T cell receptor convergence is an indicator of antigen-specific T cell response in cancer immunotherapies

Ming Yao Pan and Bo Li

Lyda Hill Department of Bioinformatics, University of Texas Southwestern Medical Center,
Dallas, TX 75390, USA

Correspondence: Bo.Li@UTSouthwestern.edu

Abstract

T cells are potent at eliminating pathogens and playing a crucial role in the adaptive immune response. T cell receptor (TCR) convergence describes T cells that share identical TCRs with the same amino acid sequences but have different DNA sequences due to codon degeneracy. We conducted a systematic investigation of TCR convergence using single-cell immune profiling and bulk TCRβ-sequence (TCR-seq) data obtained from both mouse and human samples, and uncovered a strong link between antigen-specificity and convergence. This association was stronger than T cell expansion, a putative indicator of antigen-specific T cells. By using flow sorted tetramer$^+$ single T cell data, we discovered that convergent T cells were enriched for a neoantigen-specific CD8$^+$ effector phenotype in the tumor microenvironment. Moreover, TCR convergence demonstrated better prediction accuracy for immunotherapy response than the existing TCR repertoire indexes. In conclusion, convergent T cells are likely to be antigen-specific and might be a novel prognostic biomarker for anti-cancer immunotherapy.

Keywords: T cell receptor; antigen-specificity; TCR convergence; immunotherapy; TCR sequencing

Introduction

T lymphocytes, or T cells, are one of the most important components of the adaptive immune system. T cell receptors (TCRs) are protein complexes found on the surface of T cells that can specifically recognize antigens (Marrack & Kappler, 1987; Davis & Bjorkman, 1988). Through the combinational somatic rearrangement of multiple variable (V), diversity (D) (only for β chains), joining (J) and constant (C) gene segments, theoretically, the diversity of unique TCR α and β chains pairs can reach to 2×10^{19} (Pai & Satpathy, 2021). Having such a wide variety of TCRs allows the recognition of numerous endogenous and exogenous antigens. However, due to codon degeneracy, which means multiple codons can encode the same amino acid (Trainor et al., 1984), different V(D)J rearrangements can end up encoding the same TCR proteins and this phenomenon is called TCR convergence (Looney et al., 2019). TCR convergence can be observed universally in almost every individual, but very few studies have been conducted to explain its biological significance.

While little is known about the functional impact of TCR convergence, codon degeneracy has been widely studied (Gonzalez et al., 2019; McClellan, 2000). The fact that the number of three-nucleotides codons (64) exceeds the number of encoding amino acids (20) provides the basis of codon degeneracy. The distribution of degeneracy among the 20 encoding amino acids is uneven. Arginine, leucine, and serine are among the amino acids with six corresponding codons, while others have one to four codons (Crick et al., 1961), and the origin of codon degeneracy remains the subject of debate. Some believe that codon degeneracy is the result of co-evolution in which codon assignment occurred by organisms inheriting parts of the codon set from precursor amino acids (Wong, 1975). Some have tried to explain it through stereochemical interaction, where stereochemical specificity enables codons to selectively bind to assigned amino acids (Copley et al., 2005). There are also other theories pertaining to this topic (Crick et al., 1961; Di Giulio, 2004; Koonin & Novozhilov, 2009). Apart from its elusive origins, the biological significance of codon degeneracy also varies in different scenarios. In the context of T cell immunity, the T cell repertoire is not only results from somatic recombination but also is shaped by multiple selective pressures (Jameson & Bevan, 1998). Therefore, it is possible that TCR convergence is more than the mere consequence of codon degeneracy. It may provide unique insights into the antigen-driven TCR selection process.

T cells can be activated by their cognate antigens through the binding of the TCR to the peptide/MHC complex (Smith-Garvin et al., 2009). Hence, antigen-specific T cells are crucial parts of the T cell 'army' and play dominant roles in eliminating pathogens or tumor cells. There is an increasing number of immunotherapies based on antigen-specific T cells, such as immune checkpoint blockade or neoantigen vaccine therapies (Postow, Callahan, et al., 2015; Peng et al., 2019), which further emphasizes the necessity to study T cell antigen-specificity. However, the enormous diversity of epitopes targeted by T cells and the highly polymorphic nature of MHC make it extremely challenging to identify the specificity of given T cells (Newell & Davis, 2014). Several studies have attempted to predict antigen-specific T cell responses by utilizing TCR characteristics such as diversity, clonality, or evenness scores. While there were cases where these signals were associated with the patient's survival (Nadeem Riaz et al., 2017; Tumeh et al., 2014), in most cases these signals did not correlate with the outcomes (Johnson et al., 2016; Robert et al., 2014; Amaria et al., 2018; Kidman et al., 2020). This could be attributed to the weak correlation between these parameters and the antigen selection process. Here, we speculate that TCR convergence is a better indicator of antigen-driven selection. This is because compared with a non-convergent TCR, a convergent TCR consumes more rearrangement resources, and the principle of parsimony (Stewart, 1993) implies a greater impact on the adaptive immune response for these convergent TCRs.

TCRs with similar CDR3 sequences shared similar binding structures and antigen specificity (Dash et al., 2017). This idea has led to several TCR similarity-based clustering algorithms, such as ALICE (Pogorelyy et al., 2019), TCRdist (Dash et al., 2017), GLIPH2 (Huang et al., 2020), iSMART (Zhang et al., 2020), and GIANA (Zhang et al., 2021), for studying antigen-driven T cell expansion during viral infection or tumorigenesis. While TCR convergence seems like a special case of similarity-based TCR clustering, it is conceptually different. By default, around 90% of the resulting TCR clusters by most previous algorithms are antigen-specific, while the ratio of some, like GLIPH2, is substantially lower (around 35%) (Zhang et al., 2021). Nevertheless, the fact that T cells within each convergent cluster share identical CDR3β amino acid sequence more likely represents shared antigen-specificity compared to those with mismatches (Dash et al., 2017). The exploit of multiple nucleotide combinations for the same antigen-specificity can potentially be a functional redundancy of adaptive immunity and reveal novel TCR selection process.

Over the last few years, high-throughput bulk TCRβ-sequencing (bulk TCR-seq) and single-cell immune profiling have become powerful tools for unraveling a variety of biological problems (Longchao Liu et al., 2022; Zheng et al., 2017). Previously generated datasets are excellent resources for re-analysis. Utilizing three single-cell immune profiling and eleven bulk TCR-seq datasets, we conducted a comprehensive analysis of TCR convergence. We began by investigating the basic features of convergent TCRs, including their degeneracies, publicity, and amino acid usage. Next, we explored the correlation between convergent and antigen-specific TCRs from different perspectives. Significant overlaps were found between convergent T cells and antigen-specific T cells across independent datasets. We proceeded to examine the phenotypes of convergent T cells in both mouse and human samples. As a result, convergent T cells exhibited gene signatures associated with cytotoxicity, memory, and exhaustion, related to antigen-specific T cells. Finally, we found that TCR convergence predicts improved clinical outcomes for cancer patients who received immune checkpoint blockade therapies. Our work presents a new angle to study T cell repertoire and deepens our understanding of antigen-specific T cell selection.

Results

TCR convergence is different from publicity

We defined the convergent TCRs as above described (Figure 1A) and analyzed their basic characteristics, including the degeneracies of the TCRs, sequence lengths, variable gene usage and amino acid components using a bulk TCR dataset containing 666 blood samples from healthy donors (R. O. Emerson et al., 2017). Before analysis, quality control was performed to exclude CDR3 reads that did not meet the following two standards: 1) between 8 and 23 amino acids in length; 2) starting with cysteine and ending with phenylalanine. We randomly selected 50 samples, with a total of 9,567,751 that passing quality control (QC) unique TCR used in the following analysis.

We defined the degeneracy of a TCR protein as the number of distinct clonotypes (defined by DNA sequences) encoding that TCR within one sample. In other words, a degeneracy equal to 1 means non-convergent TCRs, while greater than 1 indicates convergent TCRs. As expected, convergent TCRs only comprised a small proportion (5.40%) of the total T cell population (Figure 1B). As TCR degeneracy increased from two to four, the number of detected TCRs decreased approximately tenfold with each unit increase in degeneracy. Most convergent TCRs had a degeneracy of two (90.65%) or three (7.16%), while larger degeneracies were rare.

We studied the length distribution of the convergent TCR CDR3 regions, which peaked at 11 amino acids (AA) (Figure 1—figure supplement 1A). While TCRs with 14 to 15 AA made up the greatest percentage of all the unique TCRs (Figure 1—figure supplement 1B), the proportion of convergent TCRs continually decreased after reaching 11 AA or longer. This result suggested that shorter TCR sequences appeared to be more conducive to TCR convergence. In terms of the correlation between TCR degeneracy and CDR3 lengths, CDR3s of 9 AA long displayed the greatest degeneracy on average (~2.5) (Figure 1 — figure supplement 1C). From 11 amino acids onward, the longer the sequences, the smaller the degeneracy. After reaching 17 amino acids, the degeneracy remained constant. This is an unexpected observation, as the probability of the occurrence of codon degeneracy should increase with longer sequence lengths. We further analyzed the variable gene usage of convergent TCRs and observed that the usage of every TCRβ chain variable gene (TCRBV) superfamily was relatively even (Figure 1—figure supplement 1D). TCR degeneracy was not significantly impacted by variable gene usage either (Figure 1—figure supplement 1E).

We then calculated the average percentage of a given amino acid in TCRs with degeneracy from one to four. The first and last three AA in the CDR3 sequences were not included since they are determined by the types of V or J genes. The result showed that, generally, convergent TCR favors polar neutral amino acids and acidic amino acids, particularly tyrosine (Y), serine (S), and glutamic acid (E), with alkaline amino acids and nonpolar hydrophobic amino acid less favorable (Figure 1C). Arginine (R), leucine (L), and serine (S), each of which has 6 corresponding codons, have the largest codon degeneracy. However, leucine and arginine did not show the same level of relevance to TCR convergence as serine. While the ratio of leucine slightly increased in convergent TCR, the percentage of arginine even decreased as the TCR degeneracy increased. Tyrosine and glutamic acid, on the other hand, have only two codons each but were found in a higher percentage of convergent TCRs than non-convergent TCRs. The possibility of each above-mentioned amino acid (L, Y, S, E, R) in each degree of TCR degeneracy has been validated by two-sided binomial exact tests with p-values less than 2.2×10^{-16} (except for the difference between 'D2' and 'D1' for leucine, whose p-value is 7.018×10^{-7}). These results indicate that the level of codon degeneracy is not a determinant of TCR degeneracy. Instead, the physiochemical properties of a given amino acid might have a greater impact on TCR convergence. Therefore, even though TCR convergence is the result of codon degeneracy, it exhibits an independent distribution and may carry a biological significance that is distinct from codon degeneracy.

Public TCRs are generated from VDJ recombination biases and are shared across different individuals, which might target common antigens, such as viral epitopes (Ryan O. Emerson et al., 2017). In contrast, TCRs targeting cancer neoantigens are mostly 'private' that are unique to each individual (Madi et al., 2014). Healthy individuals are expected to be exposed to common pathogens, which might induce public T cell responses. On the other hand, cancer patients have more neoantigens due to the accumulative mutation, which drives their antigen-specific T cells to recognize these 'private' antigens. This reduces the proportion of public TCRs in antigen-specific TCRs. A higher tumor mutation burden (TMB) would indicate a higher abundance of neoantigens, resulting in a lower ratio of public TCRs. Convergent recombination was claimed to be the mechanistic basis for public TCR response in many previous studies (Quigley et al., 2010; Venturi et al., 2006). As TCR convergence and publicity are conceptually similar, we next investigated the differences between the two by comparing the fractions of public TCRs within the convergent TCRs from healthy donors and patients with different cancers (Figure 1D). As expected, we observed a high overlap (47.92%) between the two in healthy donors as well as in another independent cohort consisting of patients who recovered from COVID-19 for more than 6 weeks (47.44%). Whereas for the cancer patients, the overlap between TCR convergence and publicity decreases for cancer types with higher TMB, which was estimated from previous studies (Yarchoan et al., 2019). The fraction of public TCRs within the convergence sequences dropped from 34.11% for the low mutation burden ovarian cancer to 21.34% for the highly mutated blood cancer samples. As the public TCRs in these cohorts are consistently defined by the Emerson 2017 cohort (Ryan O. Emerson et al., 2017), this difference is unlikely caused by unknown batch effects. Together, our results indicated that TCR convergence and publicity represent two different biological processes that diverge in cancer patients by tumor mutation load.

Convergent TCRs are more likely to be antigen-specific

To explore the relationship between convergent T cells and antigen-specific T cells, we began by examining datasets with known TCR antigen specificity. To be precise, since all the T cells are specific to some antigen(s) during positive thymic selection, the term “antigen-specific” means T cells with an ongoing or memory antigen-specific immune response in this context. In 2020, 10x Genomics detected the antigen specificity of T cells using highly multiplexed peptide-MHC (pMHC) multimers (10xGenomics) from PBMC samples of four healthy individuals. We picked out all convergent TCRs within each donor and examined their overlaps with antigen-specific multimer$^+$ TCRs. Interestingly, while convergent TCRs only consisted of a small proportion of all unique clones, they were dominantly enriched for antigen-specific TCRs (Figure 2A). In these four donors, only 12.91%, 33.32%, 71.65%, and 6.24% of the unique TCRs were multimer positive, respectively. However, 88.00%, 88.28%, 93.10%, and 22.22% of convergent TCRs were antigen-specific, suggesting that convergent TCRs were much more likely to be antigen-specific than their non-convergent counterpart. We next estimated the statistical significance of this observation for each donor using Fisher's exact test (Blevins & McDonald, 1985) (Figure 2B). The odds ratio for donor 1 reached 50.16, indicating that convergent TCRs are much more likely to be antigen-specific. Similar conclusions also held for donors 2 and 3, while the p-value (0.105) of donor 4 was less significant. Considering the small percentage of antigen-specific T cells detected in donor 4, it is possible that most antigens specific to donor 4's T cells did not fall within the spectrum of tested antigens in this experiment, which caused the lower overlap. Overall, the findings from this dataset supported the hypothesis that convergent T cells are more likely to be antigen-specific than other T cells.

We further tested this hypothesis using high-throughput bulk TCRβ-seq data. TCR-seq data were downloaded from ImmunoAccess, containing TCRβ sequences from over 1,400 subjects that had been exposed to or infected with the SARS-CoV-2 virus, along with over 160,000 SARS-CoV-2-specific TCRs (Snyder et al., 2020). Similarly, we calculated the odds ratio of convergent TCRs and SARS-CoV-2-specific TCRs using Fisher's exact test for each sample. In comparison, we also investigated the overlap between clonally expanded T cells and antigen-specific T cells in these datasets, as T cell clonal expansion is a commonly used criterion for antigen-specificity in recent clinical studies (Kidman et al., 2020; Subudhi et al., 2016). In each sample, TCR sequences with the highest read numbers were selected as expanded TCRs, and their number was restricted to the same number as convergent TCRs. Overall, the odds ratio of convergent TCR was almost unanimously higher than that of the clonally expanded T cells (Figure 2C). The average odds ratio of TCR convergence was 8.94, significantly higher than TCR clonal expansion, which was only 1.74 (Figure 2D). These results confirmed that converged TCRs are more likely to be antigen-experienced, and indicated that TCR convergence is a better indicator of antigen-specificity than clonal expansion.

Convergent T cells exhibit a CD8$^+$ cytotoxic gene signature

In the next step, we studied the gene expression signatures of the convergent T cells using single cell RNA sequencing (scRNA-seq) data in the tumor samples to test the hypothesis that convergence is associated with an antigen-experienced effector phenotype. In the tumor microenvironment, antigen-specific T cells typically express cytotoxicity, memory, and/or exhaustion gene signatures (Caushi et al., 2021; Oliveira et al., 2021). We, therefore, checked the phenotypes of convergent T cells using recent single cell RNA-seq data generated from cancer studies.

The first dataset we investigated contained tetramer-labeled neoantigen-specific T cells in the MC38 tumor mouse model treated with neoantigen vaccine and anti-PD-L1 (Longchao Liu et al., 2022). The CD4$^+$ and CD8$^+$ T cells were analyzed separately using the same cell cluster annotations as previously described (Longchao Liu et al., 2022). We then defined and analyzed the convergent TCRs using the scTCR-seq data of the same T cells. T cells collected at tumor sites exhibited a much higher level of convergence than T cells in lymph nodes and spleen. Therefore, in the following description, we specifically refer to tumoral T cells. Only 1.85% (n=90) of the 4,878 passed-QC CD4$^+$ cells had convergent TCRs (Figure 3—figure supplement 1A), which were not enriched in any cluster. In contrast, 9,143 CD8$^+$ T cells passed the quality control and 15.71% (n=1,436) of them were convergent T cells (Figure 3A). Following this result, we noticed two interesting clusters: effector T cells with significant enrichment of convergent CD8$^+$ T cells and naïve T CD8$^+$ cells with none of the T cells having a convergent TCR (Figure 3A). To further investigate these two clusters, we re-clustered the CD8$^+$ cells with a higher resolution (0.3) and divided the T cells into 5 new clusters (Figure 3B). 87.5% (419 out of 479) of T cells from cluster 04 were convergent T cells, whereas none of the cluster 05 T cells (n=201) were convergent. Cluster 04 T cells expressed both effector and exhaustion markers, such as *Tox* (Bordon, 2019; Sekine et al., 2020), *Tigit* (Ostroumov et al., 2021), *Eomes* (Li et al., 2018), and inhibitory receptors like *Il10ra* (Al-Abbasi et al., 2018). On the other hand, cluster 05 mainly expressed naïve T cell gene signatures (Al-Abbasi et al., 2018), such as *Sell*, *Tcf7*, and *Ccr7* (Figure 3C), and down-regulated genes associated with effector function, like *Gzmb*, *Lgals1* (Li et al., 2020), and T cell activation or exhaustion (Saleh et al., 2020), like *Pdcd1* and *Lag3* (Figure 3C). Finally, by comparing the TCRs of cluster 04 T cells and tetramer sorted T cells specific to tumor neoantigen *Adpgk*, we observed a significant enrichment of convergent TCRs for neoantigen-specificity (OR=4.41, $p<0.001$) (Figure 3—figure supplement 1B). Based on these results, we concluded that convergent T cells are enriched for a neoantigen-activated effector phenotype in the tumor microenvironment.

To verify this observation in humans, we analyzed the second dataset of pan-cancer scRNA-seq data of nine cancer types (Zheng et al., 2021). There were samples collected from tumor tissues, normal adjacent tissues, and peripheral blood from human patients. Similar to the mouse data, convergent T cells were rare in CD4$^+$ T cells (n=52,643), accounting for 0.22% (n=118) of the CD4$^+$ population (Figure 3—figure supplement 1C), whereas 10.15% (n=7,065) of the CD8$^+$ T cells (n=69,618) were convergent (Figure 3D). 14 CD4$^+$ clusters and 15 CD8$^+$ clusters were defined using the Seurat R package, and the cluster annotations were assigned by the gene markers of each cluster shown in Supplementary File 2. Similar to the mouse data, convergent T cells were enriched in clusters CD8-01-XCL1 and CD8-05-FGFBP2, yet remained at low levels in the naïve clusters, CD8-09-IL7R and CD8-13-CCR7 (1.27% and 0.00%) (Table 1). CD8-01-XCL1 cluster expressed gene signatures of activated T cells, including *Il7r* (Seddiki et al., 2006), *ZFP36L2* (Petkau et al., 2021), *XCL1* (Ordway et al., 2007), and *CXCR6* (Wang et al., 2021) (Figure 3E), as well as proliferation and mobility markers *LDLRAD4* (Liu et al., 2017) and *CAPG* (Wei et al., 2020). CD8-05-FGFBP2 cluster exhibited a CD8$^+$ effector phenotype, upregulating genes involved in cytotoxicity and anti-tumor activity, *FGFBP2*, *GNLY*, *FCGR3A* (Zheng et al., 2017), *GZMH*, *PRF1*, *TGFBR3*, and *GZMB* (Figure 3E). In conclusion, the phenotypes of convergent T cells in human samples resembled those discovered in the mouse model. Convergent T cells were consistently enriched in the activated CD8$^+$ T cells clusters with the gene signatures of cytotoxicity, memory, and exhaustion, and demonstrated the phenotypes of antigen-specific T cells.

To validate the conclusion that TCR convergence was more common in $CD8^+$ T cells than in $CD4^+$ T cells, we used another bulk TCRβ data collected from patients with Hodgkin lymphoma (Cader et al., 2020). As a result, $CD8^+$ T cells consistently exhibited a higher level of TCR convergence throughout different collection time points (Figure 3—figure supplement 2A-2B). The results may be explained by larger cell expansions of $CD8^+$ T cells than $CD4^+$ T cells. Therefore, we calculated the number of convergent clones within $CD8^+$ T cells and $CD4^+$ T cells from the above datasets to exclude the effects of cell expansion. As a result, in the scRNA-seq mouse data, while only 1.54% of the $CD4^+$ clones were convergent, 3.76% of the $CD8^+$ clones showed convergence. Likewise, 0.17% of convergent $CD4^+$ T cell clones and 1.03% of convergent $CD8^+$ T cell clones were found in human scRNA-seq data. In the bulk TCR-seq lymphoma data, similar results were also observed, where the gap between the convergent levels of $CD4^+$ and $CD8^+$ T cells narrowed but remained significant (Figure 3 — figure supplement 2C-2D). In conclusion, these results suggest that $CD8^+$ T cells show higher levels of convergence than $CD4^+$ T cells, which substantiated our hypothesis that convergent T cells are more likely antigen-experienced. This observation has been tested using multiple datasets with diverse sequencing platforms and sequencing depth to minimize the impact of batch or other technical artifacts.

TCR convergence is associated with the clinical outcome of ICB treatment

Immune checkpoint blockade (ICB) has seen great success in treating late-stage cancer patients (Bagchi et al., 2021; Jenkins et al., 2018). While ICB treatment is effective for some patients, a significant proportion of patients remain unresponsive, and the reason for this is not completely understood. The discovery of a biomarker that can assist in the prognosis of immunotherapy remains one of the top clinical priorities. Since antigen-specific T cells play a crucial role in fighting tumor cells, we next investigated whether TCR convergence is predictive of the immunotherapy outcomes. We used bulk TCRβ-seq data generated by Snyder A, et al (Snyder et al., 2017) which included PBMC samples from 29 urothelial cancer patients collected on the first day of the anti-PD1 treatment. We observed a significant association between TCR convergence level and both the overall survival (OS) ($p=0.02$) and progression-free survival (PFS) ($p=0.00038$) (Figures 4A-4B). In the low TCR convergence group, over 90% of the patients suffered from disease progression within 75 days after initial treatment, whereas only 27% of the patients in the high TCR convergence group experienced disease progression at this time point (Figure 4B).

To validate the result, we examined another independent melanoma data set containing 30 samples collected from melanoma patients treated with sequential ICB reagents (Yusko et al., 2019). Patients were randomly divided into two arms with the flipped ordering of anti-PD1 or anti-CTLA4 treatments (Yusko et al., 2019). TCR convergence significantly predicted patient outcome ($p<0.0001$) (Figure 4C), with higher TCR convergence associated with longer survival. Responders of the ICB treatment have significantly higher levels of TCR convergence (Figure 4D). To determine whether this link is confounded by other signals, we also included three additional variables that might influence the clinical outcomes in the multivariate Cox regression model, including TCR clonality, TCR diversity, different sequential treatments, and sequencing depth. TCR convergence remained significant ($p=0.011$) while adjusted for the additional variables (Figure 4E). Similarly, Cox regression models were applied to the urothelial cancer dataset above to adjust for the effects of TCR clonality, TCR diversity, as well as sequencing depth. As a result, TCR convergence remained a significant predictor of OS ($p=0.045$) and PFS ($p=0.002$) (Figure 4 — figure supplement 1A-1B). Together, our results indicated that TCR convergence is an independent prognostic predictor for patients receiving ICB treatments.

Discussion

Functional redundancy in crucial molecular pathways, such as cell cycle, metabolism, apoptosis, etc., is one of the key mechanisms that keeps biological systems robust and stable against genetic variations and environmental changes (Duncan et al., 1999; Chambon, 1994). Redundancy may occur at both molecular and cellular levels, the latter being particularly exploited by the immune system (Dombrowski & Wright, 1992). In this study, we investigated the convergence of TCRs, which could be a potentially unreported mechanism of redundancy in the adaptive immunity. The selection of multiple clones of T cells with identical TCRs during antigen encounters might be a fail-safe mechanism to ensure the expansion of the antigen-specific T cells. This speculation is in line with our observations in this work: convergent TCRs are enriched for antigen-experienced receptors, and these T cells exhibited a neoantigen-specific, cytotoxic CD8$^+$ phenotype in the tumor microenvironment.

There are conceptual overlaps between TCR publicity and convergence, as both describe the redundancy of TCRs. However, our data suggested that these two processes could be fundamentally different. Currently, the functional consequence of V(D)J recombination bias remains unclear, but it has been speculated that this conserved process is evolved to eliminate common foreign pathogens, such as viruses and bacteria, that are frequently encountered in the environment (Huisman et al., 2022; Li et al., 2012). This process is genetically encoded and arises intrinsically. In contrast, TCR convergence might reflect the magnitude of clonal selection, which is highly contingent on the antigen landscape. Our observation that convergent and public TCRs diverge in cancers with high neoantigen load (Figure 1D) strongly supports this speculation.

Previous studies have demonstrated that CD8$^+$ and CD4$^+$ T cells may possess distinct TCRβ repertoire (Wang et al., 2010; Emerson et al., 2013), which results in the difference in their capacity to generate high avidity T cells (Nakatsugawa et al., 2016). Based on our findings in this study, CD8$^+$ T cells have a higher level of convergence than CD4$^+$ T cells on both single cell and cell clone levels. In general, CD8$^+$ T cells play a direct role in killing abnormal cells (Dustin & Long, 2010; Halle et al., 2017), whereas most activated CD4$^+$ T cells function as conventional helper T cells or T regulatory cells to facilitate and regulate the immune response (Wan, 2010; London et al., 1998). This may lead to a greater impact of antigen selection on CD8$^+$ T cells than on CD4$^+$ T cells and thus a higher convergence level in cytotoxic CD8$^+$ T cells.

This indication of antigen-driven selection makes TCR convergence an attractive biomarker for disease diagnosis and/or prognosis. Many previous clinical studies have attempted to use the TCR repertoire to monitor the response of immunotherapies (Page et al., 2016; Postow, Manuel, et al., 2015; Hopkins et al., 2018; Roh et al., 2017). TCR diversity indexes, such as Shannon's entropy, richness, clonality, etc. are routinely used in these studies, yet a strong association with the patient outcome is rarely observed. This is potentially because these indexes only reflect the overall dynamics of the repertoire, which cannot specifically capture the antigen-specific response. As a new summary statistic of the immune repertoire, TCR convergence hinges on the (neo)antigen-specific T cell responses in late-stage tumors, and achieved better prediction accuracy compared to the diversity indexes. In addition, the potential prognostic value of TCR convergence and TCR similarity-based clustering was tested in previous studies (Looney et al., 2019; Pogorelyy et al., 2019). Awaiting further clinical validation, we anticipate that TCR convergence will be a powerful new biomarker to predict ICB therapy responses.

There are also limitations of this study. First, as TCR convergence is identified by small DNA changes, it relies heavily on sequencing accuracy. Improper handling of sequencing errors may result in the overestimation of TCR convergence (Looney et al., 2019). Second, since convergent T cells constitute only a small proportion of the total population of T cells, the study of TCR convergence requires a large number of sequenced T cells from each sample. Therefore, an accurate and deep sequencing approach is required for the study of TCR convergence. Third, although the phenotypic signatures of the convergent T cells were confirmed using both mouse and human scRNA-seq/scTCR-seq data, additional neoantigen-specific single T cell datasets would be necessary to consolidate our conclusions. Finally, our observation that TCR convergence is a favored prognostic predictor for ICB therapy was based on only three cohorts, due to the lack of qualified datasets. Future clinical studies with TCR repertoire profiles will be required to confirm TCR convergence as a new biomarker.

Antigen-specific T cells are central to a T cell immune response and determine the efficacy of immunotherapy, but their identification is challenging. However, the detection of convergent T cells is an easier task that only requires the profiling of the immune repertoire. The insights regarding TCR convergence gained from this work might provide an alternative angle to study the dynamics of antigen-specific T cells and a better understanding of the TCR repertoire.

Materials and Methods

Datasets preparation

All the single cell RNA-seq (scRNA-seq) or TCR-seq datasets used were accessed from public resources. This study analyzed three single-cell immune profiling datasets and eleven bulk TCRβ sequencing datasets, which were subjected to different analyses based on the features of the datasets. All the TCR-seq samples were downloaded from the immuneACCESS online database (https://clients.adaptivebiotech.com/immuneaccess). The detailed information for all datasets is described in supplementary file 1.

Definition of TCR convergence

T cells with identical CDR3 amino acid sequences and variable genes, but different CDR3 nucleotide reads were defined as convergent T cells. Single cell immune profiling data analysis included pairing the α-β chain CDR3 sequences and variable genes of each T cell to form a complete TCR sequence. As for the bulk TCR-seq data, the β chain CDR3 sequences and their variable genes were used to represent their TCR sequences.

Calculation of public TCRs proportion among convergent TCRs in different cohorts

The TCR sequences shared by at least 5% (n=34) of different individuals within the 666 samples from the Emerson cohort (Ryan O. Emerson et al., 2017) were defined as the public TCRs. The following datasets were constructed for each group. From immunoSEQ hsTCRB-V4b Control Data (Hamm, 10/21/2020), we selected 50 healthy samples with the deepest sequencing depths. 38 samples collected from whom have recovered from COVID-19 for more than 6 weeks were selected from the Nolan cohort (Nolan et al., 2020). The kidney cancer cohort (n=19) is a combination of data from our internal database and samples from the Chow cohort (Chow et al., 2020). Ovarian cancer data was also sourced from our in-house data (n=45). The immune sequencing data for lymphoma (n=53), lung cancer (n=50), and melanoma (n=29) came from the Cader cohort (Cader et al., 2020), Reuben cohort (Reuben et al., 2020), and Riaz cohort (N. Riaz et al., 2017), respectively. By dividing convergence-publicity overlap by the number of all convergent TCRs, we calculated the proportion of public TCRs among convergent TCRs in each sample.

Definition of TCR clonality and diversity

TCR clonality was defined as 1-Pielou's evenness and was calculated as:

$$TCR\ clonality = 1 + \frac{\sum_i^n p_i log_2\ (p_i)}{log_2(n)},$$

where p_i is the proportional abundance of unique TCR clonotype I and n is the total number of TCR clonotypes within a given sample. TCR diversity was calculated as n/N, where n is the total number of TCR clonotypes and N is the total reproductive sequence reads.

Analysis of scRNA-seq data

The scRNA-seq data were analyzed primarily using R/4.1.1 and Seurat 4.0.6. Data were processed with the conventional scRNA-seq pipeline before the analysis of TCR convergence. In brief, all data first had to pass quality control, where cells with a high ratio of mitochondria and ribosomes were excluded since they were likely to be dying cells. Cells with too many or too few RNA reads were also excluded from studies because they might be doublets (a single barcode identifies more than one cell) or background noise. Data scaling and normalization were then performed. We regressed out the cell-cycle associated genes in our scaled data. Principal Component Analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) algorithm were applied to reduce the dimensions and visualize the data. Based on the markers of the cells within a cluster, cells were grouped into an appropriate number of clusters and assigned T cell types. Combining the scTCR-seq data with the scRNA-seq data allowed us to study the distribution of convergent T cells and their phenotypes.

Statistical analysis

The two-tailed exact binomial tests were performed with the built-in function of R/4.1.1 to estimate the statistical differences in amino acid composition between convergent TCRs and non-convergent TCRs. Fisher's exact tests were employed to analyze the association between convergent TCRs and antigen-specific TCRs, as well as the relationship between clonally expanded TCRs and antigen-specific TCRs. Using ggstatsplot v.0.8.0, the Fisher's exact test results of the scTCR-seq data were visualized. Ggsignif v.0.6.3 was used to calculate the statistical difference between different groups in all boxplots using Welch's t-test. In the 'box' within every boxplot, the black liner refers to the median and the red line indicates the mean value. All survival analyses were conducted using the R package survival v.3.2.13. Survival curves were derived using the Kaplan-Meier method. Cox multivariate regressions were performed to assess the association between the patients' outcomes and other variables.

Acknowledgments

This work is supported by NCI 1R01CA245318 (B.L.) and 1R01CA258524 (B.L.).

Data and code availability

The Python code related to TCR convergence calculation are available at: https://github.com/Mia-yao/TCR-convergence/tree/main. The convergent TCR sequences of each cohort are uploaded to Zenodo, with DOI: 10.5281/zenodo.7139024.

Competing interests

No competing interests.

Table 1 The Fisher's exact test result of CD8$^+$ clusters

clusters	normal adjacent tissues			tumor tissues		
	convergence ratio	odds ratio	p-value	convergence ratio	odds ratio	p-value
CD8-01-XCL1	22.95%	5.07	<2.2e-16	15.03%	5.66	<2.2e-16
CD8-02-GZMK	2.59%	0.24	1	3.48%	0.85	0.9949
CD8-03-TIGIT	3.69%	0.37	1	2.12%	0.49	1
CD8-04-ZNF683	7.43%	0.79	0.9984	3.72%	1.62	<2.2e-16
CD8-05-FGFBP2	32.44%	6.19	<2.2e-16	26.69%	12.25	<2.2e-16
CD8-06-IFNG	5.80%	0.59	1	8.71%	2.34	9.725E-15
CD8-07-IHSPA6	22.75%	3.93	<2.2e-16	4.99%	1.27	0.004411
CD8-08-PRF1	4.84%	0.5	1	3.11%	0.77	0.9404
CD8-09-IL7R	1.02%	0.1	1	1.44%	0.35	1
CD8-10-IFI6	11.11%	1.23	0.2523	5.54%	1.42	0.0003609
CD8-11-STMN1	6.07%	0.64	0.9603	3.93%	0.98	0.5834
CD8-12-KLRB1	0.48%	0.03	1	0.13%	0.03	1
CD8-13-CCR7	0.00%	0	1	0.00%	0	1
CD8-14-CXCR6	0.45%	0.01	1	5.00%	1.26	0.4325
CD8-15-CTLA4	0.00%	0	1	1.68%	0.41	0.2

Figure Legends

Figure 1 with 1 supplement

TCR convergence is different from publicity

A. Cartoon illustration of TCR convergence, created with BioRender.com. T cells with different DNA sequences encode the same TCR due to codon degeneracy.

B. log10-transformed counts of TCRs with different degrees of degeneracies. The X-axis indicates the number of distinct DNA sequences corresponding to the same TCR, or, the "degree of degeneracy". "D1" refers to the sequences without any degeneracy.

C. The proportion of twenty amino acids in TCRs with degeneracies from one to four. Each of the 20 natural amino acids is represented on the x-axis by a one-letter abbreviation.

D. Tumor mutation burden (TMB) and the proportion of public TCRs in convergent TCRs among different cancer types. The upper panel shows the TMB of ovary, kidney, lung, melanoma, and blood cancer types. Cancer types are ordered from the lowest median TMB (left) to the highest median TMB (right). Lower panel shows the ratio of public TCRs among convergent TCRs in the respective cancer types.

Figure 1 — figure supplement 1

The length and variable gene usage of convergent TCRs

A. The proportion of convergent TCRs among TCRs of a certain length. The x-axis represents the amino acid length of TCR CDR3. Each dot indicates one sample.

B. The length distribution of all unique TCR CDR3 reads.

C. The average degeneracy of convergent TCRs with different CDR3 lengths.

D. The proportion of convergent TCRs among TCRs with a certain variable gene. The x-axis refers to the TCRβ chain variable gene (TCRBV) superfamily.

E. The average degeneracy of convergent TCRs with different variable gene usages.

Figure 2

Convergent TCRs are more likely to be antigen-specific

A. The overlaps between all unique TCRs, viral antigen-specific TCRs, and convergent TCRs of four donors respectively.

B. The results of Fisher's exact tests regarding the association of convergent TCRs and viral antigen-specific TCRs of four donors.

C. Log10-scaled odds ratio of TCR convergence and expansion with respect to antigen-specificity of each sample. The samples were sorted in order of TCR convergence odds ratio from small to large. The horizontal stripes on the left represent the density of samples, colored by the TCR signal.

D. The difference of log10-transformed odds ratio between TCR convergence and TCR expansion.

Figure 3 with 2 supplements

Convergent T cells exhibit a CD8$^+$ cytotoxic gene signature

A. t-distributed stochastic neighbor embedding (t-SNE) plots showing cell type annotations and convergent T cell distribution of CD8$^+$ T cells collected from mouse tumor tissues. (Each cell type corresponds to the following clusters in the original paper: T.na: CD8-04-Ccr7; T.eff: CD8-05-Fasl; T.rm: CD8-06-Xcl1; T.ex: CD8-08-Lag3. (L. Liu et al., 2022)) Each dot represents a T cell. The area within the red dashed line indicates enrichment of convergent T cells and cells within the blue dashed line were naive T cells.

B. t-SNE displaying the re-cluster result of CD8$^+$ T cells. The red dashed line and blue dashed line highlighted the same groups of cells as in Fig 3A.

C. Volcano plots showing the differentially expressed genes of cluster 04 and cluster 05.

D. t-SNE plots showing the cell clusters and convergent T cell distribution of CD8$^+$ T cells collected from human tumors and normal adjacent tissues. An area within the red dashed line was enriched with convergent T cells and the T cells within the blue dashed line were naive T cells.

E. The gene markers of cluster 01 and cluster 05 of human pan-cancer data. (Only the up-regulated genes of each cluster are shown)

Figure 3 — figure supplement 1

TCR convergence in CD4$^+$ T cells

A. Cell type annotations and convergent T cell distribution of CD4$^+$ T cells collected from mouse tumor tissues. (Each cell type corresponds to the following clusters in the original paper: T.na: CD4-04-S1pr1; T.Treg: CD4-05-Ctla4; T.Th1: CD4-06-Bhlhe40; T.rTreg: CD4-07-Ccr5 (L. Liu et al., 2022)).

B. The results of Fisher's exact tests regarding the association of tetramer$^+$ T cells with T cells from cluster 04 or cluster 05.

C. Cell clusters and convergent T cell distribution of CD4$^+$ T cells collected from human tumors and normal adjacent tissues.

Figure 3 — figure supplement 2

The convergent T cell levels in CD4$^+$ and CD8$^+$ T cells

A. The different levels of convergent T cells in CD4$^+$ and CD8$^+$ T cells. The samples (n=17) were collected from baseline BCL patients. Each dot represents one sample.

B. The different levels of convergent T cells in CD4$^+$ and CD8$^+$ T cells. The samples (n=16) were collected from BCL patients on day 1 of anti-PD1 treatment cycle 4. Each dot represents one sample.

C. The proportion of convergent T cell clones among all CD4$^+$ and CD8$^+$ T cell clones of baseline BCL patients. Each dot represents one sample.

D. The proportion of convergent T cell clones among all CD4$^+$ and CD8$^+$ T cell clones of anti-PD1 PreCycle4 BCL patients. Each dot represents one sample.

Figure 4 with 1 supplement

TCR convergence is associated with the clinical outcome of ICB treatment

A. Kaplan-Meier overall survival curves of urothelial cancer patients (n=29) with different levels of convergent T cells. The convergence levels were defined by the number of convergent T cell clones within each sample. High, above median; low, below the median. For TCR convergence, all the survival analyses in this study followed this definition.

B. Progression-free survival curve of urothelial cancer patients (n=29) with different levels of convergent T cells. The progression of the tumor was determined by mRECIST 1.1 test.

C. Overall survival curves of melanoma patients (n=30) with different levels of convergent T cells.

D. The differences in the TCR convergence levels among patients with the distinct best overall response towards anti-PD1 immunotherapy. The number of patients with each BORs are as follows: complete response = CR (n=1), partial response = PR (n=9), progressive disease = PD (n=18), stable disease = SD (n=2).

E. Cox model multivariate analysis for the overall survival of melanomas cohort with covariate adjustments for other four variables: TCR clonality, TCR diversity, treatment method, and sequencing depth. The sequencing depth was determined by the total number of detected TCR clonotypes.

Figure 4 — figure supplement 1

Multivariate cox regression analysis for urothelial cancer cohort

A. Cox model multivariate analysis for the overall survival with covariate adjustments for other three variables: TCR clonality, TCR diversity, and sequencing depth. The sequencing depth was determined by the total number of detected TCR clonotypes.

B. Cox model multivariate analysis for the progression-free survival with covariate adjustments for other three variables: TCR clonality, TCR diversity, and sequencing depth. The sequencing depth was determined by the total number of detected TCR clonotypes.

Supplementary Table 1.

Details of datasets used in this study.

Public datasets used in this study are summarized in this table. A total of three single-cell immune profiling datasets and eleven bulk TCR-seq datasets are involved in this study. The table lists the original project information and accession numbers for each dataset.

Supplementary Table 2.

The clusters annotation in human pan-cancer data.

The T cell cluster characteristics were summarized based on the gene expression signature of each cluster. This study only included the samples sequenced in the original study (Zheng et al., 2021), additional datasets used in the original publication that were sourced from other cohorts were not included in this study.

References

[1] Marrack, P., & Kappler, J. (1987). The T-Cell Receptor. *Science*, *238*(4830), 1073-1079. https://doi.org/DOI 10.1126/science.3317824

[2] Davis, M. M., & Bjorkman, P. J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature*, *334*(6181), 395-402. https://doi.org/10.1038/334395a0

[3] Pai, J. A., & Satpathy, A. T. (2021). High-throughput and single-cell T cell receptor sequencing technologies. *Nat Methods*, *18*(8), 881-892. https://doi.org/10.1038/s41592-021-01201-8

[4] Trainor, L. E., Rowe, G. W., & Szabo, V. L. (1984). A tetrahedral representation of poly-codon sequences and a possible origin of codon degeneracy. *Journal of theoretical biology*, *108*(3), 459-468. https://doi.org/10.1016/s0022-5193(84)80046-6

[5] Looney, T. J., Topacio-Hall, D., Lowman, G., Conroy, J., Morrison, C., Oh, D., Fong, L., & Zhang, L. (2019). TCR Convergence in Individuals Treated With Immune Checkpoint Inhibition for Cancer. *Front Immunol*, *10*, 2985. https://doi.org/10.3389/fimmu.2019.02985

[6] Gonzalez, D. L., Giannerini, S., & Rosa, R. (2019). On the origin of degeneracy in the genetic code. *Interface Focus*, *9*(6), 20190038. https://doi.org/10.1098/rsfs.2019.0038

[7] McClellan, D. A. (2000). The Codon-Degeneracy Model of Molecular Evolution. *Journal of Molecular Evolution*, *50*(2), 131-140. https://doi.org/10.1007/s002399910015

[8] Crick, F. H., Barnett, L., Brenner, S., & Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature*, *192*, 1227-1232. https://doi.org/10.1038/1921227a0

[9] Wong, J. T.-F. (1975). A Co-Evolution Theory of the Genetic Code. *Proceedings of the National Academy of Sciences*, *72*(5), 1909-1912. https://doi.org/doi:10.1073/pnas.72.5.1909

[10] Copley, S. D., Smith, E., & Morowitz, H. J. (2005). A mechanism for the association of amino acids with their codons and the origin of the genetic code. *Proceedings of the National Academy of Sciences*, *102*(12), 4442-4447. https://doi.org/doi:10.1073/pnas.0501049102

[11] Di Giulio, M. (2004). The origin of the tRNA molecule: implications for the origin of protein synthesis. *Journal of*

theoretical biology, *226*(1), 89-93. https://doi.org/https://doi.org/10.1016/j.jtbi.2003.07.001

[12] Koonin, E. V., & Novozhilov, A. S. (2009). Origin and evolution of the genetic code: the universal enigma. *IUBMB life*, *61*(2), 99-111. https://doi.org/10.1002/iub.146

[13] Jameson, S. C., & Bevan, M. J. (1998). T-cell selection. *Current Opinion in Immunology*, *10*(2), 214-219. https://doi.org/https://doi.org/10.1016/S0952-7915(98)80251-3

[14] Smith-Garvin, J. E., Koretzky, G. A., & Jordan, M. S. (2009). T Cell Activation. *Annual Review of Immunology*, *27*(1), 591-619. https://doi.org/10.1146/annurev.immunol.021908.132706

[15] Postow, M. A., Callahan, M. K., & Wolchok, J. D. (2015). Immune Checkpoint Blockade in Cancer Therapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, *33*(17), 1974-1982. https://doi.org/10.1200/JCO.2014.59.4358

[16] Peng, M., Mo, Y., Wang, Y., Wu, P., Zhang, Y., Xiong, F., Guo, C., Wu, X., Li, Y., Li, X., Li, G., Xiong, W., & Zeng, Z. (2019). Neoantigen vaccine: an emerging tumor immunotherapy. *Molecular Cancer*, *18*(1), 128. https://doi.org/10.1186/s12943-019-1055-6

[17] Newell, E. W., & Davis, M. M. (2014). Beyond model antigens: high-dimensional methods for the analysis of antigen-specific T cells. *Nat Biotechnol*, *32*(2), 149-157. https://doi.org/10.1038/nbt.2783

[18] Riaz, N., Havel, J. J., Makarov, V., Desrichard, A., Urba, W. J., Sims, J. S., Hodi, F. S., Martín-Algarra, S., Mandal, R., Sharfman, W. H., Bhatia, S., Hwu, W.-J., Gajewski, T. F., Slingluff, C. L., Chowell, D., Kendall, S. M., Chang, H., Shah, R., Kuo, F., Morris, L. G. T., Sidhom, J.-W., Schneck, J. P., Horak, C. E., Weinhold, N., & Chan, T. A. (2017). Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell*, *171*(4), 934-949.e916. https://doi.org/https://doi.org/10.1016/j.cell.2017.09.028

[19] Tumeh, P. C., Harview, C. L., Yearley, J. H., Shintaku, I. P., Taylor, E. J. M., Robert, L., Chmielowski, B., Spasic, M., Henry, G., Ciobanu, V., West, A. N., Carmona, M., Kivork, C., Seja, E., Cherry, G., Gutierrez, A. J., Grogan, T. R., Mateus, C., Tomasic, G., Glaspy, J. A., Emerson, R. O., Robins, H., Pierce, R. H., Elashoff, D. A., Robert, C., & Ribas, A. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*, *515*(7528), 568-571. https://doi.org/10.1038/nature13954

[20] Johnson, D. B., Frampton, G. M., Rioth, M. J., Yusko, E., Xu, Y., Guo, X., Ennis, R. C., Fabrizio, D., Chalmers, Z. R., Greenbowe, J., Ali, S. M., Balasubramanian, S., Sun, J. X., He, Y., Frederick, D. T., Puzanov, I., Balko, J. M., Cates, J. M., Ross, J. S., Sanders, C., Robins, H., Shyr, Y., Miller, V. A., Stephens, P. J., Sullivan, R. J., Sosman, J. A., & Lovly, C. M. (2016). Targeted Next Generation Sequencing Identifies Markers of Response to PD-1 Blockade. *Cancer Immunology Research*, *4*(11), 959-967. https://doi.org/10.1158/2326-6066.Cir-16-0143

[21] Robert, L., Tsoi, J., Wang, X., Emerson, R., Homet, B., Chodon, T., Mok, S., Huang, R. R., Cochran, A. J., Comin-Anduix, B., Koya, R. C., Graeber, T. G., Robins, H., & Ribas, A. (2014). CTLA4 Blockade Broadens the Peripheral T-Cell Receptor Repertoire. *Clinical Cancer Research*, *20*(9), 2424-2432. https://doi.org/10.1158/1078-0432.Ccr-13-2648

[22] Amaria, R. N., Reddy, S. M., Tawbi, H. A., Davies, M. A., Ross, M. I., Glitza, I. C., Cormier, J. N., Lewis, C., Hwu, W.-J., Hanna, E., Diab, A., Wong, M. K., Royal, R., Gross, N., Weber, R., Lai, S. Y., Ehlers, R., Blando, J., Milton, D. R., Woodman, S., Kageyama, R., Wells, D. K., Hwu, P., Patel, S. P., Lucci, A., Hessel, A., Lee, J. E., Gershenwald, J., Simpson, L., Burton, E. M., Posada, L., Haydu, L., Wang, L., Zhang, S., Lazar, A. J., Hudgens, C. W., Gopalakrishnan, V., Reuben, A., Andrews, M. C., Spencer, C. N., Prieto, V., Sharma, P., Allison, J., Tetzlaff, M. T., & Wargo, J. A. (2018). Neoadjuvant immune checkpoint blockade in high-risk resectable melanoma. *Nature Medicine*, *24*(11), 1649-1654. https://doi.org/10.1038/s41591-018-0197-1

[23] Kidman, J., Principe, N., Watson, M., Lassmann, T., Holt, R. A., Nowak, A. K., Lesterhuis, W. J., Lake, R. A., & Chee, J. (2020). Characteristics of TCR Repertoire Associated With Successful Immune Checkpoint Therapy Responses [Review]. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/fimmu.2020.587014

[24] Stewart, C. B. (1993). The powers and pitfalls of parsimony. *Nature*, *361*(6413), 603-607. https://doi.org/10.1038/361603a0

[25] Dash, P., Fiore-Gartland, A. J., Hertz, T., Wang, G. C., Sharma, S., Souquette, A., Crawford, J. C., Clemens, E. B., Nguyen, T. H. O., Kedzierska, K., La Gruta, N. L., Bradley, P., & Thomas, P. G. (2017). Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature*, *547*(7661), 89-93. https://doi.org/10.1038/nature22383

[26] Pogorelyy, M. V., Minervina, A. A., Shugay, M., Chudakov, D. M., Lebedev, Y. B., Mora, T., & Walczak, A. M. (2019). Detecting T cell receptors involved in immune responses from single repertoire snapshots. *PLoS Biol*, *17*(6), e3000314. https://doi.org/10.1371/journal.pbio.3000314

[27] Huang, H., Wang, C., Rubelt, F., Scriba, T. J., & Davis, M. M. (2020). Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol*, *38*(10), 1194-1202. https://doi.org/10.1038/s41587-020-0505-4

[28] Zhang, H., Liu, L., Zhang, J., Chen, J., Ye, J., Shukla, S., Qiao, J., Zhan, X., Chen, H., Wu, C. J., Fu, Y. X., & Li, B. (2020). Investigation of Antigen-Specific T-Cell Receptor Clusters in Human Cancers. *Clin Cancer Res*, *26*(6), 1359-1371. https://doi.org/10.1158/1078-0432.CCR-19-3249

[29] Zhang, H., Zhan, X., & Li, B. (2021). GIANA allows computationally-efficient TCR clustering and multi-disease repertoire classification by isometric transformation. *Nat Commun*, *12*(1), 4699. https://doi.org/10.1038/s41467-021-25006-7

[30] Liu, L., Chen, J., Zhang, H., Ye, J., Moore, C., Lu, C., Fang, Y., Fu, Y.-X., & Li, B. (2022). Concurrent delivery of immune checkpoint blockade modulates T cell dynamics to enhance neoantigen vaccine-generated antitumor immunity. *Nature Cancer*. https://doi.org/10.1038/s43018-022-00352-7

[31] Zheng, C., Zheng, L., Yoo, J. K., Guo, H., Zhang, Y., Guo, X., Kang, B., Hu, R., Huang, J. Y., Zhang, Q., Liu, Z., Dong, M., Hu, X., Ouyang, W., Peng, J., & Zhang, Z. (2017). Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell*, *169*(7), 1342-1356 e1316. https://doi.org/10.1016/j.cell.2017.05.035

[32] Emerson, R. O., DeWitt, W. S., Vignali, M., Gravley, J., Hu, J. K., Osborne, E. J., Desmarais, C., Klinger, M., Carlson, C. S., Hansen, J. A., Rieder, M., & Robins, H. S. (2017). Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nature Genetics*, *49*(5), 659-665. https://doi.org/10.1038/ng.3822

[33] Hamm, D., Adaptive Biotechnologies. (10/21/2020). *immunoSEQ hsTCRB-V4b Control Data*.

[34] Nolan, S., Vignali, M., Klinger, M., Dines, J. N., Kaplan, I. M., Svejnoha, E., Craft, T., Boland, K., Pesesky, M., Gittelman, R. M., Snyder, T. M., Gooley, C. J., Semprini, S., Cerchione, C., Mazza, M., Delmonte, O. M., Dobbs, K., Carreño-Tarragona, G., Barrio, S., Sambri, V., Martinelli, G., Goldman, J. D., Heath, J. R., Notarangelo, L. D., Carlson, J. M., Martinez-Lopez, J., & Robins, H. S. (2020). A large-scale database of T-cell receptor beta (TCRβ) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2. *Research square*, rs.3.rs-51964. https://doi.org/10.21203/rs.3.rs-51964/v1

[35] Chow, J., Hoffend, N. C., Abrams, S. I., Schwaab, T., Singh, A. K., & Muhitch, J. B. (2020). Radiation induces dynamic changes to the T cell repertoire in renal cell carcinoma patients. *Proc Natl Acad Sci U S A*, *117*(38), 23721-23729. https://doi.org/10.1073/pnas.2001933117

[36] Cader, F. Z., Hu, X., Goh, W. L., Wienand, K., Ouyang, J., Mandato, E., Redd, R., Lawton, L. N., Chen, P. H., Weirather, J. L., Schackmann, R. C. J., Li, B., Ma, W., Armand, P., Rodig, S. J., Neuberg, D., Liu, X. S., & Shipp, M. A. (2020). A peripheral immune signature of responsiveness to PD-1 blockade in patients with classical Hodgkin lymphoma. *Nat Med*, *26*(9), 1468-1479. https://doi.org/10.1038/s41591-020-1006-1

[37] Reuben, A., Zhang, J., Chiou, S.-H., Gittelman, R. M., Li, J., Lee, W.-C., Fujimoto, J., Behrens, C., Liu, X., Wang, F., Quek, K., Wang, C., Kheradmand, F., Chen, R., Chow, C.-W., Lin, H., Bernatchez, C., Jalali, A., Hu, X., Wu, C.-J., Eterovic, A. K., Parra, E. R., Yusko, E., Emerson, R., Benzeno, S., Vignali, M., Wu, X., Ye, Y., Little, L. D., Gumbs, C., Mao, X., Song, X., Tippen, S., Thornton, R. L., Cascone, T., Snyder, A., Wargo, J. A., Herbst, R., Swisher, S., Kadara, H., Moran, C., Kalhor, N., Zhang, J., Scheet, P., Vaporciyan, A. A., Sepesi, B., Gibbons, D. L., Robins, H., Hwu, P., Heymach, J. V., Sharma, P., Allison, J. P., Baladandayuthapani, V., Lee, J. J., Davis, M. M., Wistuba, I. I., Futreal, P. A., & Zhang, J. (2020). Comprehensive T cell repertoire characterization of non-small cell lung cancer. *Nature Communications*, *11*(1), 603. https://doi.org/10.1038/s41467-019-14273-0

[38] Riaz, N., Havel, J. J., Makarov, V., Desrichard, A., Urba, W. J., Sims, J. S., Hodi, F. S., Martin-Algarra, S., Mandal, R., Sharfman, W. H., Bhatia, S., Hwu, W. J., Gajewski, T. F., Slingluff, C. L., Jr., Chowell, D., Kendall, S. M., Chang, H., Shah, R., Kuo, F., Morris, L. G. T., Sidhom, J. W., Schneck, J. P., Horak, C. E., Weinhold, N., & Chan, T. A. (2017). Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell*, *171*(4), 934-949 e916. https://doi.org/10.1016/j.cell.2017.09.028

[39] Emerson, R. O., DeWitt, W. S., Vignali, M., Gravley, J., Hu, J. K., Osborne, E. J., Desmarais, C., Klinger, M., Carlson, C. S., Hansen, J. A., Rieder, M., & Robins, H. S. (2017). Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nature Genetics*, *49*(5), 659-+. https://doi.org/10.1038/ng.3822

[40] Madi, A., Shifrut, E., Reich-Zeliger, S., Gal, H., Best, K., Ndifon, W., Chain, B., Cohen, I. R., & Friedman, N. (2014). T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. *Genome Res*, *24*(10), 1603-1612. https://doi.org/10.1101/gr.170753.113

[41] Quigley, M. F., Greenaway, H. Y., Venturi, V., Lindsay, R., Quinn, K. M., Seder, R. A., Douek, D. C., Davenport, M. P., & Price, D. A. (2010). Convergent recombination shapes the clonotypic landscape of the na&#xef;ve T-cell repertoire. *Proceedings of the National Academy of Sciences*, *107*(45), 19414-19419. https://doi.org/doi:10.1073/pnas.1010586107

[42] Venturi, V., Kedzierska, K., Price, D. A., Doherty, P. C., Douek, D. C., Turner, S. J., & Davenport, M. P. (2006). Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proceedings of the National Academy of Sciences*, *103*(49), 18691-18696. https://doi.org/doi:10.1073/pnas.0608907103

[43] Yarchoan, M., Albacker, L. A., Hopkins, A. C., Montesion, M., Murugesan, K., Vithayathil, T. T., Zaidi, N., Azad, N. S., Laheru, D. A., Frampton, G. M., & Jaffee, E. M. (2019). PD-L1 expression and tumor mutational burden are independent biomarkers in most cancers. *JCI insight*, *4*(6). https://doi.org/10.1172/jci.insight.126908

[44] 10xGenomics. A new way of exploring immunity: linking highly multiplexed antigen recognition to immune repertoire and phenotype.

[45] Blevins, L., & McDonald, C. J. (1985). Fisher's Exact Test: an easy-to-use statistical test for comparing outcomes. *MD Comput*, *2*(1), 15-19, 68. https://www.ncbi.nlm.nih.gov/pubmed/3842426

[46] Snyder, T. M., Gittelman, R. M., Klinger, M., May, D. H., Osborne, E. J., Taniguchi, R., Zahid, H. J., Kaplan, I. M., Dines, J. N., Noakes, M. T., Pandya, R., Chen, X., Elasady, S., Svejnoha, E., Ebert, P., Pesesky, M. W., De Almeida, P., O'Donnell, H., DeGottardi, Q., Keitany, G., Lu, J., Vong, A., Elyanow, R., Fields, P., Greissl, J., Baldo, L., Semprini, S., Cerchione, C., Nicolini, F., Mazza, M., Delmonte, O. M., Dobbs, K., Laguna-Goya, R., Carreño-Tarragona, G., Barrio, S., Imberti, L., Sottini, A., Quiros-Roldan, E., Rossi, C., Biondi, A., Bettini, L. R., D'Angio, M., Bonfanti, P., Tompkins, M. F., Alba, C., Dalgard, C., Sambri, V., Martinelli, G., Goldman, J. D., Heath, J. R., Su, H. C., Notarangelo, L. D., Paz-Artal, E., Martinez-Lopez, J., Carlson, J. M., & Robins, H. S. (2020). Magnitude and Dynamics of the T-Cell Response to SARS-CoV-2 Infection at Both Individual and Population Levels. *medRxiv*. https://doi.org/10.1101/2020.07.31.20165647

[47] Subudhi, S. K., Aparicio, A., Gao, J., Zurita, A. J., Araujo, J. C., Logothetis, C. J., Tahir, S. A., Korivi, B. R., Slack, R. S., Vence, L., Emerson, R. O., Yusko, E., Vignali, M., Robins, H. S., Sun, J., Allison, J. P., & Sharma, P. (2016). Clonal expansion of CD8 T cells in the systemic circulation precedes development of ipilimumab-induced toxicities. *Proceedings of the National Academy of Sciences*, *113*(42), 11919-11924. https://doi.org/doi:10.1073/pnas.1611421113

[48] Liu, L., Chen, J., Zhang, H., Ye, J., Moore, C., Lu, C., Fang, Y., Fu, Y. X., & Li, B. (2022). Concurrent delivery of immune checkpoint blockade modulates T cell dynamics to enhance neoantigen vaccine-generated antitumor immunity. *Nat Cancer*, *3*(4), 437-452. https://doi.org/10.1038/s43018-022-00352-7

[49] Caushi, J. X., Zhang, J., Ji, Z., Vaghasia, A., Zhang, B., Hsiue, E. H., Mog, B. J., Hou, W., Justesen, S., Blosser, R., Tam, A., Anagnostou, V., Cottrell, T. R., Guo, H., Chan, H. Y., Singh, D., Thapa, S., Dykema, A. G., Burman, P., Choudhury, B., Aparicio, L., Cheung, L. S., Lanis, M., Belcaid, Z., El Asmar, M., Illei, P. B., Wang, R., Meyers, J., Schuebel, K., Gupta, A., Skaist, A., Wheelan, S., Naidoo, J., Marrone, K. A., Brock, M., Ha, J., Bush, E. L., Park, B. J., Bott, M., Jones, D. R., Reuss, J. E., Velculescu, V. E., Chaft, J. E., Kinzler, K. W., Zhou, S., Vogelstein, B., Taube, J. M., Hellmann, M. D., Brahmer, J. R., Merghoub, T., Forde, P. M., Yegnasubramanian, S., Ji, H., Pardoll, D. M., & Smith, K. N. (2021). Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. *Nature*, *596*(7870), 126-132. https://doi.org/10.1038/s41586-021-03752-4

[50] Oliveira, G., Stromhaug, K., Klaeger, S., Kula, T., Frederick, D. T., Le, P. M., Forman, J., Huang, T., Li, S., Zhang, W., Xu, Q., Cieri, N., Clauser, K. R., Shukla, S. A., Neuberg, D., Justesen, S., MacBeath, G., Carr, S. A., Fritsch, E. F., Hacohen, N., Sade-Feldman, M., Livak, K. J., Boland, G. M., Ott, P. A., Keskin, D. B., & Wu, C. J. (2021). Phenotype, specificity and avidity of antitumour CD8(+) T cells in melanoma. *Nature*, *596*(7870), 119-125. https://doi.org/10.1038/s41586-021-03704-y

[51] Bordon, Y. (2019). TOX for tired T cells. *Nature Reviews Immunology*, *19*(8), 476-476. https://doi.org/10.1038/s41577-019-0193-9

[52] Sekine, T., Perez-Potti, A., Nguyen, S., Gorin, J.-B., Wu, V. H., Gostick, E., Llewellyn-Lacey, S., Hammer, Q., Falck-Jones, S., Vangeti, S., Yu, M., Smed-Sörensen, A., Gaballa, A., Uhlin, M., Sandberg, J. K., Brander, C., Nowak, P., Goepfert, P. A., Price, D. A., Betts, M. R., & Buggert, M. (2020). TOX is expressed by exhausted and polyfunctional human effector memory CD8+ T cells. *Science Immunology*, *5*(49), eaba7918. https://doi.org/doi:10.1126/sciimmunol.aba7918

[53] Ostroumov, D., Duong, S., Wingerath, J., Woller, N., Manns, M. P., Timrott, K., Kleine, M., Ramackers, W., Roessler, S., Nahnsen, S., immunoglobulin, tyrosiCzemmel, i., Dittrich-Breiholz, O., Eggert, T., Kühnel, F., & Wirth, T. C. (2021). Transcriptome Profiling Identifies TIGIT as a Marker of T-Cell Exhaustion in Liver Cancer. *Hepatology*, *73*(4), 1399-1418. https://doi.org/https://doi.org/10.1002/hep.31466

[54] Li, J., He, Y., Hao, J., Ni, L., & Dong, C. (2018). High Levels of Eomes Promote Exhaustion of Anti-tumor CD8+ T Cells [Original Research]. *Frontiers in Immunology*, *9*. https://doi.org/10.3389/fimmu.2018.02981

[55] Al-Abbasi, F. A., Mohammed, K., Sadath, S., Banaganapalli, B., Nasser, K., & Shaik, N. A. (2018). Computational Protein Phenotype Characterization of IL10RA Mutations Causative to Early Onset Inflammatory Bowel Disease (IBD) [Original Research]. *Frontiers in Genetics*, *9*. https://doi.org/10.3389/fgene.2018.00146

[56] Li, Y., Yang, S., Yue, H., Yuan, D., Li, L., Zhao, J., & Zhao, L. (2020). Unraveling LGALS1 as a Potential Immune Checkpoint and a Predictor of the Response to Anti-PD1 Therapy in Clear Cell Renal Carcinoma. *Pathology & Oncology Research*, *26*(3), 1451-1458. https://doi.org/10.1007/s12253-019-00710-4

[57] Saleh, R., Taha, R. Z., Toor, S. M., Sasidharan Nair, V., Murshed, K., Khawar, M., Al-Dhaheri, M., Petkar, M. A., Abu Nada, M., & Elkord, E. (2020). Expression of immune checkpoints and T cell exhaustion markers in early and advanced stages of colorectal cancer. *Cancer Immunology, Immunotherapy*, *69*(10), 1989-1999. https://doi.org/10.1007/s00262-020-02593-w

[58] Zheng, L., Qin, S., Si, W., Wang, A., Xing, B., Gao, R., Ren, X., Wang, L., Wu, X., Zhang, J., Wu, N., Zhang, N., Zheng, H., Ouyang, H., Chen, K., Bu, Z., Hu, X., Ji, J., & Zhang, Z. (2021). Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science*, *374*(6574), abe6474. https://doi.org/10.1126/science.abe6474

[59] Seddiki , N., Santner-Nanan , B., Martinson , J., Zaunders , J., Sasson , S., Landay , A., Solomon , M., Selby , W., Alexander , S. I., Nanan , R., Kelleher , A., & de St. Groth , B. F. (2006). Expression of interleukin (IL)-2 and IL-7 receptors discriminates between human regulatory and activated T cells. *Journal of Experimental Medicine*, *203*(7), 1693-1700. https://doi.org/10.1084/jem.20060468

[60] Petkau, G., Mitchell, T. J., Chakraborty, K., Bell, S. E., D'Angeli, V., Matheson, L., Turner, D. J., Salerno, F., Katsikis, P. D., & Turner, M. (2021). The timing of differentiation and potency of CD8 effector function is set by RNA binding proteins. *bioRxiv*, 2021.2006.2003.446738. https://doi.org/10.1101/2021.06.03.446738

[61] Ordway, D., Higgins, D. M., Sanchez-Campillo, J., Spencer, J. S., Henao-Tamayo, M., Harton, M., Orme, I. M., & Gonzalez Juarrero, M. (2007). XCL1 (lymphotactin) chemokine produced by activated CD8 T cells during the chronic stage of infection with Mycobacterium tuberculosis negatively affects production of IFN-γ by CD4 T cells and participates in granuloma stability. *Journal of Leukocyte Biology*, *82*(5), 1221-1229. https://doi.org/https://doi.org/10.1189/jlb.0607426

[62] Wang, B., Wang, Y., Sun, X., Deng, G., Huang, W., Wu, X., Gu, Y., Tian, Z., Fan, Z., Xu, Q., Chen, H., & Sun, Y. (2021). CXCR6 is required for antitumor efficacy of intratumoral CD8+ T cell. *Journal for ImmunoTherapy of Cancer*, *9*(8), e003100. https://doi.org/10.1136/jitc-2021-003100

[63] Liu, Z., Huo, X., Zhao, S., Yang, J., Shi, W., Jing, L., Li, W., Li, Y., Ma, L., Gao, Y., & Diao, A. (2017). Low density lipoprotein receptor class A domain containing 4 (LDLRAD4) promotes tumorigenesis of hepatic cancer cells. *Exp Cell Res*, *360*(2), 189-198. https://doi.org/10.1016/j.yexcr.2017.09.005

[64] Wei, J., Feng, L., & Wu, L. (2020). Integrated analysis identified CAPG as a prognosis factor correlated with immune infiltrates in lower-grade glioma. *Clinical and translational medicine*, *10*(2), e51-e51. https://doi.org/10.1002/ctm2.51

[65] Bagchi, S., Yuan, R., & Engleman, E. G. (2021). Immune Checkpoint Inhibitors for the Treatment of Cancer: Clinical Impact and Mechanisms of Response and Resistance. *Annu Rev Pathol*, *16*, 223-249. https://doi.org/10.1146/annurev-pathol-042020-042741

[66] Jenkins, R. W., Barbie, D. A., & Flaherty, K. T. (2018). Mechanisms of resistance to immune checkpoint inhibitors. *British*

Journal of Cancer, *118*(1), 9-16. https://doi.org/10.1038/bjc.2017.434

[67] Snyder, A., Nathanson, T., Funt, S. A., Ahuja, A., Buros Novik, J., Hellmann, M. D., Chang, E., Aksoy, B. A., Al-Ahmadie, H., Yusko, E., Vignali, M., Benzeno, S., Boyd, M., Moran, M., Iyer, G., Robins, H. S., Mardis, E. R., Merghoub, T., Hammerbacher, J., Rosenberg, J. E., & Bajorin, D. F. (2017). Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis. *PLoS Med*, *14*(5), e1002309. https://doi.org/10.1371/journal.pmed.1002309

[68] Yusko, E., Vignali, M., Wilson, R. K., Mardis, E. R., Hodi, F. S., Horak, C., Chang, H., Woods, D. M., Robins, H., & Weber, J. (2019). Association of Tumor Microenvironment T-cell Repertoire and Mutational Load with Clinical Outcome after Sequential Checkpoint Blockade in Melanoma. *Cancer Immunol Res*, *7*(3), 458-465. https://doi.org/10.1158/2326-6066.CIR-18-0226

[69] Duncan, G. S., Andrew, D. P., Takimoto, H., Kaufman, S. A., Yoshida, H., Spellberg, J., Luis de la Pompa, J., Elia, A., Wakeham, A., Karan-Tamir, B., Muller, W. A., Senaldi, G., Zukowski, M. M., & Mak, T. W. (1999). Genetic Evidence for Functional Redundancy of Platelet/Endothelial Cell Adhesion Molecule-1 (PECAM-1): CD31-Deficient Mice Reveal PECAM-1-Dependent and PECAM-1-Independent Functions. *The Journal of Immunology*, *162*(5), 3022-3030. https://www.jimmunol.org/content/jimmunol/162/5/3022.full.pdf

[70] Chambon, P. (1994). The retinoid signaling pathway: molecular and genetic analyses. *Seminars in Cell Biology*, *5*(2), 115-125. https://doi.org/https://doi.org/10.1006/scel.1994.1015

[71] Dombrowski, K. E., & Wright, S. E. (1992). Construction of a multiple mucin tandem repeat with a mutation in the tumor-specific epitope by a solid-phase gene assembly protocol. *Nucleic Acids Res*, *20*(24), 6743-6744. https://doi.org/10.1093/nar/20.24.6743

[72] Huisman, W., Hageman, L., Leboux, D. A. T., Khmelevskaya, A., Efimov, G. A., Roex, M. C. J., Amsen, D., Falkenburg, J. H. F., & Jedema, I. (2022). Public T-Cell Receptors (TCRs) Revisited by Analysis of the Magnitude of Identical and Highly-Similar TCRs in Virus-Specific T-Cell Repertoires of Healthy Individuals [Original Research]. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/fimmu.2022.851868

[73] Li, H., Ye, C., Ji, G., & Han, J. (2012). Determinants of public T cell responses. *Cell research*, *22*(1), 33-42. https://doi.org/10.1038/cr.2012.1

[74] Wang, C., Sanders, C. M., Yang, Q., Schroeder, H. W., Jr., Wang, E., Babrzadeh, F., Gharizadeh, B., Myers, R. M., Hudson, J. R., Jr., Davis, R. W., & Han, J. (2010). High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc Natl Acad Sci U S A*, *107*(4), 1518-1523. https://doi.org/10.1073/pnas.0913939107

[75] Emerson, R., Sherwood, A., Desmarais, C., Malhotra, S., Phippard, D., & Robins, H. (2013). Estimating the ratio of CD4+ to CD8+ T cells using high-throughput sequence data. *Journal of Immunological Methods*, *391*(1), 14-21. https://doi.org/https://doi.org/10.1016/j.jim.2013.02.002

[76] Nakatsugawa, M., Rahman, M. A., Yamashita, Y., Ochi, T., Wnuk, P., Tanaka, S., Chamoto, K., Kagoya, Y., Saso, K., Guo, T., Anczurowski, M., Butler, M. O., & Hirano, N. (2016). CD4+ and CD8+ TCRβ repertoires possess different potentials to generate extraordinarily high-avidity T cells. *Scientific Reports*, *6*(1), 23821. https://doi.org/10.1038/srep23821

[77] Dustin, M. L., & Long, E. O. (2010). Cytotoxic immunological synapses. *Immunol Rev*, *235*(1), 24-34. https://doi.org/10.1111/j.0105-2896.2010.00904.x

[78] Halle, S., Halle, O., & Förster, R. (2017). Mechanisms and Dynamics of T Cell-Mediated Cytotoxicity In Vivo. *Trends in Immunology*, *38*(6), 432-443. https://doi.org/https://doi.org/10.1016/j.it.2017.04.002

[79] Wan, Y. Y. (2010). Multi-tasking of helper T cells. *Immunology*, *130*(2), 166-171. https://doi.org/https://doi.org/10.1111/j.1365-2567.2010.03289.x

[80] London, C. A., Abbas, A. K., & Kelso, A. (1998). Helper T cell subsets: Heterogeneity, functions and development. *Veterinary Immunology and Immunopathology*, *63*(1), 37-44. https://doi.org/https://doi.org/10.1016/S0165-2427(98)00080-4

[81] Page, D. B., Yuan, J., Redmond, D., Wen, Y. H., Durack, J. C., Emerson, R., Solomon, S., Dong, Z., Wong, P., Comstock, C., Diab, A., Sung, J., Maybody, M., Morris, E., Brogi, E., Morrow, M., Sacchini, V., Elemento, O., Robins, H., Patil, S., Allison, J. P., Wolchok, J. D., Hudis, C., Norton, L., & McArthur, H. L. (2016). Deep Sequencing of T-cell Receptor DNA as a Biomarker of Clonally Expanded TILs in Breast Cancer after Immunotherapy. *Cancer Immunology Research*, *4*(10), 835-

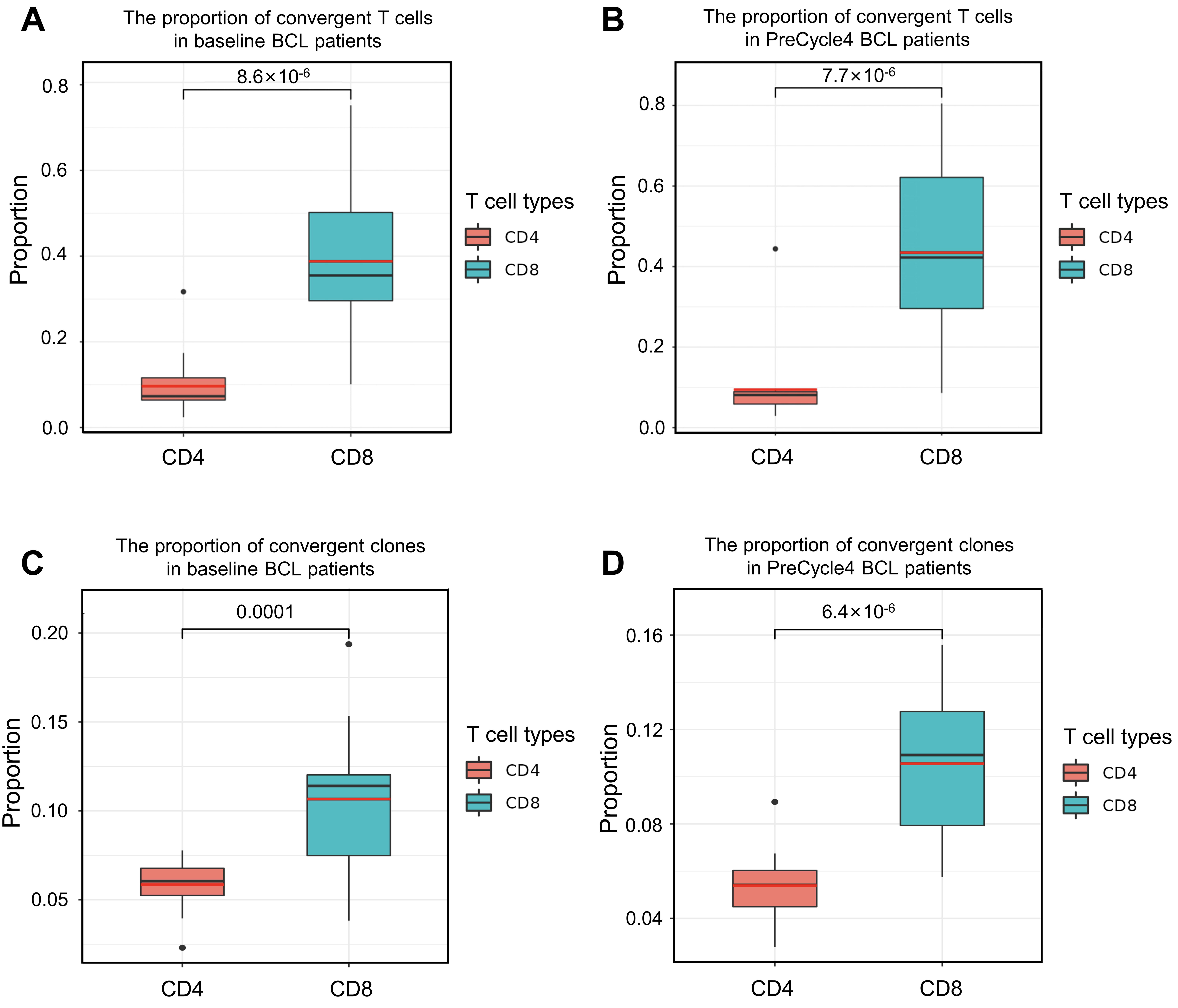
844. https://doi.org/10.1158/2326-6066.Cir-16-0013

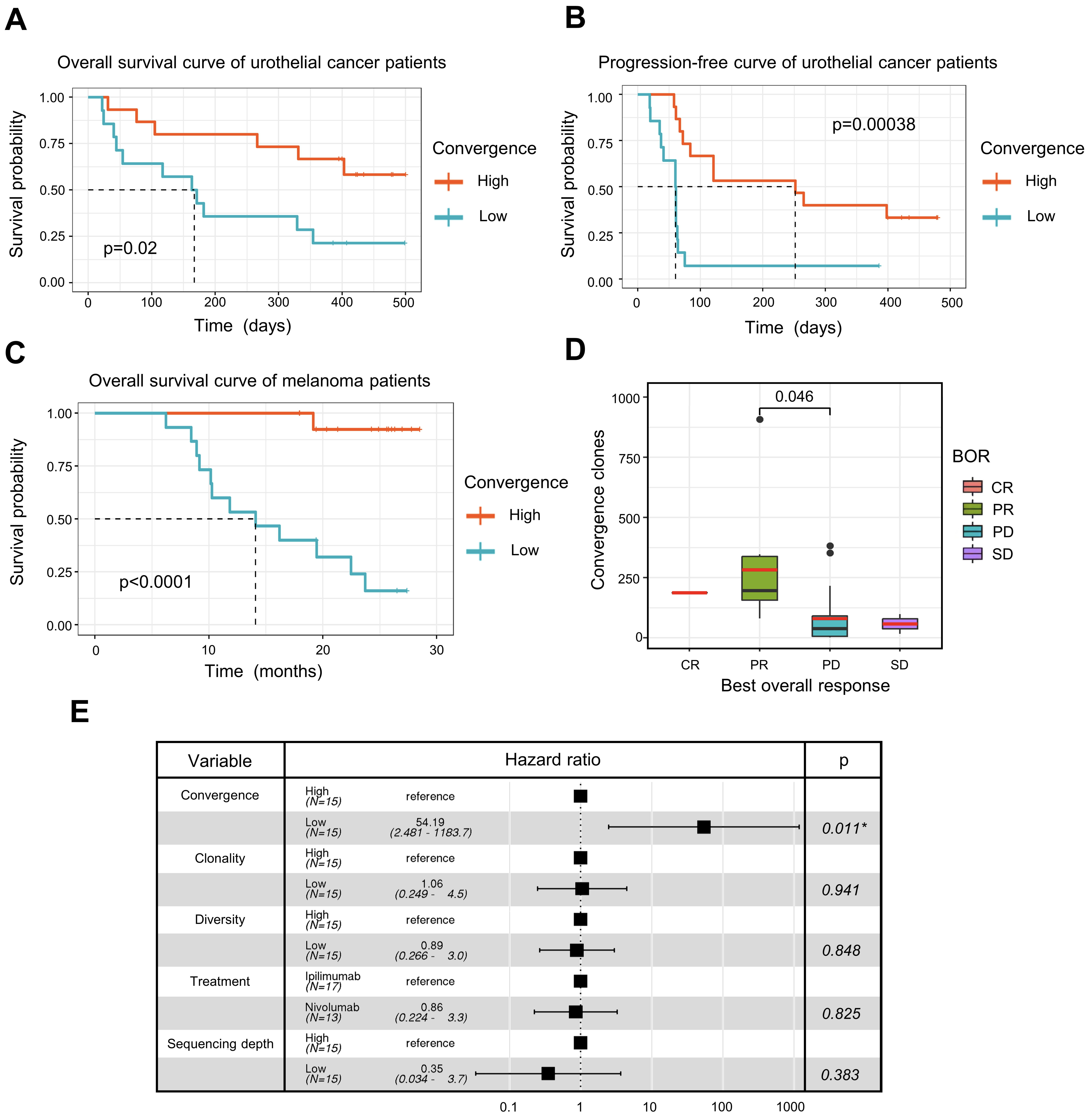
[82] Postow, M. A., Manuel, M., Wong, P., Yuan, J., Dong, Z., Liu, C., Perez, S., Tanneau, I., Noel, M., Courtier, A., Pasqual, N., & Wolchok, J. D. (2015). Peripheral T cell receptor diversity is associated with clinical outcomes following ipilimumab treatment in metastatic melanoma. *Journal for ImmunoTherapy of Cancer*, *3*(1), 23. https://doi.org/10.1186/s40425-015-0070-4

[83] Hopkins, A. C., Yarchoan, M., Durham, J. N., Yusko, E. C., Rytlewski, J. A., Robins, H. S., Laheru, D. A., Le, D. T., Lutz, E. R., & Jaffee, E. M. (2018). T cell receptor repertoire features associated with survival in immunotherapy-treated pancreatic ductal adenocarcinoma. *JCI insight*, *3*(13), e122092. https://doi.org/10.1172/jci.insight.122092

[84] Roh, W., Chen, P. L., Reuben, A., Spencer, C. N., Prieto, P. A., Miller, J. P., Gopalakrishnan, V., Wang, F., Cooper, Z. A., Reddy, S. M., Gumbs, C., Little, L., Chang, Q., Chen, W. S., Wani, K., De Macedo, M. P., Chen, E., Austin-Breneman, J. L., Jiang, H., Roszik, J., Tetzlaff, M. T., Davies, M. A., Gershenwald, J. E., Tawbi, H., Lazar, A. J., Hwu, P., Hwu, W. J., Diab, A., Glitza, I. C., Patel, S. P., Woodman, S. E., Amaria, R. N., Prieto, V. G., Hu, J., Sharma, P., Allison, J. P., Chin, L., Zhang, J., Wargo, J. A., & Futreal, P. A. (2017). Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med*, *9*(379). https://doi.org/10.1126/scitranslmed.aah3560

A
Cell type
T.na
T.Treg
T.Th1
T.rTreg
Type 1 helper
Regulatory
Recruited Treg
Naïve
tSNE_2
tSNE_1
40
20
0
-20
-40
-20
0
20
Convergence
True
False
tSNE_2
tSNE_1
B
Fisher's exact test, p-value = < 0.001
odds ratio=4.410693
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
2%
98%
9%
91%
(n = 8,664)
(n = 479)
False
True
Cluster 04
Tetramer
True
False
Fisher's exact test, p-value = 1
odds ratio=0.8164083
2%
98%
2%
98%
(n = 8,942)
(n = 201)
False
True
Cluster 05
Tetramer
True
False
C
Cluster
01
02
03
04
05
06
07
08
09
10
11
12
13
14
tSNE_2
tSNE_1
30
20
10
0
-10
-20
-20
-10
0
10
20
30
Convergence
True
False
tSNE_2
tSNE_1

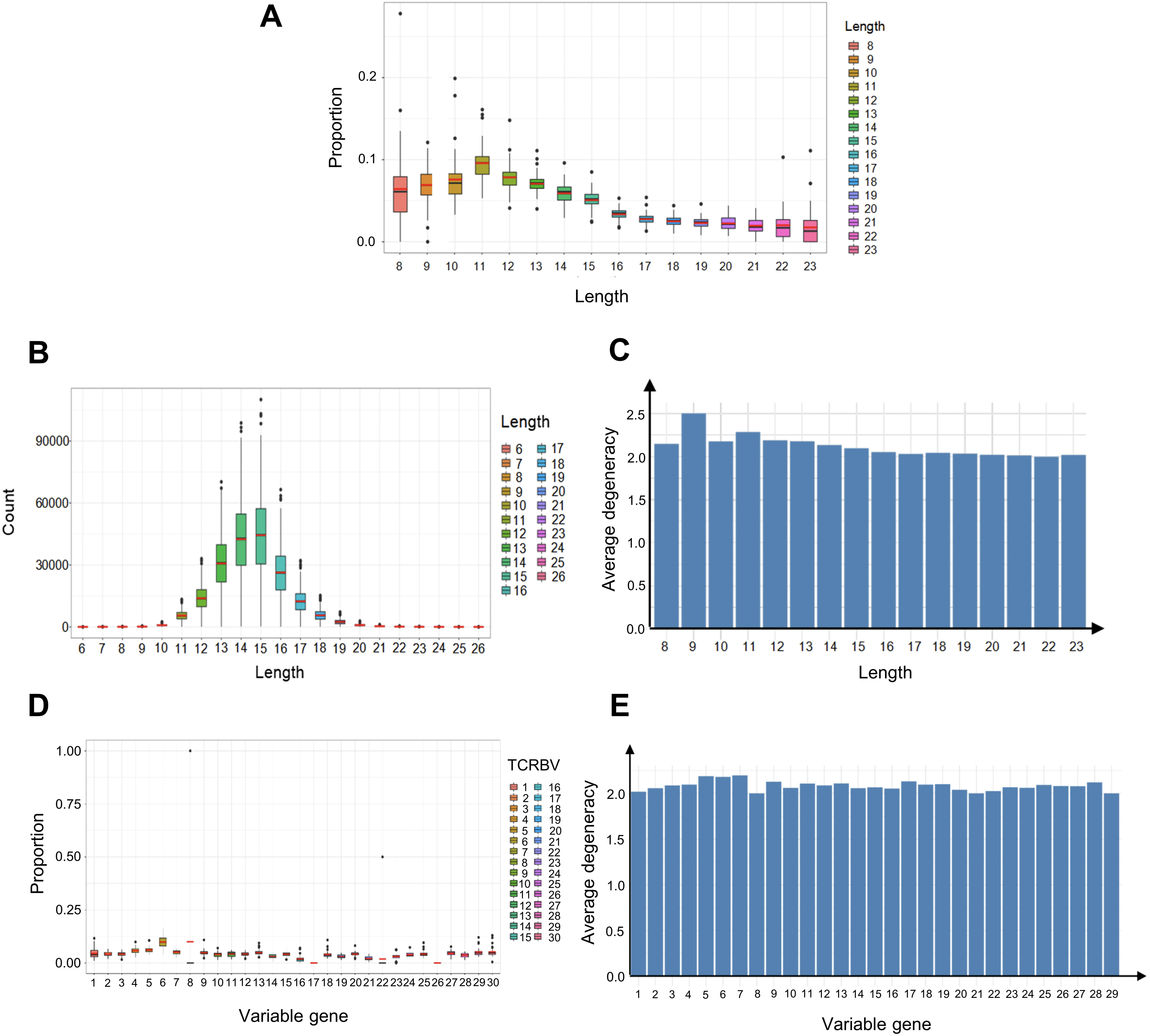

A
The proportion of convergent T cells in baseline BCL patients
8.6×10^-6
Proportion
0.0
0.2
0.4
0.6
0.8
CD4
CD8
T cell types
CD4
CD8
B
The proportion of convergent T cells in PreCycle4 BCL patients
7.7×10^-6
Proportion
0.0
0.2
0.4
0.6
0.8
CD4
CD8
T cell types
CD4
CD8
C
The proportion of convergent clones in baseline BCL patients
0.0001
Proportion
0.05
0.10
0.15
0.20
CD4
CD8
T cell types
CD4
CD8
D
The proportion of convergent clones in PreCycle4 BCL patients
6.4×10^-6
Proportion
0.04
0.08
0.12
0.16
CD4
CD8
T cell types
CD4
CD8

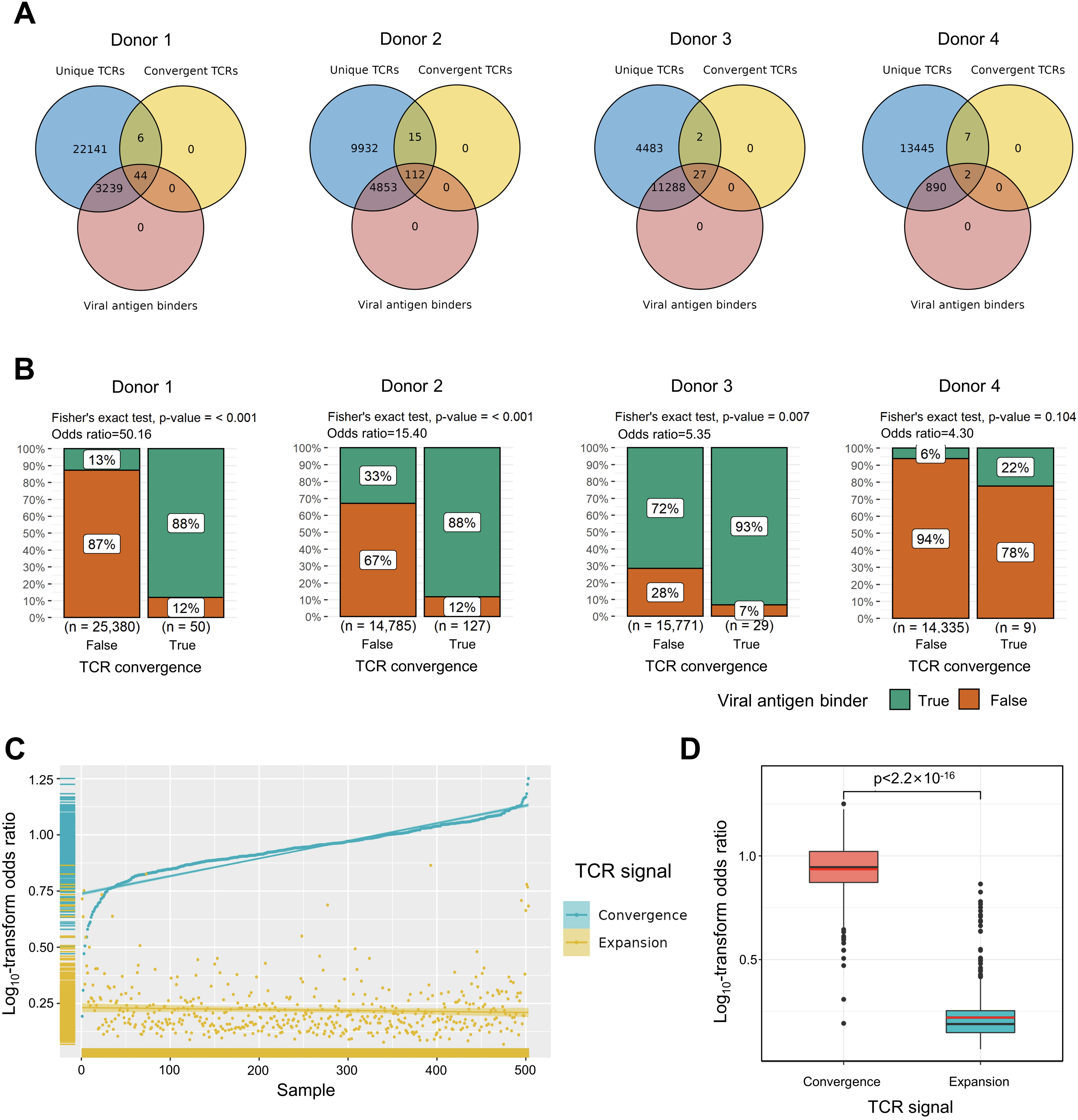


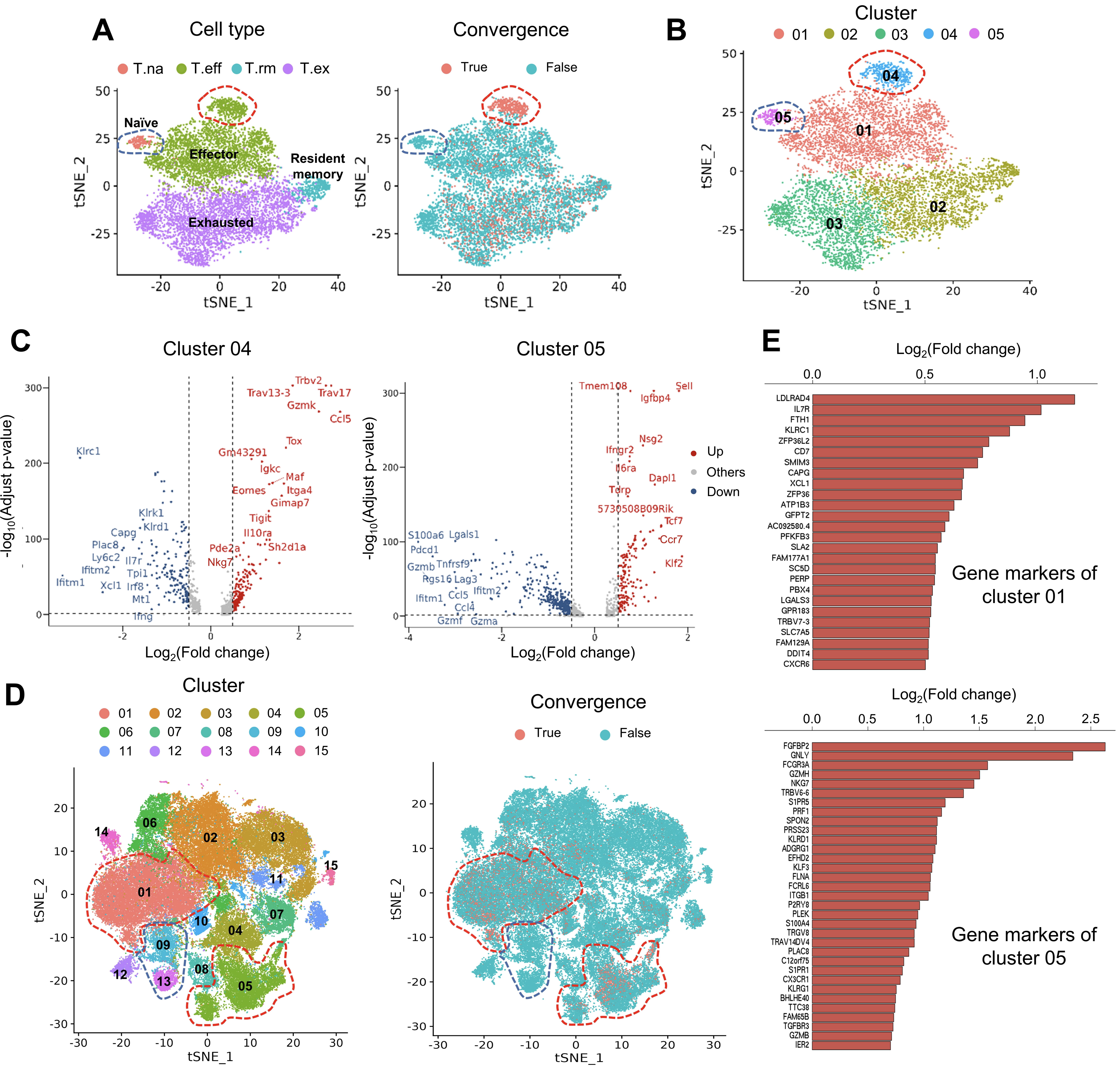
Variable	Hazard ratio		p
Convergence	High *(N=15)*	reference	
	Low *(N=15)*	54.19 *(2.481 - 1183.7)*	*0.011**
Clonality	High *(N=15)*	reference	
	Low *(N=15)*	1.06 *(0.249 - 4.5)*	*0.941*
Diversity	High *(N=15)*	reference	
	Low *(N=15)*	0.89 *(0.266 - 3.0)*	*0.848*
Treatment	Ipilimumab *(N=17)*	reference	
	Nivolumab *(N=13)*	0.86 *(0.224 - 3.3)*	*0.825*
Sequencing depth	High *(N=15)*	reference	
	Low *(N=15)*	0.35 *(0.034 - 3.7)*	*0.383*

A
Variable
Hazard ratio
p
Convergence
High (N=15)
reference
Low (N=14)
6.290 (1.0418 - 37.97)
0.045*
Clonality
High (N=15)
reference
Low (N=14)
0.055 (0.0042 - 0.71)
0.026*
Diversity
High (N=15)
reference
Low (N=14)
0.177 (0.0170 - 1.85)
0.148
Sequencing depth
High (N=15)
reference
Low (N=14)
0.281 (0.0518 - 1.52)
0.14
0.005
0.1
0.05
0.1
0.5
1
5
10
B
Variable
Hazard ratio
p
Convergence
High (N=15)
reference
Low (N=14)
17.201 (2.7394 - 108.0)
0.002**
Clonality
High (N=15)
reference
Low (N=14)
0.051 (0.0037 - 0.7)
0.026*
Diversity
High (N=15)
reference
Low (N=14)
0.109 (0.0088 - 1.3)
0.084
Sequencing depth
High (N=15)
reference
Low (N=14)
0.216 (0.0394 - 1.2)
0.078
0.01
0.1
1
10
100

A
Translation
TCR convergent cluster
T cells with different TCR nucleotide sequences
TCRs with the identical amino acid sequence
B
Log10-transform Counts
Degeneracy
D1 6.96
D2 5.67
D3 4.57
D4 3.86
D5 3.36
D6 2.98
D7 2.64
D8 2.36
D9 2.04
D10 1.82
0
2
4
6
C
Nonpolar hydrophobic amino acid
Polar neutral amino acid
Acidic amino acid
Alkaline amino acid
Proportion
0.0
0.1
0.2
G A V L I P F
Y W S T C M N Q
D E
K R H
D1 D2 D3 D4
D
TMB (per MB)
100
10
1
COVID-19 recovered
Healthy
Ovary
Kidney
Lung
Melanoma
Blood
Public TCRs Proportion
0.2
0.4
0.6
0.64
5.7×10-16


A
Proportion
0.0
0.1
0.2
Length
8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Length
8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
B
Count
0
30000
60000
90000
Length
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Length
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
C
Average degeneracy
0.0 0.5 1.0 1.5 2.0 2.5
Length
8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
D
Proportion
0.00 0.25 0.50 0.75 1.00
Variable gene
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
TCRBV
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
E
Average degeneracy
0.0 0.5 1.0 1.5 2.0
Variable gene
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29


A
Donor 1
Unique TCRs
Convergent TCRs
22141
6
0
44
3239
0
0
Viral antigen binders
Donor 2
Unique TCRs
Convergent TCRs
9932
15
0
112
4853
0
0
Viral antigen binders
Donor 3
Unique TCRs
Convergent TCRs
4483
2
0
27
11288
0
0
Viral antigen binders
Donor 4
Unique TCRs
Convergent TCRs
13445
7
0
2
890
0
0
Viral antigen binders
B
Donor 1
Fisher's exact test, p-value = < 0.001
Odds ratio=50.16
13%
87%
88%
12%
(n = 25,380)
(n = 50)
False
True
TCR convergence
Donor 2
Fisher's exact test, p-value = < 0.001
Odds ratio=15.40
33%
67%
88%
12%
(n = 14,785)
(n = 127)
False
True
TCR convergence
Donor 3
Fisher's exact test, p-value = 0.007
Odds ratio=5.35
72%
28%
93%
7%
(n = 15,771)
(n = 29)
False
True
TCR convergence
Donor 4
Fisher's exact test, p-value = 0.104
Odds ratio=4.30
6%
94%
22%
78%
(n = 14,335)
(n = 9)
False
True
TCR convergence
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Viral antigen binder
True
False
C
Log$_{10}$-transform odds ratio
1.25
1.00
0.75
0.50
0.25
0
100
200
300
400
500
Sample
TCR signal
Convergence
Expansion
D
p<2.2×10^{-16}
Log$_{10}$-transform odds ratio
1.0
0.5
Convergence
Expansion
TCR signal


A
Cell type
T.na
T.eff
T.rm
T.ex
Naïve
Effector
Resident memory
Exhausted
tSNE_2
tSNE_1
50
25
0
-25
-20
0
20
40
Convergence
True
False
B
Cluster
01
02
03
04
05
C
Cluster 04
-log10(Adjust p-value)
Log2(Fold change)
Klrc1
Klrk1
Klrd1
Capg
Plac8
Ly6c2
Il7r
Ifitm2
Tpi1
Ifitm1
Xcl1
Irf8
Mt1
Ifng
Trbv2
Trav13-3
Trav17
Gzmk
Ccl5
Tox
Gm43291
Igkc
Maf
Eomes
Itga4
Gimap7
Tigit
Il10ra
Pde2a
Sh2d1a
Nkg7
Cluster 05
Tmem108
Sell
Igfbp4
Nsg2
Ifngr2
Il6ra
Dapl1
Tdrp
5730508B09Rik
Tcf7
Ccr7
Klf2
S100a6
Lgals1
Pdcd1
Gzmb
Tnfrsf9
Rgs16
Lag3
Ifitm2
Ifitm1
Ccl5
Ccl4
Gzmf
Gzma
Up
Others
Down
300
200
100
0
-4
-2
0
2
D
Cluster
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
Convergence
True
False
-30
-20
-10
0
10
20
30
E
Log2(Fold change)
0.0
0.5
1.0
LDLRAD4
IL7R
FTH1
KLRC1
ZFP36L2
CD7
SMIM3
CAPG
XCL1
ZFP36
ATP1B3
GFPT2
AC092580.4
PFKFB3
SLA2
FAM177A1
SC5D
PERP
PBX4
LGALS3
GPR183
TRBV7-3
SLC7A5
FAM129A
DDIT4
CXCR6
Gene markers of cluster 01
Log2(Fold change)
0.0
0.5
1.0
1.5
2.0
2.5
FGFBP2
GNLY
FCGR3A
GZMH
NKG7
TRBV6-6
S1PR5
PRF1
SPON2
PRSS23
KLRD1
ADGRG1
EFHD2
KLF3
FLNA
FCRL6
ITGB1
P2RY8
PLEK
S100A4
TRGV8
TRAV14DV4
PLAC8
C12orf75
S1PR1
CX3CR1
KLRG1
BHLHE40
TTC38
FAM65B
TGFBR3
GZMB
IER2
Gene markers of cluster 05