

# B-cell Receptor Recombinations in Lung Adenocarcinoma Exome Files Correlate With a Higher Overall Survival Rate

YAPING N. TU<sup>1</sup>, WEI LUE TONG<sup>1</sup>, BLAKE M. CALLAHAN<sup>1</sup>,  
BORIS I. CHOBRUTSKIY<sup>1</sup> and GEORGE BLANCK<sup>1,2</sup>

<sup>1</sup>Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, U.S.A.;

<sup>2</sup>Department of Immunology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, U.S.A.

**Abstract.** *Background/Aim:* While there has been a rapid development in genomic data mining approaches for T-cell receptor recombinations (TcR), less emphasis has been placed on B-cell receptor (BcR) recombinations. *Materials and Methods:* We obtained lung cancer exome files from the cancer genome atlas (TCGA) and mined the files for TcR and BcR recombination reads. *Results:* There was a robust detection of BcR light chain recombination reads in lung adenocarcinoma (TCGA-LUAD) samples, and there was a correlation between the detection of light chain recombination reads and a more favorable outcome. This result was supported by analyses of the expression of B-cell markers as indicated by LUAD RNASeq files. *Conclusion:* BcR and TcR recombination reads recovered from LUAD WXS files, either alone or in combination with the human leukocyte antigen (HLA) type, are likely to have prognostic value.

Tumor immunoscore has become important for both prognosis and therapy decisions. For example, in numerous instances, high lymphocyte counts in the tumor micro-environment have been associated with a better prognosis (1-4). And, immunoscore can facilitate decisions regarding expanding tumor infiltrating lymphocytes in culture, for re-administration to the patient (5, 6); and the use of immune checkpoint blockade therapies (7-10). The use of genomics technologies for immunoscore may be more precise and cost-efficient. For example, in many settings, a whole exome sequence (WXS) is obtained from a surgically removed

tumor specimen for identifying mutations for targeted therapies. Thus, the simultaneous opportunity to use WXS files for immunoscore, *i.e.*, the recovery of immune receptor recombination reads, could be more efficient, and in some cases potentially more accurate, than using other methods (11, 12). In such cases, there is the presumption that the use of the genomics files for obtaining expression levels of immune cell markers, or for obtaining indications of immune receptor recombinations, reflects immune activity. This presumption has been extensively substantiated by correlative data (13-20), although occasional situations of tumor cell expression of immune cell markers, and even recombinant immune function genes in tumor cells, could be an issue. In short, high quality benchmarking of genomics-based immunoscore remains an important goal.

Thus, the future development of genomics-based immunoscore is dependent on continued indications that genomics files can be processed in a way to acquire important information about patients and patient tumor growth. As noted, a substantial effort has been made along these lines for TcR recombinations and RNASeq based immune biomarkers (13-20). However, BcR gene recombinations, detected by genomics approaches, have been less well-studied. Past work has focused on such recombinations in tumors with a B-cell origin, and in two recent cases, breast cancer WXS indicated a relatively high level of BcR V(D)J recombinations (14, 19), compared to several other solid cancers. Here we extend the assessment of BcR recombination reads present in a solid tumor, namely lung adenocarcinoma, with clear indications that these assessments have potential for prognoses, consistent with prognosis information related to BcR VJ recombinations for a recent pancreatic cancer study (17).

This article is freely accessible online.

*Correspondence to:* George Blanck, 12901 Bruce B. Downs Blvd., Tampa, FL 33612, U.S.A. Tel: +1 8139749585, e-mail: gblanck@usf.edu

*Key Words:* BcR gene recombinations, IGK, lung adenocarcinoma, tumor specimen exomes, CD19, CD20, CD79 B-cell markers, survival rates.

## Materials and Methods

*Processing exome files for reads representing V(D)J recombinations.* The nucleotide sequences for both V and J regions of human BcR were collected from NCBI, and additional allele sequences were obtained from IMGT/GENE -DB. Example V and

J sequences, used by the search script, are listed in Table S1 of the supplementary material. Whole exome sequences (WXS) of lung adenocarcinoma were collected from the Genomic Data Commons (TCGA; LUAD), using a key file provided for dbGaP approved project number 6300. The download-manifests for the LUAD WXS files are shown in Table S2. A `Module_Search_IgTcR` shell script (Table S3) was written to execute individual search scripts that identify candidate V and J region sequences specific for each B-cell receptor gene. The search script contains a list of unduplicated 10 nucleotide long sequences close to the 3' end of the V regions. The V region sequences used were 3, 5, 7, 9, and 11 nucleotides away from the 3' end of the V sequence to avoid losses in the search due to N-region diversity. Similarly, 10 nucleotide long sequences from 3, 5, 7, 9, and 11 nucleotides away from the 5' end of the J sequence were used (Table S1). The search script compared each read in the chromosomal segments of the TcR gene regions with the set of V and J sequences, and the reads containing a match for both V and J regions were deposited into a TSV file and were then the object of the script, `IMGTSearchTCR.php`. This php process queried the Immune GeneTics (IMGT) tool for determining whether the read was productive or unproductive and for determining how many nucleotides were an exact match to V and J regions, as well as the quality of the match. Complete match lists for detections of productive IGK and productive IGL recombination reads are shown in Tables S4 and S5. For this project, only recombinations containing greater than 20 nucleotide matches for both V and J regions were accepted as a productive or unproductive read. This relatively high standard was used to ensure the validity of the recombinations.

**RNASeq data.** The RNASeq Version 2 RNA-Seq by Expectation Maximization (RNASeq V2 RSEM) scores for the genes indicated in the Results section were downloaded directly from cBioPortal.org (21, 22) (Table S6). The LUAD barcodes (samples) outputted by the above processing steps for recovering V-J recombination reads, were grouped (using Microsoft Excel) into the following categories: (i) all barcodes (samples), (ii) barcodes with productive IGL, (iii) barcodes with productive IGK, (iv) all barcodes excluding barcodes with productive IGK, (v) barcodes with co-detection of productive IGL and IGK, (vi) all barcodes excluding barcodes with co-detection of IGL and IGK. No group was established representing barcodes without IGL because the IGL barcode group, when compared with all remaining barcodes, in preliminary analyses, did not yield significant survival rate distinctions, as further indicated in the Results section. Moreover, the number of barcodes representing recovery of IGL was substantially less than the number of barcodes representing IGK. Thus, we emphasized IGK for single chain recovery comparisons in this report. The RNASeq V2 RSEM scores for each gene of interest (further discussed in the Results section) were then matched to the barcodes in each recombination read category indicated above, using the `VLOOKUP` function if Microsoft Excel. A t-test was performed to compare the RNA expression levels of each category, for each gene of interest, to provide the p-values. Box and whisker plots (Figure S1) were created for each gene of interest when comparing RNASeq values for the following four categories, representing recovery of IGK and IGL recombination reads from the WXS files, detailed above: iii, iv, v, and vi. The data for the CD19 gene are provided as an example (Table S7). (See also Tables S8-S11 for additional raw RNASeq data used in the Results section).

Table 1. LUAD barcode count for productive and unproductive recombinations identified for IGK and IGL genes in WXS files.

IG Gene	LUAD barcodes and percentages			
	Productive		Unproductive	
IGK	143/571	24.90%	128/571	22.30%
IGL	114/571	20%	30/571	5.22%
Co-detection of IGK and IGL	40/571	7.01%	N/A	N/A

**Survival data.** Survival analyses were conducted in one of two ways, either by use of the cBioPortal.org web tool or the IBM statistical package for the social sciences (SPSS). In the case of the web tool, when needed for analyses, a subset of LUAD barcodes was compared against all remaining samples. When comparing two subsets of barcodes (Tables S8-S11) the IBM SPSS software was used. For the IBM SPSS software, the survival data for the different categories (of RNASeq values, Tables S8-S11, detailed in the previous Methods section) were downloaded from cBioPortal.org, and used as SPSS input (Table S12). Note, in most cases, where the web tool was used, the SPSS software was used for verification.

**Obtaining the human leukocyte antigen (HLA) class I types for the LUAD barcodes.** The HLA class I alleles for the LUAD barcodes were obtained from WXS files using the Optiype software (23) and results were verified with an independent analysis of several RNASeq files representing matching WXS barcodes.

## Results

Results of processing the TCGA-LUAD WXS files for light chain, VJ BcR recombination reads indicated a robust detection of B-cell infiltration in the primary tumor specimen represented by the LUAD barcodes (samples) (Table I). Examples of productive IGK and IGL recombination reads, with the N-region nucleotides and the respective V- and J-gene segment usages are shown in Figure 1. The entire collection of productive recombination reads for IGK and IGL, along with the V- and J-gene segment usages, is available in Tables S4 and S5.

To support the tentative conclusion that the recovery of BcR VJ recombination reads is a measure of B-cell infiltration, RNASeq V2 RSEM values were obtained for the following genes (HUGO symbols), representing B-cell markers: *CD19*, *C22*, *CD24*, *CD38*, *CD72*, *CD79A*, *CD79B*, *CD83*, *CD86*, *CR2*, *FCGR2*, *MS4A1*, *TNFRSF13B*, *TNFRSF13C*, and *TNFRSF17* (17, 24, 25). The expression level for each gene was then compared across these barcode groups: (i) all barcodes, (ii) barcodes representing productive IGL recombination reads, (iii) barcodes representing productive IGK recombination reads, (iv) all barcodes excluding those with productive IGK recombination reads, (v) barcodes representing co-detection of productive IGK

**A**

TCGA-05-4398



TCGA-05-4418



**B**

TCGA-05-4382



TCGA-05-4434



Figure 1. *IGK productive, V-J recombination reads and IGL productive reads from WXS LUAD files. (A) Examples of IGK productive, V-J recombination reads from WXS LUAD files. Barcode TCGA-05-4398 (top); TCGA-05-4418 (bottom). (B) Examples of IGL productive reads from WXS LUAD files. Barcode TCGA-05-4382 (top); TCGA-05-4434 (bottom).*

and IGL recombination reads, and (vi) all barcodes excluding those with co-detection of IGK and IGL recombination reads. Results indicated that RNASeq values in the IGK and IGL groups (ii), (iii), and (v) were significantly higher than the RNASeq values in the corresponding control groups (i), (iv), and (vi), for all of the above B-cell marker genes, except for CD24 and CD83 (Table II, with a subset of comparisons in Figure 2; see also Figure S1). In the case of CD24, RNASeq value averages were higher in the group of barcodes representing co-detection of IGL and IGK recombination reads and in the barcode group representing recovery of productive IGK reads. However, there was no statistically significant difference indicated for these two comparisons. In the case of CD83, although the group of barcodes representing co-detection of IGK and IGL recombination reads had a higher RNA expression average, there was not a statistically significant difference. For the group of barcodes representing recovery of IGK productive recombination reads, there was a significantly higher CD83 expression level, in comparison to the remaining barcodes ( $p < 0.033$ ) (Table II, Figure 2, Figure S1).

To determine whether the co-detection of productive IGK and IGL recombination reads correlated with a better or worse survival outcome, we employed the cBioPortal.org web tool allowing a comparison of this co-detection barcode group (v) against all remaining LUAD barcodes, group (vi), using a Kaplan–Meier (KM) analysis. The results indicated that the IGK/IGL recombination read, co-detection group (v) had a significantly better overall survival rate ( $p < 0.015$ ), compared to all remaining barcodes (Figures S2, S3). To understand the drivers for the positive survival trend, barcodes containing productive IGK were plotted similarly against the remaining population in a KM curve, also using the above cBioPortal web tool, and a significantly better survival rate was observed for this IGK barcode group ( $p < 0.018$ ). To verify the results obtained from the cBioPortal tool, the above survival rate determinations were obtained independently using the IBM Statistical Package for the Social Sciences (SPSS) (Figures S3, S4).

To confirm the survival rate increases observed for the barcode groups representing the recovery of productive BcR VJ recombination reads, survival rates were obtained for

Table II. B-cell marker mRNA expression (RNASeq V2 RSEM score) for each barcode category (with corresponding t-test p-value).

Barcode categories (number of samples)	B-cell marker gene expression averages, RSEM values														
	CD19	CD22	CD24	CD38	CD72	CD79A	CD79B	CD83	CD86	CR2	FCGR2B	BMS4A1	TNFR SF13B	TNFR SF13C	TNFR SF17
All samples (517)	91.3	197.1	8324.6	239.2	125.1	729.7	180.6	602.4	422.7	275.7	472.1	306.7	30.1	16.9	90.8
IGL productive (114)	169.2	314.0	9502.7	292.6	164.3	1265.2	310.9	660.6	508.7	531.6	616.0	565.3	55.9	27.3	148.3
IGK productive (143)	160.1	289.9	9135.0	313.3	168.6	1260.5	314.4	664.1	522.9	428.7	664.9	516.6	48.7	26.1	152.7
All samples excluding IGK productive	69.3	167.6	8066.2	215.6	111.2	560.4	137.9	582.7	390.8	226.9	410.6	239.7	24.2	13.9	71.1
Co-detection of productive IGL & IGK (40)	208.5	349.0	10064.0	310.7	178.9	1803.0	414.4	691.1	567.2	625.2	824.3	621.8	69.5	31.9	218.5
All samples excluding co-detection of IGK/IGL	83.0	186.5	8202.2	234.2	121.3	654.1	164.1	596.1	412.6	251.1	447.3	284.5	27.4	15.8	81.8
p-Value: All (no co-detection vs. co-detection IGK/IGL)	<b>0.0008</b>	<b>0.0233</b>	0.2392	<b>0.0216</b>	<b>0.0088</b>	<b>0.0002</b>	<b>0.0063</b>	0.1393	<b>0.0075</b>	<b>0.0326</b>	<b>0.0023</b>	<b>0.0114</b>	<b>0.0020</b>	<b>0.0034</b>	<b>0.0041</b>
p-Value: All (no IGK productive vs. IGK productive)	<b>2.301E-08</b>	<b>9.675E-05</b>	<b>1.555E-01</b>	<b>1.195E-04</b>	<b>1.897E-06</b>	<b>3.494E-09</b>	<b>5.599E-07</b>	<b>3.310E-02</b>	<b>5.111E-05</b>	<b>2.658E-03</b>	<b>1.431E-05</b>	<b>9.079E-06</b>	<b>5.499E-06</b>	<b>1.169E-06</b>	<b>2.410E-06</b>

Bold values show significance.

Table III. Overall survival p-value for top 20% and bottom 20% expressers.

Gene	All samples	Top 20%		Bottom 20%	
	RNA expression average	RNA expression average	p-Value against remaining population	RNA expression average	p-Value against remaining population
CD19	91.256	292.951	0.0039	4.962	0.0013
CD20	306.678	1068.433	0.0011	10.592	0.0063
CD79A	729.668	2156.784	0.01	66.414	0.0044
CD79B	180.559	509.138	0.0074	26.642	0.17

barcode groups representing RNASeq values for the B-cell markers, CD19, CD20, CD79A, and CD79B. For each gene, the 20 percent of barcodes with the highest level of RNASeq values were plotted against the remaining population in a KM curve. The KM curve analyses were repeated with the 20 percent of barcodes representing the lowest level of expression for the B-cell markers, whereby these barcodes were plotted against all remaining LUAD barcodes (Figures S5-S8). The results indicated that the barcodes representing the top 20% of RNASeq values corresponded to significantly higher overall survival rates in the case of all four genes. Conversely, the barcodes representing the bottom 20% of the RNASeq values, for the B-cell marker genes, indicated significantly reduced overall survival rates, except in the case of CD79B (Table III). KM curves were then constructed

to directly compare the barcode group associated with the top 20% to the group representing the bottom 20% of RNA expression for each of the four B-cell marker genes (Figure 3). A statistically significant difference in survival between the groups was found for all of the genes: The barcodes with high B-cell marker expression had better survival rates than the barcodes with low expression.

To determine whether additional factors are associated with better survival for LUAD patients with productive IGK detection in the WXS files, we focused on a barcode set representing an HLA-C\*702 allele and TcR-β-J2-gene segment usage, as determined by recovery of TcR-β-J2 recombination reads from the LUAD WXS files. We considered this approach because of related projects whereby significant survival differences were associated with a

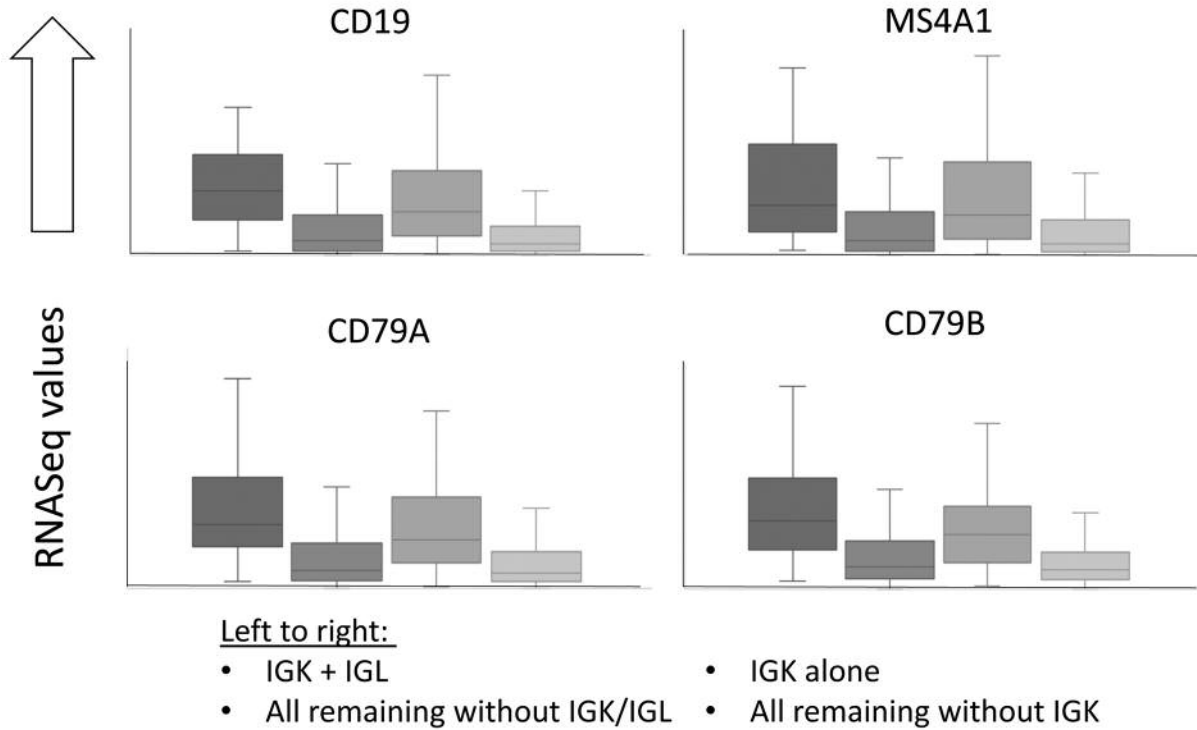


Figure 2. Box and whisker plots for expression of the CD19, MS4A1, CD79A, and CD79B genes. Levels of mRNA expression (RNASeq V2 RSEM) are compared across the categories (from left to right): group v, barcodes representing co-detection of productive IGK and IGL; group vi, all remaining barcodes without co-detection; group iii, barcodes representing productive IGK; group iv, all remaining barcodes without productive IGK.

Table IV. Overall survival represented by barcodes with HLA class I allele/TcR-β-J2 usage and recovery of productive IGK reads (See SOM for details).

Sample categories	Overall survival average
Both productive IGK; and HLA-C*702 allele/TcR-β-J2 usage	49.99 months
HLA-C*702 allele/TcR-β-J2 usage but no IGK detection in WXS files	30.48 months
Productive IGK but no HLA-C*702 allele/TcR-β-J2 usage	34.09 months
<i>p</i> -Value: Productive IGK and HLA-C*702 allele/TcR-β-J2 usage; <i>versus</i> HLA-C*702 allele/TcR-β-J2 usage and no IGK	0.018
<i>p</i> -Value: Productive IGK and HLA-C*702 allele/TcR-β-J2 usage; <i>versus</i> Productive IGK but no HLA-C*702 allele/TcR-β-J2 usage	0.036
<i>p</i> -Value: HLA-C*702 allele/TcR-β-J2 usage; <i>versus</i> all remaining barcodes, from Kaplan–Meier Curve (cBioPortal.org, Figure S9)	0.03
<i>p</i> -Value: Productive IGK and HLA-C*702 allele/TcR-β-J2 usage; <i>versus</i> all remaining barcodes, from Kaplan–Meier Curve (cBioPortal.org, Figure S9)	0.03

particular HLA class I allele in combination with either J1 or J2 usage for TcR-β (13, 15, 16, 26, 27). In particular, patients with the HLA-C\*702 allele/TcR-β-J2 usage combination have been shown to have significantly higher survival than the remaining patients ( $p < 0.031$ ; Figure S9) for LUAD. Barcodes representing both productive IGK and HLA-C\*702 allele/TcR-β-J2 usage were identified, and KM analyses results indicated that this combination group had a

significantly higher survival (Figure S9). Average survival time for this population increased compared to the barcodes with HLA-C\*702 allele/TcR-β-J2 usage detection in the absence of IGK detection. Average overall survival in HLA-C\*702 allele/TcR-β-J2 usage samples without IGK detection in the WXS files was about 30 months, and overall survival represented by barcodes with productive IGK, when the barcodes representing HLA-C\*702 allele/TcR-β-J2 usage

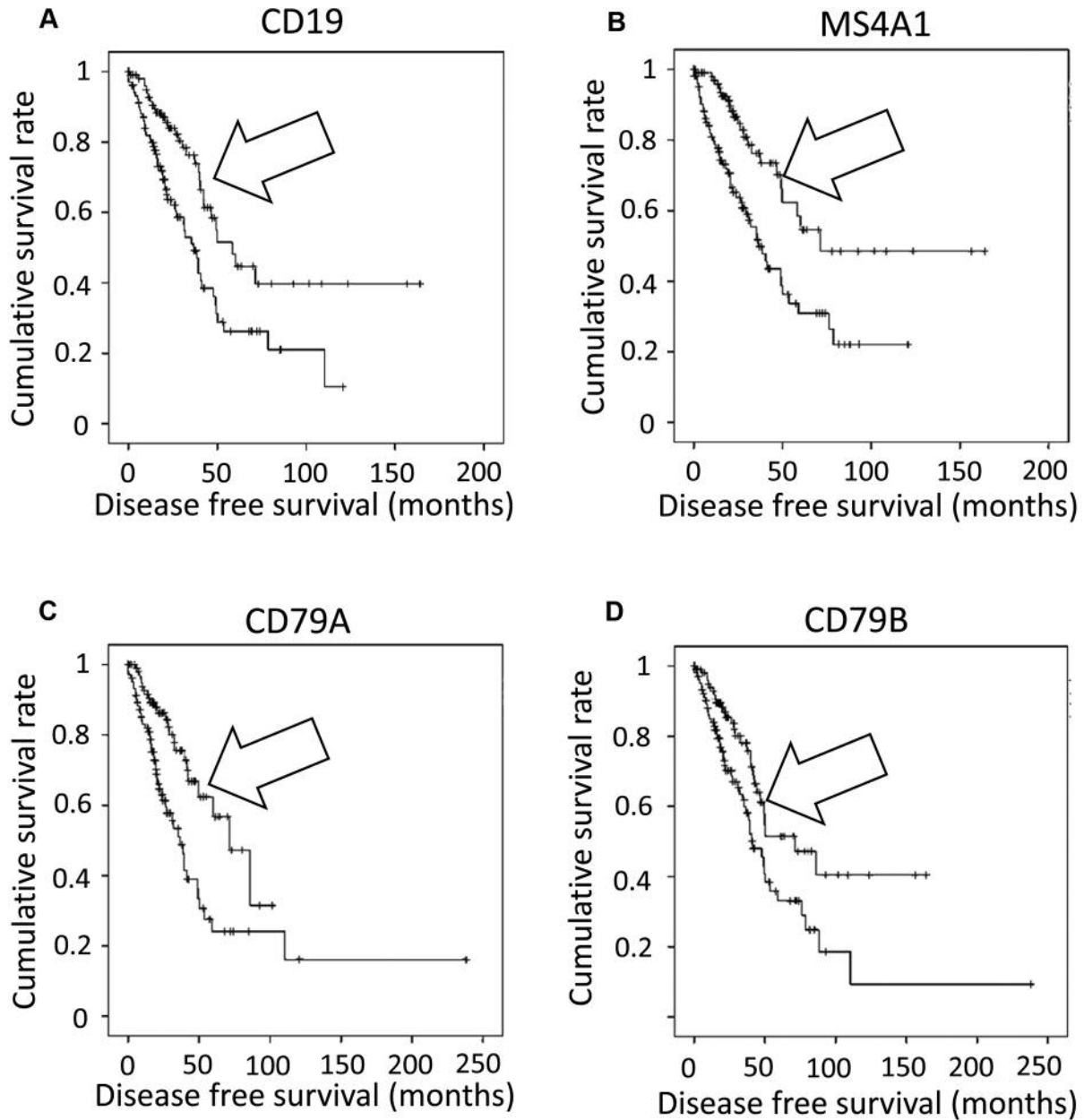


Figure 3. Kaplan–Meier (KM) curves resulting from of a direct comparison of the barcode group associated with the top 20% to the group representing the bottom 20% of RNA expression for each of the four B-cell marker genes. (A) Kaplan–Meier (KM) analysis comparing the barcodes representing the highest 20% of CD19 RNA expression (top 20, arrow) versus barcodes representing the lowest 20% of CD19 RNA expression Log Rank (Mantel–Cox)  $p \leq 0.001$ . (B) KM analysis comparing the barcodes representing the highest 20% of MS4A1 RNA expression (top 20, arrow) versus barcodes representing the lowest 20% of MS4A1 RNA expression Log Rank (Mantel–Cox)  $p \leq 0.001$ . (C) KM analysis comparing the barcodes representing the highest 20% of CD79A RNA expression (top 20, arrow) versus barcodes representing the lowest 20% of CD79A RNA expression Log Rank (Mantel–Cox)  $p < 0.001$ . (D) KM analysis comparing the barcodes representing the highest 20% of CD79B mRNA expression (top 20, arrow) versus barcodes representing the lowest 20% of CD79B mRNA expression Log Rank (Mantel–Cox)  $p = 0.005$ .

were removed, was about 34 months. Samples with recovery of IGK V-J recombination reads and the HLA-C\*702 allele/TcR- $\beta$ -J2 usage combination represented an average overall survival of about 50 months, *i.e.*, a statistically

significant difference when compared to either of the other two groups indicated above (Tables IV, S13).

Given that the cytotoxic T cell response is more commonly associated with an active immune response

component for certain cancers, KM analyses were employed to compare barcodes representing productive TcR  $\alpha$ -,  $\beta$ -, and unproductive TcR-  $\beta$  to the remaining population. No significant differences were found for any of these comparisons (Figures S10-11).

To further confirm and extend the basic findings above, we re-searched the LUAD tumor WXS files with a revised algorithm for identification of the recombination reads. In this revised, scripted algorithm, we allowed the assessment of a V-gene segment match to make use of any available germline nucleotides 3' of the second V-gene segment, conserved cysteine. This is in contrast to the first approach used above (Methods), whereby all V-gene segment matches were made only with the available nucleotides, within the candidate recombination read, on the 5' side of the conserved cysteine. This second approach was more robust and allowed more recovery of recombination reads representing IGH, where the read must have room for verifiable V- and J-gene segments, as well as the IGH D-gene segment and the remainder of the junction, non-germline amino acid sequences (28). Thus, we assessed survival rates represented by the LUAD barcodes representing recovery of all three BcR gene recombination reads in comparison to the survival rate represented by all remaining barcodes. Results indicated that the barcodes representing the recovery of IGH, IGL or IGK recombination reads, as a group, also represented a statistically significant greater survival rate (log rank  $p=0.030$ , Figure S12).

Given the robustness of this second BcR read recovery algorithm, we next searched the LUAD blood WXS files for BcR recombination reads and compared survival rates represented by barcodes also representing BcR recombination read recovery and all remaining barcodes. The results indicated that blood sample WXS files yielding IGH recombination reads also represented LUAD patients with a higher survival rate (KM log rank  $p=0.030$ , Figure S13).

## Discussion

Immunoscore, being both cost-effective and timely, has demonstrated its value in diagnosis, prognosis, and treatment of cancer. With increasing prevalence of tumor exome sequencing in both lab and clinical settings, the data available for genomics-based immunoscore is growing rapidly. This provides further opportunities to discover useful functions of immunoscore in general, as well as to better identify patient sub-populations based on parameters uniquely available from genomics approaches.

This study reports increased algorithm efficiency and result output, as well as correlations of specific B-cell receptor recombinations with statistically significant increases in survival. The B-cell receptor recombination extraction software utilized in this research was a modification of the method used in previous studies (11, 12).

The current software is capable of detecting more recombinations by increasing V and J sequence searches near the N-region diversity section of the recombination (28, 29).

The above WXS file searches yielded recovery of productive IGK recombination reads for nearly 25% of the lung adenocarcinoma barcodes, and over 22% of the WXS files contained unproductive IGK recombination reads, in addition to productive and unproductive IGL recombination reads. The presence of the IGK and IGL recombinations were consistent with the high level of RNA expression of B-cell specific genes, such as *CD19*, *CD20*, and *CD79*. Overall, there was a significantly higher level of B-cell specific gene expression for the barcodes representing recovery of IGK and IGL recombination reads.

The B-cell RNASeq data provided survival correlations that mirrored the survival correlations with the recovery of BcR IGK/IGL recombination reads, when the barcodes were distinguished by upper and lower RNASeq quintiles. This is in contrast to a recent case of pancreatic cancer, where BcR IGK/IGL recombination reads strongly correlated with a worse survival outcome and where B-cell marker RNASeq data did not reveal statistically significant associations with worse outcome (17). In this latter case there was the tentative conclusion that the recovery of the BcR recombination reads represented a higher standard for B-cell presence in the tumor and thereby provided the opportunity for the statistically significant survival correlation. In this current case of LUAD, the recovery of BcR recombination reads represents an absolute standard, whereas the comparison to RNASeq involved the arbitrary selection of the barcodes at the upper and lower quintiles. Thus, in both cases, there is an apparent extra value in using the recovery of the BcR recombination reads for the patient-survival assessments. In addition, the WXS provides a simultaneous opportunity to obtain TcR recombination reads and HLA types, which in this case, when overlapped with BcR recombination read recovery data, indicated yet another distinct, survival rate group (Figure S9).

The association between the better survival rates and recovery of the BcR recombination reads in LUAD is consistent with a previous study whereby IGKC (constant region) expression in non-small cell lung cancer also correlated with a better outcome (30). Indeed, the association between better outcomes and B-cell genomics markers may not be limited to lung cancer. Previous work has indicated that breast cancer exome files are significantly enriched in BcR recombination reads (19), and subsequent work has correlated recovery of BcR recombination reads from breast cancer exome files with better survival rates (14). And, another report has also indicated the association of IGKC with better outcomes in breast cancer (31). Overall, these results raise the question of what might distinguish cancer with B-cell infiltrates where the infiltrate represents a better

outcome, *versus* cancers where the B-cell infiltrate represents a worse outcome? Also, what antigen(s) might a BcR be targeting in the lung adenocarcinoma setting?

Finally, the above results also indicated the potential for blood exome files to be used for recovery of BcR recombination reads, followed by the correlation of such recoveries with clinical features (Figure S13). This result is reported for the first time, and raises questions about the links between the detection of the lymphocyte BcR recombination in blood exome files and the course of disease. For example, does detection of the BcR recombination reads simply represent higher numbers of B-cells present in the blood samples of certain patients, such that the preparation of a typical exome has higher likelihood of including BcR recombination DNA? And if so, what are the mechanistic connections between the higher numbers of B-lymphocytes and disease course? Regardless, these initial approaches using blood exome files indicate that, at least when conducting a sufficiently large study, it is possible to generate statistically significant correlative data, *i.e.*, with immune receptor recombinations and clinical features.

### Supplementary Material

All supplementary material (Tables S1-13 and Figures S1-13) is freely available at: <http://www.universityseminarassociates.com/media/Tu%20et%20al%20supporting%20online%20material%2020.pdf>

### Conflicts of Interest

The Authors declare that they have no conflicts of interest regarding this study.

### Authors' Contributions

Yaping N. Tu conceived of plan, conducted most of the basic analyses, and wrote several early drafts of the manuscript. Wei Lue Tong, Blake M. Callahan, and Boris I. Chobrutskiy wrote and applied software; and provided some analyses of data. George Blanck supervised the project, assisted in the data analyses, and finalized manuscript draft.

### Acknowledgements

The Authors wish to thank USF research computing and the taxpayers of the State of Florida. Blake M. Callahan was a recipient of a Bonati research stipend.

### References

- Butler SN and Blanck G: Immunoscoring by correlating MHC class II and TCR expression: High level immune functions represented by the KIRP dataset of TCGA. *Cell Tissue Res* 363(2): 491-496, 2016. PMID: 26293619. DOI: 10.1007/s00441-015-2261-1
- Curigliano G and Perez EA: Immunoscoring breast cancer: TILs remember what they target. *Ann Oncol* 25(8): 1455-1456, 2014. PMID: 24950980. DOI: 10.1093/annonc/mdl227
- Galon J, Mlecnik B, Bindea G, Angell HK, Berger A, Lagorce C, Lugli A, Zlobec I, Hartmann A, Bifulco C, Nagtegaal ID, Palmqvist R, Masucci GV, Botti G, Tatangelo F, Delrio P, Maio M, Laghi L, Grizzi F, Asslaber M, D'Arrigo C, Vidal-Vanaclocha F, Zavadova E, Chouchane L, Ohashi PS, Hafezi-Bakhtiari S, Wouters BG, Roehrl M, Nguyen L, Kawakami Y, Hazama S, Okuno K, Ogino S, Gibbs P, Waring P, Sato N, Torigoe T, Itoh K, Patel PS, Shukla SN, Wang Y, Kopetz S, Sinicrope FA, Scripcariu V, Ascierto PA, Marincola FM, Fox BA and Pages F: Towards the introduction of the 'immunosome' in the classification of malignant tumours. *J Pathol* 232(2): 199-209, 2014. PMID: 24122236. DOI: 10.1002/path.4287
- Anitei MG, Zeitoun G, Mlecnik B, Marliot F, Haicheur N, Todosi AM, Kirilovsky A, Lagorce C, Bindea G, Ferariu D, Danciu M, Bruneval P, Scripcariu V, Chevallier JM, Zinzindohoue F, Berger A, Galon J and Pages F: Prognostic and predictive values of the immunosome in patients with rectal cancer. *Clin Cancer Res* 20(7): 1891-1899, 2014. PMID: 24691640. DOI: 10.1158/1078-0432.CCR-13-2830
- Pandolfino MC, Labarriere N, Tessier MH, Cassidanius A, Bercegeay S, Lemarre P, Dehaut F, Dreno B and Jotereau F: High-scale expansion of melanoma-reactive til by a polyclonal stimulus: Predictability and relation with disease advancement. *Cancer Immunol Immunother* 50(3): 134-140, 2001. PMID: 11419180. DOI: 10.1007/PL00006683
- Rosenberg SA: IL-2: The first effective immunotherapy for human cancer. *J Immunol* 192(12): 5451-5458, 2014. PMID: 24907378. DOI: 10.4049/jimmunol.1490019
- Gruosso T, Gigoux M, Manem VSK, Bertos N, Zuo D, Perlitch I, Saleh SMI, Zhao H, Souleimanova M, Johnson RM, Monette A, Ramos VM, Hallett MT, Stagg J, Lapointe R, Omeroglu A, Meterissian S, Buisseret L, Van den Eynden G, Salgado R, Guiot MC, Haibe-Kains B and Park M: Spatially distinct tumor immune microenvironments stratify triple-negative breast cancers. *J Clin Invest* 129(4): 1785-1800, 2019. PMID: 30753167. DOI: 10.1172/JCI96313
- Park R, Winnicki M, Liu E and Chu WM: Immune checkpoints and cancer in the immunogenomics era. *Brief Funct Genomics* 18(2): 133-139, 2019. PMID: 30137232. DOI: 10.1093/bfpg/ely027
- Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A and Peters S: Development of tumor mutation burden as an immunotherapy biomarker: Utility for the oncology clinic. *Ann Oncol* 30(1): 44-56, 2019. PMID: 30395155. DOI: 10.1093/annonc/mdy495
- Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H and Trajanoski Z: Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* 18(1): 248-262, 2017. PMID: 28052254. DOI: 10.1016/j.celrep.2016.12.019
- Gill TR, Samy MD, Butler SN, Mauro JA, Sexton WJ and Blanck G: Detection of productively rearranged tcr-alpha v-j sequences in tcga exome files: Implications for tumor immunoscoring and recovery of antitumor t-cells. *Cancer Inform* 15: 23-28, 2016. PMID: 26966347. DOI: 10.4137/CIN.S35784
- Samy MD, Tong WL, Yavorski JM, Sexton WJ and Blanck G: T cell receptor gene recombinations in human tumor specimen



- exome files: Detection of t cell receptor-beta v/dj recombinations associates with a favorable oncologic outcome for bladder cancer. *Cancer Immunol Immunother* 66(3): 403-410, 2017. PMID: 27995306. DOI: 10.1007/s00262-016-1943-1
- 13 Clark KR, Tong WL, Callahan BM, Yavorski JM, Tu YN and Blanck G: Trb-j1 usage, in combination with the hla-a\*01:01 allele, represents an apparent survival advantage for uterine corpus endometrial carcinoma: Comparisons with microscopic assessments of lymphocyte infiltrates. *Int J Immunogenet* 46(1): 31-37, 2019. PMID: 30474304. DOI: 10.1111/iji.12409
  - 14 Tong WL, Callahan BM, Tu YN, Zaman S, Chobrutskiy BI and Blanck G: Immune receptor recombinations from breast cancer exome files, independently and in combination with specific hla alleles, correlate with better survival rates. *Breast Cancer Res Treat* 173(1): 167-177, 2019. PMID: 30229447. DOI: 10.1007/s10549-018-4961-1
  - 15 Callahan BM, Tong WL and Blanck G: T cell receptor-beta j usage, in combination with particular hla class ii alleles, correlates with better cancer survival rates. *Immunol Res* 66(2): 219-223, 2018. PMID: 29594966. DOI: 10.1007/s12026-018-8990-y
  - 16 Callahan BM, Yavorski JM, Tu YN, Tong WL, Kinskey JC, Clark KR, Fawcett TJ and Blanck G: T-cell receptor-beta v and j usage, in combination with particular hla class i and class ii alleles, correlates with cancer survival patterns. *Cancer Immunol Immunother* 67(6): 885-892, 2018. PMID: 29508024. DOI: 10.1007/s00262-018-2139-7
  - 17 Kinskey JC, Tu YN, Tong WL, Yavorski JM and Blanck G: Recovery of immunoglobulin vj recombinations from pancreatic cancer exome files strongly correlates with reduced survival. *Cancer Microenviron* 11(1): 51-59, 2018. PMID: 29404962. DOI: 10.1007/s12307-018-0205-5
  - 18 Mai AT, Tong WL, Tu YN and Blanck G: Tcr-alpha recombinations in renal cell carcinoma exome files correlate with an intermediate level of t-cell exhaustion biomarkers. *Int Immunol* 30(1): 35-40, 2018. PMID: 29361059. DOI: 10.1093/intimm/dxx074
  - 19 Tong WL, Tu YN, Samy MD, Sexton WJ and Blanck G: Identification of immunoglobulin v(dj) recombinations in solid tumor specimen exome files: Evidence for high level b-cell infiltrates in breast cancer. *Hum Vaccin Immunother* 13(3): 501-506, 2017. PMID: 28085544. DOI: 10.1080/21645515.2016.1246095
  - 20 Tu YN, Tong WL, Samy MD, Yavorski JM, Kim M and Blanck G: Assessing microenvironment immunogenicity using tumor specimen exomes: Co-detection of tcr-alpha/beta v(dj) recombinations correlates with pd-1 expression. *Int J Cancer* 140(11): 2568-2576, 2017. PMID: 28256716. DOI: 10.1002/ijc.30675
  - 21 Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C and Schultz N: Integrative analysis of complex cancer genomics and clinical profiles using the cbiportal. *Sci Signal* 6(269): p11, 2013. PMID: 23550210. DOI: 10.1126/scisignal.2004088
  - 22 Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C and Schultz N: The cbio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2(5): 401-404, 2012. PMID: 22588877. DOI: 10.1158/2159-8290.CD-12-0095
  - 23 Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M and Kohlbacher O: Optitype: Precision hla typing from next-generation sequencing data. *Bioinformatics* 30(23): 3310-3316, 2014. PMID: 25143287. DOI: 10.1093/bioinformatics/btu548
  - 24 Iglesia MD, Parker JS, Hoadley KA, Serody JS, Perou CM and Vincent BG: Genomic analysis of immune cell infiltrates across 11 tumor types. *J Natl Cancer Inst* 108(11), 2016. PMID: 27335052. DOI: 10.1093/jnci/djw144
  - 25 Chu PG and Arber DA: Cd79: A review. *Appl Immunohistochem Mol Morphol* 9(2): 97-106, 2001. PMID: 11396639. DOI: 10.1097/00129039-200106000-00001
  - 26 Roca AM, Chobrutskiy BI, Callahan BM and Blanck G: T-cell receptor v and j usage paired with specific hla alleles associates with distinct cervical cancer survival rates. *Hum Immunol* 80: 237-242, 2019. PMID: 30684568. DOI: 10.1016/j.humimm.2019.01.005
  - 27 Tong WL, Callahan BM, Tu YN, Zaman S, Chobrutskiy BI and Blanck G: Immune receptor recombinations from breast cancer exome files, independently and in combination with specific hla alleles, correlate with better survival rates. *Breast Cancer Res Treat* 173: 167-177, 2019. PMID: 30229447. DOI: 10.1007/s10549-018-4961-1
  - 28 Chobrutskiy BI, Zaman S, Tong WL, Diviney A and Blanck G: Recovery of t-cell receptor v(dj) recombination reads from lower grade glioma exome files correlates with reduced survival and advanced cancer grade. *J Neurooncol* 140(3): 697-704, 2018. PMID: 30382482. DOI: 10.1007/s11060-018-03001-1
  - 29 Chobrutskiy BI, Zaman S, Diviney A, Mihyu MM and Blanck G: T-cell receptor-alpha cdr3 domain chemical features correlate with survival rates in bladder cancer. *J Cancer Res Clin Oncol* 145(3): 615-623, 2019. PMID: 30539280. DOI: 10.1007/s00432-018-2815-1
  - 30 Schmidt M, Hellwig B, Hammad S, Othman A, Lohr M, Chen Z, Boehm D, Gebhard S, Petry I, Lebrecht A, Cadenas C, Marchan R, Stewart JD, Solbach C, Holmberg L, Edlund K, Kultima HG, Rody A, Berglund A, Lambe M, Isaksson A, Botling J, Karn T, Muller V, Gerhold-Ay A, Cotarelo C, Sebastian M, Kronenwett R, Bojar H, Lehr HA, Sahin U, Koelbl H, Gehrman M, Micke P, Rahnenfuhrer J and Hengstler JG: A comprehensive analysis of human gene expression profiles identifies stromal immunoglobulin kappa c as a compatible prognostic marker in human solid tumors. *Clin Cancer Res* 18(9): 2695-2703, 2012. PMID: 22351685. DOI: 10.1158/1078-0432.CCR-11-2210
  - 31 Whiteside TL and Ferrone S: For breast cancer prognosis, immunoglobulin kappa chain surfaces to the top. *Clin Cancer Res* 18(9): 2417-2419, 2012. PMID: 22442060. DOI: 10.1158/1078-0432.CCR-12-0566

Received January 27, 2020

Revised February 24, 2020

Accepted March 4, 2020