Published in final edited form as:

J Immunol Methods. 2014 October ; 412: 24–34. doi:10.1016/j.jim.2014.06.013.

A NOVEL METHOD FOR ANALYSIS OF HUMAN T CELL REPERTOIRES BY REAL-TIME PCR

Peter J. Wettstein^a, Nancy D. Borson^{a,b}, and Neil E. Kay^c

^a Departments of Surgery and Immunology, Guggenheim 5, Mayo Clinic, 200 First St. SW, Rochester, MN 55902 USA

^b Retired. nancyborson@aol.com

^c Division of Hematology, Stabile 6, Mayo Clinic, 200 First St. SW, Rochester, MN 55902 USA
kay.neil@mayo.edu

Abstract

T lymphocyte responses to challenges with multiple pathogens depend on the diversity of their T cell receptors (TcRs) that are heteroduplexes of alpha and beta chains. The regions of alpha and beta chains that define TcR specificity are encoded by rearranged variable (V) and joining (J) genes that are separated by variable numbers of nucleotides that encode the complementarity determining region 3 (CDR3). The assumption that a “healthy” T cell compartment exhibits broad TcR and CDR3 diversity has driven development of methods to evaluate diversity of TcR beta transcripts expressed by T lymphocyte populations and subpopulations in inflammatory sites and human malignancies. To that end, we have developed the BV:BJ matrix assay that uniquely generates a single statistic that describes TcR repertoire diversity and improves identification of beta transcripts expressed by expanded T cell clonotypes. The BV:BJ matrix uses rigorously selected primers specific for individual V and J genes to amplify beta transcripts in real-time PCRs driven by 533 BV:BJ primer pairs. The quantitative control of real-time PCRs produces Shannon entropy estimates of diversity that are reproducible over a range of template amounts and amenable to statistical analyses that have been difficult to perform with existing methods of repertoire analysis.

Keywords

T cell receptor; Real-time PCR; Complementarity-determining region 3; Shannon entropy

© 2014 Elsevier B.V. All rights reserved.

Corresponding author: Peter J. Wettstein Phone: 507-284-9654 Fax: 507-284-3757 wettstein.peter@mayo.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Thymus-derived (T) lymphocytes recognize complexes of major histocompatibility complex (MHC) gene products and bound, processed peptides through T cell receptors (TcRs) that are heteroduplexes of alpha and beta subunits. Healthy T lymphocyte populations are characterized by diverse TcR repertoires that have been hypothesized to be essential for effective T lymphocyte responses to large numbers of pathogens. The alpha and beta subunits are encoded by constant region (C) genes and rearranged variable (V) and joining (J) genes that map within extensive multi-gene families (Chien et al., 1984). In humans, families of alpha variable (AV) and beta variable genes (BV) include 45 and 48 expressed members, respectively. Likewise, there are 50 alpha joining (AJ) and 13 beta joining genes (BJ) (Lefranc et al., 2005). In addition, two beta diversity (BD) genes contribute to the diversity at the sites of BV:BJ rearrangement. Diversity is further increased by the insertion of random nucleotides at the sites of V:J rearrangements that contribute to the complementarity-determining region 3s (CDR3s) that play direct roles in recognition of MHC-bound peptides (McHeyzer-Williams and Davis, 1995; McHeyzer-Williams, 1999; Kedzierska et al., 2005).

The importance of diversity in T lymphocyte populations has driven the development of methods to quantitate TcR repertoire diversity. The currently most utilized method is spectratyping of human and mouse alpha and beta transcripts that involves RT-PCR amplification with primers specific for variable and constant regions coupled with analysis of CDR3 length distributions (Pannetier et al., 1993; Pannetier et al., 1997) with more limited use of paired variable and joining region primers (Ria et al., 2001). Application of this method in a large number of studies has generated important data on diversity of T cell populations (Guillet et al., 2002) as well as identification of TcR transcripts that are expressed by expanded T cell clonotypes (Caignard et al., 1994; DePalma and Gorsky, 1995; Johnston et al., 1997; Puisieux et al., 1994; Musette et al., 1996). Increased productivity with this method is restrained by difficulties in (1) amplifying transcripts with single V gene resolution due to the high homology between V genes and the single stage of RT-PCR amplification and (2) accurate quantification of the diversity of transcript populations with RT-PCR products that are resolved by gel electrophoresis.

Lengths of CDR3s have been shown to be selectable characteristics of T cell responses (McHeyzer-Williams et al., 1999) so diversity of CDR3 lengths reflects the effects of antigen-driven selection pressure without necessarily revealing the underlying diversity of beta chain transcripts that carry CDR3s with single lengths. In order to more directly analyze the complete extent of beta transcript diversity, we have developed a method that subdivides beta transcript repertoires on the basis of expression of segments of beta transcripts that, in comparison to CDR3s, have relatively limited involvement in determining antigenic specificity. The BV and BJ genes encode parts of beta chains that determine beta chain structure as well as their orientation toward and contact with MHC gene products with limited contact with bound peptides (Garcia et al., 1996). Accordingly, the usage of specific BV and BJ genes in T cell responses should be under reduced selective pressure in these responses relative to CDR3s. Therefore, we have approached the analysis of beta transcript

diversity through rigorous estimation of frequencies of combinations of BV and BJ genes that are expressed in T cell populations.

The novel BV:BJ matrix assay that we devised for analysis of murine TcR repertoires involves amplification in real-time PCRs using 252 combinations of BV- and BJ-specific primer pairs that derive from 21 BV and 12 BJ genes in mice (Wettstein et al., 2008). The templates for these reactions are generated by RT-PCRs that use pools of BV-specific primers that are paired with a constant region (BC)-specific primer. The use of real-time PCRs in this fashion facilitates both estimation of diversity and statistical analysis with this assay. In this report, we describe the extension of this method to the analysis of human TcR beta transcripts. We have used an early prototype of the human BV:BJ matrix to identify over-represented beta chain transcripts in small populations of peptide-specific T lymphocytes (Guillaume et al., 2010). However, the method required considerably more detailed development before it could be used to dependably analyze TcR repertoire diversity in entire populations of T lymphocytes. In addition, the extension to more complex human TcR repertoires required rigorous selection of primers in order to achieve resolution at each of the increased number of BV:BJ combinations as well as technical modifications to improve reproducibility when amplifying populations of more diverse beta chain transcripts. This report describes in detail the advances in the original matrix method we have reported and can be applied to studies of human T cell diversity.

2. Experimental

2.1. Materials

2.1.1. RNA Template—A 36 yr old healthy human donor was anonymously recruited and consented at the Mayo Clinic Blood Bank in accordance with Mayo Clinic Institutional Review Board policy. Lymphocytes were collected in a buffy coat, and T lymphocytes were enriched from peripheral blood lymphocytes (PBLs) by positive selection with anti-CD3 paramagnetic particles according to the manufacturer's protocol (Miltenyi, Auburn, CA). RNA was extracted with an RNeasy Protect MiniKit (Qiagen) according to the manufacturer's instructions. Residual genomic DNA was removed from extracted RNA samples using an RNase-Free DNase Set (Qiagen), and RNA samples were stored at -80°C . Total RNA was then diluted to 2 ng/ μl in water immediately prior to use in RT-PCRs.

2.1.2. Primers—The RT-PCRs and real-time PCRs required the following primers: (i) a reverse BC region primer that was biotinylated at the 5'-end, (ii) a pair of nested forward primers specific for each of 41 BV genes and (iii) a set of reverse primers specific for the 13 individual BJ genes (Tables 1 and 2). BV genes are cited according to classification by the international ImMunoGeneTics information system (Lefranc et al., 2005). Sequences of 37 forward, outer BV primers are homologous to sequences within the CDR1 regions of BV genes, and these primers were divided into eight primer pools based on relative homologies. Forty-one nested BV primers are based on sequences within the CDR2 regions. All primers were synthesized by the Invitrogen (Carlsbad, CA) Supply Center located in the Mayo Clinic Primer Core Facility. The biotinylated BC region reverse primer was diluted to 20 μM in water. Outer, forward BV primers for RT-PCRs were mixed in eight pools of four to five

primers. The concentration of each primer in a pool was adjusted to 6.6 μM in water. BV- and BJ-specific primers for real-time PCRs were diluted to 10 μM in water.

2.2. Methods

2.2.1. Pooled RT-PCRs—RNA templates were thawed and incubated at room temperature for 30 minutes (min). RNA concentration was estimated, and an aliquot was removed for amplification. The RNA aliquot was denatured at 75°C for 4 min and placed on ice. The denaturation and chilling were repeated immediately prior to the addition of RNA template to RT-PCR reactions. Eight pooled RT-PCRs were performed in 27 μl volumes using (i) a master mix from a one-step RT-PCR Kit (Qiagen; 20 μl), (ii) 10 ng of denatured total RNA (5 μl), (iii) 20 pmol of a 5'-biotinylated BC primer (1 μl) and (iv) 6.6 pmol of each of the pooled BV primers (1 μl of pooled primers). RT-PCRs were performed on a PTC-225 Peltier Thermal Cycler (MJ Research, Waltham, MA, USA) as follows. cDNA was synthesized at 50°C for 32 min followed by incubation at 95°C for 15 min to inactivate the reverse transcriptase. Subsequent PCR parameters were 1 min at 94°C, 30 sec at 60°C, and 1 min at 72°C for 25 cycles. A final extension cycle was performed for 6 min at 72°C. RT-PCR products were separated from residual primers and amplification reagents using a QIAquick PCR Purification Kit (Qiagen) and eluted with 50 μl of elution buffer.

2.2.2. Enrichment of Biotinylated PCR Products—Biotinylated RT-PCR products were enriched with magnetic My One™ Streptavidin C1 Dynabeads (Invitrogen) using the manufacturer's protocol. For each RT-PCR product, 50 μl of Dynabeads were washed twice in 50 μl of 2x washing and binding buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 2 M NaCl). The washed beads were resuspended in 100 μl of 2x washing and binding buffer and added to 50 μl of sterile water and the RT-PCR product (50 μl). The amplicon-bound bead suspensions were gently shaken at room temperature for 15 min. The product-bound beads were washed twice with 100 μl of 1x washing and binding buffer and then resuspended in 100 μl of 10 mM Tris-HCl, pH 8.5. Suspensions of amplicon-bound beads were diluted 1:10 immediately prior to use as templates in real-time PCR reactions.

2.2.3. Real-Time PCRs—A total of 533 individual real-time PCRs (41 BV and 13 BJ primers) were performed in 10 μl volumes in 384-well Clear Optical Reaction Plates with Optical Adhesive Covers (Applied Biosystems, Foster City, CA). BV and BJ primers (10 pmol each) were robotically loaded into wells in 10 μl 3% sucrose in water; primer solutions were lyophilized and stored at -80°C. The components of reactions were (i) 5 μl of Power SYBR Green PCR Master Mix (2x) (Applied Biosystems), (ii) 1 μl of the respective amplicon-bound bead suspension, and (iii) 4 μl of nuclease-free water (Qiagen). Cycling was performed on an ABI Prism 7900HT Sequence Detection System at the Mayo Clinic AGTC Microarray Shared Resource Core Facility using SYBR Green detection. Cycling parameters were as follows: (i) an initial incubation at 50°C for 2 min, (ii) a 10 min incubation at 95°C to activate the DNA polymerase and (iii) 40 cycles of 15 sec at 95°C followed by 1 min at 60°C. Dissociation curves were generated by (i) incubating the amplicons at 95°C for 15 sec, (ii) reducing the temperature to 60°C for 15 sec and (iii) increasing the temperature to 95°C over a dissociation time of 20 min. Data were analysed with the 7900HT Sequence Detection System (SDS) Version 2.3 software (Applied

Biosystems) to estimate cycle threshold (Ct) values and dissociation curves to estimate the optimal melting temperatures for products from all reactions. Ct values are fractional cycle numbers at which fluorescence equals the threshold level (designated by a horizontal line in Ct plots), that is automatically set to be within the exponential region of the amplification curve where there is a linear relationship between the log of change in fluorescence and cycle number. Dissociation curves are generated by plotting rising temperature versus the change in fluorescence/change in temperature.

2.2.4 Sequencing of PCR Products—Real-time PCR products were cleaned using a QIAquick PCR Purification Kit prior to sequencing with 2 pmol of the respective, nested BV primers. Sequencing was performed by the Mayo Clinic Molecular Biology Core Facility using a Big Dye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) prior to sequence analysis with a 96-capillary ABI PRISM™ 3730 XL DNA Analyzer (Perkin Elmer Applied Biosystems, Foster City, CA) by the Mayo Clinic Molecular Biology Core Facility.

2.2.5. Statistical Analysis—The relative abundance of individual BV:BJ combinations was defined by the observed Ct values. Dissociation curves were used to confirm the presence of amplicons from beta transcripts by excluding (i) primer-dimers that had relatively low melting temperatures and (ii) amplicons with dissociation peak heights that did not exceed a threshold of 0.07 (change in fluorescence/change in temperature). This threshold was selected experimentally due to the inability to sequence amplicons that were below this value. Amplicons with either or both of these characteristics were assigned Ct values of 40 cycles. Importantly, the diversities of the 533 BV:BJ combinations within individual RNA templates were estimated by Shannon entropy (Shannon 1949) that has been used for estimating variability at individual amino acid positions in immunoglobulin and TcR V gene products (Stewart et al., 1997). An estimate of scaled entropy (H) was calculated for each BV:BJ matrix by the equation $H = \sum(P \log_2 P) / \log_2(1/533)$ where P was the probability of abundance calculated for each BV:BJ combination by the equation $P = 2^{-y} / \sum 2^{-y}$ where y was the Ct value for the BV:BJ primer pair and $P = 0$ when $Ct > 40$ cycles. Scaled entropy ranges from 0 to 1. One represents maximal repertoire diversity when all BV:BJ combinations are expressed equally, and 0 represents minimum diversity exhibited by a monoclonal T cell population.

3. Results

3.1. Conceptual Basis for Beta Transcript Amplification and Analysis

Previously developed approaches to analysis of the diversity of T cell repertoires have been hindered by a general lack of quantitative and rigorous methods that yield single, uncomplicated statistics that describe repertoire diversity. The BV:BJ matrix was designed to overcome this limitation through quantitative amplification of beta transcripts with resolution at the level of individual BV:BJ combinations. The relatively high homology between the 48 expressed human BV genes required nested, BV-specific primers for two levels of amplification in (1) RT-PCRs with pools of BV forward primers plus a constant region primer and followed by (2) real-time PCRs with fully nested BV forward and BJ

reverse primers (Figure 1). Real-time PCRs generate estimates of cycle thresholds (Ct) and melting temperatures that quantitatively describe the speed and specificity of amplification, respectively, for each BV:BJ pair of primers. Ct values are then used to calculate two statistics that characterize the diversity of beta transcript diversity: (1) the percentage of BV:BJ primer pairs that amplify products and (2) an estimate of Shannon entropy that describes the diversity of beta transcripts and is determined by the representations of all BV:BJ combinations.

Shannon entropy was chosen as the descriptor for beta transcript diversity on the basis of its long-term use as a measure of level of uncertainty of a random variable (Shannon 1949). Entropy was originally derived for information theory to describe the information content in messages and is solely determined by the diversity and frequency of characters, including letters and numbers. Maximum entropy is associated with random characters that have equal frequencies, and minimum entropy is associated with repetitive use of only a single character. Shannon entropy continues to be the method of choice for analysis of information content in communications due to its relatively low coefficient of variation and use of all characters regardless of their frequencies with resulting increases in precision. Entropy has also been applied in immunology for (1) the analysis of variability at single amino acid positions in proteins to identify regions of low variability that may be useful as immunogens (Khan et al., 2006) and conversely (2) the analysis of immunoglobulin and TcR sequences to identify highly variable regions that may function in antigen binding (Stewart et al., 1997).

We have hypothesized that a TcR repertoire includes a distribution of rearranged beta transcripts that constitute a message that describes the capacity of an individual to respond to a range of foreign peptides in the context of self MHC. The information content of such “messages” can be estimated by the frequencies of individual beta transcripts at the highest level of resolution. However, at a lesser and more approachable level of resolution, these beta transcripts can be categorized according to individual BV:BJ combinations, and frequencies of BV:BJ combinations can then be used to estimate repertoire diversity. Such frequencies can be estimated using the Ct values from individual BV:BJ-specific amplifications relative to the total number of Ct values from all amplifications by using the equation described above that is a derivation of the original Shannon entropy equation (Shannon 1949). As with characters in a message, over-representation by single BV:BJ combinations reduces “information content” in repertoires resulting in increased recognition of a limited set of antigens with associated reduction in diversity of recognition of other immunogens.

3.2. Technical Overview of BV:BJ Matrix Method

The experimental approach was developed to amplify beta transcripts with resolution at the level of individual BV:BJ combinations (Figure 1). Briefly, beta transcripts are reverse-transcribed from total RNA with a biotinylated BC region reverse primer and amplified with pools of BV-specific forward primers. The resulting amplicons are mixed with streptavidin-coated magnetic beads to enrich products that carry the biotinylated BC region primer. The bead-enriched products are delivered to microtiter wells for amplification in real-time PCRs using nested BV:BJ primer pairs. The selection of BV-specific primers for the RT-PCRs and

real-time PCRs was complicated by the relatively high homologies of human BV genes (Figure 2). Primer locations were focused on nucleotide sequences that encode CDR1 (RT-PCRs) and CDR2 regions (real-time PCRs). Primers were identified to amplify the 48 BV genes that have been shown to be expressed, and the BV6-2 and BV6-3 genes were amplified together with a single primer in real-time PCRs. Preliminary assays showed that the BV6-8, BV6-9, BV7-4, BV10-1, BV10-2, and BV16 BV genes were either very weakly expressed or could not be reliably amplified when they were rearranged with specific BJ genes. Primers that were specific for these six BV genes were omitted resulting in a final matrix of 533 BV:BJ primer pairs. Sequencing of amplicons showed that all BV primers were specific for their respective BV genes (data not shown).

3.3. Amplification of TcR Beta Transcripts

CD3⁺ T cells were enriched from peripheral blood lymphocytes donated by a 36 year old healthy donor. RNA was extracted and amplified with the BV:BJ matrix method. Ct values were estimated for each BV:BJ combination and these estimates are represented in a heat map and summarized in a histogram (Figures 3 and 4). Ct values of >35 cycles were grouped with 40 cycle values that indicated lack of detectable amplification. Ct values ranged from 9.1-35 cycles, and 98% of BV:BJ primer pairs amplified products with a median Ct of 19.3 cycles. Diversity of beta transcripts was estimated as Shannon entropy with a value of 0.68. Although Ct values of 40 cycles may stand out prominently in heat maps, they have minimal effects on Shannon entropy that is most strongly reduced by low Ct values that are indicative of amplification of highly abundant transcripts.

Although the distribution of Ct values is presented on a linear scale in Figure 4, amplification in PCRs is exponential, and a single cycle shift in Ct value indicates a two-fold change in the abundance of the target beta transcripts. The effects of low Ct values on diversity are exemplified by using Ct values from individual BV:BJ-specific reactions to estimate the contribution of each BV:BJ combination to the estimated total number of beta transcripts that is based on the total sum of Ct values for a single template (Figure 5). This interpretation of the data first shows that the reactions that are characterized by relatively high Ct values contribute minimally to the total number of beta chain transcripts. On the other end of the spectrum, only 14 BV:BJ gene pairs are included in ≈50% of the total number of beta chain transcripts in T cells from this subject. These results highlight the importance of using quantitative assays that generate statistics that accurately describe the diversity of TcR repertoires.

3.4. Analysis of Reproducibility

The combined stages of pooled RT-PCRs and real-time PCRs must amplify templates to generate reproducible estimates of beta transcript diversity. Replicate assays were performed to evaluate reproducibility in (1) the combined stages of pooled RT-PCRs and 533 BV:BJ-specific real-time PCRs and (2) only the real-time PCR stage. The RNA extracted from the 36 yr subject was amplified a second time by the pooled RT-PCRs and real-time PCRs. The distribution of Ct values was comparable to that of the first assay of this RNA template (Figure 6). Likewise, estimates of entropy (0.76), number of BV:BJ primer pairs that amplified (98%), and median Ct (20.1 cycles) were comparable to those estimated in the

first assay. The variability in Ct values for each BV:BJ primer pair was estimated (Figure 7 for variation grouped by BV gene). The median difference for individual BV:BJ primer pairs was 1.1 cycles (interquartile range of 0.48-2.13 cycles); only 4.9% of the total BV:BJ primer pairs exhibited differences that exceeded five cycles.

The reproducibility of amplification in the real-time PCR stage was evaluated in two additional matrices of the 533 BV:BJ primer pairs with bead-enriched templates from the first BV:BJ matrix assay described in Figures 3 and 4. The results (entropy, percent of amplified BV:BJ pairs, and median Ct) were virtually identical to those generated by the original assay (Figure 8). The variability in Ct values for each BV:BJ primer pair was smaller than the variability observed for the assays that duplicated both RT-PCRs and real-time PCRs (Figure 9). In this comparison, the median difference for individual primer pairs was 0.3 cycles (interquartile range of 0.11-0.77 cycles) with only 2% of the BV:BJ primer pairs having differences greater than five cycles. These results indicate that the greatest variability in results from BV:BJ matrices is in the RT-PCR stage as it includes both reverse transcription of RNA and amplification of cDNA.

Reproducibility was further assessed by amplifying beta transcripts from RNA extracted from a second, separate sample of CD3⁺ T cells from the 36 yr donor in duplicate BV:BJ matrix assays. The results that are described in Figure 10 show that the statistics that quantify diversity (entropy, percent of amplified BV:BJ pairs, and median Ct) were comparable to the respective statistics from duplicate analysis of the first RNA sample described in Figures 4 and 6. Combining the four replicate assays, the mean entropy was 0.70 (± 0.05), and the mean median Ct was 19.7 cycles (± 0.38). The percentages of BV:BJ pairs that yielded amplification were identical for all four assays. The results of replicate assays that were run to assess reproducibility suggest that the BV:BJ matrix method yields dependable estimates of beta transcript diversity even when using independently extracted RNA templates.

3.5. Effect of Template Titration

The effects of diluting bead-enriched template on results from real-time PCRs were investigated to determine if reduction in template altered estimates of diversity that were generated in the real-time PCRs. Real-time PCRs were performed with bead-enriched template that was diluted 1/40 rather than the 1/10 dilution in the standard assay. The results indicate that the 1/4 dilution changed neither the entropy estimate of diversity nor the number of BV:BJ primer pairs that yielded amplification (Figure 11) when compared with the results from the first assay that had used those bead-enriched templates (Figure 4). The four-fold dilution did result in an average 2.6 cycle increase in median Ct which is comparable to the two cycle shift that is predicted by two-fold decreases in amplicons with each increasing cycle of real-time PCRs. These results are important because they show that the entropy estimate of beta transcript diversity is not dependent on the amount of template within at least the range used in these experiments. This is the expected result since entropy estimates are sensitive to the distributions of Ct estimates rather than the median Ct estimates.

3.6. Sequencing of Over-Represented Transcripts

The use of 533 BV:BJ primer pairs increases the efficiency of identifying beta transcripts expressed by expanded T cell clonotypes over that exhibited by previous methods. Dissociation curves that were generated by amplification in real-time PCRs described in Figure 4 were examined in order to identify products with relatively tight dissociation curves. We have previously shown that products amplified from single transcripts exhibit relatively tight dissociation curves although such dissociation curves are not always predictive of single products (Wettstein et al., 2008). Ten products were selected and sequenced using the respective BV gene primers. A single product yielded single-copy sequence that extended from the BV gene through the CDR3 and into the BJ gene. However, this sequence was clearly the product of a non-productive rearrangement since the nucleotide sequence that encoded the CDR3 included a stop codon. The other nine products yielded clear BV gene sequences up to the segment that encodes the CASS element that is at the carboxy end of the BV region, but no discernable CDR3 and BJ sequences were obtained for the remainders of the products indicating that the nine products included multiple sequences.

4. Discussion and conclusion

Healthy T lymphocyte populations must maintain sufficient diversity to respond to challenges with broad arrays of pathogens that have the capacity to change their expressed antigens in attempts to avoid T cell recognition. T lymphocytes must also continuously control chronic infections, and the strength of the responses to these two classes of pathogens is strongly dependent on the diversity of TcR repertoires. The BV:BJ matrix was originally developed to quantify the diversity of murine T lymphocytes (Wettstein et al., 2008), and this communication describes our extension of the method to human T lymphocytes which included more rigorous selection of BV gene-specific primers due to the increased number of human BV genes. The increased inter-genic homology and number of BV:BJ pairs made the preliminary amplification in RT-PCRs even more essential for resolution of the maximal number of BV:BJ combinations and analysis of biological samples that ranged from circulating T cell populations to biopsies of tumors and inflammatory sites.

The two stages of RT-PCR and real-time PCR involve a total of 65 cycles of amplification that could result in considerable variability in amplification. The data presented here suggest that the results from the BV:BJ matrix are reproducible with replicate assays generating comparable estimates of diversity. Not unexpectedly, the greatest variability appears to be in the pooled RT-PCR stage that includes both reverse transcription and amplification. The vast majority of BV:BJ primer pairs amplified products with relatively low variation in Ct values, but there were BV:BJ primer pairs that yielded more variable Ct values in replicate assays. The basis for the increased levels of variability may lie in the diverse transcripts in RNA templates. Unlike RNA transcripts from single genes that are generally amplified by PCR methods, beta chain transcripts that carry a single BV gene include multiple BJ genes as well as diverse CDR3s. This complexity would be expected to result in diverse secondary structures, some of which may be difficult to denature prior to reverse transcription. A

modification that was specifically required for the human BV:BJ matrix was an additional denaturation step that was necessary for increased reproducibility. Even with this modification, six BV genes were excluded from analysis due to highly variable amplification when paired with specific BJ genes. Working under the assumption that this variability is due to diverse secondary structures, we are presently varying denaturation times and temperatures to increase the number of BV genes that can be analyzed in the matrix without sacrificing the efficiency of amplification with other BV:BJ primer pairs.

The BV:BJ matrix is a significant advance in the evolution of methods for analysis of TcR repertoire diversity even though it and other methods may ultimately be replaced by deep sequencing approaches (Robins et al., 2009). Amplification with 533 BV:BJ primer pairs has the power to identify and sequence unprecedented numbers of beta chain transcripts that are expressed by expanded T cell clonotypes. This level of resolution significantly increases the amount of information that can be gained from the analysis of T cell subpopulations that are specific for single immunogenic peptides or that infiltrate tumors and sites of inflammation. The BV:BJ matrix is also distinguished from previous methods of repertoire analysis by its use of real-time PCR amplification to estimate relative frequencies of individual BV:BJ combinations that are then used to generate a single statistic that numerically describes diversity of beta transcripts expressed by single T cell populations. The calculation of a single statistic greatly simplifies the statistical analysis of estimates of repertoire diversity exhibited by different experimental groups of T cell populations. Such statistical analysis has been particularly difficult with spectratyping results that are essentially compilations of BV-specific distributions of CDR3 lengths without rigorous quantitation of amplification (Pannetier 1993).

Shannon entropy was chosen as the single statistic to describe beta transcript diversity when the BV:BJ matrix method was designed. Shannon entropy was originally derived for information theory to represent the expected value of information that is contained in a message (Shannon 1949). The frequencies of individual characters comprise the sole input for estimation of the total information that is contained within a message or, in other words, the level of unpredictability in the message. We reasoned that repertoires of TcR beta transcripts are analogous to communication messages in that they express variable distributions of BV:BJ gene combinations that can be considered as “characters” whose frequencies and effects on “information content” can be quantified. Accordingly, Shannon entropy as a repertoire descriptor should exhibit the characteristics of its application to communication: (1) high frequency characters have the greatest effects on predictability/information content and (2) use of all characters to estimate entropy without regard to their individual frequencies which yields continuous variables with increased precision and reduced coefficient of variation (Shannon 1949).

The data presented in this communication demonstrate that Shannon entropy estimates for TcR repertoire diversity respond to variation in BV:BJ frequencies as do entropy estimates for communication messages. First, relatively high frequency BV:BJ pairings, as indicated by relatively low Ct estimates, have the greatest effects on entropy estimates of diversity since they have the greatest effects on lowering diversity of beta transcripts in the same way that repeating characters reduce the information content of messages. This effect can be seen

by modifying the data represented in Figure 10 (Assay 1). Reducing the lowest Ct value of 9.9 cycles to 8 cycles, that would be expected from an ≈ 4 -fold increase in transcript number, resulted in a decrease in entropy from 0.73 to 0.65. An additional decrease to 7 cycles resulted in an even greater decrease in entropy from 0.65 to 0.56 demonstrating the effects on entropy of exponentially increasing numbers of transcripts carrying only a single BV:BJ combination. Comparably low Ct values for single BV:BJ combinations have been observed with patients with chronic lymphocytic leukemia in preliminary studies with the expected reduction in entropy estimates (data not shown). Second, unlike median and mean Ct statistics, entropy estimates of TcR repertoire diversity are consistent over a range of amounts of RNA template since frequencies of BV:BJ pairs are calculated using both individual Ct values and total Ct values from all real-time PCRs. The use of Shannon entropy estimates as the single descriptors of TcR repertoire diversity also facilitates the statistical analysis of repertoire diversity through traditional, statistical methods.

The sensitivity of Shannon entropy to over-represented BV:BJ combinations is important for meaningful functional analysis of T cell repertoires in that it highlights and quantifies expanded T cell clonotypes that can potentially dominate antigen recognition by T cell populations. The importance of these dominant T cell clonotypes for human responses to chronic infections is well-documented for a number of infectious agents (Weekes et al., 1999; Levitsky 1998). Over-represented T cell clonotypes are also significant for their consequential reduction of the frequencies of other T cell clonotypes that reduces their abilities to respond to newly encountered antigens. The combination of BV:BJ-specific amplification in real-time PCRs and Shannon entropy analysis to quantitate diversity of T cell repertoires has the potential to increase our knowledge of the diversity of T cell repertoires and rigorously evaluate the effects of over-represented T cell clonotypes on those T cell populations.

Acknowledgments

The authors gratefully acknowledge the technical assistance of Ms. Minzhi Zhang, Ms. Bharati Sanyal, and Mr. Michael Strausbauch and the secretarial assistance of Ms. DeAnn Frederixon. PJW is supported by NIH grant UL1TR000135, and NEK is supported by NIH grant CA95241 and by CLL Global Foundation.

Abbreviations

TcR	T cell receptor
BV	beta variable
BJ	beta joining
BC	beta constant
CDR3	complementarity-determining region 3
Ct	cycle threshold

References

- Caignard A, Dietrich P, Morand V, Lim A, Pannetier C, Leridant AM, Hercend T, Even J, Kourilsky P, Triebel F. Evidence for T-cell clonal expansion in a patient with squamous cell carcinoma of the head and neck. *Cancer Research*. 1994; 54:1292–1297. [PubMed: 8118818]
- Chien Y-H, Gascoigne RJ, Kavalier J, Lee NE, Davis MM. Somatic recombination in a murine T-cell receptor gene. *Nature*. 1984; 309:322–326. [PubMed: 6328306]
- DePalma R, Gorsky J. Restricted and conserved T-cell repertoires involved in allorecognition of class II major histocompatibility complex. *Proc. Natl. Acad. Sci. USA*. 1995; 92:8836–8840. [PubMed: 7568027]
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucl. Acids Res*. 2008; 36:W465–W469. [PubMed: 18424797]
- Garcia K, Degano M, Stanfield R, Brunmark A, Jackson M, Peterson P, Teyton L, Wilson I. An $\alpha\beta$ T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science*. 1996; 274:209–219. [PubMed: 8824178]
- Guillaume P, Baumgaertner P, Neff L, Ruff N, Wettstein PJ, Speiser DE, Luescher IF. Novel soluble HLA-A2/MELAN-A complexes selectively stain a differentiation defective subpopulation of CD8⁺ T cells in patients with melanoma. *Intl. J. Cancer*. 2010; 127:910–23.
- Guillet M, Brouard S, Gagne K, Sebille F, Cuturi MC, Delsuc MA, Souillou JP. Different qualitative and quantitative regulation of V TCR transcripts during early acute allograft rejection and tolerance induction. *J. Immunol*. 2002; 168:5088–5095. [PubMed: 11994461]
- Johnston SL, Borson ND, Wettstein PJ. Spectratyping of TCRs expressed by CTL infiltrating minor histocompatibility antigen-disparate allografts. *J. Immunol*. 1997; 159:5233–5245. [PubMed: 9548462]
- Kedzierska K, La Gruta NL, Davenport MP, Turner SJ, Doherty PC. Contribution of T cell receptor affinity to overall avidity for virus-specific CD8⁺ T cell responses. *Proc. Natl. Acad. Sci. USA*. 2005; 102:11432–11437. [PubMed: 16061802]
- Khan AM, Miotto O, Heiny AT, Salmon J, Srinivasan KN, Nascimento E, Marques ET, Brusic V, Tan TW, August JT. A systematic bioinformatics approach for selection of epitope-based vaccine targets. *Cell Immunol*. 2006; 244:141–147. [PubMed: 17434154]
- Lefranc MP, Giudicelli V, Kaas Q, Duprat E, Jabado-Michaloud J, Scaviner D, Ginestoux C, Clement O, Chaume D, Lefranc G. IMGT, the international ImMunoGeneTics information system. *Nucl. Acids Res*. 2005; 33:D593–D597. [PubMed: 15608269]
- Levitsky V, de Campos-Lima PO, Frisan T, Masucci MG. The clonal composition of a peptide-specific oligoclonal CTL repertoire selected in response to persistent EBV infection is stable over time. *J. Immunol*. 1998; 161:594–601. [PubMed: 9670932]
- McHeyzer-Williams LJ, Panus JF, Mikszta JA, McHeyzer-Williams MG. Evolution of antigen-specific T cell receptors in vivo: preimmune and antigen-driven selection of preferred complementarity-determining region 3 (CDR3) motifs. *J. Exp. Med*. 1999; 189:1823–1837. [PubMed: 10359586]
- McHeyzer-Williams MG, Davis MM. Antigen-specific development of primary and memory T cells in vivo. *Science*. 1995; 268:106–111. [PubMed: 7535476]
- Musette P, Bequet D, Delarbre C, Gachelin G, Kourilsky P, Dormont D. Expansion of a recurrent V beta 5.3⁺ T-cell population in newly diagnosed and untreated HLA-DR2 multiple sclerosis patients. *Proc. Natl. Acad. Sci. USA*. 1996; 93:12461–6. [PubMed: 8901604]
- Pannetier C, Cochet M, Darche S, Casrouge A, Zoller M, Kourilsky P. The sizes of the CDR3 hypervariable regions of the murine T-cell receptor α chains vary as a function of the recombined germ-line segments. *Proc. Natl. Acad. Sci. USA*. 1993; 90:4319–4323. [PubMed: 8483950]
- Pannetier, C.; Levraud, JP.; Lim, A.; Even, J.; Kourilsky, P. The immunoscope approach for the analysis of T cell repertoires.. In: Oksenberg, JR., editor. *The antigen T cell receptor: Selected protocols and applications*. R.G. Landes Company; Georgetown, TX: 1997. p. 287-325. Section 9
- Puisieux I, Even J, Pannetier C, Jotereau F, Favrot M, Kourilsky P. Oligoclonality of tumor-infiltrating lymphocytes from human melanomas. *J. Immunol*. 1994; 153:2807–2818. [PubMed: 8077684]

- Ria F, van den Elzen P, Madakamutil LT, Miller JE, Maverakis E, Sercarz EE. Molecular characterization of the T cell repertoire using immunoscope analysis and its possible implementation in clinical practice. *Current Molecular Medicine*. 2001; 1:297–304. [PubMed: 11899078]
- Robins HS, Campregher PV, Srivastava S, Wachter A, Turtle CJ, Kabsai O, Riddell SR, Warren EH, Carlson CS. Comprehensive assessment of T-cell receptor α -chain diversity in T cells. *Blood*. 2009; 114:4099–4107. [PubMed: 19706884]
- Shannon, CE. *The mathematical theory of communication*. University of Illinois Press; 1949.
- Stewart JJ, Lee CY, Ibrahim S, Watts P, Shlomchik M, Weigert M, Litwin S. A Shannon entropy analysis of immunoglobulin and T cell receptor. *Mol. Immunol*. 1997; 34:1067–1082. [PubMed: 9519765]
- Weekes MP, Wills MR, Mynard K, Carmichael AJ, Sissons JG. The memory cytotoxic T-lymphocyte (CTL) response to human cytomegalovirus infection contains individual peptide-specific CTL clones that have undergone extensive expansion in vivo. *J. Virol*. 1999; 73:2099–2108. [PubMed: 9971792]
- Wettstein PJ, Strausbauch M, Therneau TM, Borson NB. The application of real-time PCR to the analysis of T cell repertoires. *Nucl. Acids Res*. 2008; 36(21):e140. PMID 18835849. [PubMed: 18835849]

Highlights

- A method for high resolution analysis of T lymphocyte repertoires is described.
- Real-time PCRs amplify all pairs of variable and joining genes for beta chains.
- The amplifications produce a single statistic that describes repertoire diversity.
- The method increases identification of beta chains expressed by expanded T cells.

Figure 1. Schematic diagram of the BV:BJ matrix method with pooled RT-PCRs and 533 BV:BJ-specific real-time PCRs.

Figure 2.

Phylogenetic tree of expressed human BV genes. BV nucleotide sequences were retrieved from the IMGT database (Lefranc et al., 2005). The sequences were aligned, and the unrooted phylogenetic tree was constructed by Maximum-Likelihood with published software (Dereeper et al., 2008).

Figure 3.

Representation of Ct values from amplifications with 533 BV:BJ primer pairs and RNA from T lymphocytes from a 36 yr old human donor. Total RNA was extracted from CD3⁺ T lymphocytes and amplified with the BV:BJ matrix method. Ct values were estimated for all real-time PCRs, and Ct values are figuratively shown by grayscale infill ranging from black (5-10 cycles) to white (36-40 cycles) in 5 cycle increments.

Figure 4.

Distribution of Ct values for BV:BJ-specific amplifications of beta transcripts expressed by T lymphocytes. The Ct values presented in Figure 3 were placed in two cycle bins in the histogram, and the entropy value, median Ct, and percentage of BV:BJ primer pairs that yielded amplification (Ct < 36 cycles) were calculated.

Figure 5.

The majority of beta transcripts in the healthy human test subject carry only a minority of the 533 BV:BJ gene combinations. The Ct values from individual BV:BJ-specific reactions were used to estimate the contribution (in percentage) of each BV:BJ combination to the estimated total number of beta transcripts that was based on the total sum of Ct values for the single template. These percentages were then ranked from low to high and sequentially added to equal a total of 100% in a “cumulative” curve.

Figure 6.

Distribution of Ct values for BV:BJ-specific amplifications in a duplicate BV:BJ matrix assay. The RNA template used for Figure 4 was amplified in a replicate BV:BJ matrix. The Ct values were placed in two cycle bins in the histogram, and the entropy value, median Ct, and percentage of BV:BJ primer pairs that yielded amplification (Ct < 36 cycles) were calculated.

Figure 7.

Variation in Ct values for each BV:BJ combination in the duplicate BV:BJ matrix assays. The differences between Ct values for each BV:BJ primer pair that are presented in Figures 4 and 6 were calculated, and the absolute values of those differences were grouped according to BV gene and plotted. The horizontal bars denote interquartile ranges. The median difference was 1.1 cycles and the interquartile range was 0.48-2.13 cycles for the 533 BV:BJ primer pairs.

Figure 8.

Distributions of Ct values and statistics for amplifications in duplicate real-time PCRs using a single set of templates from pooled RT-PCRs. These real-time PCRs used the bead-enriched templates from the pooled RT-PCRs that are described in Figure 4. The plots of Ct values and the calculations of entropy values, median Cts, and percentages of BV:BJ primer pairs that yielded amplification are as described for Figure 4.

Figure 9.

Variation in Ct values for each BV:BJ combination in duplicate real-time PCRs. The differences between Ct values for each BV:BJ primer pair that are presented in Figure 8 were calculated, and the absolute values of those differences were grouped according to BV gene and plotted. The horizontal bars denote interquartile ranges. The median difference was 0.3 cycles and the interquartile range was 0.11-0.77 cycles for the 533 BV:BJ primer pairs.

Figure 10.

Distributions of Ct values and statistics for duplicate BV:BJ matrix analyses of a second RNA template. A second aliquot of CD3⁺ T cells was enriched from the PBLs donated by the 36 yr subject, and total RNA was extracted for analysis in two BV:BJ matrices. The plots of Ct values and the calculations of entropy values, median Cts, and percentages of BV:BJ primer pairs that yielded amplification are as described for Figure 4.

Figure 11.

The effect of diluting bead-bound templates on the distribution of Ct values and statistics from BV:BJ-specific real-time PCRs. Bead-bound templates from pooled RT-PCRs used for the assay described by Figure 4 were diluted 1/40 for amplification in real-time PCRs rather than the routine 1/10 dilution. The plots of Ct values and the calculations of entropy values, median Cts, and percentages of BV:BJ primer pairs that yielded amplification are as described for Figure 4.

Table 1**BV Primer Sequences and Distributions in Pooled RT-PCRs**

BV Gene	Sequence	RT-PCR Pool
BV2	GACAGGAAGTGATCTTGCGC	3
BV3-1	GACAAGTCCATTAAATGTGAACAAAA	4
BV4-1	AGAAGTCTTTGAAATGTGAACAACATA	4
BV4-2,3	AACAACATCTGGGGCATAACG	2
BV5-1	CCCTATCTCTGGGCATAGGAG	1
BV5-4	CTTCTCAGTCTGGGCACAACAC	2
BV5-5	TCTCCTATCTCTGGGCACAAGAG	3
BV5-6	TCTCCTAAGTCTGGGCATGACA	1
BV5-8	CCTATCTCTGGGCACACCAGT	6
BV6-1	TGCCCAGGATATGAACCATAACT	2
BV 6-2,3,5	GCCCAGGATATGAACCATGAA	4
BV6-4	AGATGTACCCAGGATATGAGACATAAT	1
BV6-6	TGTACCCAGGATATGAACCATAACTA	7
BV7-2,3	GGTGTGATCCAATTTTCAGGTCATA	6
BV7-6,8	CCAATTTTCGGGTCATGTATCC	3
BV7-7	GATCCAATTTTCGAGTCATGCAA	7
BV7-9	ATCCAATTTCTGAACACAACCG	8
BV9	TGCTCCCCTAGGTCTGGAGAC	5
BV10-3	ACCAGACTGAGAACCACCGC	3
BV11-1	TGGCTTTTTGGTGTGATCCTAT	5
BV11-2	GGCTTTTTGGTGCAATCCTATA	6
BV11-3	GGCTTTTTGGTGCAATCCTATT	8
BV12-3	ACCAATTTTCAGGCCACAACCTC	4
BV12-4	GTAAACCAATTTTCAGGACACGACTA	5
BV12-5	CAGCCAATTTTAGGCCACAATAC	1
BV13	CCACTCTGAAATGCTATCCTATCC	3
BV14	GACCCAATTTCTGGACATGATAAT	7
BV15	GTTCTCAGACTTTGAACCATAACGT	7
BV18	CCCAATGAAAGGACACAGTCAT	2
BV19	TGAACAGAATTTGAACCACGATG	1
BV20	CTGTGAAGATCGAGTGCCGTT	6
BV24	TGTTCTCAGACTAAGGGTCATGATAG	4
BV25	CACTCTGGAATGTTCTCAAACCA	8
BV27	TTGTTCTCAGAATATGAACCATGAGTAT	5
BV28	GGAATGTGTCCAGGATATGGAC	8
BV29	AAGTCGATAGCCAAGTCACCAT	7
BV30	GTGGAGGGAACATCAAACCC	8
Constant	Bio-GGCTCAAACACAGCGACCTC	All

Table 2

Sequences of BV and BJ Primers for Real-Time PCRs

Gene	Sequence
BV2	AATCTTGGGGCAGAAAGTCG
BV3-1	GATAATGTTTAGCTACAATAATAAGGAGC
BV4-1	TCATGTTTGTCTACAGCTATGAGAAA
BV4-2	GAGCTCATGTTTGTCTACAACTTTAAA
BV4-3	AGCTCATGTTTGTCTACAGTCTTGAA
BV5-1	TTCTCTTTGAATACTTCAGTGAGAC
BV5-4	CAGTTTATCTTTCAGTATTATAGGGAGG
BV5-5	CCCAGTTTATCTTTCAGTATTATGAGAA
BV5-6	CCAGTTTATCTTTCAGTATTATGAGGAG
BV5-8	CCTTTGGTATGACGAGGGTG
BV6-1	GATTTATTACTCAGCTTCTGAGGGT
BV6-2,3	GCTGATTCATTACTCAGTTGGTGAG
BV6-4	CTAAGGCTCATCCATTATTCAAATAC
BV6-5	CTGATTCATTACTCAGTTGGTGCT
BV6-6	GGGGCTGAAGCTGATTATTAT
BV7-2	TTTAAATTTACTTCCAAGGCAACA
BV7-3	TTCTAATTTACTTCCAAGGCACG
BV7-6	TGACTTACTTCAATTATGAAGCCC
BV7-7	CCCAGAGTTTCTGACTTACTTCAATTA
BV7-8	CTGACTTATTTCCAGAATGAAGCTC
BV7-9	GGGCCCAGAGTTTCTGACTTAC
BV9	CAGTTCCTCATTAGTATTATAATGGA
BV10-3	GATCCATTACTCATATGGTGTTAAAGA
BV11-1	CCCGGAGCTTCTGGTTCAA
BV11-2	CCAAAGCTTCTGATTAGTTTCA
BV11-3	GATTCGATATGAGAATGAGGAAGC
BV12-3	ATTTACTTTAACAACAACGTTCCG
BV12-4	ATTTACTTTAACAACAACGTTCCG (same as BV12-3)
BV12-5	ACTTCCGCAACCGGGCT
BV13	ATTTTCGTTTTATGAAAAGATGCAG
BV14	TCTGTTACATTTTGTGAAAGAGTCTAAA
BV15	AAAGCTGCTGTTCCTACTACTATGA
BV18	ATGGTTTATCTCCAGAAAGAAAATATC
BV19	ATTGATCTACTACTCACAGATAGTAAATGAC
BV20	CAACTTCCAATGAGGGCTCC
BV24	GCCTACGGTTGATCTATTACTCCT
BV25	CTCATCCACTATTCCTATGGAGTTAA
BV27	AGATCTACTATTCAATGAATGTTGAGG
BV28	GCTGATCTATTTCTCATATGATGTTAAAT

Gene	Sequence
BV29	TGACACTGATCGCAACTGCAA
BV30	CTTCTACTCCGTTGGTATTGGC
BJ1.1	TGCCTTGTCCAAAGAAAGCT
BJ1.2	CGAACCGAAGGTGTAGCCA
BJ1.3	AACTTCCCTCTCCAAAATATATGGT
BJ1.4	TTCCACTGCCAAAAACAGTTT
BJ1.5	ACCAAAATGCTGGGGCTG
BJ1.6	CCCAAAGTGGAGGGGTGAA
BJ2.1	CCCGAAGAACTGCTCATTGT
BJ2.2	CCTTCTCCAAAAACAGCTCC
BJ2.3	CTGGGCCAAAATACTGCGTA
BJ2.4	CGGCGCCGAAGTACTGAAT
BJ2.5	TGGCCCGAAGTACTGGGTC
BJ2.6	GCCCCGAAAGTCAGGACG
BJ2.7	GGCCCGAAGTACTGCTCGTA