Breast Cancer Research and Treatment (2019) 173:167–177
https://doi.org/10.1007/s10549-018-4961-1

EPIDEMIOLOGY

Immune receptor recombinations from breast cancer exome files, independently and in combination with specific HLA alleles, correlate with better survival rates

Wei Lue Tong[1] · Blake M. Callahan[1] · Yaping N. Tu[1] · Saif Zaman[1] · Boris I. Chobrutskiy[1] · George Blanck[1,2]

Received: 2 July 2018 / Accepted: 6 September 2018 / Published online: 18 September 2018

Abstract

Purpose Immune characterizations of cancers, including breast cancer, have led to information useful for prognoses and are considered to be important in the future of refining the use of immunotherapies, including immune checkpoint inhibitor therapies. In this study, we sought to extend these characterizations with genomics approaches, particularly with cost-effective employment of exome files.

Methods By recovery of immune receptor recombination reads from the cancer genome atlas (TCGA) breast cancer dataset, we observed associations of these recombinations with T-cell and B-cell biomarkers and with distinct survival rates.

Results Recovery of TRD or IGH recombination reads was associated with an improved disease-free survival ($p=0.047$ and 0.045, respectively). Determination of the HLA types using the exome files allowed matching of T-cell receptor V- and J-gene segment usage with specific HLA alleles, in turn allowing a refinement of the association of immune receptor recombination read recoveries with survival. For example, the TRBV7, HLA-C*07:01 combination represented a significantly worse, disease-free outcome ($p=0.014$) compared to all other breast cancer samples. By direct comparisons of distinct TRB gene segment usage, HLA allele combinations revealed breast cancer subgroups, within the entire TCGA breast cancer dataset with even more dramatic survival distinctions.

Conclusions In sum, the use of exome files for recovery of adaptive immune receptor recombination reads, and the simultaneous determination of HLA types, has the potential of advancing the use of immunogenomics for immune characterization of breast tumor samples.

Keywords Breast cancer · Immune receptor recombinations · V- and J-gene segment usage · HLA alleles

Abbreviations

BCR	B-cell receptor
BRCA	Breast cancer
DFS	Disease-free survival
ER	Estrogen receptor
GDC	Genomic data commons
HLA	human leukocyte antigen
IGH	Immunoglobulin heavy gene
IGK	Immunoglobulin kappa gene
IGL	immunoglobulin lambda gene
IMGT	ImmunoGeneTics organization
KM	Kaplan–Meier
OS	Overall survival
PR	Progesterone receptor
TCGA	The cancer genome atlas
TIL	Tumor-infiltrating lymphocyte
TCR	T-cell receptor
TNBC	Triple-negative breast cancer (negative for ER, PR, and HER2)
TRA	T-cell receptor alpha gene
TRB	T-cell receptor beta gene
TRG	T-cell receptor gamma gene
TRD	T-cell receptor delta gene
WXS	Whole exome file

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10549-018-4961-1) contains supplementary material, which is available to authorized users.

✉ George Blanck
gblanck@health.usf.edu

[1] Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, USA

[2] Department of Immunology, H. Lee Moffitt Cancer Center and Research Institute, 12901 Bruce B. Downs Blvd., Tampa, FL 33612, USA

Introduction

Immune characterization of cancer specimens has the potential of guiding therapies or informing prognoses. The role of humoral immunity in suppression of immunogenic breast cancers has been previously established, for example, with correlations between increased presence of IGK and improved prognosis [1–3].

However, the potential correlation of the presence of T-cell receptor δ (TRD) and γ (TRG) with survival rates for breast cancer remains undetermined. In other cancers, the presence of tumor-infiltrating lymphocytes (TILs) expressing γδ TCRs has correlated with increased survival, for example, in bladder cancer and melanoma [4], although in prostate cancer the presence of γδ TCRs correlated with decreased survival [5]. In breast cancer, both better and worse outcomes have been reported in the cases of infiltration of γδ TILs [4]. In one case, Vδ1 + γδ TILs have been postulated to contribute to the arrest of tumor rejection via IL-10 signaling to dendritic cells and T lymphocytes in mouse models [4]. Thus, in this report, we take advantage of an immunogenomics approach for the analysis of the very large cancer genome atlas (TCGA) breast cancer (BRCA) dataset to further shed light on the role of γδ TILs.

In addition, we recently demonstrated that the usage of T-cell receptor-β (TRB) J sub-segments (i.e., segments from either the TRBJ1 or TRBJ2 groups), in combination with particular HLA class II alleles, or TRB V-segments in combination with particular HLA alleles, correlate with either better or worse cancer survival rates in studies that did not include BRCA [6, 7]. BRCA TRBV and TRBJ usage, in combination with particular HLA alleles, is addressed in this report. Thus, in addition to providing new information regarding γδ TILs, the large number of TCGA-BRCA samples has facilitated a statistically significant alignment of survival rates with TRB gene segment usage, either alone or in combination with specific HLA alleles.

Methods

Identification of reads representing V(D)J recombinations in WXS files

File slices from a total of 1065 whole exome sequence (WXS) files for breast carcinoma were downloaded from the Genomic Data Commons (GDC) (TCGA-BRCA) under dbGaP approval #6300, using a manifest file generated at the GDC web site. A full list of barcodes from this manifest is in Table S1. Each WXS file slice represented a five million base pair (bp) region surrounding each of the seven immune receptor coding sections of the genome, based on the hg38 reference genome. The nucleotide sequences (used for searching the WXS files) for the V and J regions of human immunoglobulin heavy chain, kappa chain, and light chain genes (IGH, IGK, IGL) represented the IMGT repertoire. Similar sequences were used for the TCR α, β, γ, and δ chain genes (TRA, TRB, TRG, TRD). For IGH and TRG, the positive strand sequence was used for the search, i.e., representing the reverse complement of the coding regions for these genes. For each V sequence, a total of five individual 10 bp length sequences were generated, at 3, 5, 7, 9, and 11 nucleotides away from the 3′ end of the V sequence to account for N-region diversity and rare polymorphisms. Likewise, for each J sequence, five individual 10 bp length sequences were generated, at 3, 5, 7, 9, and 11 nucleotides away from the 5′ end of the J sequence. A shell script (Module_Search_IgTcR) was written to execute individual search scripts (Search_Reads.sh) via GNU Parallel in a Unix shell running on an Xen platform optimized for distributed computing [8, 9]. The search script compared each read in the five million base pair region encompassing each immune receptor genome section to the list of 10 bp V and J sequences. Each read that matched at least one 10 bp V sequence and at least one 10 bp J sequence was written to a tab separated values (TSV) file. For each of the seven immune receptors, the TSV files were combined into a single fasta file via tsv2fasta.php. The fasta file was broken up into individual fasta files up to 500,000 sequences in size if the primary output files represented more than 500,000 sequences. Each fasta file was submitted to IMGT/HighV-QUEST for secondary analysis. The output from IMGT/HighV-QUEST was used to generate individual "matchlist" files, via imgtHVQuest2matchlist.php, representing a table with each read and the IMGT result for assessment of read parameters using Excel. For this project, only recombinations containing at least 15 matching nucleotides for both the V- and J-gene segments were included in subsequent data analyses, to reduce the impact of false positives. See also ref [10]. All computer codes, including the above V and J search strings, are available upon request by email to corresponding author.

RNASeq data

The RNASeq Version 2 Remote Scanning Electron Microscopy (RNASeq V2 RSEM) scores for the genes indicated in **Results** were downloaded from cBioPortal.org for the TCGA-BRCA (Provisional) dataset [11–13]. For each immune receptor sequence, two groups of patient barcodes were created for comparison: a group of patient barcodes

whose WXS file contained a productive recombination for the immune receptor sequence and a group of all remaining patient barcodes. For each group, Microsoft Excel 2016 was used to calculate a simple average for RNA expression for each gene. A *t* test was performed on the two groups of barcodes to generate a *p* value for determination of statistical significance ($p < 0.05$) (Tables S2, S3).

HLA data

The HLA class I (HLA-A, HLA-B, HLA-C) and class II (HLA-DPB1, -DQB1, -DRB1) alleles for each of the BRCA WXS files were determined via xHLA as obtained from Github and described in Xie et al [14] and Callahan et al [15]. Alleles expressed in a large number of patients were used to create groups for further data analysis. (The HLA types for the entire set BRCA barcodes are available upon request by email to corresponding author.)

Survival data

For survival analyses, the cBioPortal web tool was used for preliminary analyses. Each set of patient barcodes was compared against all remaining patient barcodes, as detailed in **Results**. Then, each group of barcodes with statistically significant ($p < 0.05$) survival distinctions was independently confirmed using GraphPad Prism (Tables S4–S13). For analysis of the HLA data, GraphPad Prism software was also used to compare the survival between two subsets of barcodes, as the cBioPortal web tool provides only the opportunity to compare one subset of barcodes against all remaining barcodes.

Results

As detailed in **Methods** and refs. [5, 16–18], we searched the TCGA set of BRCA WXS files for immune receptor recombination reads representing minimum criteria, for example, a minimum number of nucleotides used to establish a match to a given V or J region gene segment. A summary of the recombination reads recovered from the WXS files containing productive recombinations, as well as the number of BRCA barcodes representing productive and unproductive recombinations, is provided in Table 1 (Tables S14, S15). To support the indication that recovery of immune receptor recombination reads represents the presence of TILs in the tumor samples, we analyzed RNASeq data for T- and B-lymphocyte markers. For almost all of the lymphocyte RNASeq data, the T- and B-lymphocyte biomarkers were consistent with the recovery of the TCR and BCR recombination reads (Tables 2, 3; Tables S2, S3).

We next examined the RNASeq values for several proteins involved in immune modulation and antigen presentation, for an attempted correlation with barcodes representing WXS files found to have productive TRB receptor recombinations. Detection of productive TRB-VDJ recombinations was found to correlate with approximately twice the level of PD-L1 (CD274) RNA expression (Table 4). Likewise, statistically significant increases in RNA expression were found for CD33 (reflective of myeloid cells) [19], CIITA (a positive regulator of HLA class II gene transcription) [20], CTLA4 (an immune checkpoint receptor) [21], and CTSL in barcodes representing productive TRB recombinations (Table 4).

We next examined the potential correlations between the presence of specific TCR and BCR recombination reads and survival rates. Detection of productive or unproductive TRD recombinations among the entire WXS set correlated with an improved disease-free survival (DFS) ($p = 0.035$) (Table 5; Fig. 1a), and detection of productive TRD recombination reads alone correlated with improved DFS ($p = 0.047$) (Table 5; Fig. 1b). Detection of TRG productive recombinations trended towards improved DFS for invasive ductal carcinoma subset only ($p = 0.073$, unpublished observations).

Co-detection of productive and unproductive IGK recombination reads, i.e., co-detection within one WXS file, representing a B-cell infiltrate [18], trended towards better overall

Table 1 Summary of immune receptor recombination reads obtained from TCGA-BRCA WXS files (Tables S13, S14)

Immune receptor gene	Productive recombinations	Unproductive recombinations	Barcodes with productive recombinations	Barcodes with unproductive recombinations
IGH	283	62	155	54
IGK	305	388	176	177
IGL	690	279	234	127
TRA	4673	73,788	883	1045
TRB	1519	351	519	217
TRD	261	144	162	104
TRG	2602	6135	602	789
TRA and TRB	ND	ND	476	217

Table 2 Summary of RNASeq values for T-cell biomarkers

	CD3D	CD4	CD8A	ICOS	IFNG
Mean for barcodes w/productive TRA recombination reads	148.95	1447.72	311.32	53.52	451.82
Mean for all remaining barcodes	58.41	1028.14	141.52	18.80	328.82
T test *p* value	5.21E-19	4.74E-07	3.64E-16	4.97E-23	5.16E-07
Mean for barcodes w/productive TRB recombination reads	206.62	1801.88	426.29	73.77	13.79
Mean for all remaining barcodes	65.66	978.54	148.60	23.38	3.63
T test *p* value	1.33E-27	5.12E-38	1.06E-30	1.18E-28	3.56E-13
Mean for barcodes w/productive TRD recombination reads	261.29	2085.76	554.68	100.72	21.16
Mean for all remaining barcodes	112.12	1256.19	236.45	38.67	6.37
T test *p* value	2.24E-07	8.46E-12	3.21E-09	3.40E-10	3.45E-07
Mean for barcodes w/productive TRG recombination reads	187.44	1667.86	394.98	68.67	12.80
Mean for all remaining barcodes	65.07	1003.99	138.90	20.85	3.06
T test *p* value	1.20E-25	5.08E-27	3.10E-32	1.09E-31	6.14E-15

Table 3 Summary of RNASeq values for B-cell markers

	CD19	CD22	CD38	CD72	CD79A	CD79B	CR2	MS4A1
Mean for barcodes w/productive IGH recombination reads	173.84	488.50	189.89	174.04	1062.66	424.62	483.95	980.51
Mean for all remaining barcodes	18.64	158.77	48.66	75.57	159.37	88.95	41.31	90.24
T test *p* value	4.55E-08	8.50E-05	3.71E-14	1.48E-08	5.81E-12	2.33E-08	1.09E-03	1.25E-05
Mean for barcodes w/productive IGK recombination reads	138.95	415.88	179.60	160.64	888.08	339.55	362.63	757.78
Mean for all remaining barcodes	21.72	165.02	47.19	75.79	171.64	97.53	54.44	112.40
T test *p* value	1.28E-06	3.96E-04	3.22E-14	3.35E-08	2.21E-10	1.67E-06	7.01E-03	1.94E-04
Mean for Barcodes w/IGL Productive	124.23	366.58	168.59	150.04	789.18	318.61	331.96	658.23
Mean for all remaining barcodes	17.79	161.62	41.18	72.93	150.15	86.76	41.85	95.99
T test *p* value	1.08E-08	1.34E-04	3.30E-17	1.75E-10	1.50E-12	1.34E-08	1.10E-03	1.40E-05

Table 4 Summary of RNASeq expression data for immune checkpoint genes

	CD274	CD33	CIITA	CTLA4	CTSL
Mean for barcodes w/productive TRB recombination reads	44.30	85.18	473.69	71.73	2295.74
Mean for all remaining barcodes	22.61	58.11	221.45	26.04	1835.55
T test *p* value	2.89E-13	1.03E-15	2.81E-28	4.06E-28	6.43E-09

survival (OS) and DFS ($p = 0.06$ and 0.052, respectively) (Table 5). While neither detection of productive IGH nor productive IGK recombinations separately, statistically significantly correlated with survival, the combination of these two sets of WXS files (i.e., either productive IGH or IGK) correlated with increased DFS ($p = 0.047$)(Fig. 2). It should also be noted that recovery of IGH or IGK recombination reads, as independent barcodes sets, did trend towards increased DFS (Table 5).

We next examined the survival distinctions represented by the HLA alleles in the BRCA dataset. A statistically significant survival distinction was found, for barcodes representing HLA-DRB1*04:01 (Table 6), the occurrence of which correlated with decreased DFS compared with all remaining barcodes ($p = 0.033$). However, for barcodes representing the HLA-DRB1*13:01 allele, there was a positive outcome for OS ($p = 0.037$) as well as for DFS ($p = 0.025$) (Table 6).

We next considered the possibility of identifying V or J usage, HLA allele combinations that correlated with survival rates, as indicated by analyses of other cancer datasets [6, 7]. A total of three combinations of specific TRB V or J sub-segment usage (i.e., J1 segments or J2 segments) and HLA alleles were identified as showing a statistically

Table 5 Survival associations with groups of barcodes representing the recovery of recombination reads from BRCA WXS files, compared to all remaining barcodes

	Overall survival		Disease-free survival	
	Survival benefit	*p* value	Survival benefit	*p* value
TRD (productive recombination) (Table S4)	Equivocal	0.88	Positive	*0.047*
TRD (barcodes representing either productive or unproductive recombinations) (Table S5)	Equivocal	0.64	Positive	*0.035*
IGK (productive recombinations)	Equivocal	0.80	Positive	0.09
IGK (unproductive recombinations)	Positive	0.08	Positive	0.09
IGK (barcodes representing either productive or unproductive recombinations)	Equivocal	0.68	Positive	0.08
IGK (barcodes representing the recovery of both productive and unproductive recombinations within one WXS file)	Positive	0.06	Positive	0.052
IGH or IGK (barcodes representing productive recombinations) (Table S6)	Positive	0.28	Positive	*0.047*

$p < 0.05$ in italics

All $p < 0.05$ were confirmed with the cBioPortal web tool and GraphPad Prism

GraphPad Prism output for $p < 0.05$ is available in the indicated SOM tables, left column

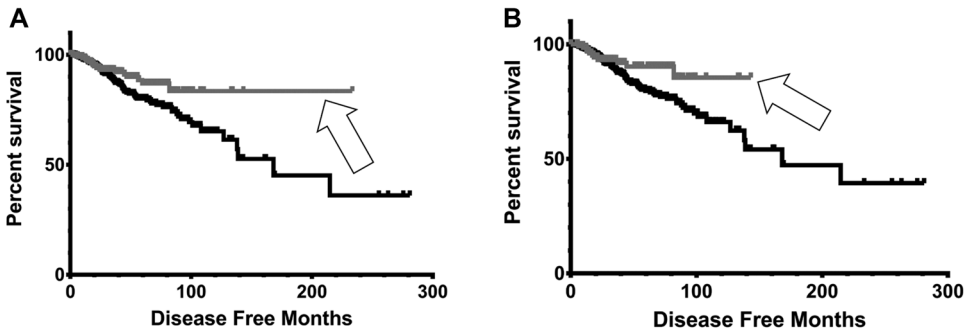

Fig. 1 Survival outcomes associated with the recovery of TRD recombination reads. **a** Kaplan–Meier (KM) DFS curve for barcodes representing either TRD productive or unproductive recombinations (gray, arrow) versus all remaining barcodes (black) ($p = 0.035$). **b** KM DFS curve for barcodes representing TRD productive recombinations (gray, arrow) versus all remaining barcodes (black) ($p = 0.047$)

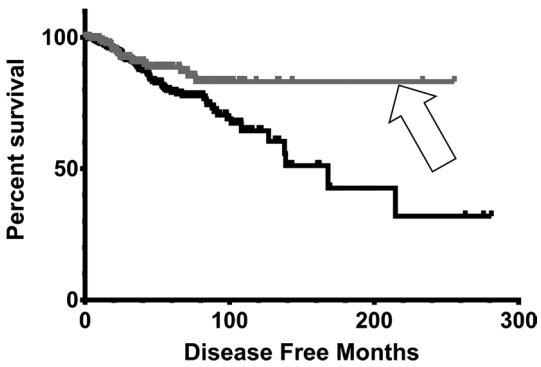

Fig. 2 Survival outcomes associated with the recovery of BCR recombination reads. KM DFS curve for barcodes representing either IGH productive or IGK productive recombinations (gray, arrow) versus all remaining samples (black) ($p = 0.047$)

significant associations with survival rates (Table 6). Specifically, in these cases, neither the V or J usage alone, nor the HLA allele alone, represented a statistically significant survival distinction. Kaplan–Meier (KM) survival curves were generated (and statistically analyzed using GraphPad Prism) for barcodes representing the gene segment usage, HLA allele combinations compared to all remaining barcodes. To verify the results, independent analyses of survival rates were performed using the KM web tool at cBioPortal.org [11–13].

In particular, for barcodes representing the TRBV5, HLA-B*07:02 combination, neither TRBV5 usage alone, nor HLA-B*07:02 allele alone, represented survival distinctions for OS or DFS. However, barcodes with the indicated TRB gene segment, HLA allele combination had a statistically significant reduced DFS ($p = 0.026$, Table 6). The TRBV7, HLA-C*07:01 combination was associated with reduced DFS ($p = 0.014$, Table 6), and the TRBJ2, HLA-DQB1*06:03 combination was associated with increased

Table 6 Survival patterns of BRCA barcode groups representing TRB-VDJ recombinations in combination with specific HLA class I and II alleles

TRB gene segment, HLA class I or II allele combinations	Number of barcodes representing left column combination	Overall survival in comparison to all remaining barcodes, *p* value	Disease-free survival in comparison to all remaining barcodes, *p* value	Overall survival outcome: P positive; N negative, E equivocal	Disease-free survival outcome: P positive; N negative, E equivocal
*TRBV5, HLA-B*07:02 (Table S7)*	25	0.86	*0.026*	E	N
TRBV6, HLA-C*07:02	67	0.055	0.32	N	E
*TRBV7, HLA-C*07:01 (Table S8)*	25	0.69	*0.014*	E	N
*TRBJ2, HLA-DQB1*06:03 (Table S9)*	53	0.058	*0.021*	E	P
TRBJ1, HLA-DRB1*04:01 *NOT an independently, statistically significant combination; see two rows below. See also* Fig. 3C, D *(Table S10)*	36	0.048	0.007	N	N
TRBJ2, HLA-DRB1*13:01 *NOT an independently, statistically significant combination; see two rows below See also* Fig. 3C, D *(Table S11)*	50	0.055	0.021	E	P
*HLA-DRB1*04:01 (Table S12)*	64	0.20	*0.033*	N	N
*HLA-DRB1*13:01 (Table S13)*	63	*0.037*	*0.025*	P	P

$p < 0.05$ in italics

All $p < 0.05$ were confirmed with the cBioPortal web tool and GraphPad Prism

GraphPad Prism output for $p < 0.05$ is available in the indicated SOM tables, left column

DFS ($p = 0.021$, Table 6), with a trend towards increased overall survival ($p = 0.058$).

KM survival curves were generated to compare the relatively large survival differences between the above combinations associated with positive outcomes versus combinations associated with negative outcomes (Figs. 3a–d, 4a–e, *p* values indicated in the figure legends).

To assess for the possibility of clinical characteristics with known prognostic value overlapping the above survival distinctions, TCR and BCR groups found to have a significant correlation with DFS were stratified by clinical characteristics available from TCGA: expression of estrogen receptor (ER), progesterone receptor (PR), HER2, clinical stage, and breast cancer type. Neoadjuvant chemotherapy was only performed on two samples, and thus not included in this analysis. Due to the limited sample sizes, KM curves could not be generated for subgroup analysis (e.g., DFS of barcodes with TRD productive reads and expressing ER versus DFS of barcodes with TRB productive reads and not expressing ER). Thus, the DFS ratio (number of barcodes with DFS at a time point divided by total barcodes) was employed as a proxy for KM analysis. An arbitrary time point of 100 months was selected. The percentage of barcodes in each group expressing the clinical characteristic, as well as the DFS ratio of the subgroup at 100 months was calculated. Extensive data are shown in Table S6 in the SOM. No overt bias is noted in the percentage of each recombination read recovery group having a particular clinical characteristic compared to all samples. Moreover, the DFS ratio for each recombination read recovery subgroup remained higher

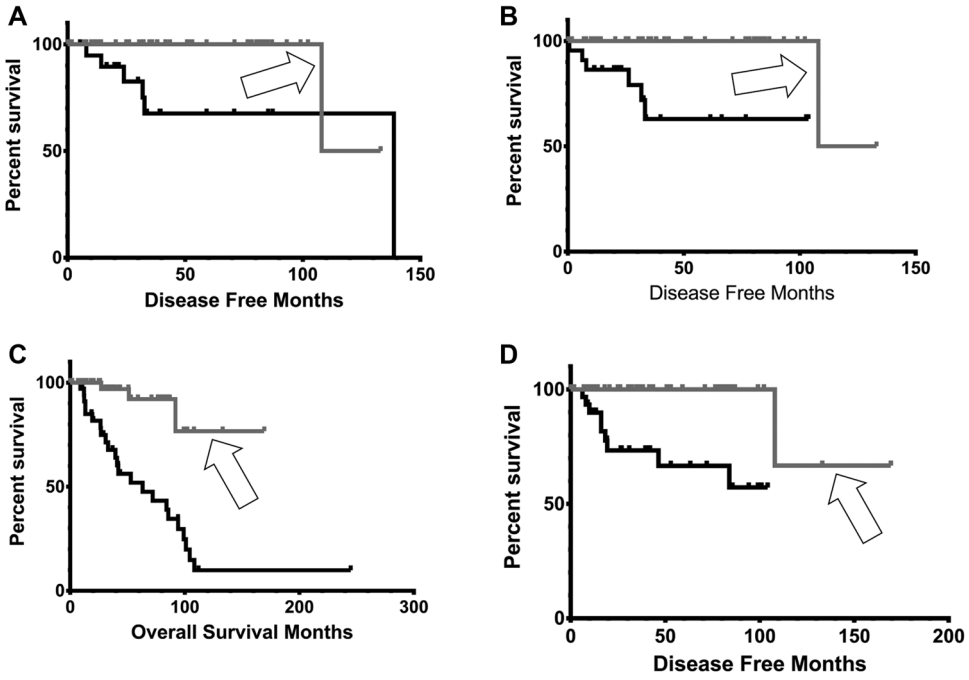

Fig. 3 Contrasting survival outcomes for specific TRBV or TRBJ, HLA combinations. **a** KM DFS curve for barcodes representing the TRBJ2, HLA-DQB1*06:03 combination (gray, arrow) versus the TRBV5, HLA-B*07:02 combination (black) ($p=0.021$). **b** KM DFS curve for barcodes representing the TRBJ2, HLA-DQB1*06:03 combination (gray, arrow) versus the TRBV7, HLA-C*07:01 combination (black) ($p=0.014$). **c** KM OS curve for barcodes representing the TRBJ2, HLA-DRB1*13:01 combination (gray, arrow) versus the TRBJ1, HLA-DRB1*04:01 combination (black) ($p=0.001$). See also Table 6, for additional TRBJ2, HLA-DRB1*13:01 information and additional TRBJ1, HLA-DRB1*04:01 information. **d** KM DFS curve for barcodes representing the TRBJ2, HLADRB1*13:01 combination (gray, arrow) versus the TRBJ1-DRB1*04:01 combination (black) ($p=0.0002$)

than the DFS ratio of all barcodes following stratification, indicating that the results seen in Table 5 do not represent 100% overlaps with known clinical prognostic features.

Further assessment of the results from the DFS ratios and clinical correlations was performed specifically for the DFS ratio of triple-negative breast cancer (TNBC). As shown in Table 7, barcodes with recovery of IGK productive reads do not significantly increase the DFS ratio for TNBC. However, there is a significant increase in the DFS ratio for TNBC for barcodes with recovery of TRD productive reads, from a DFS ratio of 0.68 for all TNBC barcodes to 0.93 (population proportions test, $p<0.002$). KM analyses was consistent with this distinction, although not enough TNBC samples were available to achieve statistical significance ($p=0.07$, unpublished observations).

Discussion

The above data reinforce the value of the exome as a source of cancer immunology information, particularly when dealing with a large cohort of patients. Specifically, extensive work, prior to this report, established the very high likelihood of immune receptor recombination reads being reflective of lymphocytes, representing relatively dominant TCR and BCR recombinations, present in the tumor and related to tumor growth or reduced growth, depending on the case. For example, a recent report has indicated that higher levels of bacteria in pancreatic cancer correlate with a worse outcome [22], and befitting the Th2 response that would be expected in such cases, high levels of BCR recombination read recoveries are also associated with a worse pancreatic cancer outcome [23]. In fact, recoveries of the BCR recombination reads from the exome files led to a much stronger correlation with outcomes than did all other studied, B-cell markers [23]. Many other sets of exome files, yielding TRB and BCR recombination reads, have reflected the usefulness and power of the pancreatic cancer example [5, 16–18, 24–29].

There were three basic conclusions regarding breast cancer from this study. First, the recovery of immune receptor recombination reads representing the γ/δ T-cell receptor was associated with a better outcome. Unfortunately, the number of samples, and the number of reads recovered, did not permit focusing on T-cell or NKT-cell subsets that might provide further insight into the possible mechanisms of the effect of longer survival. Also, in comparison to α/β T-cells, the number of previously described biomarkers uniquely

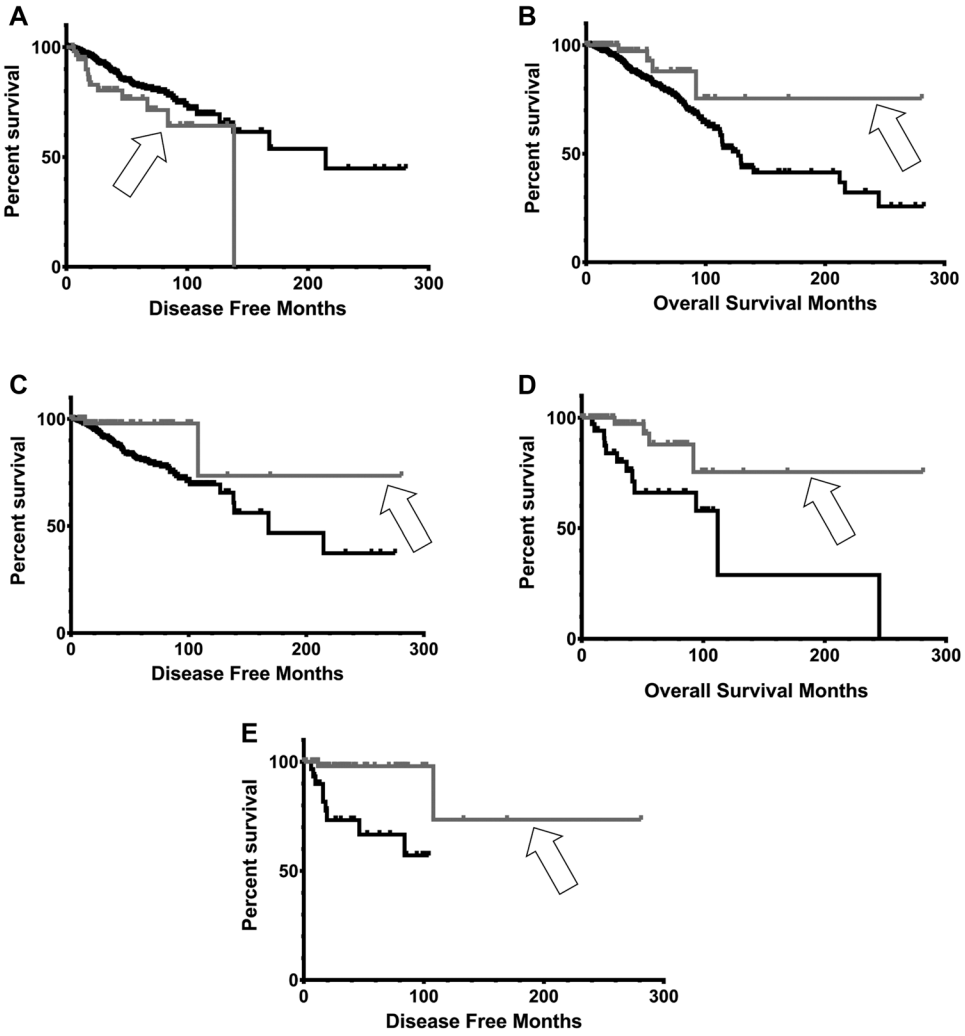

Fig. 4 Contrasting survival outcomes associated with the HLA-DRB1*13:01 and HLA-DRB1*04:01 alleles. **a** KM DFS curve for barcodes with HLA-DRB1*04:01 allele (gray, arrow) versus all remaining barcodes (black) ($p = 0.033$, Table 6). **b** KM OS curve for barcodes with HLA-DRB1*13:01 allele (gray, arrow) versus all remaining barcodes (black) ($p = 0.037$, Table 6). **c** KM DFS curve for barcodes with HLA-DRB1*13:01 allele (gray, arrow) versus all remaining barcodes (black) ($p = 0.025$, Table 6). **d** KM OS curve for barcodes with HLA-DRB1*13:01 allele (gray, arrow) versus HLA-DRB1*04:01 allele (black) ($p = 0.003$). **e** KM DFS curve for barcodes with HLA-DRB1*13:01 allele (gray, arrow) versus HLA-DRB1*04:01 allele (black) ($p = 0.006$)

Table 7 Recovery of TRD recombination reads from BRCA WXS files where samples where characterized as negative for the estrogen receptor, progesterone receptor, and HER2 (TNBC)

	Percentage of barcodes identified as TNBC	TNBC DFS ratio at 100 months: ratio of barcodes with immune receptor recombination read recovery to total barcodes for TNBC
All barcodes	15	0.68[a]
IGK productive	27	0.72
IGH or IGK productive	26	0.72
TRD productive	22	0.93
TRD productive or unproductive	20	0.94[a]

[a]n-proportion test p value < 0.002

reflecting T-cell subsets with TRG or TRD recombinations is minimal at best [30]. Thus, the above data regarding breast cancer and γ/δ T-cell receptors will not resolve any of the controversy referred to in the **Introduction**, regarding γ/δ T-cells with specific V or J usage. However, it is clear that the mining of immune receptor recombinations from exome files does offer an assay opportunity for future studies that hopefully will, in aggregate, lead to a resolution to this issue or possibly to a definition of different states of disease that have distinct responses to distinct subsets of γ/δ T-cells.

The second conclusion, that breast cancer has a significant B-cell component, substantiates previous work, and in particular, substantiates the proposal that the presence of B-cells has a protective effect [1, 31]. Thus, the above work

provides a monitoring opportunity for clinical trials, among many other opportunities, whereby exomes can be mined for BCR recombination reads, particularly in cases where other methods for obtaining BCR recombination reads are not available, to assess therapy impacts on B-cell presence in breast cancer.

The third conclusion provided by the data above is that TRB V and J usage, in combination with specific HLA alleles, associates with survival outcomes. This result is consistent with TRB V and J usage, HLA allele combinations in other immune settings, such as infection and autoimmunity [32–34]; is consistent with related results associating TRB V and J usage with HLA class type [35]; and is consistent with similar results obtained for other cancer datasets [6, 7]. The limitation of these results, however, is the lack of a replicative dataset. In fact, this may turn out to be a very long-term or indefinite limitation, due to the improbability of assembling two very large groups of patients with the same HLA allele distributions, for two copies of six HLA genes; and with an apparently similar collection of dominant TRB recombinations, among persons with breast cancer. However, the TRB V and J usage combinations indicated here, in combination with particular HLA alleles, can be individually re-evaluated as the opportunities arise, either with specific contrasts to other TRB V and J usage, HLA combinations in focused clinical trials, or via statistically significant results from other large datasets. Regardless, it must be noted that, given the basic mechanisms of function of the adaptive immune response and the understanding of cancer as interacting with the adaptive immune response, such TRB V and J usage, HLA allele combinations associating with cancer progression, one way or another, are inevitable. Also, continuing with mechanistic aspects of the here reported associations, it should be noted that the above TRBJ, HLA allele associations do not distinguish the individual TRB J segments within the J1 and J2 families. Having said that there are amino acid (AA) sequence distinctions between these two families, such that certain members of the J1 family have AA sequences that do not appear in the J2 family and vice versa [6, 7]. Thus, if the reported TRBJ2, HLA-DQB1*06:03 combination (Table 6) functions by efficient interaction of J2 family members with the HLA-DQB1*06:03 allelic variant of the HLA-DQB1 molecule, then presumably the J2 family members, with the four AA that specifically distinguish these J2 family members from J1 family members, would be likely candidates for further resolution of the TRBJ2, HLA-DQB1*06:03 associations.

The above analyses do not have the power to rule out the possibility that recovery of immune receptors from the WXS files overlaps with or distinguishes other prognostic variables. The stratification by clinical characteristics did not reveal any certain confounders (Table S16), possibly due to the relatively small sample size, although data in this report provide preliminary evidence that recovery of TRD recombination reads may in fact be an independent prognostic indicator.

In TNBC, TILs present at diagnosis have been associated with increased DFS [36]. Recovery of TRD recombination reads was significantly associated with increased DFS for TNBC (Table 7), possibly consistent with the association of the presences of T-cells with better TNBC DFS.

A recent systematic review [37] showed that the level of stromal TILs was an independent predictive and prognostic marker for better survival outcomes in HER2+ breast cancer patients receiving Trastuzumab therapy. Likewise, other studies [38] found that in patients with HER2+ disease, increased TILs were associated with increased response to high-dose anthracyclines, another highly immunogenic therapy. Loi et al. [36] demonstrated that using the amount of lymphocyte infiltration as stratification for guiding clinical decision making for selection of immunogenic therapies for an individual patient improves outcomes. However, in many clinical settings, quantification of TILs has been challenging, as many have reported poor repeatability of the quantification of intra-tumoral TILs by pathologists due to issues with random sampling, limiting quantification to stromal TILs [39]. The methods outlined above using next-generation sequencing of samples represent an alternative method, perhaps a superior method.

Acknowledgements Authors gratefully acknowledge the support of USF research computing and the taxpayers of the State of Florida. This study is dedicated to Frances.

Compliance with ethical standards

Conflict of interest Authors have nothing to declare.

Research involving human participants and/or animals Not applicable. (non-human subjects research).

Informed consent Not applicable.

References

1. Whiteside TL, Ferrone S (2012) For breast cancer prognosis, immunoglobulin kappa chain surfaces to the top. Clinical Cancer Res 18:2417–2419
2. Kato T, Park JH, Kiyotani K, Ikeda Y, Miyoshi Y, Nakamura Y (2017) Integrated analysis of somatic mutations and immune microenvironment of multiple regions in breast cancers. Oncotarget 8:62029–62038
3. 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, Gehrmann M, Micke P, Rahnenfuhrer J, Hengstler JG (2012) A comprehensive analysis of

human gene expression profiles identifies stromal immunoglobulin kappa C as a compatible prognostic marker in human solid tumors, Clinical Cancer Res 18: 2695–2703
4. Fournie JJ, Sicard H, Poupot M, Bezombes C, Blanc A, Romagne F, Ysebaert L, Laurent G (2013) What lessons can be learned from gammadelta T cell-based cancer immunotherapy trials? Cell Mol Immunol 10:35–41
5. Tu YN, Tong WL, Yavorski JM, Blanck G (2018) Immunogenomics: a negative prostate cancer outcome associated with TcR-gamma/delta recombinations. Cancer Microenviron 11(1):41–49
6. Callahan BM, Yavorski JM, Tu YN, Tong WL, Kinskey JC, Clark KR, Fawcett TJ, Blanck G (2018) 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
7. Callahan BM, Tong WL, Blanck G (2018) T cell receptor-beta J usage, in combination with particular HLA class II alleles, correlates with better cancer survival rates. Immunol Res 66:219–223
8. Malaspinas AS, Tange O, Moreno-Mayar JV, Rasmussen M, DeGiorgio M, Wang Y, Valdiosera CE, Politis G, Willerslev E, Nielsen R. (2014) bammds: a tool for assessing the ancestry of low-depth whole-genome data using multidimensional scaling (MDS), Bioinformatics, 30 2962–2964
9. Tange O (2011) Gnu parallel-the command-line power tool. USENIX Mag 36:42–47
10. Govindan S, Choi J, Urgaonkar B, Sivasubramaniam A, Baldini A (2009) Statistical profiling-based techniques for effective power provisioning in data centers. Proceedings of the 4th ACM European conference on Computer systems, ACM, pp. 317–330
11. Ping Z, Siegal GP, Almeida JS, Schnitt SJ, Shen D (2014) Mining genome sequencing data to identify the genomic features linked to breast cancer histopathology. J Pathol Inf 5:3
12. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6:pl1
13. 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, Schultz N (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2:401–404
14. Xie C, Yeo ZX, Wong M, Piper J, Long T, Kirkness EF, Biggs WH, Bloom K, Spellman S, Vierra-Green C, Brady C, Scheuermann RH, Telenti A, Howard S, Brewerton S, Turpaz Y, Venter JC (2017) Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. Proc Natl Acad Sci USA 114:8059–8064
15. Callahan BM, Patel JS, Fawcett TJ, Blanck G (2018) Cytoskeleton and ECM tumor mutant peptides: increased protease sensitivities and potential consequences for the HLA class I mutant epitope reservoir. Int J Cancer 142:988–998
16. Gill TR, Samy MD, Butler SN, Mauro JA, Sexton WJ, Blanck G (2016) 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
17. Mai AT, Tong WL, Tu YN, Blanck G (2018) T-cell receptor-alpha recombinations in renal cell carcinoma exome files correlate with an intermediate level of T-cell exhaustion biomarkers. Int immunol 30:35–40
18. Tong WL, Tu YN, Samy MD, Sexton WJ, Blanck G (2017) Identification of immunoglobulin V(D)J recombinations in solid tumor specimen exome files: evidence for high level B-cell infiltrates in breast cancer. Human Vaccin Immunother 13:501–506
19. Christiansson L, Soderlund S, Mangsbo S, Hjorth-Hansen H, Hoglund M, Markevarn B, Richter J, Stenke L, Mustjoki S, Loskog A, Olsson-Stromberg U (2015) The tyrosine kinase inhibitors imatinib and dasatinib reduce myeloid suppressor cells and release effector lymphocyte responses. Mol Cancer Ther 14:1181–1191
20. Steimle V, Otten LA, Zufferey M, Mach B (1993) Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome), Cell 75:135–146
21. Chae YK, Arya A, Iams W, Cruz MR, Chandra S, Choi J, Giles F (2018) Current landscape and future of dual anti-CTLA4 and PD-1/PD-L1 blockade immunotherapy in cancer; lessons learned from clinical trials with melanoma and non-small cell lung cancer (NSCLC). J Immunother Cancer 6:39
22. Geller LT, Barzily-Rokni M, Danino T, Jonas OH, Shental N, Nejman D, Gavert N, Zwang Y, Cooper ZA, Shee K, Thaiss CA, Reuben A, Livny J, Avraham R, Frederick DT, Ligorio M, Chatman K, Johnston SE, Mosher CM, Brandis A, Fuks G, Gurbatri C, Gopalakrishnan V, Kim M, Hurd MW, Katz M, Fleming J, Maitra A, Smith DA, Skalak M, Bu J, Michaud M, Trauger SA, Barshack I, Golan T, Sandbank J, Flaherty KT, Mandinova A, Garrett WS, Thayer SP, Ferrone CR, Huttenhower C, Bhatia SN, Gevers D, Wargo JA, Golub TR, Straussman R (2017) Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. Science 357:1156–1160
23. Kinskey JC, Tu YN, Tong WL, Yavorski JM, Blanck G (2018) Recovery of immunoglobulin VJ recombinations from pancreatic cancer exome files strongly correlates with reduced survival, Cancer Microenviron 11(1):51–59
24. Li B, Li T, Pignon JC, Wang B, Wang J, Shukla SA, Dou R, Chen Q, Hodi FS, Choueiri TK, Wu C, Hacohen N, Signoretti S, Liu JS, Liu XS (2016) Landscape of tumor-infiltrating T cell repertoire of human cancers. Nat Genet 48:725–732
25. Levy E, Marty R, Garate Calderon V, Woo B, Dow M, Armisen R, Carter H, Harismendy O (2016) Immune DNA signature of T-cell infiltration in breast tumor exomes. Sci Rep 6:30064
26. Iglesia MD, Parker JS, Hoadley KA, Serody JS, Perou CM, Vincent BG (2016) Genomic analysis of immune cell infiltrates across 11 tumor types, J Natl Cancer Inst, https://doi.org/10.1093/jnci/djw144
27. Brown SD, Raeburn LA, Holt RA (2015) Profiling tissue-resident T cell repertoires by RNA sequencing. Genome Med 7:125
28. Samy MD, Tong WL, Yavorski JM, Sexton WJ, Blanck G (2017) T cell receptor gene recombinations in human tumor specimen exome files: detection of T cell receptor-beta VDJ recombinations associates with a favorable oncologic outcome for bladder cancer. Cancer Immunol Immunother 66:403–410
29. Tu YN, Tong WL, Fawcett TJ, Blanck G (2017) Lung tumor exome files with T-cell receptor recombinations: a mouse model of T-cell infiltrates reflecting mutation burdens. Lab Invest 97:1516–1520
30. In TSH, Trotman-Grant A, Fahl S, Chen ELY, Zarin P, Moore AJ, Wiest DL, Zuniga-Pflucker JC, Anderson MK (2017) HEB is required for the specification of fetal IL-17-producing gammadelta T cells. Nat Commun 8:2004
31. Arias-Pulido H, Cimino-Mathews A, Chaher N, Qualls C, Joste N, Colpaert C, Marotti JD, Foisey M, Prossnitz ER, Emens LA, Fiering S (2018) The combined presence of CD20 + B cells and PD-L1 + tumor-infiltrating lymphocytes in inflammatory breast cancer is prognostic of improved patient outcome, Breast Cancer Res Treat, 171(2):273–282
32. Lang Kuhs KA, Lin SW, Hua X, Schiffman M, Burk RD, Rodriguez AC, Herrero R, Abnet CC, Freedman ND, Pinto LA, Hamm D, Robins H, Hildesheim A, Shi J, Safaeian M (2018) T cell receptor repertoire among women who cleared and failed to clear cervical human papillomavirus infection: an exploratory proof-of-principle study. PLoS ONE 13:e0178167

33. Zeng G, Huang Y, Huang Y, Lyu Z, Lesniak D, Randhawa P (2016) Antigen-specificity of T cell infiltrates in biopsies with T cell-mediated rejection and BK polyomavirus viremia: analysis by next generation sequencing. Am J Transplant 16:3131–3138
34. Di Sante G, Tolusso B, Fedele AL, Gremese E, Alivernini S, Nicolo C, Ria F, Ferraccioli G (2015) Collagen specific T-cell repertoire and HLA-DR alleles: biomarkers of active refractory rheumatoid arthritis, EBioMedicine, 2:2037–2045
35. Klarenbeek PL, Doorenspleet ME, Esveldt RE, van Schaik BD, Lardy N, van Kampen AH, Tak PP, Plenge RM, Baas F, de Bakker PI, de Vries N (2015) Somatic variation of T-cell receptor genes strongly associate with HLA class restriction. PLoS ONE 10:e0140815
36. Loi S, Michiels S, Salgado R, Sirtaine N, Jose V, Fumagalli D, Kellokumpu-Lehtinen PL, Bono P, Kataja V, Desmedt C, Piccart MJ, Loibl S, Denkert C, Smyth MJ, Joensuu H, Sotiriou C (2014) Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. Ann Oncol 25:1544–1550
37. Xu T, He BS, Liu XX, Hu XX, Lin K, Pan YQ, Sun HL, Peng HX, Chen XX, Wang SK (2017) The predictive and prognostic role of stromal tumor-infiltrating lymphocytes in HER2−positive breast cancer with trastuzumab-based treatment: a meta-analysis and systematic review. J Cancer 8:3838–3848
38. Loi S (2013) Tumor-infiltrating lymphocytes, breast cancer subtypes and therapeutic efficacy, Oncoimmunology, 2:e24720
39. Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, Rouas G, Francis P, Crown JP, Hitre E, de Azambuja E, Quinaux E, Di Leo A, Michiels S, Piccart MJ, Sotiriou C (2013) Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02–98. J Clin Oncol 31:860–867