Original Research

Haplotype Analysis of the T-Cell Receptor Beta (TCRB) Locus by Long-amplicon TCRB Repertoire Sequencing

Timothy J Looney, Dzifa Y Duose[1], Geoffrey Lowman, Elizabeth Linch, Joud Hajjar[2], Denise Topacio-Hall, Mingxuan Xu[1], Jianping Zheng, Anas Alshawa[1], Coya Tapia[1], Bettzy Stephen[1], Linghua Wang[1], Funda Meric-Bernstam[1], Lauren Miller, Alexander Glavin, Lifeng Lin, Jing Gong[1], Jeffrey Conroy[3,4], Carl Morrison[3,4], Fiona Hyland, Aung Naing[1]

Thermo Fisher Scientific, Austin, TX, USA, [1]The University of Texas MD Anderson Cancer Center, Houston, TX, USA, [2]Department of Pediatrics, Section of Immunology, Allergy, and Retrovirology, Texas Children's Hospital, Baylor College of Medicine, Houston, TX, USA, [3]OmniSeq, Inc, Buffalo, NY, USA, [4]Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA

Address for correspondence:
Dr. Timothy J Looney, Thermo Fisher Scientific, Austin, TX, USA.
E-mail: timothy.looney@thermofisher.com

Access this article online

Website: www.jipoonline.org

DOI: 10.4103/JIPO.JIPO_16_19

Quick Response Code:

Abstract

Background: Polymorphism within the human T-cell receptor beta variable (TRBV) gene has been proposed as a risk factor for autoimmune disease and immune-related adverse events (IRAEs) during immunotherapy. Previous efforts to evaluate TRBV polymorphism by whole genome sequencing have been hampered by the repetitive nature of the T-cell receptor beta (TCRB) locus. We present a novel long-amplicon TCRB repertoire sequencing approach to enable TRBV haplotype analysis from peripheral blood. **Methods:** Peripheral blood leukocyte total RNA from 81 Caucasians was used for sequencing of TCRB chains via the Oncomine TCRB-LR assay (amplicon spanning CDR1, 2 and 3) and the Ion Gene Studio S5. VDJ rearrangements were annotated by comparison to the IMGT database, then mined to construct TRBV allele profiles for each individual including, where detected, novel alleles not present in the ImMunoGeneTics (IMGT) database. Finally, TRBV allele profiles were subjected to principal component analysis and k-means clustering to identify TRBV allele haplotypes. **Results:** Clustering analysis revealed the presence of six major sets of coincident TRBV alleles, which we term haplotype groups. Allelic diversity varied markedly across haplotype groups, with approximately one third of the cohort showing limited TRBV allelic diversity and few uncommon alleles compared to members of other groups. Analysis revealed 37 putatively novel TRBV alleles that are absent from the IMGT database. **Conclusion:** We demonstrate a straightforward and cost-efficient method for TRBV haplotype analysis from long-amplicon TCRB sequencing data.

Keywords: *Immune-related adverse events, immunotherapy, polymorphism, T-cell receptor*

Introduction

Checkpoint blockade immunotherapy (CPI) can elicit anticancer T-cell responses mediating durable progression-free survival but may also promote T-cell destruction of healthy tissue to elicit immune-related adverse events (IRAEs). Efforts to identify germline variants associated with adverse events using whole-genome sequencing (WGS) or microarrays have yet to reveal markers predictive of adverse events following immunotherapy.[1] Identifying such biomarkers could allow for personalized drug selection and dosing to enable safer and more effective immunotherapy, particularly in light of the increasing use of combination CPI regimens having a significant incidence of severe adverse events.[2,3] Despite previous efforts, three lines of reasoning support the notion that germline-encoded T-cell receptor beta variable (TRBV) polymorphism could be a key determinant of adverse events during CPI. First, the T-cell receptor (TCR) locus is repetitive and structurally complex, impeding the measurement of variation by traditional short-read WGS or microarray-based methods;[4] second, single amino acid substitutions within the framework or CDR 1 and 2 regions of the rearranged TCRB chain are known to significantly alter TCR affinity for HLA;[5-7] and third, adverse events during immunotherapy may manifest as acute versions of chronic autoimmune diseases that have been separately linked to TRBV and HLA polymorphism.[8-17] These ranges from common and typically manageable cutaneous IRAEs such as eczema, pruritus, and vitiligo, where the evidence supports a role for autoreactive resident memory T cells, to rare and severe IRAEs such as fulminant type 1 diabetes, where comparatively less is understood with respect to the potentially causative T-cell populations. Of note, germline-encoded

How to cite this article: Looney TJ, Duose DY, Lowman G, et al. Haplotype analysis of the T-cell receptor beta (TCRB) locus by long-amplicon TCRB repertoire sequencing. J Immunother Precis Oncol 2019;2:137-43.

Submission: 10-May-2019 **Accepted:** 1-Aug-2019 **Published:** 27-Sep-2019

polymorphism in the cytotoxic T-lymphocyte-associated antigen 4 and programmed death-1 pathways has already been linked to autoimmune disease,[18,19] highlighting the potential mechanistic link between CPI-mediated IRAEs and chronic autoimmune disease. To circumvent the challenge in measuring TRBV polymorphism by WGS, we developed a method for the detection of TRBV polymorphism by next-generation sequencing (NGS) of rearranged TCR beta (TCRB) chains from peripheral blood leukocytes (PBL). To the best of our knowledge, this represents the first NGS-based method to permit haplotype-level resolution of the TRB locus.

Materials and Methods

Collection and preservation of peripheral blood

Peripheral blood samples were collected in ethylenediaminetetraacetic acid (EDTA) tubes and immediately frozen at −80 prior to RNA extraction. This protocol requires RNA that is moderate-to-well preserved (RIN > 4). We, therefore, recommend minimizing the time between sample collection and cryopreservation. Where possible, we suggest immediate isolation and cryopreservation of the buffy coat pellet for future RNA extraction. In cases, where this is not possible, EDTA or PAXgene (Beckton Dickenson, Franklin Lake, NJ) tubes are preferred. In our experience, the maintenance of whole blood in Streck tubes (Strek, La Vista, NE) at room temperature for several days will lead to a degradation in the RNA that makes it less suitable for this technique.

RNA extraction

RNA was extracted from either 500uL of frozen whole blood using the RiboPure RNA Purification Kit, blood (Thermo Fisher Scientific, Waltham, MA, Cat. No. AM1928) or from cryopreserved buffy coat straws using the Qiagen RNeasy Midi Kit (Qiagen, Venlo, Netherlands Cat. No. 75144), depending on the sample availability. Extractions were performed over multiple days at two different sites.

Sample qualification, quantitation, and reverse transcription

Purified RNA samples were quantified using Qubit RNA HS Assay Kit (Thermo Fisher Scientific Cat. No. Q32852). The Agilent 2100 Bioanalyzer (Santa Clara, CA) and Agilent RNA 6000 Nano Kit were used to quantify and evaluate RNA integrity. 25 ng of total RNA were reverse transcribed using SuperScript IV VILO Master Mix (Thermo Fisher Scientific Cat. No. 11756050).

Library preparation

For each sample, 25 ng cDNA was amplified using the Oncomine TCR Beta-LR Assay (Thermo Fisher Scientific Cat. No. A35386) and Ion AmpliSeq HiFi Mix for 23 cycles. The resulting amplicons were partially digested as directed in the protocol, followed by ligation of Ion Select Barcode adapters using library reagents supplied in the Ion AmpliSeq Kit Plus (Thermo Fisher Scientific Cat. No. 4488990) and protocol as described in the Oncomine TCR Beta Assay User Guide MAN0017438 Revision A.0. Libraries were purified with Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, Cat. No. A63880), washed with 70% ethanol, and eluted in 50uL Low TE buffer.

Library quantitation by quantitative polymerase chain reaction

Resulting library samples were diluted 1:100 and quantified using the Ion Library Quantitation Kit (Thermo Fisher Scientific Cat. No. 4468802), then diluted to 25 pM with Low TE buffer. Equal volumes from eight samples were pooled together for sequencing on one chip.

Sequencing

25uL of the 25 pM library pool (8-plex) were template using the Ion 510 and Ion 520 and Ion 530 Kit-Chef (Thermo Fisher Scientific Cat. No. A364461). Ion 530 Chips were sequenced on the Ion S5 Sequencing System with 850 flows. All results were analyzed using the Ion Torrent Suite Software and the Ion Reporter Software version 5.6.

Principal component and tSNE-based dimension reduction

The matrix of variable gene allele profiles was used as input for principal component analysis (PCA)-based dimension through the "prcomp" function in R. tSNE-based dimension reduction was accomplished in R using the "Rtsne" function from the "Rtsne" library and the following parameters: dims = 2, max_iter = 20000, and perplexity values ranging from 5 to 20.

K-means clustering

The top two principal components derived from PCA analysis of the matrix of variable gene allele profiles were used to cluster samples in R through the "kmeans" function from the "class" library and the following parameters: nstart = 50, iter. max = 1000, algorithm = "Lloyd," and centers set to the number of expected clusters.

Analysis by other clustering methods

To further explore the structure of the data, we evaluated alternative methods for identifying the optimal number of clusters using the NbDist package in R version 3.2.3 (https://www.r-project.org). This analysis revealed six clusters to be a valid choice for the optimal number of clusters but also indicated other cluster numbers could be valid. For example, applying the function NbClust with the parameters distance = "euclidean," min. nc = 2, max. nc = 10, method = "ward. D2" and index = "cindex" revealed an optimal number of 6 clusters, consistent analysis through k-means clustering and the elbow method.

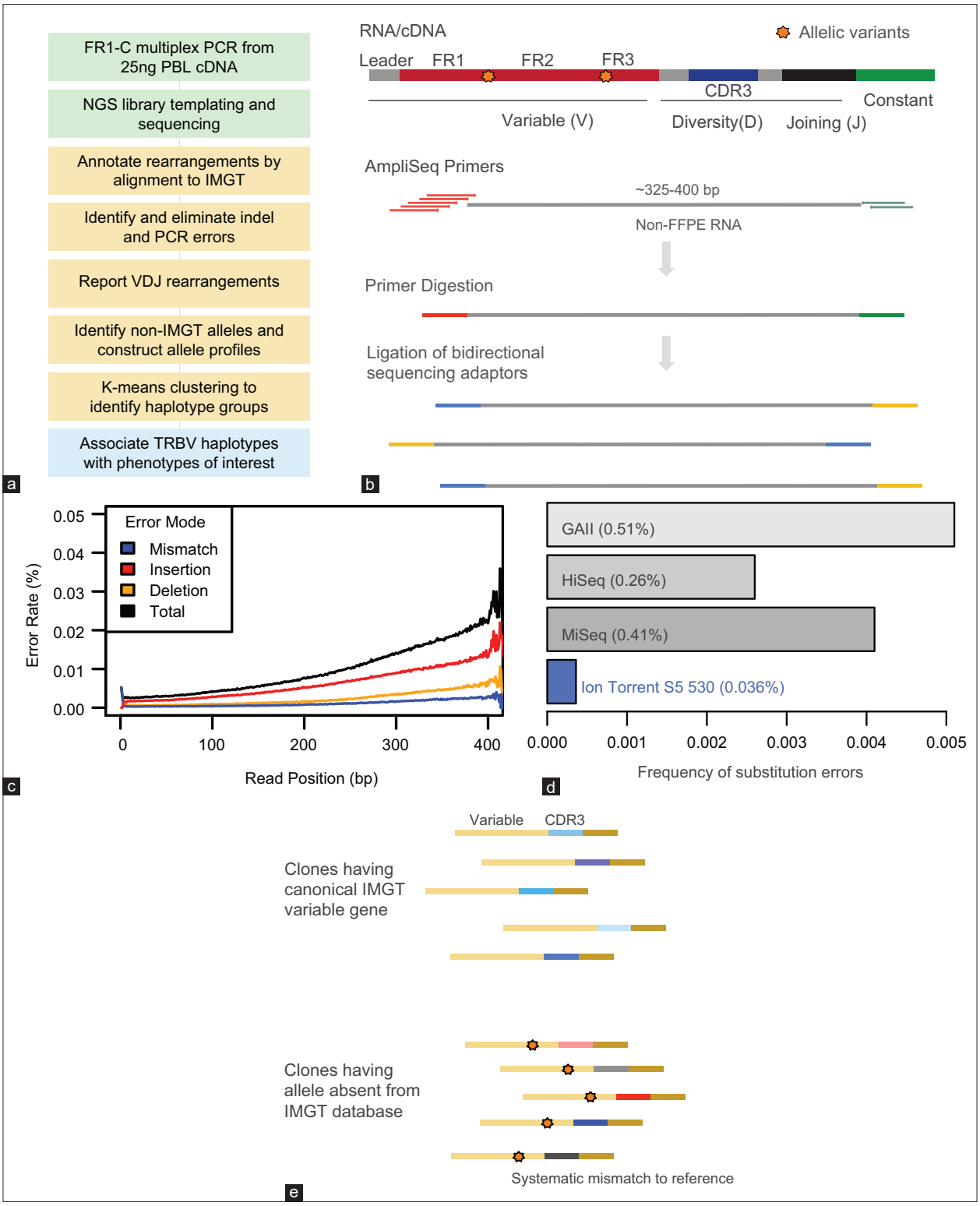

Figure 1: (a) Overview of workflow. The reagents and informatics methodology have been incorporated into the publicly available Oncomine TCR Beta-LR assay and accompanying Ion Reporter analysis software. (b) Method for library preparation and sequencing. Our strategy utilizes multiplex polymerase chain reaction (PCR) through framework 1 and constant gene primers followed by bidirectional next-generation sequencing (NGS) of amplicon libraries. (c) Error profile of the S5 530 chip. The average substitution error rate for the S5 530 chip was calculated by sequencing of the *E. coli* dh1b genome and mapping sequence reads to reference through the Torrent accuracy plugin for Torrent Suite. (d) Comparison of substitution error rates across three Illumina sequencers and the Ion Torrent S5 530 chip. To account for differences in reading lengths between the Illumina and Ion Torrent platforms, the error rate was calculated as the average over the bases 10–100 of the read. The first 10 base pairs were excluded given that alignment scores cannot accurately distinguish indel errors from substitution errors at the read ends. Error rates for the Illumina platform are taken from Schirmer et al., 2016.[21] (e) Strategy for identification of non-IMGT variable gene alleles. Bona fide novel alleles will present as systematic mismatches to IMGT across a plurality of clones, each possessing a distinct CDR3 nucleotide sequence. IGMT: ImMunoGeneTics; TRBV: T-cell receptor beta variable gene; FFPE: Formalin-fixed paraffin-embedded.

However, applying NbClust with the same parameters but index = "hartigan" or index = "ratkowsky" revealed 3 and 9 clusters to be optimal, respectively, suggesting that a range of cluster numbers may be valid. Ultimately, the optimal number of clusters will depend on the sample size, given that a small sample size will not capture rare haplotypes.

Results

Our workflow begins with the cDNA derived from PBL total RNA, which is used as input for multiplex polymerase chain reaction (PCR) through framework 1 (FR1) and constant gene primers to generate ~330 bp TCRB amplicons spanning the three beta chain CDR regions. This extended amplicon enables detection of polymorphism within the germline-encoded framework and CDR1 and two regions in addition to the highly variable CDR3 region [Figure 1a and b]. To minimize primer-primer interaction and amplification bias, we generated primer sequences in accordance with AmpliSeq design principles[20] to cover known population variants and ensure robust and reproducible amplification of all alleles. Following amplification, primers are digested and resultant amplicons ligated to oligomer adaptors for bidirectional NGS (as detailed in materials and methods). We choose to employ Ion Torrent sequencing through the S5 530 chip, which is favorable for this application owing to the long read lengths (up to 600 bp) and low substitution error rate;[20] our estimates from sequencing of bacterial genomes indicate that the substitution error rate may be up to a magnitude lower than that of commonly used Illumina machines [Figure 1c and d].[21] As discussed in depth below, substitution errors pose a key challenge to immune repertoire analysis given that they may mimic the natural variation in the repertoire.

Following sequencing, our workflow maps sequence reads to variable, diversity, and joining genes from the IMGT[22] database and eliminates reads representing off-target products or those that do not span the entire FR1-constant region of the amplicon. The Ion Torrent instruments have an error mode dominated by homopolymer errors manifesting as a single-based expansion or reduction of an existing homopolymer tract. To identify sequences with indel errors, we translate VDJ sequences to protein space and evaluate the productivity of the rearrangement, i.e. whether the V and J gene is in frame and there are no premature stop codons. Indel errors manifest as frameshift mutations that make rearrangements appear unproductive. Owing to nonsense-mediated decay, T-cell allelic exclusion and thymic selection, it is expected that unproductive rearrangements should be exceedingly rare in RNA and thus one may infer that unproductive sequences contain a sequencing error. In some cases, an indel error may occur within the variable gene portion of the sequence such that alignment to the variable gene IMGT reference indicates a gap within the homopolymer tract containing the indel error. Re-evaluation of the initial base calling at such position often leads to correction of the error to produce an ostensibly sequencing error-free read representing a productive rearrangement. After indel error correction, we eliminate PCR-derived errors by evaluating edit distances between VDJ rearrangements, taking into account the frequency of the rearrangement and whether the rearrangement is supported by forward strand and reverse strand reads. Finally, the pipeline reports the sequence and frequency of IMGT-annotated VDJ rearrangements detected in the sample.

We next use the set of annotated VDJ rearrangements to determine the variable gene allele profile of an individual. Although the IMGT database is considered to be gold standard, evidence[23] suggests that it is incomplete. To identify instances where an individual possesses a non-IMGT (putatively novel) variable gene allele, we leverage the fact that clones utilizing a non-IMGT allele will present as having a systematic mismatch to the IMGT database. Given that each clone is readily distinguishable from one another in sequence space owing to the diversity of the CDR3 region, the number of clones having a particular systematic mismatch is indicative of the minimum number of unique template molecules supporting a putative non-IMGT allele. Bona fide novel alleles will be found on a plurality of clones, each possessing a distinct CDR3 nucleotide sequence, while mismatches owing to random PCR error or sequencing error will not be found on multiple clones within a repertoire [Figure 1e]. To report an allele for downstream analysis, either a putative novel allele or canonical IMGT allele, we require that the allele is present on a minimum of 5 clones and makeup at least 5% of the sequences obtained for that variable gene. We allow for up to two alleles of a particular variable gene to be detected in a single sample. In the hypothetical case, where more than two potential alleles are detected for a particular variable gene, only the two alleles having the greatest clone support are reported.

Owing to genetic linkage and population structure, we expect to observe sets of co-inherited variable gene alleles (i.e. allele haplotypes) within human populations. To identify such haplotypes, we applied our method to obtain TRBV allele profiles for 81 Caucasians, then combined the profiles into a matrix such that each row of the matrix represents a different individual, and each column represents a different variable gene allele, where "1" indicates presence of allele and "0" indicates allele absence [Supplemental Table 1]. Our pipeline identified 37 non-IMGT variable gene alleles in the sample set, of which 19 are found in less stringent databases such as Lym1k[23] or the NCBI NR archive. The remaining 18 variable gene alleles appear novel to literature; sequences of these alleles are provided in Supplemental Table 2. We next performed PCA of the matrix of variable gene allele profiles to extract the two largest components contributing to differences

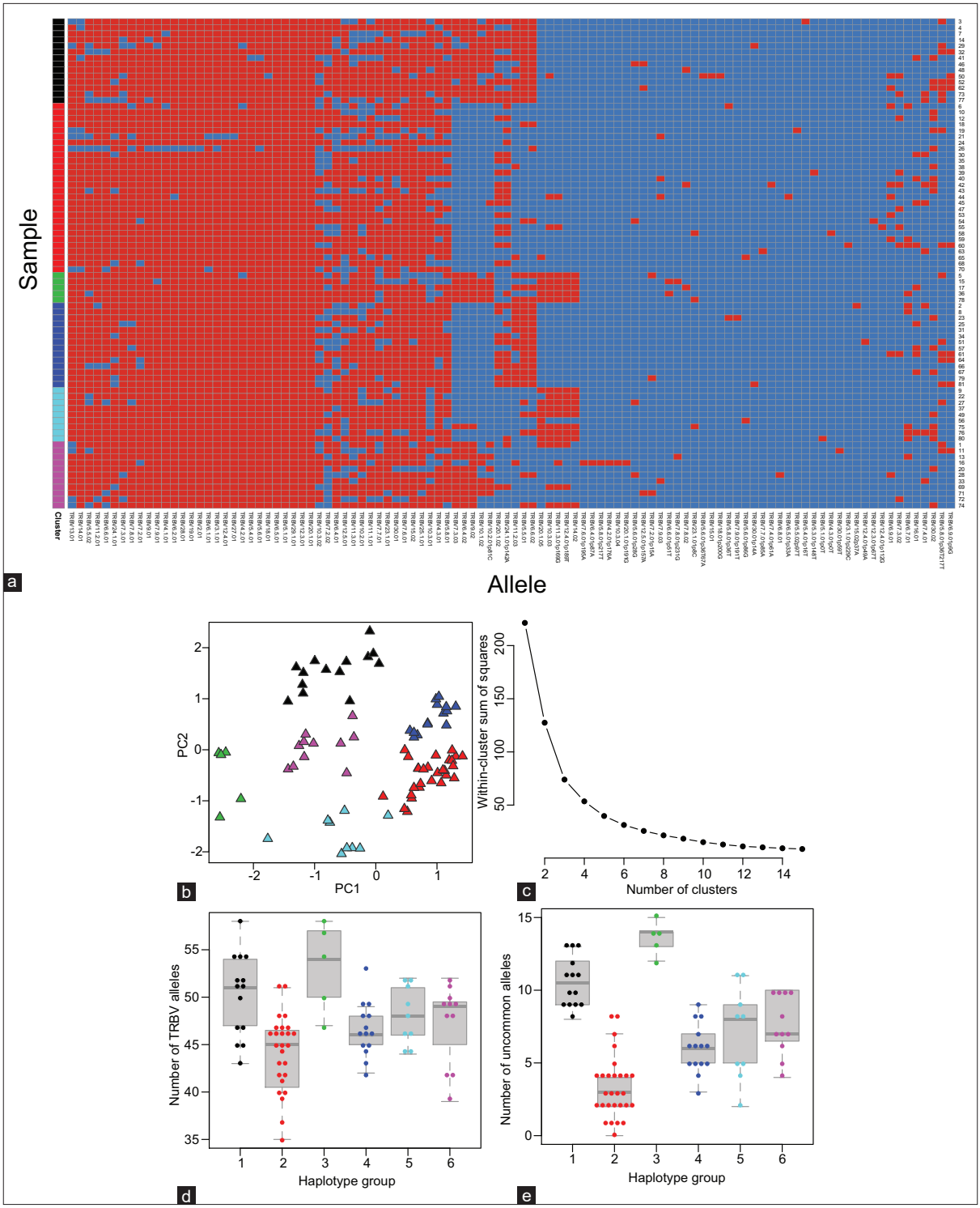

Figure 2: (a) Heatmap of T-cell receptor beta variable (TRBV) allele profiles for 81 Caucasians. T-cell receptor beta (TCRB) repertoires were used to construct variable gene allele profiles for each individual. The sets of alleles detected for each individual are displayed in heatmap form, where each row represents a different individual, and each column a different variable gene allele. Red tiles indicate that an allele was detected in an individual while blue tile indicates allele absence. Columns are arranged through hierarchical clustering, while rows are arranged according to haplotype group classification produced by k-means clustering. IMGT allele names are displayed along the X-axis; alleles having a lowercase 'p' in the name correspond to putative novel alleles absent from the IMGT database. To the left, Cluster column indicates the haplotype group classification. (b) Principal component (PC) analysis of allele profiles. Samples are displayed according to the two largest principal components derived from analysis of the TRBV allele profile matrix. Samples are colored according to haplotype group classification identified through k-means analysis. (c) Within cluster sum of squares following subdivision of data into 2 to 15 clusters through k-means analysis of sample set. (d) Number of distinct alleles detected across haplotype groups. The number of distinct alleles detected in each individual was calculated then displayed according to haplotype group. (e) Number of uncommon alleles detected across haplotype groups. Uncommon alleles are defined as those found in <50% of this Caucasian cohort. IMGT: ImMunoGeneTics.

in allelic representation among the sample set. K-means clustering of principal components revealed six major allele profiles within the dataset, which we term haplotype groups [Figure 2a and b]. The optimal number of clusters was determined using the "elbow" method and plotting the within-cluster sum of squares over cluster centers from 1 to 15 [Figure 2c], though we note that cluster center values ranging from 3 to 8 are potentially valid choices given the structure of the data and analysis by other common clustering methods (methods). For context, we present the cluster assignments of the data using k-means clustering and from 3 to 8 cluster centers [Supplemental Figure 1a].

One limitation of PCA-based cluster analysis is the challenge in visualizing nonlinear relationships between variables.[24] To evaluate whether nonlinear relationships might play a significant role in this dataset, we projected the data into two dimensions using tSNE dimension reduction (methods) and asked whether the organization of the data was consistent with PCA-based projection. Overlaying the cluster assignments from k-means analysis onto tSNE projections revealed that samples assigned to the same cluster by k-means analysis also colocalized within the tSNE projections irrespective of perplexity value [Supplemental Figure 1b]. These results suggest that nonlinear relationships between variables do not play a prominent role in the structure of these data.

We finally asked whether there were allele profile features that distinguish the haplotype groups from one another. We found that haplotype group 2, members have fewer distinct alleles and fewer uncommon alleles (defined as those present in <50% of the sample set) than members of other haplotype groups, indicating that members of this group tend to be homozygous for a common Caucasian allele haplotype, while members of other groups have elevated TRBV allele heterozygosity and carry TRBV haplotypes that are less common in Caucasians [Figure 2d and e; P = 1.7E-4 and 3.6E-13 for number of distinct alleles and uncommon alleles, respectively, in group 2 versus all other groups combined, by Student's t-test].

Discussion

Evidence from literature suggests that polymorphism within the TRB locus may play a role in autoantigen recognition in the context of autoimmune disease or IRAEs following CPI, although research has been limited owing to the technical challenge in obtaining accurate measurements of genetic variation in this locus. Here, we sought to develop a straightforward method for assessing TRBV polymorphism from long-amplicon TCRB chain sequencing data. In this proof of concept study, we demonstrate the use of this approach to partition a set of 81 Caucasian samples into six major haplotype groups through principal component analysis of TRBV allele profiles. Our focus on a single population group for this initial study was motivated by the likelihood that TRBV haplotypes, and the association between certain haplotypes and autoimmunity, may differ across population groups, thus complicating analysis. Future studies applying this technique to other ethnic groups are warranted and would likely reveal additional TRBV allele diversity.

Given that haplotyping of the TRB locus is not commonly employed, it is not yet known whether particular TRBV haplotypes may be linked to autoantigen recognition in the context of CPI. In evaluating TRBV allele features across the six identified haplotype groups, we note that haplotype group 2 members, comprising approximately one-third of the cohort, have reduced allelic diversity and fewer uncommon alleles than members of other haplotype groups. This finding may be significant given proposals regarding the existence of balanced, functional polymorphisms within antigen receptor genes, which posit that lower frequency alleles may have a greater tendency toward autoreactivity than the most common alleles in human populations.[25] Extending on this notion, haplotype group 2 members may be at a lower risk of autoimmune disease and IRAEs than members of other haplotype groups. Testing this hypothesis would require TRBV haplotype analysis of larger validation sets of individuals with graded IRAEs, and subsequently proof of concept samples from patients with autoimmune diseases versus patients experiencing IRAEs during immunotherapy, compared with normal controls.

While variable gene allele discovery was not a primary objective of this study, it deserves mention that this work revealed evidence for 37 non-IMGT alleles, of which 18 are neither found in the NCBI-NR nor Lym1k databases. The Lym1k database of putative immune receptor gene alleles was built using short-read WGS data generated from the 1000 genomes project. As noted by the authors of that study,[23] one limitation of the approach is that allele discovery is not possible for TRBV genes that are not included in the genome assembly. Following the publication of this database, others have noted potential pitfalls in the use of short-read WGS data for allele discovery within the repetitive TRB locus, namely the increased likelihood of detecting artifactual alleles arising from incorrectly mapped reads.[4] In this work, we identified 13 TRBV alleles that are reported in Lym1k but not found in the IMGT database. We view this finding as evidence that the Lym1k database contains many bona fide alleles, thus confirming the utility of this approach for novel allele discovery. Of the 18 alleles that are not found in the Lym1k database, a number derived from variable genes not included in the genome assembly, highlighting the advantage of this TCRB repertoire-based approach to allele discovery. Together, these results illustrate the benefit of using orthogonal methodologies to identify and validate novel TRBV alleles.

Conclusion

In summary, we have developed a cost-efficient and rapid method for routine analysis of germline-encoded

polymorphism within the TRB locus to enable high-resolution studies of genetic variation and its potential link to IRAEs, chronic autoimmune disease, and response to infectious disease. The reagents and informatics methodology described herein have been incorporated into the publicly available Oncomine TCR Beta-LR assay and accompanying Ion Reporter analysis software. The current and future research will clarify the potential role of TRBV polymorphism in IRAEs and chronic autoimmune disease.

Supplemental Material

The supplemental material is available with the article online at jipoonline.org.

Financial support and sponsorship

The authors declare that this study received funding from Thermo Fisher Scientific. The funder is or was the employer of TL, GL, EL, DT, JZ, LM, AG, LL, FH, and supported the procurement and sequencing of samples used in this study. The funder was not involved in the study design, collection, analysis, interpretation of data, nor the writing of this article. The decision to submit this work for publication was agreed upon by all authors.

Conflicts of interest

The authors disclosed no conflicts of interest related to this article.

References

1. Gowen MF, Giles KM, Simpson D, et al. Baseline antibody profiles predict toxicity in melanoma patients treated with immune checkpoint inhibitors. *J Transl Med* 2018;16:82.
2. Larkin J, Chiarion-Sileni V, Gonzalez R, et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N Engl J Med* 2015;373:23-34.
3. Stucci S, Palmirotta R, Passarelli A, et al. Immune-related adverse events during anticancer immunotherapy: Pathogenesis and management. *Oncol Lett* 2017;14:5671-80.
4. Watson CT, Matsen FA 4th, Jackson KJ, et al. Comment on "A database of human immune receptor alleles recovered from population sequencing data". *J Immunol* 2017;198:3371-3.
5. Gras S, Chen Z, Miles JJ, et al. Allelic polymorphism in the T cell receptor and its impact on immune responses. *J Exp Med* 2010;207:1555-67.
`6. Robbins PF, Li YF, El-Gamil M, et al. Single and dual amino acid substitutions in TCR CDRs can enhance antigen-specific T cell functions. *J Immunol* 2008;180:6116-31.
7. Robbins PF, Morgan RA, Feldman SA, et al. Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. *J Clin Oncol* 2011;29:917-24.
8. Patel AB, Pacha O. Skin reactions to immune checkpoint inhibitors. *Adv Exp Med Biol* 2018;995:117-29.
9. Boehncke WH, Brembilla NC. Autoreactive T-lymphocytes in inflammatory skin diseases. *Front Immunol* 2019;10:1198.
10. Riding RL, Harris JE. The role of memory CD8+ T cells in vitiligo. *J Immunol* 2019;203:11-9.
11. Pierce BG, Eberwine R, Noble JA, et al. The missing heritability in T1D and potential new targets for prevention. *J Diabetes Res* 2013;2013:737485.
12. Hockertz MK, Paty DW, Beall SS. Susceptibility to relapsing-progressive multiple sclerosis is associated with inheritance of genes linked to the variable region of the TcR beta locus: Use of affected family-based controls. *Am J Hum Genet* 1998;62:373-85.
13. Maksymowych WP, Gabriel CA, Luyrink L, et al. Polymorphism in a T-cell receptor variable gene is associated with susceptibility to a juvenile rheumatoid arthritis subset. *Immunogenetics* 1992;35:257-62.
14. McDermott M, Kastner DL, Holloman JD, et al. The role of T cell receptor beta chain genes in susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1995;38:91-5.
15. Gaudy C, Clévy C, Monestier S, et al. Anti-PD1 pembrolizumab can induce exceptional fulminant type 1 diabetes. *Diabetes Care* 2015;38:e182-3.
16. Hughes J, Vudattu N, Sznol M, et al. Precipitation of autoimmune diabetes with anti-PD-1 immunotherapy. *Diabetes Care* 2015;38:e55-7.
17. Okamoto M, Okamoto M, Gotoh K, et al. Fulminant type 1 diabetes mellitus with anti-programmed cell death-1 therapy. *J Diabetes Investig* 2016;7:915-8.
18. Prokunina L, Castillejo-López C, Oberg F, et al. A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat Genet* 2002;32:666-9.
19. Michot JM, Bigenwald C, Champiat S, et al. Immune-related adverse events with immune checkpoint blockade: A comprehensive review. *Eur J Cancer* 2016;54:139-48.
20. Merriman B, Ion Torrent R, Rothberg JM. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis* 2012;33:3397-417.
21. Schirmer M, D'Amore R, Ijaz UZ, et al. Illumina error profiles: Resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics* 2016;17:125.
22. Lefranc MP. IMGT, the international imMunoGeneTics information system. *Cold Spring Harb Protoc* 2011;2011:595-603.
23. Yu Y, Ceredig R, Seoighe C. A database of human immune receptor alleles recovered from population sequencing data. *J Immunol* 2017;198:2202-10.
24. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;9:2579-605.
25. Dean M, Carrington M, O'Brien SJ. Balanced polymorphism selected by genetic versus infectious human disease. *Annu Rev Genomics Hum Genet* 2002;3:263-92.