Immune receptor CDR3 chemical features that preserve sequence information are highly efficient in reflecting survival distinctions: A pan-cancer analysis

BROOKE E. MCBREAIRTY[1], BORIS I. CHOBRUTSKIY[1], ANDREA CHOBRUTSKIY[1], ETIENNE C. GOZLAN[1], MICHAEL J. DIAZ[1] and GEORGE BLANCK[1,2]

[1]Department of Molecular Medicine, Morsani College of Medicine, University of South Florida; [2]Department of Immunology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA

Received March 25, 2022; Accepted May 9, 2022

DOI: 10.3892/br.2022.1551

Abstract. Adaptive immune receptor (IR) chemical features have been used as signatures of an immune response for numerous medical conditions, raising the question of whether certain approaches to assessing the IR chemical features are more robust than others? In the cancer setting, a very large dataset of IR complementarity determining region-3 (CDR3) amino acid (AA) sequences has become available via the mining of cancer specimen and blood genomics files for IR recombination reads. The IR CDR3 AA sequences have been evaluated for chemical features, and survival rates have been correlated with distinct chemical features. Two common approaches have been i) to assign a single value to the CDR3, representing a chemical attribute, such as aromaticity; or ii) to reduce the actual CDR3 AA sequence to a chemical sequence motif, which merges similar CDR3 chemistries represented by distinct AA sequences but preserves potential functional aspects of the order of the AAs in the sequence. While a controlled comparison of the two approaches is not possible, the application of the two approaches to the same clinical datasets offers the opportunity to appreciate a trend with regard to the overall potential in distinguishing survival probabilities. We demonstrate that application of the chemical sequence motif approach is more likely to identify survival distinctions within cancer datasets, for both tumor specimen and blood sourced, adaptive IR CDR3 AA sequences.

Correspondence to: Dr George Blanck, Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, 12901 Bruce B. Downs Bd. MDC7, Tampa, FL 33612, USA
E-mail: gblanck@usf.edu

Abbreviations: AA, amino acid; BRCA, breast cancer; CDR3, complementarity determining region-3; IR, (adaptive) immune receptor; KM, Kaplan-Meier; OS, overall survival; OV, ovarian cancer; TCGA, The Cancer Genome Atlas; TIL, tumor-infiltrating lymphocyte; TRA, T-cell receptor α; TRB, T-cell receptor β; TRG, T-cell receptor gamma; WXS, whole exome sequence

Key words: immune receptors, complementarity determining region-3, blood and tumor samples, physicochemical features, amino acid motifs

Introduction

Over the last several years a huge volume of adaptive immune receptor (IR) recombinations has been obtained and analyzed, due to the increased use of genomic approaches and computational opportunities. In some cases, these recombinations have been obtained through an immune repertoire approach whereby PCR-based techniques lead to the amplification of a large number of clonotypes (1,2), although in several cases, such an approach indicates only a few relatively dominant clonotypes. Other sources of adaptive IR recombination reads have been genomics files representing tissues that included T-cells and B-cells, such as exome (WXS) and RNAseq files (3-6). The obvious limitation of mining such files is the lower level of IR read recoveries in comparison to a PCR-based approach specific for IRs, i.e., lack of depth. However, the advantage of the study of IR recombination reads from genomics files has been the opportunity for breadth, including the opportunity to integrate the IR recombination read data with other data platforms, especially gene expression and clinical data (7-10) over very large patient sets.

Analysis of the IR functional implications, based on the recovery of IR recombination reads from genomics files, has involved two strategies. First, a summary (single value) physicochemical assessment of the complementarity determining region-3 (CDR3) amino acid (AA) sequences has been correlated with survival distinctions. This approach is based on the simple idea that the CDR3 is the most important region of the IR for antigen recognition (11), and that, particularly in a big data setting, a dominant physicochemical feature could represent a minimum for successful antigen binding and immune mediated reductions in tumor cell numbers (12). From the outset, algorithms were developed to exploit the physicochemical features, and their correlations with data represented by other platforms, such as survival and immune biomarker expression, to identify possible antigen partners (9,12-16). For example, single value, physicochemical assessments of tumor resident

breast cancer B-cell receptor CDR3s revealed not only a strong survival distinction when those assessments included an assessment of the chemical interaction potential of same-patient TP53 mutants but also revealed clear associations of immune markers with the higher surviving, chemical complementarity group. More recently, the chemical sequence motif approach, which preserves the presumed AA sequence functional role in the physicochemical assignments and which has already been employed in immune repertoire approaches (17), has been applied to IR CDR3s, and to CDR3-antigen matching algorithms, where the IR CDR3s have been identified in large genomics file sets (10,18). For example, an algorithm for establishing basic chemical complementarity over a large dataset of CDR3s and mutants, that employed a chemical sequence motif approach, identified correlations between TRG CDR3, AA sequence-PIK3CA mutant chemical complementarity and survival rates and immune marker expression (10). This latter approach involved a series of alignments of mutant peptides throughout the breast cancer dataset with all of the same-patient, tumor resident TRG CDR3s, to establish in each case the highest possible chemical complementarity score indicated by the algorithm developed and employed in that report. Patients representing the highest chemical complementarity scores also represented the highest survival. In another case, the chemical sequence motif, survival distinctions were established with blood resident, TRG CDR3 AA chemical sequence motifs in the absence of any potential assessments of chemically complementary antigens (18).

Given the high likelihood of continued characterizations of immune receptor CDR3 chemical features for either antigen identification or basic correlative work, it became of interest to make a comparison of the potential of a relatively simple, single value physicochemical parameter to represent the CDR3 AA sequence, which is efficient in a big data setting, compared with the value of the chemical sequence motif approach, which is more processing intensive for all subsequent correlative or chemical interaction approaches and which may be overly restrictive for correlative studies and biomarker discovery. Thus, we undertook an accounting of each approach for indicating survival distinctions among The Cancer Genome Atlas (TCGA) cancer datasets, to evaluate one specific area of use for these CDR3 AA sequences, with results strongly indicating that the chemical sequence motif assessments were more productive than a single parameter physicochemical CDR3 AA assessment.

Materials and methods

IR recombination read recoveries. The approach to the recovery of the IR recombination reads has been extensively described (5,19,20). The software for using WXS files as input is freely available at https://github.com/bchobrut-USF/blanck_group, including a readme file; and as a container version at https://hub.docker.com/r/bchobrut/vdj, including a readme file. Briefly, the approach involved a pre-screen of the WXS files with 10-mers matching the V- and J-gene segments for all seven of the adaptive human IRs, followed by a high stringency validation of the V- and J-gene IDs (Table SI). All data used in this project were obtained under the NIH dbGaP project approval no. 6300.

The single value, physicochemical determinations and AA chemical sequence motif representations. The above GitHub and docker links include IR recombination read processing software for establishing the single value, physicochemical characters for the CDR3 AA sequences, e.g., net charge per residue or aromaticity (Table SII). These software modules follow the precedents of the Pappu lab (21). The CDR3 AA chemical sequence motifs were established for each CDR3 according to reference (18) and Table SIII; and the python code used to make the symbol replacements to establish the chemical sequence motifs is freely available at https://github.com/bchobrut-USF/brca_motif.

Survival distinctions. Survival distinctions based on the single value, physicochemical features of the CDR3 AA sequences were assessed as described (12), with an example output in Table SIV; and survival distinctions based on the AA chemical sequence motif approach were assessed as previously indicated (18), with example outputs in Tables SV-SXII, representing Table I in the results section below.

Overview of the comparison of the survival distinctions indicated by the single value, physicochemical feature compared with the AA chemical motif sequence approach. Given the above survival distinction outputs, exemplified by Table SIV, for the single value, physicochemical approach, and Tables SV-SXII for the CDR3 AA chemical sequence motif approach, we totaled all survival distinctions representing a Kaplan-Meier (KM) logrank P-value of less than 0.05, for any of the above values or motifs, for either TRA or TRB CDR3s. The exact same set of TRA, TRB CDR3s (exact same input) was used to obtain the survival distinctions for the single value, physicochemical and the chemical sequence motif approaches, respectively, for each of the datasets.

Results

The adaptive IR recombination reads from every blood and tumor WXS file in the TCGA database and in several publications were obtained (5,12,20,22). For numerous cancer datasets, it became apparent that it was possible to observe survival distinctions based on single value, physicochemical features of the CDR3 AA sequences (Table SIV) (23), with related approaches taken by others (17). For the purpose of this report, the physicochemical features approach, exemplified by reference (23), will be referred to as the single value, physicochemical feature(s), and in all cases, the approach of reference (23) represents a single number for characterizing a single physicochemical parameter of the CDR3 AA sequence, for example, the net charge per residue (Table SII). More recently, we have observed survival distinctions with an AA chemical sequence motif approach, whereby the actual CDR3 AA sequence is reduced to a sequence representing physicochemical features of similar AAs (Table SIII) (18), for example NNP, for 'negative, negative, positive' to represent the R-groups, at physiological pH, of the E,E,K AA sequence. To appreciate the different potential of the two approaches, we tabulated the opportunity to note survival distinctions, based on a KM analysis with a resulting P-value of less than 0.05. In the case of the single value, physicochemical feature(s)

Table I. Detection of overall survival distinctions via either the AA chemical sequence motif approach or the single value, physicochemical features approach, for the IR CDR3 AAs recovered from tumor WXS files representing eight cancer datasets.

Cancer dataset	TRA CDR3 AA chemical sequence motif approach (Tables SV-SXII)	TRA single value, physicochemical features approach (Table SIV)	TRB CDR3 AA chemical sequence motif approach (Tables SV-SXII)	TRB single value, physicochemical features approach (Table SIV)
SKCM	5	0	4	5
BRCA	8	0	1	0
STAD	0	3	0	3
LUAD	0	0	7	0
LUSC	6	0	2	1
UCEC	1	0	0	1
HNSC	0	1	0	0
OV	0	1	0	1

AA, amino acid; CDR3, complementarity determining region-3; TRA, T-cell receptor alpha; TRB, T-cell receptor-β.

approach, 22 distinct parameters were assessed for their potential to indicate a survival distinction, whereas, in the case of the AA chemical sequence motif approach, the minimum length in the assessment was four AAs.

Keeping in mind the aspects of the two approaches described in the preceding paragraph, we assessed overall survival (OS) distinctions using all TRA and TRB CDR3 AA sequences, respectively, available from all TCGA cancer datasets, with a basic result being that more such distinctions could be had using the CDR3 AA chemical sequence motif approach. For example, eight cancer types representing TCGA tumor WXS files demonstrated 20 OS distinctions for the AA chemical sequence motif approach, representing the TRA CDR3s, vs. 5 OS distinctions for the single value physicochemical features approach (Table I). For the TRB CDR3s, the eight cancer types assessed represented 14 and 11 OS distinctions for the AA chemical sequence motif and single value, physicochemical features approaches, respectively (Table I). In the case of the OS distinctions represented by the CDR3 AA chemical sequence motif approach, the OS distinction was between TCGA cases representing the AA chemical sequence motif vs. those without the chemical motif, but with a recovery of the corresponding IR recombination read, e.g., TRA. In the case of the OS distinctions representing the single value, physicochemical features, the distinctions also always represented only cases with a recovery of at least one read representing the indicated IR recombination. In the case of the single value, physicochemical features approach, this was out of necessity, as this approach compares the top and bottom 50th percentiles for the given physicochemical feature.

More recently, the AA chemical sequence motif and the single value, physicochemical features approaches, respectively, have been applied to IR CDR3 AA sequences obtained from the blood WXS files of cancer patients and have indicated survival distinctions (8,16,18), not surprising considering the relatively limited clonality of blood-borne IRs at the typical older ages of cancer patients; and considering more precise work in single patient settings indicating the detection of blood-borne, anti-cancer antigen IRs (24). Thus, we considered a comparison of the opportunity to detect survival distinctions for the AA sequence chemical motif vs. the single value, physicochemical features approach using IR recombination reads obtained from the cancer patient, blood WXS files. The same eight cancer types were compared, as were compared using only tumor specimen WXS files, and the comparison for the blood WXS files revealed 155 AA chemical sequence motif-based OS distinctions vs. 5 such distinctions for the single value, physicochemical features approach, for the TRA CDR3s (Table II). Example KM analyses representing the AA chemical sequence motif and the single parameter, physicochemical motif approaches applied to detect the OS distinctions, respectively, using the IR recombination reads recovered from the blood WXS files, are shown in Fig. 1. For the TRB CDR3 AA sequences, the comparison revealed 33 vs. 13 OS distinctions, for the AA chemical motif and single value, physicochemical features approaches, respectively (Table II).

Given the robustness of the AA chemical sequence motif approach in assessing whether an OS distinction could be had, we assessed the opportunity to distinguish OS rates using the AA chemical sequence motif vs. all remaining cases, rather than simply all remaining cases with an IR recombination read recovery, as was done for the data in Tables I and II. In sum, the analyses indicated a robust opportunity to identify OS distinctions, for TRA and TRB CDR3s, and for CDR3s recovered from both tumor and blood WXS files, when defining cases based on the recovery of a specific CDR3 AA chemical sequence motif (Tables III and IV).

Discussion

Over the last several years there has been an extensive effort to group IR CDR3s based on sequence homologies and based on chemical features, to develop CDR3 fingerprints of various conditions and to bioinformatically attempt to link the CDR3s to antigens. We have observed two basic approaches in the literature, the single value, physicochemical assessment approach and the chemical sequence approach, as detailed above. As

Table II. Detection of overall survival distinctions via either the AA chemical sequence motif approach or the single value, physicochemical features approach, for the IR CDR3 AAs recovered from blood WXS files representing eight cancer datasets.

Cancer dataset	TRA CDR3 AA chemical sequence motif approach	TRA single value, physicochemical features approach	TRB CDR3 AA chemical sequence motif approach	TRB single value, physicochemical features approach
SKCM	6	0	7	0
BRCA	46	1	7	0
STAD	1	0	0	1
LUAD	1	0	3	3
LUSC	8	0	2	6
UCEC	63	0	10	0
HNSC	20	1	1	1
OV	10	3	3	2

AA, amino acid; CDR3, complementarity determining region-3; TRA, T-cell receptor-alpha; TRB, T-cell receptor-β.

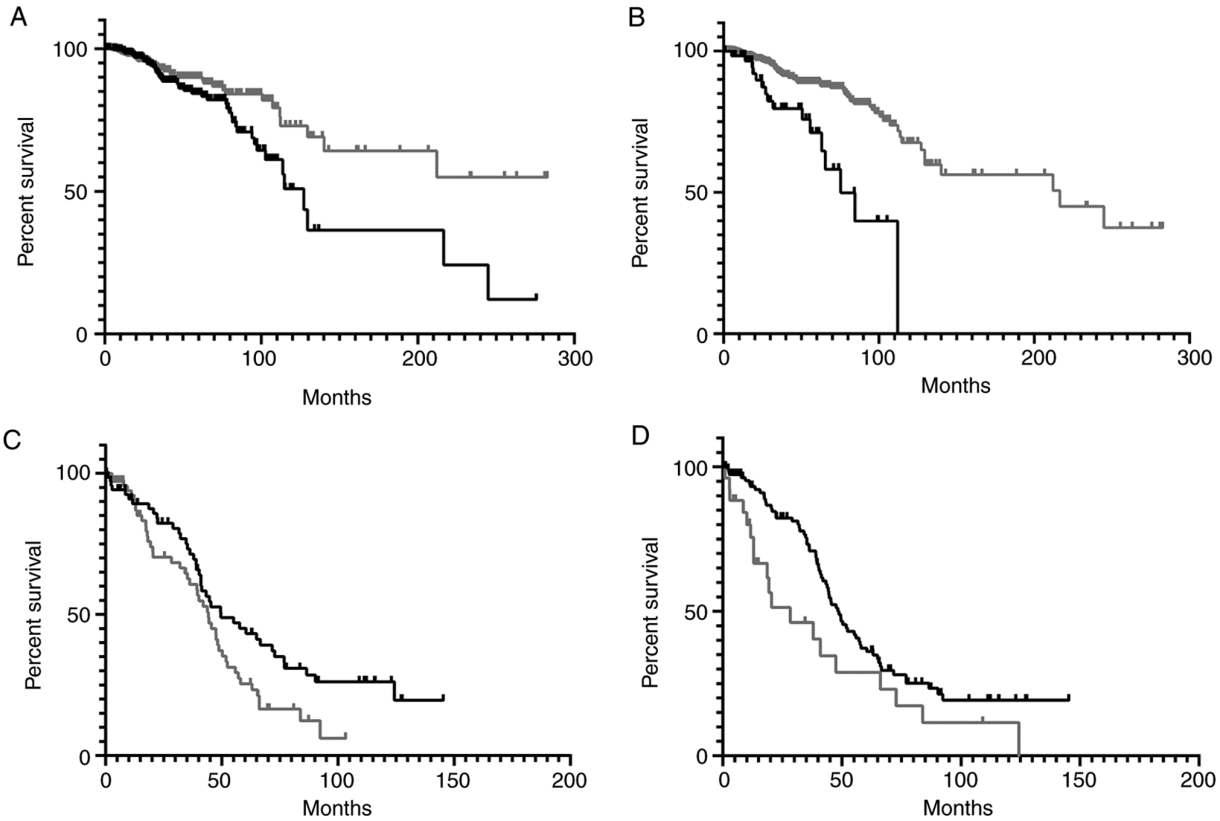

Figure 1. KM OS analyses representing TCGA-BRCA and -OV case IDs with TRA CDR3 recoveries from blood WXS files. (A) KM analysis for BRCA case IDs representing TRA CDR3s with the top (black) and bottom (gray) 50th percentiles for the single value, physiochemical parameter, aromaticity. Median OS time for case IDs representing the top 50%, 127.2 months; median OS time for case IDs representing the bottom 50%, undefined. Logrank P-value=0.0155. (B) KM analysis for BRCA case IDs representing the AXXXHX motif in the TRA CDR3 AA sequence (black), compared to all remaining case IDs with TRA but without the AXXXHX motif (gray). Median OS time for case IDs representing the TRA AXXXHX motif, 75.43 months; median OS time for all remaining case IDs with TRA but without the AXXXHX motif, 216.6 months. Logrank P-value<0.0001. (C) KM analysis for OV case IDs representing TRA CDR3s with the top (black) and bottom (gray) 50th percentiles for the single value, physiochemical parameter, κ. Median OS time for case IDs representing the top 50%, 49.64 months; median OS time for case IDs representing the bottom 50%, 44.28 months. Logrank comparison P-value=0.0283. (D) KM analysis for OV case IDs representing the XXXXA motif in the TRA CDR3 AA sequence (black), compared to all remaining case IDs with TRA but without the XXXXA motif (gray). Median OS time for case IDs representing the TRA XXXXA motif, 48.75 months; median OS time for all remaining case IDs with TRA but without the XXXXA motif, 28.35 months. Logrank comparison P-value=0.0208. KM, Kaplan-Meier; OS, overall survival; TCGA, The Cancer Genome Atlas; BRCA, breast cancer; OV, ovarian cancer; TRA, T-cell receptor-α; CDR3, complementarity determining region-3.

with the above datasets, these approaches have been applied in the cancer setting, by our group (8-10,12-16,18,23,25), by Ostmeyer *et al* (17), and by others (26,27). In addition, CDR3 AA sequence homology assessments have been applied in

Table III. Detection of OS distinctions via the AA chemical sequence motif approach via a comparison of specific AA chemical sequence motifs vs. all remaining samples, regardless of the recovery, or not, of an IR recombination read. The results represent IR recombination reads obtained from tumor specimen WXS files.

Cancer dataset	Specific TRA CDR3 AA chemical sequence motifs representing an OS distinction	Specific TRB CDR3 AA chemical sequence motifs representing an OS distinction
SKCM	26	12
BRCA	4	0
STAD	0	0
LUAD	0	6
LUSC	5	2
UCEC	1	0
HNSC	0	1
OV	0	0

OS, overall survival; AA, amino acid; CDR3, complementarity determining region-3; TRA, T-cell receptor-α; TRB, T-cell receptor-β; IR, immune receptor.

Table IV. Detection of OS distinctions via the AA chemical sequence motif approach via a comparison of specific AA chemical sequence motifs vs. all remaining samples, regardless of the recovery, or not, of an IR recombination read. The below results represent IR recombination reads obtained from blood WXS files.

Cancer dataset	Specific TRA CDR3 AA chemical motifs representing an OS distinction	Specific TRB CDR3 AA chemical motifs representing an OS distinction
SKCM	3	2
BRCA	60	7
STAD	4	0
LUAD	10	6
LUSC	26	8
UCEC	103	40
HNSC	18	1
OV	6	0

OS, overall survival; AA, amino acid; CDR3, complementarity determining region-3; TRA, T-cell receptor-α; TRB, T-cell receptor-β; IR, immune receptor.

the SARS-CoV-2 and other viral settings (28,29), in cases of allergies (30), in Alzheimer's disease and in other neurological disorders (31,32). A key issue in this development is the focus on the CDR3 AA sequences in the big data setting. Also, this focus is based on the CDR3 being highly important in IR antigen binding and has allowed for the efficient establishment of distinctions related to diseases. This is in contrast to a goal of appreciating the role of the entire IR in antigen binding, whether by *in silico* approaches, such as molecular modeling, or with *in vitro* approaches. The focus on the CDR3, and the patient distinctions provided by the different CDR3-based parameters, represents a goal of efficiency, particularly in meeting the longer-term goal of identifying patients in particular risk groups.

It is not possible to make a controlled comparison of the two CDR3-based approaches considered in this report, but it is possible to assemble the record of results over a very large dataset, particularly when these assessments are based on IR recoveries from genomics files. This goal is important, again, in the interest of maximizing efficiency, for answering the simple question, is there a role for IR-antigen binding specificity in a given condition or disease? And, the single value physicochemical assessment represents another step, likely only applicable in the big data setting, away from unnecessarily intensive approaches to processing. In most cases, a very large amount of data regarding the single value physicochemical features of CDR3s, integrated with clinical information, can be processed with common skills on a common laptop. However, the comparison in this report does indicate that a negative result with a single value physicochemical representing the CDR3s could not rule out the possibility that an AA sequenced based, homology motif approach would uncover patient-based risks.

As noted above, these IR recoveries have been obtained from tens of thousands of genomics files, which are available in conjunction with other data platforms, such as survival data. Thus, the recovery of these IR CDR3s, followed by their single value physicochemical or chemical sequence motif assessments, can be linked to distinct survival rates in many cases. Again, keeping in mind that a final assessment of which approach is more successful is not possible, what can be said is that the two approaches that have been applied in the past do differ in the quantity of their output for the same mega-dataset, namely eight cancer datasets from TCGA. Thus, there are two limitations to understanding the value of the CDR3-limited, chemical sequence motif approach over the CDR3-limited, single value physicochemical feature assessment. First, it is possible that with a PCR-based, immune repertoire approach, the vastly increased numbers of CDR3s would be sufficient to provide as many, or almost as many patient-distinction opportunities using a single value physicochemical approach as is apparently provided by the chemical sequence motif approach. In particular, even in the above work, there was no attempt to render the chemical sequence motifs back to single value physicochemical features, which may have led to a convergence of the many chemical sequence motif results. Having noted that, it is less likely that chemical sequence motifs that are 'outliers', that cannot be summarized by a single value physicochemical feature, would be missed when using the chemical sequence motif approach from the start, in categorizing patient risks. The second limitation in this report is the fact that all distinctions based on the single value physicochemical or chemical sequence motif approach were cancer survival distinctions. Thus, it remains to be learned whether the apparent, increased value of grouping the chemical sequence motif for interrogating clinical data, compared with providing only a single value physicochemical assessment, would be maintained in a viral infection setting, for example, a setting that could be much more immunologically simpler than cancer, particularly with regard to patient clinical features. In other words, would a particular single value physicochemical feature shifted in a certain direction always reflect the infection of a specific virus, especially when other clinical parameters related to a suspected viral infection are known?

However, if a comparison of the approaches in this report remains consistent over additional datasets, with additional investigators, a reasonable interpretation would be that, by preserving sequence information, i.e., the order of the AAs in the CDR3, it is more likely that groups of CDR3s can be identified as being associated with given clinical features. With the single value, physicochemical approach, it is likely that many CDR3s, which happen to group with parameter-distinctive CDR3s because of a tally of a chemical feature across multiple AAs, without regard to AA positioning in the sequence, will be specifically irrelevant to that parameter distinction. Depending on the goal, that may not make the single value, physicochemical assessment pointless, but it is possible that a lack of a result with a single value, physicochemical feature would then lead to a second effort with the chemical sequence motif approach to have a desired, clinically relevant CDR3 fingerprint.

Acknowledgements

Authors wish to thank Ms. Corinne Walters (University of South Florida) for administrative support related to NIH dataset access.

Funding

AC, BIC were RISE fellowship recipients. BIC was an American Hematology Association fellowship recipient.

Availability of data and materials

Parts of the article represents datasets available from previously published supporting online material files. Any additional raw data is available via email to the corresponding author.

Authors' contributions

BEM and GB conceived the study, performed the analyses, and wrote and edited the manuscript. BIC, AC, ECG and MJD performed the analyses. All authors have read and approved the final manuscript. BIC and GB confirm authenticity of all the raw data.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Corrie BD, Marthandan N, Zimonja B, Jaglale J, Zhou Y, Barr E, Knoetze N, Breden FM, Christley S, Scott JK, *et al*: iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. Immunol Rev 284: 24-41, 2018.
2. Chiffelle J, Genolet R, Perez MA, Coukos G, Zoete V and Harari A: T-cell repertoire analysis and metrics of diversity and clonality. Curr Opin Biotechnol 65: 284-295, 2020.
3. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, *et al*: The immune landscape of cancer. Immunity 48: 812-830.e14, 2018.
4. Brown SD, Raeburn LA and Holt RA: Profiling tissue-resident T cell repertoires by RNA sequencing. Genome Med 7: 125, 2015.
5. Gill TR, Samy MD, Butler SN, Mauro JA, Sexton WJ and Blanck G: Detection of productively rearranged TcR-α V-J sequences in TCGA exome files: Implications for tumor immunoscoring and recovery of antitumor T-cells. Cancer Inform 15: 23-28, 2016.
6. 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-β VDJ recombinations associates with a favorable oncologic outcome for bladder cancer. Cancer Immunol Immunother 66: 403-410, 2017.
7. Huda TI, Mihyu M, Gozlan EC, Arndt MF, Diaz MJ, Zaman S, Chobrutskiy BI and Blanck G: Specific HLA alleles, paired with TCR V- and J-gene segment usage, link to distinct multiple myeloma survival rates. Leuk Lymphoma 62: 1711-1720, 2021.

8. Gozlan EC, Chobrutskiy BI, Zaman S, Yeagley M and Blanck G: Systemic adaptive immune parameters associated with neuroblastoma outcomes: The significance of gamma-delta T cells. J Mol Neurosci 71: 2393-2404, 2021.
9. Hsiang M, Chobrutskiy BI, Diaz M, Huda TI, Creadore S, Zaman S, Cios KJ, Gozlan EC and Blanck G. Chemical complementarity between immune receptors and cancer mutants, independent of antigen presentation protein binding, is associated with increased survival rates. Transl Oncol 14: 101069, 2021.
10. Chobrutskiy BI, Chobrutskiy A, Zaman S, Yeagley M, Huda TI and Blanck G: High-throughput, sliding-window algorithm for assessing chemical complementarity between immune receptor CDR3 domains and cancer mutant peptides: TRG-PIK3CA interactions and breast cancer. Mol Immunol 135: 247-253, 2021.
11. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, Roulland S, Kasbekar M, Young RM, Shaffer AL, *et al*: Genetics and pathogenesis of diffuse large B-cell lymphoma. N Engl J Med 378: 1396-1407, 2018.
12. Chobrutskiy BI, Zaman S, Diviney A, Mihyu MM and Blanck G: T-cell receptor-α CDR3 domain chemical features correlate with survival rates in bladder cancer. J Cancer Res Clin Oncol 145: 615-623, 2019.
13. Yeagley M, Chobrutskiy BI, Gozlan EC, Medikonda N, Patel DN, Falasiri S, Callahan BM, Huda T and Blanck G: Electrostatic complementarity of T-cell receptor-alpha CDR3 domains and mutant amino acids is associated with better survival rates for sarcomas. Pediatr Hematol Oncol 38: 251-264, 2021.
14. Arturo JF, Chobrutskiy BI, Yeagley M, Patel DN, Falasiri S, Patel JS and Blanck G: Electrostatic complementarity of B-cell receptor CDR3s and TP53-mutant amino acids in breast cancer is associated with increased disease-free survival rates. Cell Mol Immunol 17: 776-778, 2020.
15. Chobrutskiy BI, Yeagley M, Diviney A, Zaman S, Gozlan EC, Tipping P, Koohestani DM, Roca AM and Blanck G: A scoring system for the electrostatic complementarities of T-cell receptors and cancer-mutant amino acids: Multi-cancer analyses of associated survival rates. Immunology 159: 373-383, 2020.
16. Chobrutskiy BI, Yeagley M, Tipping P, Zaman S, Diviney A, Patel DN, Falasiri S, Uversky VN and Blanck G: Chemical complementarity between immune receptor CDR3s and IDH1 mutants correlates with increased survival for lower grade glioma. Oncogene 39: 1773-1783, 2020.
17. Ostmeyer J, Lucas E, Christley S, Lea J, Monson N, Tiro J and Cowell LG: Biophysicochemical motifs in T cell receptor sequences as a potential biomarker for high-grade serous ovarian carcinoma. PLoS One 15: e0229569, 2020.
18. Chobrutskiy A, Chobrutskiy BI, Zaman S, Hsiang M and Blanck G: Chemical features of blood-borne TRG CDR3s associated with an increased overall survival in breast cancer. Breast Cancer Res Treat 185: 591-600, 2021.
19. Tong WL, Tu YN, Samy MD, Sexton WJ and Blanck G: Identification of immunoglobulin V(D)J recombinations in solid tumor specimen exome files: Evidence for high level B-cell infiltrates in breast cancer. Hum Vaccin Immunother 13: 501-506, 2017.
20. Chobrutskiy BI, Zaman S, Tong WL, Diviney A and Blanck G: Recovery of T-cell receptor V(D)J recombination reads from lower grade glioma exome files correlates with reduced survival and advanced cancer grade. J Neurooncol 140: 697-704, 2018.
21. Holehouse AS, Das RK, Ahad JN, Richardson MO and Pappu RV: CIDER: Resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. Biophys J 112: 16-21, 2017.
22. 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.
23. Diaz MJ, Chobrutskiy BI, Zaman S and Blanck G: Immunogenomics of colorectal adenocarcinoma: Survival distinctions represented by immune receptor, CDR3 chemical features and high expression of BTN gene family members. Cancer Treat Res Commun 24: 100196, 2020.
24. Gros A, Parkhurst MR, Tran E, Pasetto A, Robbins PF, Ilyas S, Prickett TD, Gartner JJ, Crystal JS, Roberts IM, *et al*: Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. Nat Med 22: 433-438, 2016.
25. Patel DN, Yeagley M, Arturo JF, Falasiri S, Chobrutskiy BI, Gozlan EC and Blanck G: A comparison of immune receptor recombination databases sourced from tumour exome or RNAseq files: Verifications of immunological distinctions between primary and metastatic melanoma. Int J Immunogenet 48: 409-418, 2021.
26. Thorsélius M, Kröber A, Murray F, Thunberg U, Tobin G, Bühler A, Kienle D, Albesiano E, Maffei R, Dao-Ung LP, *et al*: Strikingly homologous immunoglobulin gene rearrangements and poor outcome in VH3-21-using chronic lymphocytic leukemia patients independent of geographic origin and mutational status. Blood 107: 2889-2894, 2006.
27. Stamatopoulos K, Belessi C, Moreno C, Boudjograh M, Guida G, Smilevska T, Belhoul L, Stella S, Stavroyianni N, Crespo M, *et al*: Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. Blood 109: 259-270, 2007.
28. Wong MK, Liu JT, Budylowksi P, Yue FY, Li Z, Rini JM, Carlyle JR, Zia A, Ostrowski M and Martin A: Convergent CDR3 homology amongst Spike-specific antibody responses in convalescent COVID-19 subjects receiving the BNT162b2 vaccine. Clin Immunol 237: 108963, 2022.
29. Babel N, Brestrich G, Gondek LP, Sattler A, Wlodarski MW, Poliak N, Bethke N, Thiel A, Hammer MH, Reinke P and Maciejewski JP: Clonotype analysis of cytomegalovirus-specific cytotoxic T lymphocytes. J Am Soc Nephrol 20: 344-352, 2009.
30. Smith NP, Ruiter B, Virkud YV, Tu AA, Monian B, Moon JJ, Love JC and Shreffler WG: Identification of antigen-specific TCR sequences based on biological and statistical enrichment in unselected individuals. JCI Insight 6: e140028, 2021.
31. Joshi C, Sivaprakasam K, Christley S, Ireland S, Rivas J, Zhang W, Sader D, Logan R, Lambracht-Washington D, Rosenberg R, *et al*: CSF-derived CD4(+) T-cell diversity is reduced in patients with Alzheimer clinical syndrome. Neurol Neuroimmunol Neuroinflamm 9: e1106, 2021.
32. Gate D, Saligrama N, Leventhal O, Yang AC, Unger MS, Middeldorp J, Chen K, Lehallier B, Channappa D, De Los Santos MB, *et al*: Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. Nature 577: 399-404, 2020.