



# A Pan-Cancer Analysis of Tumor-Infiltrating B Cell Repertoires

Katharine Yu<sup>1,2</sup>, Akshay Ravor<sup>1</sup>, Núria Malats<sup>3</sup>, Silvia Pineda<sup>3†‡</sup> and Marina Sirota<sup>1,2\*†</sup>

<sup>1</sup> Bakar Computational Health Sciences Institute, University of California, San Francisco (UCSF), San Francisco, CA, United States, <sup>2</sup> Department of Pediatrics, University of California, San Francisco (UCSF), San Francisco, CA, United States, <sup>3</sup> Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre (CNIO), and Centro de Investigación Biomédica en Red Cáncer (CIBERONC), Madrid, Spain

## OPEN ACCESS

### Edited by:

Mohammad Hojjat-Farsangi,  
Karolinska Institutet (KI), Sweden

### Reviewed by:

Yaron Carmi,  
Tel Aviv University, Israel  
Sergei Mangul,  
University of Southern California,  
United States

### \*Correspondence:

Marina Sirota  
Marina.Sirota@ucsf.edu

### †Present Address:

Silvia Pineda,  
Department of Statistics and Data  
Science, Complutense University of  
Madrid, Madrid, Spain

‡These authors have contributed  
equally to this work and share  
senior authorship

### Specialty section:

This article was submitted to  
Cancer Immunity  
and Immunotherapy,  
a section of the journal  
Frontiers in Immunology

**Received:** 06 October 2021

**Accepted:** 06 December 2021

**Published:** 05 January 2022

### Citation:

Yu K, Ravor A, Malats N,  
Pineda S and Sirota M (2022)  
A Pan-Cancer Analysis of Tumor-  
Infiltrating B Cell Repertoires.  
*Front. Immunol.* 12:790119.  
doi: 10.3389/fimmu.2021.790119

Tumor-infiltrating B cells can play an important role in anti-tumor responses but their presence is not well understood. In this study, we extracted the B cell receptor repertoires from 9522 tumor and adjacent non-tumor samples across 28 tumor types in the Cancer Genome Atlas project and performed diversity and network analysis. We identified differences in diversity and network statistics across tumor types and subtypes and observed a trend towards increased clonality in primary tumors compared to adjacent non-tumor tissues. We also found significant associations between the repertoire features and mutation load, tumor stage, and age. Our V-gene usage analysis identified similar V-gene usage patterns in colorectal and endometrial cancers. Lastly, we evaluated the prognostic value of the repertoire features and identified significant associations with survival in seven tumor types. This study warrants further research into better understanding the role of tumor-infiltrating B cells across a wide range of tumor types.

**Keywords:** B cell repertoire, immune repertoire, tumor microenvironment, tumor infiltration, TCGA

## 1 INTRODUCTION

While B cells are well-established as an integral part of the adaptive immune system, only recently studies began to elucidate their role in cancer (1, 2). The number of studies on tumor-infiltrating B cells is vastly eclipsed by the number of studies on tumor-infiltrating T cells, the latter of which have largely been the focus of researchers and play a central role in modern immunotherapies such as checkpoint inhibitors. However, B cells hold great potential for the development of new immunotherapies and as biomarkers for immunotherapy response.

A main function of B cells is to recognize specific antigens with the immunoglobulins (Ig), or B cell receptors (BCR), on their cell surface. These Ig are made up of two heavy chains (IGH) and two light chains, the kappa ( $\kappa$ ) chains (IGK) or the lambda ( $\lambda$ ) chains (IGL). Ig are generated through a process called somatic recombination where variable (V), diversity (D), and joining (J) gene segments are randomly combined to create a diverse collection of antigen receptors which can recognize a wide range of antigens. Additionally, B cells undergo a process called somatic hypermutation (SHM) upon antigen binding which introduces additional mutations into the variable regions of the Ig genes, further diversifying the receptors.

The collection of diverse B cell receptors within an individual, or the B cell repertoire, can be interrogated using high-throughput technologies. B-cell receptor sequencing (BCR-seq) is commonly used to study the B cell repertoire as it offers greater sensitivity compared to

unselected RNA-seq by targeting the BCR region rather than the entire transcriptome. However, the amount of publicly available RNA-seq datasets is much greater than the number of BCR-seq datasets. Tools such as MiXCR (3), ImReP (4), and TRUST4 (5) have been developed to extract BCR reads from bulk RNA-seq data and align them to the V, D, and J gene segments, allowing for the characterization of the immune repertoire from sequencing data. These tools have been especially useful in mining publicly available datasets to extract insight into the adaptive immune system (6, 7).

The Cancer Genome Atlas (TCGA) is the largest publicly available dataset of molecularly characterized human tumors (8). The data generated by the TCGA includes clinical, transcriptomic, methylation, mutation, copy number, and proteomics data. This dataset has greatly advanced our understanding of tumor biology and has led to improvements in cancer diagnosis, treatment, and prevention (9–12). More recently, studies have leveraged the TCGA dataset to investigate the role of the immune system in cancer (7). However, the previous analysis of the B cell repertoires tends to be limited in scope and lacks tumor subtype stratification. For example, Thorsson et al. only analyzed the impact of TCR diversity on prognostic associations and their BCR analysis was restricted to comparing BCR diversity across tumor types and the immune subtypes identified in the paper.

Characterization of the tumor microenvironment is vital for understanding cancer biology and developing new immunotherapies as well as predicting which patients will respond to immunotherapies. B cells, in particular, can play an important role in the antitumor immune response. They can produce antibodies which can drive antibody-dependent cellular cytotoxicity and phagocytosis of tumor cells (13) and they can also present antigens to T cells and may be involved in the formation of tumor-associated tertiary lymphoid structures (14, 15). However, the presence of tumor-infiltrating B cells has also been associated with poor outcome in renal cell carcinoma (16), bladder cancer (17), prostate cancer (18), suggesting that B cells play a complex role in the tumor microenvironment. Further studies are needed to better understand how tumor-infiltrating B cells behave in different tumor contexts.

We extracted the B cell repertoires from 28 tumor types in the TCGA dataset for nearly 10,000 samples and performed diversity and network analysis to investigate the immunological differences and commonalities across tumor types and subtypes, and between tumors and adjacent non-tumor tissue. We then compared these B cell repertoire features to host, clinical, and molecular features and found significant associations with age, tumor stage, and mutation load, respectively. In our V-gene analysis, we found similar V-gene usage patterns in colorectal and endometrial cancers. Lastly, we investigated the prognostic value of each repertoire feature and found significant associations with survival in a subset of tumor types.

## 2 METHODS

### 2.1 Data Acquisition

We used the GDC Data Transfer Tool (<https://gdc.cancer.gov/access-data/gdc-data-transfer-tool>) to download every available

TCGA RNA-Seq FASTQ file from the GDC Legacy Archive (<https://portal.gdc.cancer.gov/legacy-archive/search/f>). We then used MiXCR to extract the reads that align to the VDJ region of the IGH, IGK, IGL, TRA, TRB, TRD, and TRG chains using the MiXCR pipeline for processing RNA-seq and non-targeted genomic data (<https://mixcr.readthedocs.io/en/master/rnaseq.html>). We filtered out reads with missing CDR3 sequences. We used the R package GenomicDataCommons (19) to annotate the samples with their TCGA barcodes and extracted the sample type from the TCGA barcode. We then filtered for primary tumor samples and adjacent non-tumor samples for all the tumor types except for SKCM, where we included metastatic samples as well. If there were multiple vials from the same tumor sample, we selected the first vial (e.g. -01A).

We downloaded the TCGA clinical data from the TCGA Pan-Cancer Atlas Hub hosted by the UCSC Xena platform (<https://pancanatlas.xenahubs.net>). We downloaded the leukocyte fraction data and mutation load data from the PanCanAtlas Publications page on GDC for The Immune Landscape of Cancer (<https://gdc.cancer.gov/about-data/publications/panimmune>). We used the R package TCGAblinks (20) to download the TCGA subtype data and we used the “Subtype\_Selected” column for the subtype information if there were multiple subtype classifications.

### 2.2 Expression and Diversity Analysis

We calculated BCR expression by dividing the number of reads that align to each IGH, IGK, or IGL chain by the total number of reads in each sample. We defined clones as groups of reads that share the same V and J genes, the same CDR3 length, and at least 90% shared nucleotide identity. As it is not possible to perform paired heavy and light chain analysis with bulk RNA-seq data, we analyzed the IGH, IGK, and IGL chains separately for the clonal analysis. To quantify the clonal diversity, we calculated Shannon entropy ( $H$ ) using the following formula:

$$H = -\sum_{i=1}^N p_i \log_2 p_i$$

$N$  is the number of unique clones in the sample and  $p_i$  is the proportion of clone  $i$  in the sample. Shannon entropy can range from 0, for samples with only one clone, to  $\log_2 N$ , for samples with a uniform distribution of clones. We then calculated the evenness of each sample using Pielou’s evenness index, which is:

$$J = \frac{H}{H_{max}}$$

where  $H$  is the Shannon entropy and  $H_{max}$  is the maximum possible value for  $H$ . Evenness is constrained between 0 and 1 and a higher evenness value indicates a more even distribution of clones.

### 2.3 Network Analysis

We generated networks for each sample using a previously published method (21, 22). Each vertex in the network is a unique BCR sequence and the size of the vertex corresponds to the number of reads with that sequence. Edges are drawn

between vertexes with the same V and J genes, the same CDR3 length, and at least 90% sequence similarity (our clone definition). We used the R package *igraph* to generate network plots for each sample.

We used the Gini index to quantify different repertoire network parameters. The Gini index measures the inequality in a frequency distribution and it ranges from 0, which indicates complete equality, to 1, which indicates complete inequality. We quantified clonal expansion by calculating the Gini index using the distribution of vertex sizes for each sample (vertex Gini index). This measures the unevenness in the number of unique BCR sequences and having a higher vertex Gini index indicates more clonal expansion in a sample. We quantified clonal diversification by calculating the Gini index using the distribution of the number of vertexes in each cluster (cluster Gini index). Having a higher cluster Gini index indicates a sample with expanded cluster sizes.

For the downsampling analysis, we randomly sampled 500 reads each for IGH, IGK, and IGL and then calculated the vertex Gini index and the cluster Gini index using these subsamples. We repeated this procedure 10 times and then took the mean for the final analysis presented in the paper.

## 2.4 V-Gene Analysis

To analyze V-gene usage, we first filtered out samples with fewer than 100 clones. We then calculated V-gene usage as the percent of clones in each sample which uses a particular V-gene. We used PCA to reduce the dimensionality of the data and we visualized the samples in PC1 and PC2 to identify clusters of interest. Next, we applied an elastic net model to identify the genes that are associated with these clusters of interest. We used 5-fold cross validation to test different alpha values (0.1-0.9) and selected the alpha value with the lowest mean cross validated error for the final elastic net model.

## 2.5 Statistical Analysis

We used R version 4.3 to perform the statistical analysis and generate the figures in this paper. For the association analysis between B cell repertoire features and clinical and tumor features, we used Spearman's correlation for continuous variables and the Wilcoxon rank-sum test for categorical variables. We used the R package *survival* to perform the Cox regression analysis and we used the R package *survminer* to generate the Kaplan-Meier plots. For analysis involving multiple B cell repertoire features for each tumor type, we adjusted for multiple comparisons within tumor types using the Benjamini-Hochberg procedure.

## 3 RESULTS

### 3.1 Study Overview

We analyzed the B cell repertoires across the TCGA tumor samples corresponding to 28 tumor types with a total of 8854 tumor and 688 adjacent non-tumor samples (**Figure 1A** and **Supplementary Table 1**). We used MiXCR (3) to extract BCR

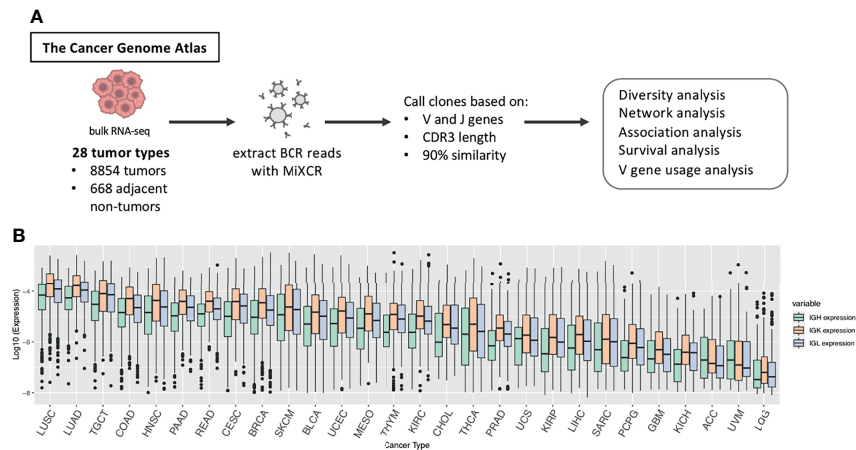
sequences from RNA-seq data and align the sequences to the VDJ region of the IGH, IGK and IGL. As a quality control check, we verified that the number of immune repertoire reads extracted by MiXCR did not have a strong correlation with the total sequencing depth of each sample across tumor types (mean  $r = 0.148$  (SD = 0.088) for tumor samples, mean  $r = 0.179$  (SD = 0.232) for adjacent non-tumor samples, **Supplementary Figures S1A, B**). We defined expression of IGH, IGK, and IGL as the number of reads aligned to each chain divided by the total number of sequenced reads in each sample. **Figure 1B** and **Supplementary Table 1** shows the sequencing summary of BCR reads for all tumor types.

Many of the tumor types that have the highest IGH, IGK, and IGL expression such as lung squamous cell carcinoma (LUAD), lung adenocarcinoma (LUAD), head and neck squamous cell carcinoma (HNSC), and skin cutaneous melanoma (SKCM) (**Figure 1B**), are also the tumor types that have mutational burden as well as high leukocyte fractions (**Supplementary Figure 1C**), which was estimated by Thorsson et al. using methylation data (7), and are most responsive to checkpoint inhibitors (7). Likewise, the tumor types with the lowest expression of IGH, IGK, and IGL, such as uveal melanoma (UVM) and adrenocortical carcinoma (ACC), tend to have low leukocyte fractions and poor responses to immunotherapies (23, 24).

We also found that the expression derived from IGK are more abundant than IGH and IGL across nearly all of the tumor types (**Figure 1B**). This is similar to a previous study which analyzed Ig repertoires across 53 human tissues and found that CDR3 sequences account for 54% of the entire B-cell population on average (4).

### 3.2 Shannon Entropy and Evenness of BCR Repertoires Differ Across Tumor Types and Tend to Be Higher in Adjacent Non-Tumor Samples

We defined clones as groups of reads that have the same V and J gene, the same CDR3 length, and at least 90% nucleotide similarity as in previous publications (21, 25). In order to quantify the diversity of Ig clones within each sample, we calculated the Shannon entropy within each Ig chain. Shannon entropy reflects both the number of clones as well as the frequency of the clones in each sample. We found that LUAD and LUSC have the highest Shannon entropy compared to the other tumor types (**Figure 2A**), which was unsurprising given the overall high Ig expression in these two tumor types. ACC, LGG, and UVM had the lowest Shannon entropy, which likely reflects the low expression of Ig in these tumor types. Overall, Shannon entropy was positively correlated with expression across all tumor types in the IGH, IGK, and IGL chains (**Supplementary Figure 2**). Interestingly, the correlation between Shannon entropy and expression was higher in IGH compared to IGK and IGL across tumor types. For example, the correlation (Spearman's rho) between entropy and expression in LUAD is 0.69 for IGH but 0.27 and 0.35 for IGK and IGL, respectively. This suggests that there may be more uneven distributions of



**FIGURE 1** | Study overview. **(A)** BCR reads were extracted from TCGA RNA-seq data across 28 tumor types using MiXCR and we called clones based on sequences having the same V and J gene, the same CDR3 length, and at least 90% sequence similarity. We then performed diversity analysis, network analysis, association analysis, survival analysis, and V gene usage analysis to investigate differences in the B cell immune repertoire across tumor types and between tumor and adjacent non-tumor samples. **(B)** Boxplots showing the log<sub>10</sub> expression of IGH, IGK, and IGL. Expression is defined as the number of reads for each chain divided by the total number of reads in the sample. The box plot depicts the median as well as the upper and lower quartiles, and the whiskers depict 1.5 times the interquartile range.

clones in IGK and IGL compared to IGH, which is reflected in IGK and IGL having lower Shannon entropy despite having higher expression than IGH.

We also observed that the tumor types that have higher diversity (LUAD, LUSC, BRCA, HNSC, KIRC, PAAD, READ, SKCM and TGCT) have a lower correlation with their expression values.

We then compared the Shannon entropy of primary tumor samples to adjacent non-tumor samples to better understand the tumor microenvironment. We analyzed the 14 tumor types (BLCA, BRCA, COAD, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, PAAD, READ, THCA, UCEC) that had at least 10 adjacent non-tumor samples and found that Shannon entropy was significantly higher in adjacent non-tumor samples compared to tumor samples in 4/14 (BRCA, COAD, LIHC, READ) tumor types for IGH, 6/14 (BRCA, COAD, HNSC, LIHC, LUSC, READ) tumor types for IGK, and 6/14 (BRCA, COAD, HNSC, LIHC, LUSC, READ) tumor types for IGL (**Figure 2B**). This trend could reflect a higher number of clones or a more even distribution of clones in the adjacent non-tumor samples compared to the tumor samples for these tumor types. Conversely, Shannon entropy was higher in tumor samples compared to adjacent non-tumor samples in only 2/14 (KIRC, LUAD) tumor types for IGH, none for IGK, and 1/14 for IGL (KIRC).

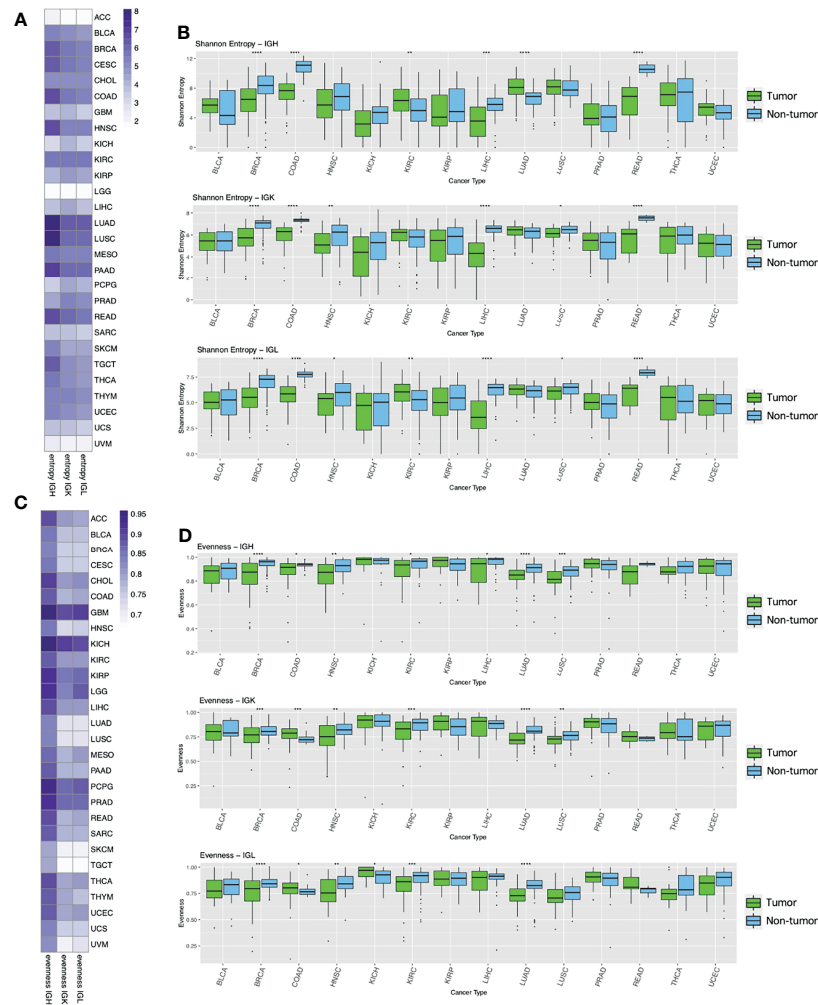
We then calculated Pielou's evenness index for each chain type, which reflects the evenness of the clone distributions within each sample (**Figure 2C**). This evenness index is calculated by dividing Shannon entropy by the maximum possible Shannon entropy index, essentially normalizing the Shannon entropy index by the number of unique clones in each sample. GBM, PCPG, and KICH have the highest evenness in all three chains while SKCM and TGCT have the lowest evenness compared to the other tumor types.

Next, we compared evenness between primary tumors and adjacent non-tumor samples across the Ig chains (**Figure 2D**). We found that evenness was significantly higher in adjacent non-tumor samples compared to tumors in 7/14 (BRCA, COAD, HNSC, KIRC, LIHC, LUAD, LUSC) tumor types for IGH, 5/14 (BRCA, HNSC, KIRC, LUAD, LUSC) for IGK, and 4/14 (BRCA, HNSC, KIRC, LUAD) for IGL. While evenness was consistently higher in adjacent non-tumor samples for IGH, we observed higher evenness in tumor samples compared to non-tumors in IGK and IGL for COAD.

### 3.3 Network Analysis Reveals Differences in Clonal Expansion and Diversification Across Tumor Types and Between Tumor and Adjacent Non-Tumor Samples

We generated networks for each sample using a previously published method (21, 22, 26) (**Figure 3A**). Each vertex in the network is a unique Ig sequence and the size of the vertex corresponds to the number of reads with that sequence. Edges connect vertexes that have the same V and J genes, the same CDR3 length, and at least 90% sequence similarity, and clusters are groups of connected vertexes. Clonal expansion of unique Ig sequences can be measured by calculating the Gini index of the vertex sizes. A high vertex Gini index indicates clonal expansion of unique Ig sequence(s) and a low Gini index indicates a more even distribution of vertex sizes. Clonal diversification can be measured by calculating the Gini index of the cluster sizes, which are the number of vertexes in each cluster. This measurement reflects the amount of diversification of B cell clones from SHM. A high cluster Gini index indicates a sample with unequal cluster sizes, which suggests that some clones are highly diversified, and a low cluster Gini index indicates a sample with even sized clusters.

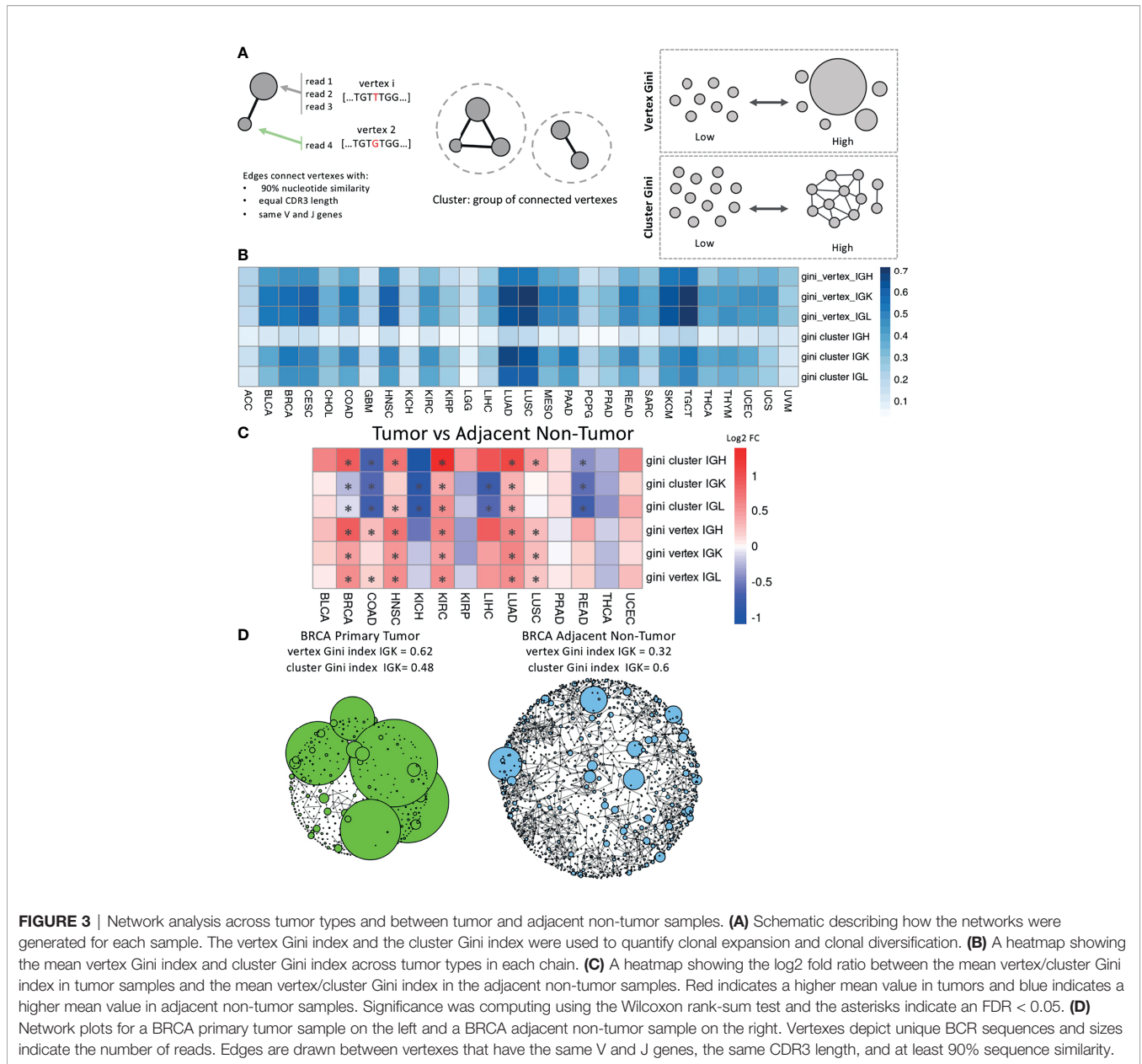




**FIGURE 2** | Entropy and evenness analysis across tumor types and between tumor and adjacent non-tumor samples. **(A)** The heatmap depicts the mean Shannon entropy value for each tumor type in each chain. **(B)** The boxplots show the Shannon entropy indexes for tumors (green) and adjacent non-tumor samples (blue) for the 14 tumor types with at least 10 adjacent non-tumor samples. Statistical significance was calculated using the Wilcoxon rank-sum test. Significant p-values are indicated by symbols above the box plots with one star corresponding to p-value <= 0.05, two stars corresponding to p-value <= 0.01, three stars corresponding to p-value <= 0.001, and four stars corresponding to p-value <= 0.0001. **(C)** The heatmap shows the mean Pielou's evenness index for each tumor type in each chain. **(D)** The boxplots show the evenness index for tumors (green) and adjacent non-tumor samples (blue) and statistical significance was calculated as described above.

We calculated the vertex and cluster Gini indexes on each sample's network across the tumor types (**Figure 3B**). LUAD, LUSC, TGCT, and SKCM had the highest mean vertex Gini indexes across the Ig chains, indicating higher levels of clonal expansion in these tumor types. LUAD, LUSC, and TGCT also had the highest mean cluster Gini indexes across the Ig chains, suggesting high levels of clonal diversification in these tumor types. Interestingly, the cluster Gini indexes are lower in IGH compared to IGK and IGL across the tumor types, suggesting that IGH has lower clonal diversification. We also observed that ACC, GBM, LGG, and KICH have consistently lower Gini vertex and cluster indexes and little difference between the IGH, IGK and IGL chains, suggesting that the B cell repertoire may not be particularly active in these tumor types.

We then compared the vertex and cluster Gini indexes between tumor samples and adjacent non-tumor samples for the tumor types with at least 10 adjacent non-tumor samples (**Figure 3C**). The vertex Gini index was significantly higher in tumor samples in 6/14 tumor types for IGH, 5/14 for IGK, and 5/14 for IGL, and none of the tumor types had significantly higher vertex Gini indexes in the adjacent non-tumor samples compared to the tumor samples. This suggests that the tumor samples generally have higher clonal expansion compared to the adjacent non-tumor samples. Similarly, the cluster Gini indexes was higher in tumor samples compared to adjacent non-tumor samples in 5/14 tumor types for IGH. However, adjacent non-tumor samples had higher cluster Gini indexes in 5/14 tumor types for IGK and 5/14 tumor types for IGL. This suggests that there is a trend towards higher clonal



**FIGURE 3** | Network analysis across tumor types and between tumor and adjacent non-tumor samples. **(A)** Schematic describing how the networks were generated for each sample. The vertex Gini index and the cluster Gini index were used to quantify clonal expansion and clonal diversification. **(B)** A heatmap showing the mean vertex Gini index and cluster Gini index across tumor types in each chain. **(C)** A heatmap showing the log2 fold ratio between the mean vertex/cluster Gini index in tumor samples and the mean vertex/cluster Gini index in the adjacent non-tumor samples. Red indicates a higher mean value in tumors and blue indicates a higher mean value in adjacent non-tumor samples. Significance was computed using the Wilcoxon rank-sum test and the asterisks indicate an FDR < 0.05. **(D)** Network plots for a BRCA primary tumor sample on the left and a BRCA adjacent non-tumor sample on the right. Vertexes depict unique BCR sequences and sizes indicate the number of reads. Edges are drawn between vertexes that have the same V and J genes, the same CDR3 length, and at least 90% sequence similarity.

diversification in tumor samples for IGH and a trend towards higher clonal diversification in adjacent non-tumor samples for IGK and IGL.

We generated plots for each sample and for each chain to visualize the networks. Example plots for BRCA tumor sample and an BRCA adjacent non-tumor sample are shown in **Figure 3D** with data for the IGK chain. The BRCA tumor sample has a higher vertex Gini index and more clonal expansion of individual reads, which can be seen in the large vertexes in the plot. The BRCA adjacent non-tumor sample has a higher cluster Gini index, which can be seen in the increased connectivity of some of the clusters in the plot.

Since the Gini indexes may be affected by differences in sequencing depth, we carried out a downsampling analysis similar to previous studies to confirm that our network analysis results are not driven by sequencing depth variability (21, 27, 28). We downsampled to 500 IGH, IGK, and IGL reads respectively and recalculated the vertex and cluster Gini indexes for each chain. Samples were removed if they did not have at least 500 reads in each chain, which removed a significant number of samples with low infiltration (**Supplementary Figure 3A**). The original analysis and the downsampled analysis were highly correlated for both vertex ( $\rho = 0.72-0.75$ ) and cluster ( $\rho = 0.56-0.81$ ) Gini indexes (**Supplementary Figure 3B**). The downsampled analysis comparing tumor and

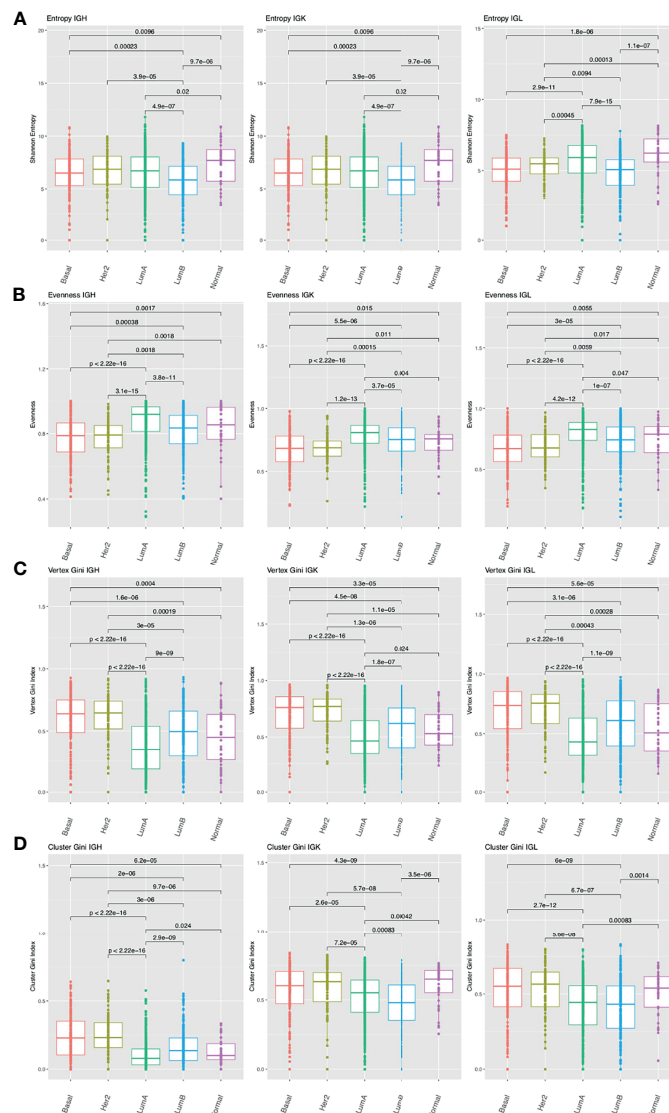
adjacent non-tumor samples also held the same general trends as the original analysis (**Supplementary Figure 3C**).

### 3.4 Case Study: Diversity and Network Analysis Across BRCA Subtypes

While our previous analysis made comparisons between tumor types, cancer is an incredibly heterogeneous disease and each tumor type can often be divided into subtypes with different molecular characteristics and prognosis. We were interested in investigating the differences between tumor subtypes and we performed subtype-specific analysis for the following 19 tumor types with subtype information curated by TCGAbiolinks (20): ACC (29), BLCA (30), BRCA (31), COAD (32), GBM (33), HNSC (34),

KICH (35), KIRC (36), KIRP (37), LGG (33), LIHC (38), LUAD (39), LUSC (40), PAAD (41), PCPG (42), PRAD (43), SKCM (44), THCA (45), UCEC (46). While 15/19 tumor types had significant differences between their subtypes (**Supplementary Figure 4**), we present the results for BRCA in the main text.

Previous studies have shown that breast cancer can be divided into subtypes with different treatment responses and outcomes based (47, 48). These subtypes include: luminal A, luminal B, HER2-enriched, basal, and normal-like. Luminal A and normal-like breast cancers tend to have higher entropy compared to the other subtypes across the chain types (**Figure 4A**). The basal and HER2-enriched subtypes have lower evenness compared to the luminal A, luminal B, and normal-like subtypes across the chain



**FIGURE 4** | Differences in B cell repertoire features across breast cancer subtypes. **(A)** Boxplots depict the Shannon entropy index in each BRCA subtype for IGH, IGK, and IGL. Brackets indicate significant comparisons using the Wilcoxon rank-sum test and p-values are placed above each bracket. **(B)** Boxplots depict Pielou's evenness index in each BRCA subtype. **(C)** Boxplots depict the vertex Gini index in each BRCA subtype. **(D)** Boxplots depict the cluster Gini index in each BRCA subtype.

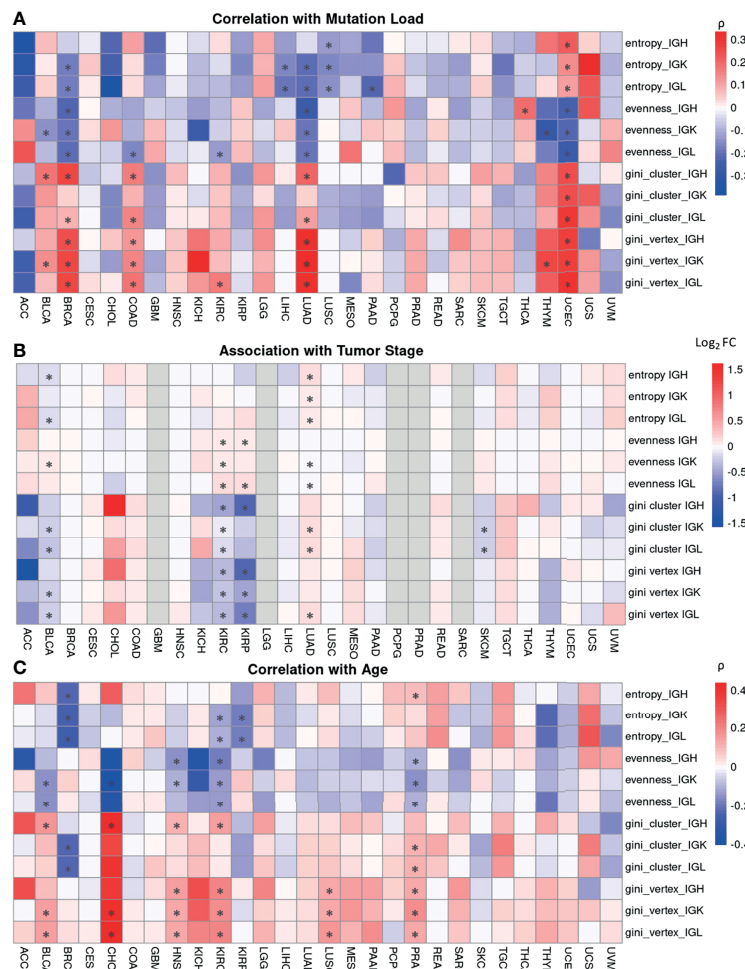
types (**Figure 4B**). In the network analysis, the basal and the HER2-enriched subtypes have higher vertex Gini indexes across the chain types, indicating higher clonal expansion in these subtypes (**Figure 4C**). The basal and HER2-enriched subtypes also have higher cluster Gini indexes compared to the luminal subtypes across the chain types, indicating that there may be higher clonal diversification in these subtypes (**Figure 4D**). This analysis revealed differences in diversity, evenness, and network features across the BRCA subtypes.

### 3.5 B Cell Receptor Repertoire Features Associated With Host and Clinical Features and Mutation Load

We were interested in investigating associations between the B cell repertoire features and mutation load, which was defined as

the number of non-silent mutations per megabase, as well as other host and clinical features available in TCGA (7, 49).

First, we correlated the B cell repertoire features with mutation load and found that mutation load was not significantly correlated with immune features in a majority of the tumor types analyzed. We also observed that the tumor types with significant correlations between their repertoire features and mutation load were not necessarily the tumor types with the highest mutational loads overall (**Supplementary Figure 5A**). However, in the tumor types with significant correlations, mutation load seems to be largely negatively correlated with entropy and evenness (**Figure 5A**). This suggests that having a more diverse, even B cell repertoire seems to be associated with tumors that have lower mutation load. The exceptions were UCEC, which had a positive correlation between mutation load and entropy, and THCA, which had a positive



**FIGURE 5** | Associations between B cell repertoire features and tumor and clinical characteristics. **(A)** A heatmap depicting Spearman’s correlation coefficient for mutation load, which is the number of non-silent mutations per megabase, and each B cell repertoire feature. Significant correlations (FDR < 0.05) are marked by an asterisk. **(B)** A heatmap showing the log2 fold ratio between the mean of the stage I-II tumors and the mean of the stage III-IV tumors. Significance was computed using the Wilcoxon rank-sum test and the asterisks indicate an FDR < 0.05. **(C)** A heatmap depicting Spearman’s correlation coefficient between patient’s age at diagnosis and each B cell repertoire feature. Significant correlations (FDR < 0.05) are marked by an asterisk.



correlation between mutation load and evenness. Similarly, when comparing mutation load to the vertex and cluster Gini indexes, few tumor types had significant correlations. However, among the tumor types with significant correlations, mutation load was positively correlated with the vertex and cluster Gini indexes. This suggests that higher clonal expansion and higher clonal diversification is associated with higher mutation load, perhaps because tumors with higher mutation loads can generate more neoantigens which can drive a better immune response. Indeed, previous studies have shown that a higher non-synonymous mutation burden in tumors was associated with improved response to immunotherapies (50, 51).

We were also interested in investigating associations between the B cell repertoire features and tumor stage. We compared the lower stage tumors (Stage I-II) to higher stage tumors (III-IV) and found that there was no significant difference in a majority of the tumor types (**Figure 5B**). In the 5 tumor types with significant associations, having a higher tumor stage was associated with higher vertex and cluster Gini indexes in 4/5 tumor types, suggesting that there may be slightly increased clonal expansion and clonal diversification in tumors with higher stages.

Next, we correlated age at diagnosis with the B cell repertoire features and found significant associations in 8 tumor types (**Figure 5C**). Age at diagnosis was negatively correlated with Shannon entropy in BRCA, KIRC, and KIRP, similar to a previous study (52). We also found a negative correlation between age and evenness and a positive correlation between age and the vertex Gini indexes in BLCA, HNSC, KIRC, and PRAD. The overall correlation strengths were relatively low, suggesting a possible slight increase of clonal expansion in older patients.

Lastly, we compared the B cell repertoire features between sexes (**Supplementary Figure 5B**) and found very few significant associations across tumor types and repertoire features. Females have significantly higher entropy than males in BLCA but the log fold difference between the mean female entropy value and the mean male entropy value was relatively small.

Overall, the B cell repertoire features were not significantly associated with mutation load or clinical features in a majority of the tumor types. However, there did seem to be some consistent trends among the tumor types with significant associations, suggesting that there may be a subtle signal in these tumor types.

### 3.6 B Cell Repertoire Features Are Prognostic in Select Tumor Types

We built Cox proportional hazard models for each B cell repertoire feature to investigate associations with survival while adjusting for age, gender, and tumor stage (**Figure 6A** and **Supplementary Table 2**). We selected tumor types with at least 40 events to have at least 10 events per predictor variable (53). After FDR correction, we found significant associations (FDR < 0.1) in 7/17 of the tumor types analyzed.

In the six tumor types that have significant associations with Shannon entropy (BRCA, CESC, GBM, HNSC, SARC, SKCM), having a higher entropy value was associated with improved survival in all tumor types except GBM. In the five tumor types that have significant associations with evenness, having a higher evenness was associated with improved survival in SARC and GBM

for a subset of chains. However, a higher evenness was associated with decreased survival in the IGH chain for HNSC, in the IGL chain for UCEC, and across all chains in SKCM as seen in a previous study (54) (**Figure 6B**). This suggests that B cells may be playing different roles in these tumor types. Vertex and cluster Gini indexes were significantly associated with survival in at least one chain type in BRCA, CESC, GBM, HNSC, SARC, SKCM, and UCEC. In BRCA, CESC, HNSC, SARC, SKCM, and UCEC, having a higher vertex and cluster Gini indexes was associated with improved survival, suggesting clonal expansion may be beneficial in these tumor types. However, having a higher vertex Gini index was associated with worse survival in GBM, suggesting that clonal expansion may be detrimental in this tumor type.

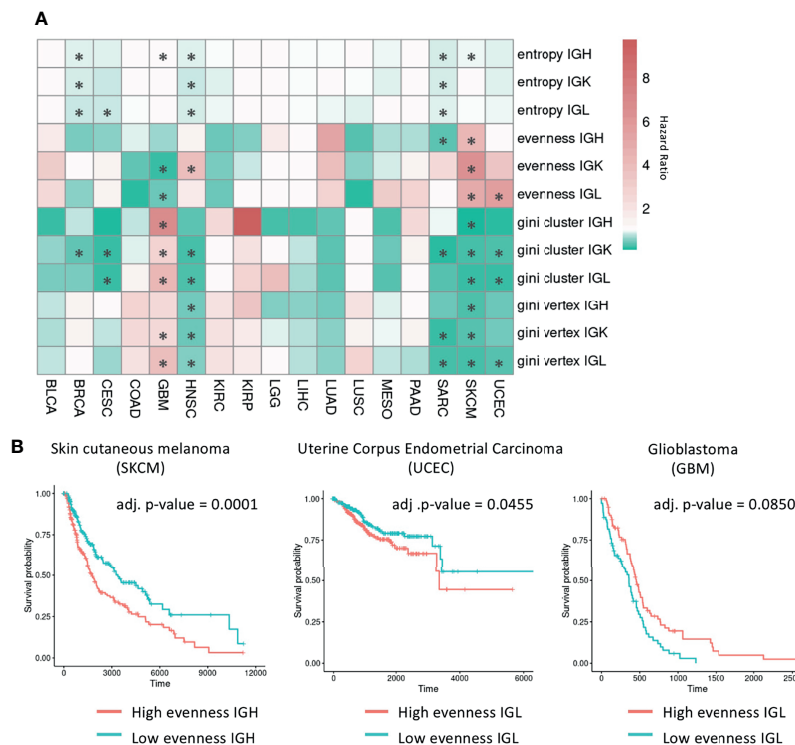
We also stratified the tumors by subtype and repeated the analysis to see if specific subtypes reveal different behaviors (**Supplementary Figure 6**). While five tumor types (HNSC, KIRC, LGG, LUSC, SKCM) had at least one subtype with significant associations with survival, we did not observe differences between subtypes from the same tumor type.

### 3.7 V-Gene Usage Reveals Similarities in COAD, READ, and UCEC Repertoires

Previous studies have shown that V-gene usage may differ in tumor tissues (55). We wanted to investigate differences in V-gene usage across the tumor types analyzed in this study. We defined V-gene usage here as the percent of clones in each sample that use a particular V-gene. We then used principal component analysis (PCA) to reduce the dimensionality of the V-gene usage data and plotted the first two principal components to visualize the data for each Ig chain (**Figures 7A–C**).

For the IGH analysis, we used k-means clustering to identify two clusters in the data after dimensionality reduction with PCA (**Supplementary Figure 7A**). One cluster was primarily comprised of COAD, READ, and UCEC samples (**Supplementary Figure 7B**), suggesting that these samples have a similar V gene usage pattern. We used an elastic net model to identify the V genes that are associated with this cluster and identified 45 V-genes (**Supplementary Figure 7C**). We also performed hierarchical clustering on the IGHV gene usage data (**Figure 7A**) and found a cluster of four V genes that have relatively high usage compared to the others (IGHV3-21, IGHV3-23, IGH30-30, IGHV1-18), which is consistent with previous studies (56, 57).

Next, we performed a similar analysis for the IGH and IGL chains. We identified a group of 19 THYM samples that separated out from the other tumor samples in PC1 for both IGH and IGL (**Figures 7B, C**). These samples had significantly lower V-gene usage (Wilcoxon rank-sum test IGH p-value = 2.8e-06, IGL p-value = 3.9e-05) compared to the other samples, although we could not find significant associations with clinical features such as tumor site or having a history of myasthenia gravis, which is an autoimmune neuromuscular disease found in 50% of cortical thymoma patients (58). We used an elastic net model to identify the V-genes that are associated with these THYM samples and we identified 24 IGH V-genes (**Supplementary Figure 7D**) and 44 IGL V-genes (**Supplementary Figure 7E**). After performing hierarchical clustering on the IGHV gene usage data, we identified a cluster of eight IGHV genes (IGHV3-20, IGHV1-5,



**FIGURE 6** | Survival analysis using B cell repertoire features. **(A)** Heatmap showing the hazard ratio from a Cox proportional hazards model for each B cell repertoire feature adjusted for age, gender, and tumor stage. Red indicates a hazard ratio greater than 1 and green indicates a hazard ratio less than 1. Significant associations (FDR < 0.1) are marked by an asterisk. **(B)** Kaplan-Meier curves for samples with high and low evenness with the median evenness value used as the cutoff. The adjusted p-values are from the multivariate Cox regression models adjusted for age, gender, and tumor stage with the Benjamini-Hochberg procedure used to control for multiple comparisons.

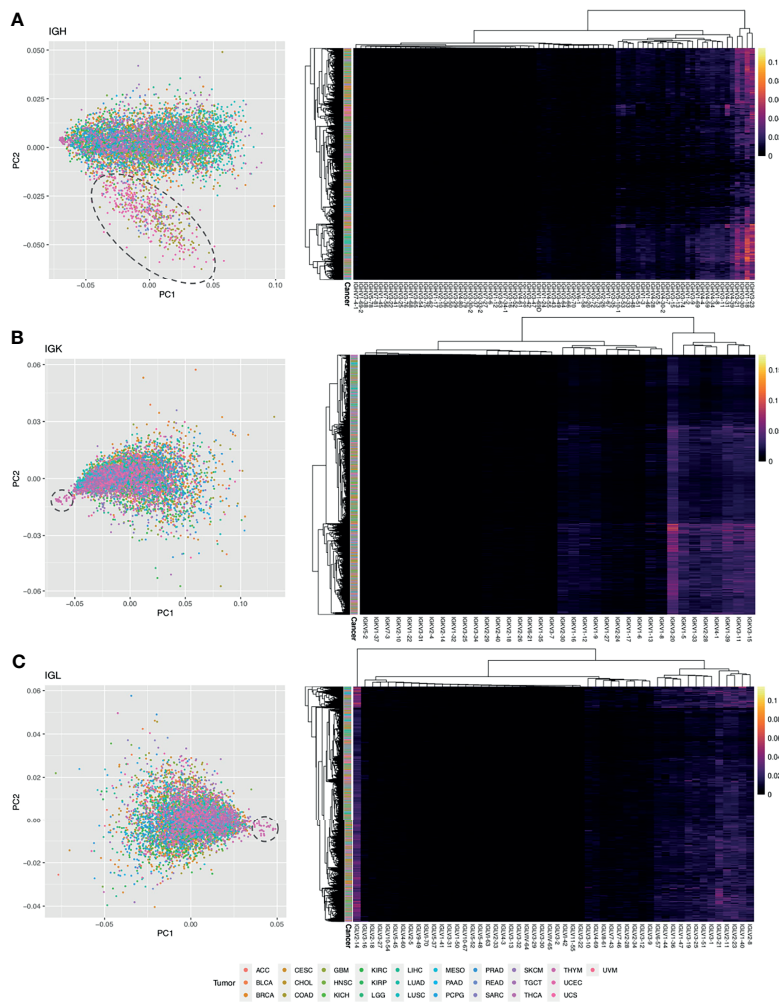
IGKV1-33, IGKV2-28, IGKV4-1, IGKV1-39, IGKV3-11, IGKV3-15) with relatively high usage compared to the other V-genes. Similarly, the hierarchical clustering results for IGL identified a cluster of 13 IGL V-genes with relatively high usage (IGLV6-57, IGLV1-44, IGLV1-36, IGLV1-47, IGLV3-19, IGLV3-25, IGLV1-51, IGLV3-1, IGLV3-21, IGLV2-11, IGLV2-23, IGLV1-40, IGLV2-8). Interestingly, the IGLV2-14 V-gene formed its own cluster separate from every other IGL V-gene and seems to have relatively high usage across many tumor samples (**Figure 7C**).

## DISCUSSION

Many studies have established the importance of T cells in immunosurveillance and immunotherapy response in cancer. However, the role of B cells has not been as well studied and tumor-infiltrating B cells have been shown to have both protumor and antitumor effects (1). Current bioinformatic tools allowed us to interrogate the composition of B cell repertoires from RNA-seq data through a pan-cancer approach, offering more detailed insights into the B cell response to tumors. For example, Thorsson et al. extracted and analyzed the TCR and BCR repertoires from the TCGA RNA-seq dataset, but their analysis did not include tumor subtype stratification, comparisons between the BCR repertoires in tumor

versus adjacent non-tumor tissue, or associations between BCR repertoire features and clinical features. In this study, we analyzed the B cell repertoires of 9,522 tumor and adjacent non-tumor samples across 28 tumor types and their subtypes using TCGA RNA-seq and clinical data.

All tumor samples were assessed for immune repertoire features, including Ig expression, Shannon entropy, clonal expansion and clonal diversification, which revealed large differences among the different tumor types. We found the highest expression of IGH, IGK, and IGL chains in LUSC and LUAD, which is similar to previous studies which found an abundant and diverse B cell population in non-small cell lung cancers (59). Many of the tumor types with the highest Ig chain expression also have the highest overall leukocyte fraction and are most responsive to checkpoint inhibitors (7), suggesting that B cells may help promote response to immunotherapies. Indeed, several studies have shown that an enrichment of B cells in tertiary lymphoid structures was predictive of response to immune checkpoint inhibitors in melanoma, soft-tissue sarcoma, and renal cell carcinoma (60, 61). Moreover, we also found that the tumor types with low IGH, IGK and IGL expression, such as UVM and ACC, also have low overall leukocyte fraction and poor responses to immunotherapies (23, 24). These patterns were also observed in the diversity analysis



**FIGURE 7 |** Analysis of V gene usage. **(A)** PCA plot using IGH V gene usage data. Each point is a sample and the color of the point corresponds to a tumor type. A dashed circle is drawn around a cluster of COAD, READ, and UCEC samples. On the right is a heatmap with hierarchical clustering performed on the samples and the IGH V gene usage data. The intensity of the heatmap corresponds to the percent of clones using each V gene. **(B)** PCA plot using IGK V gene usage data. A dashed circle is drawn around a cluster of THYM samples. On the right is a heatmap with hierarchical clustering performed on the samples and the IGK V gene usage data. **(C)** PCA plot using IGL V gene usage data. A dashed circle is drawn around a cluster of THYM samples. On the right is a heatmap with hierarchical clustering performed on the samples and the IGL V gene usage data.

using Shannon entropy, which was unsurprising given the overall Ig expression in these tumors, and the vertex and cluster Gini index indicating a clear clonal expansion and diversification in the tumors with high overall Ig expression and entropy.

We also found a significant positive correlation between expression and Shannon entropy in all the tumor types analyzed, similar to previous studies (54). Interestingly, we found that the Shannon entropy indexes in IGK and IGL had lower correlations with expression compared to IGH and that the cluster Gini indexes for IGH were lower than IGK and IGL. This suggests that there may be more uneven distribution of clones in IGK and IGL compared to IGH, which is reflected in IGK and IGL having lower Shannon entropy despite having higher expression than IGH. Indeed, IGK and IGL are less diverse

since they are produced by recombination of only the V and J genes and they likely need to go through clonal expansion and somatic hypermutation after exposure to an antigen.

Differences in B cell repertoire between tumor and adjacent non-tumor samples may be due to response to tumor specific antigens. The tumor samples tend to have lower Shannon entropy and vertex Gini indexes compared to primary tumor samples, suggesting that the increased clonal expansion in the tumor samples may be the result of B cells in the tumor microenvironment reacting to tumor neoantigens.

In our tumor subtype analysis, we analyzed 19 tumor types and found significant differences between subtypes in 15 tumor types. We decided to focus on BRCA as a case study in the main text. The basal and HER2-enriched subtypes have lower evenness and higher vertex Gini indexes and cluster Gini indexes, suggesting

more clonal expansion in these subtypes. Interestingly, previous studies have shown that the basal and HER2-enriched subtypes tend to have high immune infiltration (62, 63) and were the only subtypes where increased expression of B cell signatures was associated with metastasis-free survival (64).

Although we found large differences among the tumor types with their B cell repertoire, we found that few tumor types had significant associations between their B cell repertoire features and mutation load, tumor stage, and age. Among the tumor types with significant associations, we found that the repertoire features associated with higher clonal expansion and clonal diversification were positively correlated with mutation load. This is in line with previous studies which have shown that a higher mutation burden is associated with improved immunotherapy responses (50, 51). We also found that age tends to be negatively correlated with evenness and positively correlated with vertex and cluster Gini indexes, suggesting that older patients have greater clonal expansion than younger patients. This reflects previous observations of decreased B cell diversity and increased clonal expansion in normal aging (65). Overall, the B cell repertoire features were not significantly associated with mutation load or clinical features in a majority of the tumor types. However, there did seem to be some consistent trends among the tumor types with significant associations, suggesting that there may be a subtle signal in these tumor types.

Next, we built Cox proportional hazard models for each B cell repertoire feature to investigate the prognostic value of each feature. After adjusting for age, gender, and tumor stage, we were unable to find significant associations for a majority of the tumor types analyzed. However, in the tumor types with significant associations, we found some opposing trends such as evenness being associated with decreased survival in SKCM and UCEC but increased survival in GBM and SARC. Previous studies have shown that B cells can differentiate into plasmablast-like cells in SKCM (66) while they may act primarily as antigen-presenting cells in GBM (67), supporting the idea that B cells play different roles in these tumor types.

Finally, our V-gene usage analysis revealed that a subset of COAD, READ, and UCEC samples have similar IGH V gene usage patterns. Interestingly, a recent study that predicted tumor type from BCR sequences found that COAD samples were likely to be predicted as UCEC in their model, supporting the idea that these tumor types may have similar B cell repertoires (68). We also identified a subset of THYM patients with overall low V-gene usage, although we could not find significant associations between this subset of patients and any clinical variables. In our hierarchical clustering analysis, we identified clusters of V-genes in each chain that had higher overall usage across the tumor types which were largely consistent with previous studies (56, 57). For the IGL V-gene analysis, we found that the IGLV2-14 V-gene formed its own cluster and seems to have relatively high usage across many tumor samples. Interestingly, IGLV2-14 has been previously associated with chronic lymphocytic leukemia (69), multiple melanoma (70), and it is the most common IGLV gene in human cord blood (71).

There are several limitations that should be noted. One limitation of our study is the use of RNA-seq data rather than targeted sequencing data (e.g. BCR-seq). While RNA-seq data has

lower sequencing depth compared to targeted sequencing data, we chose to use RNA-seq data because we wanted to leverage the large number of tumor samples in the TCGA dataset. Additionally, it is not possible to perform paired heavy and light chain analysis using short-read bulk RNA-seq data. Therefore, results from this study warrants validation with targeted sequencing data. Another limitation of this study is the limited number of adjacent non-tumor samples and the lack of true healthy samples, as previous studies have shown that adjacent non-tumor samples tend to have more inflammatory-associated cell types compared to healthy samples (72). Additional analysis using datasets with both tumor and healthy samples would be informative for validating the results of this study. Moreover, the TCGA dataset, like many adaptive immune receptor repertoire sequencing (AIRR-seq) datasets (73), is primarily comprised of individuals with European ancestry (74). Including more individuals with non-European ancestry in immunogenomic studies is critical to understanding population differences in the adaptive immune system and improving precision immunodiagnostics and therapeutics. Lastly, our study is unable to perform analysis within individual types of B cells, which single-cell sequencing would allow, or to analyze the localization of B cell populations within the tumor, which new technologies such as spatial transcriptomics would allow. However, the amount of data generated using these newer technologies is limited compared to the amount of publicly available RNA-seq data currently available, making it more feasible for future studies.

In summary, our study characterizes the B cell repertoire of 28 tumor types and reveals differences across tumors and tumor subtypes, as well as between adjacent non-tumor and tumor samples. These results help further our understanding of the role of B cells in the tumor microenvironment with implications for the development of novel B cell immunotherapies, therapeutic strategies, and patient stratification.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://portal.gdc.cancer.gov/legacy-archive/search/f>.

## AUTHOR CONTRIBUTIONS

KY and SP performed the data analysis. KY and AR extracted the data. KY, SP, and MS wrote the paper. KY, SP, MS, and NM reviewed and edited the paper. SP, MS, and NM supervised the work. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the 2019 AACR-AstraZeneca Immuno-oncology Research Fellowship (19-40-12-PINE). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.



## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.790119/full#supplementary-material>

## REFERENCES

- Largeot A, Pagano G, Gonder S, Moussay E, Paggetti J. The B-Side of Cancer Immunity: The Underrated Tune. *Cells* (2019) 8:449. doi: 10.3390/cells8050449
- Sharonov GV, Serebrovskaya EO, Yuzhakova DV, Britanova OV, Chudakov DM. B Cells, Plasma Cells and Antibody Repertoires in the Tumour Microenvironment. *Nat Rev Immunol* (2020) 20:294–307. doi: 10.1038/s41577-019-0257-x
- Bolotin DA, Poslavsky S, Davydov AN, Frenkel FE, Fanchi L, Zolotareva OI, et al. Antigen Receptor Repertoire Profiling From RNA-Seq Data. *Nat Biotechnol* (2017) 35:908–11. doi: 10.1038/nbt.3979
- Mandric I, Rotman J, Yang HT, Strauli N, Montoya DJ, van der Wey W, et al. Profiling Immunoglobulin Repertoires Across Multiple Human Tissues Using RNA Sequencing. *Nat Commun* (2020) 11:3126. doi: 10.1038/s41467-020-16857-7
- Song L, Cohen D, Ouyang Z, Cao Y, Hu X, Liu XS. TRUST4: Immune Repertoire Reconstruction From Bulk and Single-Cell RNA-Seq Data. *Nat Methods* (2021) 18:627–30. doi: 10.1038/s41592-021-01142-2
- Selitsky SR, Mose LE, Smith CC, Chai S, Hoadley KA, Dittmer DP, et al. Prognostic Value of B Cells in Cutaneous Melanoma. *Genome Med* (2019) 11:36. doi: 10.1186/s13073-019-0647-5
- Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, et al. The Immune Landscape of Cancer. *Immunity* (2018) 48:812–30.e14. doi: 10.1016/j.immuni.2018.03.023
- Weinstein JN, Collisson EA, Mills GB, Shaw KM, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat Genet* (2013) 45:1113–20. doi: 10.1038/ng.2764
- Killock D. Molecular Classification of Glioma. *Nat Rev Clin Oncol* (2015) 12:502–2. doi: 10.1038/nrclinonc.2015.111
- Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas | *NEJM*. Available at: [https://www.nejm.org/doi/10.1056/NEJMoa1402121?url\\_ver=Z39.88-2003&rfr\\_id=ori:rid:crossref.org&rfr\\_dat=cr\\_pub%20%200www.ncbi.nlm.nih.gov](https://www.nejm.org/doi/10.1056/NEJMoa1402121?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200www.ncbi.nlm.nih.gov).
- Zetia N, Agoston AT, Han HS, Mullen JT, Duda DG, Clark JW, et al. A Protein and mRNA Expression-Based Classification of Gastric Cancer. *Mod Pathol Off J U S Can Acad Pathol Inc* (2016) 29:772–84. doi: 10.1038/modpathol.2016.55
- Activating mTOR Mutations in a Patient With an Extraordinary Response on a Phase I Trial of Everolimus and Pazopanib | *Cancer Discovery*. Available at: <https://cancerdiscovery.aacrjournals.org/content/4/5/546>.
- Monitoring the Systemic Human Memory B Cell Compartment of Melanoma Patients for Anti-Tumor IgG Antibodies. Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019330>.
- Sautès-Fridman C, Petitprez F, Calderaro J, Fridman WH. Tertiary Lymphoid Structures in the Era of Cancer Immunotherapy. *Nat Rev Cancer* (2019) 19:307–25. doi: 10.1038/s41568-019-0144-6
- Zhu W, Germain C, Liu Z, Sebastian Y, Devi P, Knockaert S, et al. A High Density of Tertiary Lymphoid Structure B Cells in Lung Tumors is Associated With Increased CD4+ T Cell Receptor Repertoire Clonality. *Oncoimmunology* (2015) 4:12. doi: 10.1080/2162402X.2015.1051922
- Genomic Analysis of Immune Cell Infiltrates Across 11 Tumor Types. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5241901/>.
- Tumor Microenvironment B Cells Increase Bladder Cancer Metastasis via Modulation of the IL-8/Androgen Receptor (AR)/MMPs Signals | *Oncotarget*. Available at: <https://www.oncotarget.com/article/4569/text/>.
- Woo JR, Liss MA, Muldong MT, Palazzi K, Strasner A, Ammirante M, et al. Tumor Infiltrating B-Cells Are Increased in Prostate Cancer Tissue. *J Transl Med* (2014) 12:30. doi: 10.1186/1479-5876-12-30
- Morgan M, Davis S. *GenomicDataCommons: NIH/NCI Genomic Data Commons Access. (Bioconductor Version: Release (3.12), (2021)*. doi: 10.18129/B9.bioc.GenomicDataCommons
- Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAAbioblinks: An R/Bioconductor Package for Integrative Analysis of TCGA Data. *Nucleic Acids Res* (2016) 44:e71–1. doi: 10.1093/nar/gkv1507
- Pineda S, Sigdel TK, Liberto JM, Vincenti F, Sirota M, Sarwal MM. Characterizing Pre-Transplant and Post-Transplant Kidney Rejection Risk by B Cell Immune Repertoire Sequencing. *Nat Commun* (2019) 10:1906. doi: 10.1038/s41467-019-09930-3
- Bashford-Rogers RJM, Palser AL, Huntly BJ, Rance R, Vassiliou GS, Follows GA, et al. Network Properties Derived From Deep Sequencing of Human B-Cell Receptor Repertoires Delineate B-Cell Populations. *Genome Res* (2013) 23:1874–84. doi: 10.1101/gr.154815.113
- Algazi AP, Tsai KK, Shoushtari AN, Munhoz RR, Eroglu Z, Piulats JM, et al. Clinical Outcomes in Metastatic Uveal Melanoma Treated With PD-1 and PD-L1 Antibodies. *Cancer* (2016) 122:3344–53. doi: 10.1002/cncr.30258
- Cosentini D, Grisanti S, Volta AD, Laganà M, Fiorentini C, Perotti P, et al. Immunotherapy Failure in Adrenocortical Cancer: Where Next? *Endocr Connect* (2018) 7:E5–8. doi: 10.1530/EC-18-0398
- IgH Sequences in Common Variable Immune Deficiency Reveal Altered B Cell Development and Selection | *Science Translational Medicine*. Available at: <https://stm.sciencemag.org/content/7/302/302ra135>.
- Bashford-Rogers RJM, Bergamaschi L, McKinney EF, Pombal DC, Mescia F, Lee JC, et al. Analysis of the B Cell Receptor Repertoire in Six Immune-Mediated Diseases. *Nature* (2019) 574:122–6. doi: 10.1038/s41586-019-1595-3
- Strauli NB, Hernandez RD. Statistical Inference of a Convergent Antibody Repertoire Response to Influenza Vaccine. *Genome Med* (2016) 8:60. doi: 10.1186/s13073-016-0314-z
- Bashford-Rogers R, Bergamaschi L, McKinney E, Pombal D, Mescia F, Lee J, et al. B Cell Receptor Repertoire Analysis in Six Immune-Mediated Diseases. *Nature* (2019) 574:122–6. doi: 10.1038/s41586-019-1595-3
- Zheng S, Cherniack AD, Deval N, Moffitt RA, Danilova L, Murray BA, et al. Comprehensive Pan-Genomic Characterization of Adrenocortical Carcinoma. *Cancer Cell* (2016) 29:723–36. doi: 10.1016/j.ccell.2016.04.002
- Weinstein JN, Akbani R, Broom BM, Wang W, Verhaak RGW, McConkey D, et al. Comprehensive Molecular Characterization of Urothelial Bladder Carcinoma. *Nature* (2014) 507:315–22. doi: 10.1038/nature12965
- Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Verzei J, McMichael JF, et al. Comprehensive Molecular Portraits of Human Breast Tumours. *Nature* (2012) 490:61–70. doi: 10.1038/nature11412
- Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Song S, et al. The Consensus Molecular Subtypes of Colorectal Cancer. *Nat Med* (2015) 21:1350–6. doi: 10.1038/nm.3967
- Ceccarelli M, Barthel FP, Malta TM, Sabetts TS, Salama SR, Murray BA, et al. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* (2016) 164:550–63. doi: 10.1016/j.cell.2015.12.028
- Lawrence MS, Sougnez C, Lichtenstein L, Cibulskis K, Lander E, Gabriel SB, et al. Comprehensive Genomic Characterization of Head and Neck Squamous Cell Carcinomas. *Nature* (2015) 517:576–82. doi: 10.1038/nature14129
- Davis CF, Ricketts CJ, Wang M, Yang L, Cherniack AD, Shen H, et al. The Somatic Genomic Landscape of Chromophobe Renal Cell Carcinoma. *Cancer Cell* (2014) 26:319–30. doi: 10.1016/j.ccr.2014.07.014
- Creighton CJ, Morgan M, Gunaratne PH, Wheeler DA, Gibbs RA, Gordon Robertson A, et al. Comprehensive Molecular Characterization of Clear Cell Renal Cell Carcinoma. *Nature* (2013) 499:43–9. doi: 10.1038/nature12222
- Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, Davis C, et al. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med* (2016) 374:135–45. doi: 10.1056/NEJMoa1505917
- Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, Dhalla N, et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* (2017) 169:1327–41.e23. doi: 10.1016/j.cell.2017.05.046
- Collisson EA, Campbell JD, Brooks AN, Berger AH, Lee W, Chmielecki J, et al. Comprehensive Molecular Profiling of Lung Adenocarcinoma. *Nature* (2014) 511:543–50. doi: 10.1038/nature13385

40. Hammerman PS, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, et al. Comprehensive Genomic Characterization of Squamous Cell Lung Cancers. *Nature* (2012) 489:519–25. doi: 10.1038/nature11404
41. Raphael BJ, Hruban RH, Aguirre AJ, Moffitt RA, Yeh JJ, Stewart C, et al. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* (2017) 32:185–203.e13. doi: 10.1016/j.ccell.2017.07.007
42. Fishbein L, Leshchiner I, Walter V, Danilova L, Robertson AG, Johnson AR, et al. Comprehensive Molecular Characterization of Pheochromocytoma and Paraganglioma. *Cancer Cell* (2017) 31:181–93. doi: 10.1016/j.ccell.2017.01.001
43. Abeshouse A, Ahn J, Akbani R, Ally A, Amin S, Andry CD, et al. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* (2015) 163:1011–25. doi: 10.1016/j.ccell.2015.10.025
44. Akbani R, Akdemir KC, Aksoy BA, Albert M, Ally A, Amin SB, et al. Genomic Classification of Cutaneous Melanoma. *Cell* (2015) 161:1681–96. doi: 10.1016/j.ccell.2015.05.044
45. Levine DA. Integrated Genomic Characterization of Endometrial Carcinoma. *Nature* (2013) 497:67–73. doi: 10.1038/nature12113
46. Levine DA. Integrated Genomic Characterization of Endometrial Carcinoma. *Nature* (2013) 497:67–73.
47. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular Portraits of Human Breast Tumours. *Nature* (2000) 406:747–52. doi: 10.1038/35021093
48. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene Expression Patterns of Breast Carcinomas Distinguish Tumor Subclasses With Clinical Implications. *Proc Natl Acad Sci* (2001) 98:10869–74. doi: 10.1073/pnas.191367098
49. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* (2018) 173:400–16.e11. doi: 10.1016/j.ccell.2018.02.052
50. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational Landscape Determines Sensitivity to PD-1 Blockade in Non-Small Cell Lung Cancer. *Science* (2015) 348:124–8. doi: 10.1126/science.aaa1348
51. Johnson DB, Frampton GM, Rioth MJ, Yusko E, Xu Y, Guo X, et al. Targeted Next Generation Sequencing Identifies Markers of Response to PD-1 Blockade. *Cancer Immunol Res* (2016) 4:959–67. doi: 10.1158/2326-6066.CIR-16-0143
52. Gibson KL, Wu Y-C, Barnett Y, Duggan O, Vaughan R, Kondeatis E, et al. B-Cell Diversity Decreases in Old Age and Is Correlated With Poor Health Status. *Aging Cell* (2009) 8:18–25. doi: 10.1111/j.1474-9726.2008.00443.x
53. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of Events Per Independent Variable in Proportional Hazards Regression Analysis II. Accuracy and Precision of Regression Estimates. *J Clin Epidemiol* (1995) 48:1503–10. doi: 10.1016/0895-4356(95)00048-8
54. Selitsky SR, Mose LE, Smith CC, Chai S, Hoadley KA, Dittmer DP, et al. Prognostic Value of B Cells in Cutaneous Melanoma. *Genome Med* (2019) 11:36. doi: 10.1186/s13073-019-0647-5
55. Cui J-H, Lin K-R, Yuan S-H, Jin Y-B, Chen X-P, Su X-K, et al. TCR Repertoire as a Novel Indicator for Immune Monitoring and Prognosis Assessment of Patients With Cervical Cancer. *Front Immunol* (2018) 9:2729. doi: 10.3389/fimmu.2018.02729.
56. Laserson U, Vigneault F, Gadala-Maria D, Yaari G, Uduman M, Heiden JAV, et al. High-Resolution Antibody Dynamics of Vaccine-Induced Immune Responses. *Proc Natl Acad Sci* (2014) 111:4928–33. doi: 10.1073/pnas.1323862111
57. Zhang J-A, Zhou X-Y, Huang D, Luan C, Gu H, Ju M, et al. Development of an Immune-Related Gene Signature for Prognosis in Melanoma. *Front Oncol* (2021) 10:602555. doi: 10.3389/fonc.2020.602555.
58. Romi F. Thymoma in Myasthenia Gravis: From Diagnosis to Treatment. *Autoimmune Dis* (2011) 2011:474512.
59. Stankovic B, Bjørhovde HAK, Skarshaug R, Aamodt H, Frafjord A, Müller E, et al. Immune Cell Composition in Human Non-Small Cell Lung Cancer. *Front Immunol* (2019) 9:3101. doi: 10.3389/fimmu.2018.03101.
60. Helmink BA, Reddy SM, Gao J, Zhang S, Basar R, Thakur R, et al. B Cells and Tertiary Lymphoid Structures Promote Immunotherapy Response. *Nature* (2020) 577:549–55. doi: 10.1038/s41586-019-1922-8
61. Petitprez F, de Reyniès A, Keung EZ, Chen TW-W, Sun C-M, Calderaro J, et al. B Cells Are Associated With Survival and Immunotherapy Response in Sarcoma. *Nature* (2020) 577:556–60. doi: 10.1038/s41586-019-1906-8
62. Tekpli X, Lien T, Røssevold AH, Nebdal D, Borgen E, Ohnstad HO, et al. An Independent Poor-Prognosis Subtype of Breast Cancer Defined by a Distinct Tumor Immune Microenvironment. *Nat Commun* (2019) 10:5499. doi: 10.1038/s41467-019-13329-5
63. Zhu B, Tse LA, Wang D, Koka H, Zhang T, Abubakar M, et al. Immune Gene Expression Profiling Reveals Heterogeneity in Luminal Breast Tumors. *Breast Cancer Res* (2019) 21:147. doi: 10.1186/s13058-019-1218-9
64. Iglesia MD, Vincent BG, Parker JS, Hoadley K, Carey LA, Perou CM, et al. Prognostic B-Cell Signatures Using mRNA-Seq in Patients With Subtype-Specific Breast and Ovarian Cancer. *Clin Cancer Res Off J Am Assoc Cancer Res* (2014) 20:3818–29. doi: 10.1158/1078-0432.CCR-13-3368
65. Weksler ME, Szabo P. The Effect of Age on the B-Cell Repertoire. *J Clin Immunol* (2000) 20:240–9. doi: 10.1023/a:1006659401385
66. Griss J, Bauer W, Wagner C, Simon M, Chen M, Grabmeier-Pfistershammer K, et al. B Cells Sustain Inflammation and Predict Response to Immune Checkpoint Blockade in Human Melanoma. *Nat Commun* (2019) 10:4186. doi: 10.1038/s41467-019-12160-2
67. Candolfi M, Curtin JF, Yagiz K, Assi H, Wibowo MK, Alzadeh GE, et al. B Cells Are Critical to T-Cell-Mediated Antitumor Immunity Induced by a Combined Immune-Stimulatory/Conditionally Cytotoxic Therapy for Glioblastoma. *Neoplasia N Y N* (2011) 13:947–60. doi: 10.1593/neo.11024
68. Hu X, Liu XS. DeepBCR: Deep Learning Framework for Cancer-Type Classification and Binding Affinity Estimation Using B Cell Receptor Repertoires. *bioRxiv [Preprint]* (2019). doi: 10.1101/731158v1.full
69. Stamatopoulos K, Belessi C, Hadzidimitriou A, Smilevska T, Kalagiakou E, Hatziki K, et al. Immunoglobulin Light Chain Repertoire in Chronic Lymphocytic Leukemia. *Blood* (2005) 106:3575–83. doi: 10.1182/blood-2005-04-1511
70. Hadzidimitriou A, Stamatopoulos K, Belessi C, Lalayianni C, Stavroyianni N, Smilevska T, et al. Immunoglobulin Genes in Multiple Myeloma: Expressed and non-Expressed Repertoires, Heavy and Light Chain Pairings and Somatic Mutation Patterns in a Series of 101 Cases. *Haematologica* (2006) 91:781–7.
71. Prabakaran P, Chen W, Singarayan MG, Stewart CC, Streeker E, Feng Y, et al. Expressed Antibody Repertoires in Human Cord Blood Cells: 454 Sequencing and IMGT/HighV-QUEST Analysis of Germline Gene Usage, Junctional Diversity, and Somatic Mutations. *Immunogenetics* (2012) 64:337–50. doi: 10.1007/s00251-011-0595-8
72. Aran D, Camarda R, Odegaard J, Paik H, Oskotsky B, Krings G, et al. Comprehensive Analysis of Normal Adjacent to Tumor Transcriptomes. *Nat Commun* (2017) 8:1077. doi: 10.1038/s41467-017-01027-z
73. Peng K, Safonova Y, Shugay M, Popejoy AB, Rodriguez OL, Breden F, et al. Diversity in Immunogenomics: The Value and the Challenge. *Nat Methods* (2021) 18:588–91. doi: 10.1038/s41592-021-01169-5
74. Carrot-Zhang J, Chambwe N, Damrauer JS, Knijnenburg TA, Robertson AG, Yau C, et al. Comprehensive Analysis of Genetic Ancestry and Its Molecular Correlates in Cancer. *Cancer Cell* (2020) 37:639–54.e6. doi: 10.1016/j.ccell.2020.04.012

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yu, Ravoore, Malats, Pineda and Sirota. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.