with ICB therapy primarily acting on pre-existent clones, whilst the proportion of HLA-matched clones frequently changed, no patients were found to develop HLA-matched clones post-treatment who did not already have pre-treatment matched clones. Moreover, for 127/130 (97.6%) patients one or more HLA-matched clones were present across both timepoints. Notably, analysis of these patients again showed that carriage of HLA matched clones at anytime point was associated with improved OS (P=0.0229) (**Figure 6**), with a similar effect observed when analysis was confined to those with metastatic melanoma (P=0.0553, **Supplementary Figure 6b**).



**Figure 5.** Association between V-gene usage and HLA matching status. In each subplot, tumour reactivity score is plotted against HLA matching status (0-unmatched, 1-matched) with each dot representing one cell. Each subplot represents a CD8<sup>+</sup> cell subset, showing that the largest difference occurs within CD8 T Effector Memory (TEM) cells. The subplots are annotated with anova P values comparing the model with HLA matching status with the null model without HLA matching status.



**Figure 6. Relationship between overall survival and presence of HLA-matched clones.** Kaplan Meier plot denoting that overall survival is better in patients who have a HLA-matched clone either before or after first cycle of ICB treatment

## DISCUSSION

We conducted a genome-wide and MHC-focused association study with TCR repertoire in cancer patients undergoing ICB, revealing significant associations in the *cis* regions of V-genes as well as *trans* associations within the MHC region. Many of the associated V-genes and HLA alleles have been shown to be associated with cancer prognosis and treatment response in other studies.

This suggests that germline genetics may affect clinical outcomes through the TCR repertoire. There are several mechanisms by which germline variation could influence TCR repertoire. Variants *cis* to the TRAV and TRBV genes could affect V(D)J recombination probabilities or could be located in regulatory elements affecting gene transcription. The *trans* effects of HLA alleles on TCR could occur via selection either centrally or peripherally. The central hypothesis suggests that HLA risk alleles may influence thymic selection to decrease the frequency of TCRs reactive to autoantigens [32] while the peripheral hypothesis suggests that this takes place in the periphery when the TCR encounters antigen-presenting cells [33]. Since our association is stronger in CD8 TEMRA cells compared to CD8 Naive cells, this would favour the peripheral hypothesis.

Our dataset is unique in that we have pre-treatment as well as post-treatment samples in the same individuals. As such, we are able to evaluate the effect of ICB on HLA V-gene selected cells. Our results demonstrate that HLA-matched cells are more tumour reactive compared to HLA-unmatched cells and this difference is even more significant after ICB treatment. This suggests that ICB enhances the effect of these cells against tumour antigens. Furthermore, the presence of HLA-matched cells confers a survival advantage, suggesting that one of the ways through which HLA influences cancer survival outcomes is through its selection effects on the TCR.

However, one puzzling aspect of our results are that TCR which are HLA selected are less long-lived than their non-HLA selected counterparts and less likely to persist through ICB treatment despite being more tumour reactive after ICB treatment. This could suggest that TCR persistence through time and ICB treatment could be driven by other selection processes, for example tumour antigens. During tumorigenesis, numerous genetic abnormalities accumulate in cells, producing mutated peptides, some of which can activate T cells if successfully presented by the MHC. While the HLA matched cells are more tumour reactive, the tumour antigens which escape their detection may stimulate the expansion of other TCRs which are not HLA-matched, providing a mechanism through which the tumour escapes HLA surveillance. This is supported by the fact that before ICB treatment, the association between HLA-matching and TRS is stronger in cells belonging to a clone, but after ICB treatment, this association is stronger in unique cells. This suggests that these unique cells may have been HLA selected by the tumour antigen after ICB treatment. Our study is limited in that we do not have sequencing data from the tumour that could corroborate this hypothesis.

Despite this limitation, this is still a highly significant observation and has crucial implications in the field of immune oncology as it implies that the patient's germline genetics, especially the HLA type, should be taken into account when designing T cell-mediated therapy such as chimeric T cell immunotherapy.

In conclusion, this is the first study of the germline genetic association of TCR repertoire within cancer

patients, showing significant associations at the MHC region with genes that are cancer relevant. These associations impact the survival outcomes of patients and are altered by ICB treatment.

## **Acknowledgements**

We are very grateful to all patients who contributed samples and participated in the study. We thank all the staff of the Day Treatment Unit, Oxford Cancer Centre, and The Brodley Centre at the Horton General Hospital. We are also grateful to all the staff of the Oxford University Hospitals NHS Foundation Trust - particularly Dr Miranda Payne, Dr Nick Coupe and Dr Rubeta Matin in the cancer centre who aid with patient recruitment, as well as the staff of the Oxford Radcliffe Biobank and Churchill Hospital Sample Handling Lab.

## **Funding**

This work was funded by a Wellcome Career Development Award to BPF (no. 201488/Z/ 16/Z). BPF is supported by the NIHR Oxford Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. YL is supported by a Kennedy Trust KTRR Senior Research Fellowship (KENN202109).

## **Declaration of interests**

ESN, OT, CT, RW, BS, GM, SM, JG, ML, YL – no competing interests. BPF – received conference support from BMS and performed consultancy for UCB.

# **Methods**

## **Participants**

Patients were recruited from Oxford University Hospital when they were referred to receive ICB as therapy for melanoma, renal cell carcinoma and colorectal cancer. All patients provided written informed consent to donate samples to the Oxford Radcliffe Biobank (ORB) (Oxford Centre for Histopathology Research ethical approval reference 19/SC/0173, project nos. 16/A019, 18/A064, 19/A114) and allow access to their clinical data. Patients received either combined ICB (ipilimumab plus nivolumab 3 weekly for  $\leq 4$  treatment cycles, followed by maintenance nivolumab) or single ICB consisting of either nivolumab monthly, pembrolizumab three-weekly or pembrolizumab six-weekly. Patient demographic and clinical characteristics were collected from the electronic health record system.

## **Sample collection**

30-50 mL blood was collected into EDTA tubes (BD vacutainer system) just before administration of the first cycle of ICB. The blood was centrifuged to obtain peripheral blood mononuclear cells (PBMCs) and plasma by density centrifugation (Ficoll Paque). CD8<sup>+</sup> T cells were isolated by CD8 positive selection using magnetic separation (Miltenyi).

## **RNA extraction**

Cells were resuspended in 350  $\mu$ L of RLTplus buffer supplemented with beta-mercaptoethanol or Dithio-threitol. Qiashredder (Qiagen) was used to homogenise the sample, and Allprep DNA/RNA/miRNA kit (Qiagen) was then used for DNA/RNA extraction. RNA was then eluted into 34  $\mu$ L of RNase-free water with concentration quantified by Qubit and DNA was eluted into 54  $\mu$ L elution buffer. Both RNA and DNA samples were subsequently stored at -80°C until sequencing and genotyping.

## **Bulk RNA sequencing**

RNA was thawed on ice prior to mRNA isolation using Poly(A) mRNA Magnetic Isolation Module kits (NEBNext). Up to 600ng of RNA was then used to generate dsDNA libraries using NEBNext Ultra II Directional RNA Library Prep Kits as previously described [34]. Samples were then sequenced on either an Illumina HiSeq4000 (75bp paired-end) or a NovaSeq6000 (150bp paired-end). We performed TCR analysis using the MiXCR package [35] with settings as previously described [34, 2].

## **Single-cell RNA and TCR sequencing**

Single-cell RNA sequencing data was acquired from two different experiments. The first utilized fresh peripheral blood mononuclear cells (PBMCs) and has been previously published [2]. The second – obtained from cryopreserved PBMCs – was processed using two separate protocols. In the first protocol, dead cells were first incubated with live:dead magnetic beads (Miltenyi dead cell removal kit) and run over a MACS LD column to remove dead cells. 60,000 live PBMCs were then loaded into the partitioning reaction and processed using library kits (10X Genomics, Pleasanton, CA) following manufacturer

protocols. In the second protocol, CD8<sup>+</sup> T cells were first isolated by incubating with live:dead and CD8<sup>+</sup> T cell negative selection beads and run over a magnetic column. 60,000 CD8<sup>+</sup> T cells were taken from the flow-through and processed as above.

Sequencing of single-cell libraries were performed on an Illumina NovaSeq 6000 S4 flow cell (150 base pairs paired end). For each pool there were five library samples - PBMC gene expression (GEX) libraries, CD8<sup>+</sup> T cell GEX libraries, PBMC origin TCR libraries, PBMC origin B cell repertoire libraries, CD8<sup>+</sup> T cell origin TCR libraries. Batch one of sequencing was the first four pools which were sequenced across one lane of an S4 flow cell. This was deliberately under-sequenced in order to establish sequencing requirements across the rest of the samples. Following this sequencing run, there were a median of 18,000 reads/cell for GEX libraries (target 27,500) and 4,800 reads per cell for V(D)J libraries (target 7,000). We therefore sequenced the remaining seven pools plus additional sequencing of libraries from the first four pools that were under-sequenced, across three further lanes of an S4 flow cell. Across both runs we achieved a median of 28,500 reads/cell for the GEX libraries, and 7100 reads/cell for the TCR libraries.

Single-cell sequencing data were aligned against the GRCh38 human reference genome using Cellranger (v6.0.1) for GEX libraries and Cellranger VDJ for V(D)J libraries. For samples which had been sequenced across multiple lanes, the FASTQ files were inputted for alignment on the same Cellranger run. Empty-Drops [36] was used to identify empty droplets (false discovery rate (FDR) 0.01). Barcodes were excluded only if they were called as empty by both Cellranger and EmptyDrops . CellsNP-lite [37] was used to reconstruct genotype data from reads and then Vireo was used to compare this to patient-level genotype in order to de-multiplex samples from pools. Doublet identification was performed during sub-clustering by identification of mixed transcriptomes of canonical markers. Cells with <300 transcripts and >20% of mitochondrial-encoded genes were removed.

### Single-cell sequencing integration and annotation

scVI [38] was used for the integration of the two single cell experiments. We first selected 4000 highly variable genes, from the whole PBMC dataset in Scanpy [39]. An scVI model was then trained using the pool label as the batch variable and the following model parameters; number of latent variables = 30, and batch size = 1024. Once the core model was trained, datasets containing enriched cell types or higher mitochondrial proportions (10- 20%) were further referenced mapped using scArches [40]. Next, clustering was implemented on a nearest neighbour graph using the 30 latent dimensions that were obtained from the scVI and scArches output. Here, the number of neighbours was set to k=30 and distance metric set to ‘cosine’. We then performed coarse leiden clustering on the graph with resolution r=0.03. For each of the resulting level 1 clusters, we calculated a new neighbour graph using scVI’s 30 latent dimensions, with the number of neighbours again set to k=30. Based on the new neighbour graph, each cluster was clustered into smaller ‘level 2’ clusters with leiden clustering at resolution r=0.3. Level 2 clusters, were then annotated based on differentially expressed genes into broad cell type categories: “CD8NK”, “CD4”, “Bcells” and “Myeloid Platelets”.

To fully capture cell type heterogeneity, we remodelled each broad ‘level 2’ subset. Per ‘level 2’ cell type, we calculated 4000 highly variable genes, re-trained scVI models using the same parameters and calculated new neighbour graphs. To guide clustering, we first performed automated cell-type annotations

on scVI embeddings with Celltypist [41]. Cluster-specific marker genes were identified by performing differential expression analysis in scanpy on a given cell type compared to the rest of the cells from the same level annotation. Where clusters were highly similar, we merged them based on hierarchical dendograms generated from gene expressions. For final CD8+ T cell annotations, cross-validation was performed based on assigned annotations from the published dataset [2].

### Quality control (QC) and HLA imputation

Genotyping was performed on the Illumina Global Screening Array 24 v3 (Illumina). For our sample QC, we used the following thresholds - heterozygosity outside 3.5 standard deviations, missing > 10% of SNPs, identity-by-descent pihat of > 0.25 and being ancestry outlier (**Supplementary Figure 4**). A total of 250 individuals were included including 231 patients with melanoma, 17 with renal cell cancer and 2 with colorectal cancer (**Supplementary Table 1**). For SNP QC, we used the following thresholds - MAF of 0.01, Hardy Weinberg equilibrium 0.000001 and present in at least 90% of individuals. There were a total of 486,469 SNPs tested for the GWAS. For HLA imputation, we used the Michigan Imputation Server version 2 multi-ethnic panel resulting in a total of 2212 amino acids, 265 classical HLA alleles and 17,952 SNPs imputed. The same QC thresholds were applied to the HLA imputed data.

### Association between TCR and genetic germline variation

Bulk TCR sequencing data was processed using MixCR [35] and features were extracted, namely V-gene usage, clone counts and CDR3 amino acid sequences. Only productive clones were included. To obtain the V-gene usage phenotype, the number of unique clones with that V-gene were counted and then normalisation was performed by Trimmed Mean of M-values (TMM) function in edgeR (version 3.18) [42] followed by batch correction using removeBatchEffect function. After that, inverse normal rank transformation (INRT) was performed for each phenotype across the samples. A linear regression model was fitted in Plink (version 2.0) [43] to test each V-gene phenotype against genome wide SNPs, HLA classical alleles, amino acids/SNPs in the HLA region. Covariates were 2 genetic principal components (PCs), 2 TCR PCs, age, gender and cancer type.

$$V\text{-gene usage} \sim \text{genetic variant} + 2 \text{ genetic PCs} + 2 \text{ TCR PCs} + \text{age} + \text{gender} + \text{cancer type}$$

### Permutation analysis

To determine an appropriate p-value threshold for significance, we permuted it 1000 times scrambling the phenotype while keeping the same V-gene within each individual. The 5% percentile of p value was then chosen as the significance threshold. In addition to testing each HLA allele separately, we also conducted an omnibus test for each amino acid position where there are more than 1 amino acid present in the samples to investigate whether there are particular amino acid positions which are more associated to Vgenes more strongly than classical HLA alleles.

### K-mer analysis

To analyse the CDR3 amino acid sequences, K-mers of length 7 were constructed using a sliding window approach across the sequence. We chose to use 7 amino acids because the Immune-receptor-Tyrosine-

based-Activation-Motif (ITAM) that is critical for the initiation of signaling following ligand engagement has a sequence of 6-8 amino acids in its centre [44].

This approach has been used successfully in other studies on TCRs [45][46]. Only sequences between length 12 and 18 were included for this analysis. The K-mers were also batch corrected and then TMM normalised followed by INRT normalised by the same procedure as the V-gene phenotype. The same linear model was fitted against HLA alleles as with the V-gene phenotype.

### **Comparing pre- and post- treatment clones**

Focusing on the 179 individuals for which paired samples (before and after first cycle of ICB) were available, we sub-divided clones into three groups: those only observed pre-treatment (*Unstable*), those observed post-treatment only (*Novel*), and those present across both samples (*Persistent*).

Using V-genes significantly associated with variants in the MHC region, we extracted the top classical HLA allele associated with each V-gene. HLA-matched clones were defined as clones which had this classical HLA-V-gene pairing. HLA-unmatched cells were defined as all other clones. We then calculated the proportion of HLA-matched clones for each individual within each clone group (Unstable, Persistent and Novel).

To investigate whether ICB alters the proportion of HLA-matched clones, we performed a paired Wilcoxon test on proportions of HLA-matched clones between unstable and novel groups.

For each of these 3 groups, we performed a V-gene vs HLA association analysis as described above for the significantly associated V-gene HLA pair.

### **Single-cell TCR sequencing analysis**

There were a total of 59 individuals for which we had this data as well as genotyping. Cells were included only if they had either 1  $\alpha$  and 1  $\beta$  chain or 2  $\alpha$  and 1  $\beta$  chain. Clones were collapsed to include only counts of unique clones and then normalised by total number of clones for the cell type being studied using the same method as bulk data. A variable representing experimental protocol was constructed. Because the experimental protocol was highly correlated with the TCR PCs, experimental protocol was regressed from the first 2 PCs and the residuals extracted. Then a linear model was fitted between V-gene usage and HLA alleles with age, sex, cancer type, 2 genetic PCs, the residuals of 2 TCR PCs and experimental protocol being the covariates. This was done individually for each cell type.

Single-cell gene expression sequencing was processed using Scanpy V1.10 [39] for the purpose of calculating TRS. Cells from each experimental protocol were normalised, log transformed and then scaled. To calculate tumour reactivity score for each cell, we took the sum of the 20 genes [30] and transformed it using INRT. We then divided the cells into 2 groups, those that were HLA selected and those that were not. Using V-genes significantly associated with variants in the MHC region, we extracted the top classical HLA allele associated with each gene. HLA-matched cells were defined as cells which had this classical HLA-V-gene pairing. HLA-matched cells were defined as all other cells. To determine whether TRS was higher or lower in HLA matched cells, we performed regression of TRS score against HLA-matching status with experimental protocol as random effect and did an anova test with the null model excluding

HLA-matching status. Further to that, we did this analysis separately in unique cells and cells belonging to clones (cells having the same TRA and TRB CDR3).

*H1 - TRS ~ HLA-matching status + experimental protocol*

*H0- TRS ~ experimental protocol*

## References

- [1] A. C. Huang, M. A. Postow, R. J. Orlowski, R. Mick, B. Bengsch, S. Manne, W. Xu, S. Harmon, J. R. Giles, B. Wenz, *et al.*, “T-cell invigoration to tumour burden ratio associated with anti-pd-1 response,” *Nature*, vol. 545, no. 7652, pp. 60–65, 2017.
- [2] R. A. Watson, O. Tong, R. Cooper, C. A. Taylor, P. K. Sharma, A. V. de Los Aires, E. A. Mahé, H. Ruffieux, I. Nassiri, M. R. Middleton, *et al.*, “Immune checkpoint blockade sensitivity and progression-free survival associates with baseline cd8+ t cell clone size and cytotoxicity,” *Science immunology*, vol. 6, no. 64, p. eabj8825, 2021.
- [3] L. Robert, J. Tsoi, X. Wang, R. Emerson, B. Homet, T. Chodon, S. Mok, R. R. Huang, A. J. Cochran, B. Comin-Anduix, R. C. Koya, T. G. Graeber, H. Robins, and A. Ribas, “Ctla4 blockade broadens the peripheral t-cell receptor repertoire,” *Clin Cancer Res*, vol. 20, no. 9, pp. 2424–32, 2014.
- [4] A. Knight, L. Karapetyan, and J. M. Kirkwood, “Immunotherapy in melanoma: Recent advances and future directions,” *Cancers (Basel)*, vol. 15, Feb 2023.
- [5] J. Larkin, V. Chiarioti-Silenti, R. Gonzalez, J.-J. Grob, P. Rutkowski, C. D. Lao, C. L. Cowey, D. Schadendorf, J. Wagstaff, R. Dummer, *et al.*, “Five-year survival with combined nivolumab and ipilimumab in advanced melanoma,” *New England journal of medicine*, vol. 381, no. 16, pp. 1535–1546, 2019.
- [6] M. M. Davis and P. J. Bjorkman, “T-cell antigen receptor genes and t-cell recognition,” *Nature*, vol. 334, no. 6181, pp. 395–402, 1988.
- [7] M. Krosgaard and M. M. Davis, “How t cells’ see’antigen,” *Nature immunology*, vol. 6, no. 3, pp. 239–245, 2005.
- [8] G. Lythe, R. E. Callard, R. L. Hoare, and C. Molina-París, “How many tcr clonotypes does a body maintain?,” *J Theor Biol*, vol. 389, pp. 214–24, Jan 2016.
- [9] T. Dupic, Q. Marcou, A. M. Walczak, and T. Mora, “Genesis of the  $\alpha\beta$  t-cell receptor,” *PLoS computational biology*, vol. 15, no. 3, p. e1006874, 2019.
- [10] J. Lu, F. Van Laethem, A. Bhattacharya, M. Craveiro, I. Saba, J. Chu, N. C. Love, A. Tikhonova, S. Radaev, X. Sun, *et al.*, “Molecular constraints on cdr3 for thymic selection of mhc-restricted tcrs from a random pre-selection repertoire,” *Nature communications*, vol. 10, no. 1, p. 1019, 2019.
- [11] A. Digre and C. Lindskog, “The human protein atlas-spatial localization of the human proteome in health and disease,” *Protein Sci*, vol. 30, no. 1, pp. 218–233, 2021.
- [12] Z. Wang, Y. Zhong, Z. Zhang, K. Zhou, Z. Huang, H. Yu, L. Liu, S. Liu, H. Yang, J. Zhou, J. Fan, L. Wu, and Y. Sun, “Characteristics and clinical significance of t-cell receptor repertoire in hepatocellular carcinoma,” *Front Immunol*, vol. 13, p. 847263, 2022.

- [13] Y. Qiu, L. Cui, Y. Lin, B. Gao, J. Li, X. Zhao, X. Zhu, S. Hu, and L. Lin, “Development and validation of a robust immune prognostic signature for head and neck squamous cell carcinoma,” *Front Oncol*, vol. 10, p. 1502, 2020.
- [14] Y. Guo, J. Dong, T. Ji, X. Li, S. Rong, and H. Guan, “A risk score for the prognosis prediction of the muscle-invasive bladder cancer patients who received gemcitabine plus cisplatin chemotherapy,” *Aging (Albany NY)*, vol. 14, no. 23, pp. 9715–9729, 2022.
- [15] M. L. Russell, A. Souquette, D. M. Levine, S. A. Schattgen, E. K. Allen, G. Kuan, N. Simon, A. Balmaseda, A. Gordon, P. G. Thomas, F. A. Matsen, 4th, and P. Bradley, “Combining genotypes and t cell receptor distributions to infer genetic loci determining v(d)j recombination probabilities,” *Elife*, vol. 11, 2022.
- [16] C. J. Ryu, B. B. Haines, D. D. Draganov, Y. H. Kang, C. E. Whitehurst, T. Schmidt, H. J. Hong, and J. Chen, “The t cell receptor  $\beta$  enhancer promotes access and pairing of d $\beta$  and j $\beta$  gene segments during v (d) j recombination,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 23, pp. 13465–13470, 2003.
- [17] S. Simon, Z. Wu, J. Cruard, V. Vignard, A. Fortun, A. Khammari, B. Dreno, F. Lang, S. J. Rulli, and N. Labarriere, “Tcr analyses of two vast and shared melanoma antigen-specific t cell repertoires: Common and specific features,” *Front Immunol*, vol. 9, p. 1962, 2018.
- [18] B. C. Healy, M. Liguori, D. Tran, T. Chitnis, B. Glanz, C. Wolfish, S. Gauthier, G. Buckle, M. Houtchens, L. Stazzone, S. Khoury, R. Hartzmann, M. Fernandez-Vina, D. A. Hafler, H. L. Weiner, C. R. G. Guttmann, and P. L. De Jager, “Hla b\*44: protective effects in ms susceptibility and mri outcome measures,” *Neurology*, vol. 75, no. 7, pp. 634–40, 2010.
- [19] N. Isobe, A. Keshavan, P.-A. Gourraud, A. H. Zhu, E. Datta, R. Schlaeger, S. J. Caillier, A. Santaniello, A. Lizée, D. S. Himmelstein, S. E. Baranzini, J. Hollenbach, B. A. C. Cree, S. L. Hauser, J. R. Oksenberg, and R. G. Henry, “Association of hla genetic risk burden with disease phenotypes in multiple sclerosis,” *JAMA Neurol*, vol. 73, pp. 795–802, Jul 2016.
- [20] I. M. S. G. Consortium, “Class ii hla interactions modulate genetic risk for multiple sclerosis,” *Nature genetics*, vol. 47, no. 10, pp. 1107–1113, 2015.
- [21] D. Chowell, L. G. T. Morris, C. M. Grigg, J. K. Weber, R. M. Samstein, V. Makarov, F. Kuo, S. M. Kendall, D. Requena, N. Riaz, B. Greenbaum, J. Carroll, E. Garon, D. M. Hyman, A. Zehir, D. Solit, M. Berger, R. Zhou, N. A. Rizvi, and T. A. Chan, “Patient hla class i genotype influences cancer response to checkpoint blockade immunotherapy,” *Science*, vol. 359, no. 6375, pp. 582–587, 2018.
- [22] A. Chhibber, L. Huang, H. Zhang, J. Xu, R. Cristescu, X. Liu, D. V. Mehrotra, J. Shen, P. M. Shaw, M. D. Hellmann, *et al.*, “Germline hla landscape does not predict efficacy of pembrolizumab monotherapy across solid tumor types,” *Immunity*, vol. 55, no. 1, pp. 56–64, 2022.
- [23] M. M. Davis and P. J. Bjorkman, “T-cell antigen receptor genes and t-cell recognition,” *Nature*, vol. 334, no. 6181, pp. 395–402, 1988.

- [24] M. Lotem, L. Kadouri, S. Merims, I. Ospovat, A. Nissan, I. Ron, S. Frankenburg, A. Machlenkin, S. Israel, H. Steinberg, T. Hamburger, and T. Peretz, “Hla-b35 correlates with a favorable outcome following adjuvant administration of an hla-matched allogeneic melanoma vaccine,” *Tissue Antigens*, vol. 78, pp. 203–7, Sep 2011.
- [25] G. Chitadze, M. Lettau, J. Bhat, D. Wesch, A. Steinle, D. Fürst, J. Mytilineos, H. Kalthoff, O. Janssen, H.-H. Oberg, *et al.*, “Shedding of endogenous mhc class i-related chain molecules a and b from different human tumor entities: heterogeneous involvement of the “a disintegrin and metalloproteases” 10 and 17,” *International journal of cancer*, vol. 133, no. 7, pp. 1557–1566, 2013.
- [26] S. Wang, G. C. Onyeaghala, N. Pankratz, H. H. Nelson, B. Thyagarajan, W. Tang, F. L. Norby, C. Ugoji, C. E. Joshu, C. R. Gomez, *et al.*, “Associations between mica and micb genetic variants, protein levels, and colorectal cancer: Atherosclerosis risk in communities (aric),” *Cancer Epidemiology, Biomarkers & Prevention*, vol. 32, no. 6, pp. 784–794, 2023.
- [27] X. Wen, S. Chen, J. Li, Y. Li, L. Li, Z. Wu, H. Yuan, X. Tian, F. Zhang, and Y. Li, “Association between genetic variants in the human leukocyte antigen-b/mica and takayasu arteritis in chinese han population,” *International journal of rheumatic diseases*, vol. 21, no. 1, pp. 271–277, 2018.
- [28] V. Groh, S. Bahram, S. Bauer, A. Herman, M. Beauchamp, and T. Spies, “Cell stress-regulated human major histocompatibility complex class i gene expressed in gastrointestinal epithelium.” *Proceedings of the National Academy of Sciences*, vol. 93, no. 22, pp. 12445–12450, 1996.
- [29] S. Bauer, V. Groh, J. Wu, A. Steinle, J. H. Phillips, L. L. Lanier, and T. Spies, “Activation of nk cells and t cells by nkg2d, a receptor for stress-inducible mica,” *science*, vol. 285, no. 5428, pp. 727–729, 1999.
- [30] M. Yan, J. Hu, Y. Ping, L. Xu, G. Liao, Z. Jiang, B. Pang, S. Sun, Y. Zhang, Y. Xiao, and X. Li, “Single-cell transcriptomic analysis reveals a tumor-reactive t cell signature associated with clinical outcome and immunotherapy response in melanoma,” *Front Immunol*, vol. 12, p. 758288, 2021.
- [31] T. Willinger, T. Freeman, H. Hasegawa, A. J. McMichael, and M. F. C. Callan, “Molecular signatures distinguish human central memory from effector memory cd8 t cell subsets,” *J Immunol*, vol. 175, pp. 5895–903, Nov 2005.
- [32] M. Simmonds and S. Gough, “The hla region and autoimmune disease: associations and mechanisms of action,” *Current genomics*, vol. 8, no. 7, pp. 453–465, 2007.
- [33] J. Gebe, E. Swanson, and W. W. Kwok, “Hla class ii peptide-binding and autoimmunity,” *Tissue antigens*, vol. 59, no. 2, pp. 78–87, 2002.
- [34] B. P. Fairfax, C. A. Taylor, R. A. Watson, I. Nassiri, S. Danielli, H. Fang, E. A. Mahé, R. Cooper, V. Woodcock, Z. Traill, *et al.*, “Peripheral cd8+ t cell characteristics associated with durable responses to immune checkpoint blockade in patients with metastatic melanoma,” *Nature medicine*, vol. 26, no. 2, pp. 193–199, 2020.

- [35] D. A. Bolotin, S. Poslavsky, I. Mitrophanov, M. Shugay, I. Z. Mamedov, E. V. Putintseva, and D. M. Chudakov, “Mixcr: software for comprehensive adaptive immunity profiling,” *Nat Methods*, vol. 12, no. 5, pp. 380–1, 2015.
- [36] A. T. Lun, S. Riesenfeld, T. Andrews, T. P. Dao, T. Gomes, P. in the 1st Human Cell Atlas Jamboree, and J. C. Marioni, “Emptydrops: distinguishing cells from empty droplets in droplet-based single-cell rna sequencing data,” *Genome biology*, vol. 20, pp. 1–9, 2019.
- [37] X. Huang and Y. Huang, “Cellsnp-lite: an efficient tool for genotyping single cells,” *Bioinformatics*, vol. 37, no. 23, pp. 4569–4571, 2021.
- [38] A. Gayoso, R. Lopez, G. Xing, P. Boyeau, V. Valiollah Pour Amiri, J. Hong, K. Wu, M. Jayasuriya, E. Mehlman, M. Langevin, *et al.*, “A python library for probabilistic analysis of single-cell omics data,” *Nature biotechnology*, vol. 40, no. 2, pp. 163–166, 2022.
- [39] F. A. Wolf, P. Angerer, and F. J. Theis, “Scanpy: large-scale single-cell gene expression data analysis,” *Genome biology*, vol. 19, pp. 1–5, 2018.
- [40] M. Lotfollahi, M. Naghipourfar, M. D. Luecken, M. Khajavi, M. Büttner, M. Wagenstetter, Ž. Avsec, A. Gayoso, N. Yosef, M. Interlandi, *et al.*, “Mapping single-cell data to reference atlases by transfer learning,” *Nature biotechnology*, vol. 40, no. 1, pp. 121–130, 2022.
- [41] C. Domínguez Conde, C. Xu, L. Jarvis, D. Rainbow, S. Wells, T. Gomes, S. Howlett, O. Suchanek, K. Polanski, H. King, *et al.*, “Cross-tissue immune cell analysis reveals tissue-specific features in humans,” *Science*, vol. 376, no. 6594, p. eab15197, 2022.
- [42] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, “edger: a bioconductor package for differential expression analysis of digital gene expression data,” *bioinformatics*, vol. 26, no. 1, pp. 139–140, 2010.
- [43] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, M. J. Daly, *et al.*, “Plink: a tool set for whole-genome association and population-based linkage analyses,” *The American journal of human genetics*, vol. 81, no. 3, pp. 559–575, 2007.
- [44] P. E. Love and S. M. Hayes, “Itam-mediated signaling by the t-cell antigen receptor,” *Cold Spring Harbor perspectives in biology*, vol. 2, no. 6, p. a002485, 2010.
- [45] M. Mark, S. Reich-Zeliger, E. Greenstein, D. Reshef, A. Madi, B. Chain, and N. Friedman, “A hierarchy of selection pressures determines the organization of the t cell receptor repertoire,” *Frontiers in immunology*, vol. 13, p. 939394, 2022.
- [46] Y. Katayama and T. J. Kobayashi, “Comparative study of repertoire classification methods reveals data efficiency of k-mer feature extraction,” *Frontiers in Immunology*, vol. 13, p. 797640, 2022.