Clonotype pattern in T-cell lymphomas map the cell of origin to immature lymphoid precursors

Aishwarya Iyer,[1] Dylan Hennessey,[1] and Robert Gniadecki[1-3]

[1]Division of Dermatology, Department of Medicine, and [2]Department of Oncology, Cross Cancer Institute, University of Alberta, Edmonton, AB, Canada; and [3]Department of Dermatology, Bispebjerg Hospital, University of Copenhagen, Denmark

Key Points

- The majority of TCLs are clonotypically oligoclonal, although the patterns of oligoclonality are varied.

- TCR rearrangements in TCL show a biased usage of V and J segments and presentation of public CDR3 sequences shared across unrelated patients.

Mature T-cell lymphomas (TCLs) are rare, clinically heterogeneous hematologic cancers with high medical need. TCLs have an inferior prognosis which is attributed to poor understanding of their pathogenesis. On the basis of phenotypic similarities between normal and neoplastic lymphocytes, it has been assumed that TCLs develop in the periphery, directly from various subtypes of normal T cells. To address the debated question of the cell of origin in TCLs, we attempted to identify the highly variable complementarity-determining regions (CDRs) of T-cell receptors (TCRs) to trace the clonal history of the T cells. We have collected previously published whole-genome, whole-exome, and whole-transcriptome sequencing data from 574 patients with TCL. TCR clonotypes were identified by de novo assembly of CDR3 regions of TCRα, TCRβ, and TCRγ. We have found that the vast majority of TCLs are clonotypically oligoclonal, although the pattern of oligoclonality varied. Anaplastic large-cell lymphoma was the most diverse comprising multiple clonotypes of TCRα, TCRβ, and TCRγ, whereas adult TCL or leukemia and peripheral TCLs often showed monoclonality for TCRβ and TCRγ but had diverse TCRα clonotypes. These patterns of rearrangements indicated that TCLs are initiated at the level of the lymphoid precursor. In keeping with this hypothesis, TCR rearrangements in TCLs resembled the pattern seen in the human thymus, which showed biased usage of V (variable) and J (joining) segments of high combinatorial probability resulting in recurrent public CDR3 sequences shared across unrelated patients and different clinical TCL entities. Clonotypically diverse initiating cells may seed target tissues that are then responsible for disease relapses after therapy.

Introduction

T-cell lymphomas (TCLs) are a heterogenous group of 29 malignancies that make up ~10% of non-Hodgkin lymphomas, and their incidence is about 6000 cases per million.[1,2] Most TCLs are classified as mature TCLs because malignant cells phenotypically resemble mature T cells and harbor rearranged T-cell receptor (TCR) variable (V), diversity (D), and joining (J) genes; many express the TCR $\alpha\beta$ hetero-dimer.[3,4] Striking similarities between subsets of normal T cells and the malignant cells in TCLs (eg, to T-regulatory cells in adult TCL or leukemia or central memory T cells to Sezary syndrome [SS]) led to the

---

The full-text version of this article contains a data supplement.

general acceptance of the theory that different subsets of mature T cells are the cells of origin in mature lymphomas.[5]

TCR gene sequences are excellent markers of T-cell lineage, because TCRγ, TCRβ, and TCRα loci are sequentially rearranged during different stages of intrathymic maturation of T cells from a diverse pool of V(D)J genes. The TCRβ locus contains 47 V (TRBV), 2 D (TRBD), and 13 J (TRBJ) segments, whereas the TCRα locus comprises 42 TRAV and 61 TRAJ segments, which recombine to yield unique DNA sequences that are retained in genomes of all daughter cells.[6] This diversity of the unique DNA sequences is enhanced by the insertion of random palindromic sequences during the recombination step of V(D)J or VJ. Analysis of the complementarity determining region 3 (CDR3), the most diverse fragment of the TCR chain coded by the V(D)J joining region and involved in antigen recognition, has long been used in the molecular diagnosis of TCLs and leukemias. The repertoire of normal human TCRs is in the range of 10^6 to 10^7 clonotypes (unique CDR3 sequences defining a T-cell clone),[7-9] whereas in mature TCLs, it is dominated by a single clonotype, representing the clonally expanded malignant cells (ie, malignant clonotypes).[10]

Clonotypic analyzes of TCLs and leukemias are not only useful for diagnostic purposes but might also provide important clues to the pathogenesis. Transformation of the mature T cell would result in perfect monoclonality in all TCR loci because TCR rearrangement does not occur in mature T cells, and TCRα and TCRβ clonotypes would be inherited from the cell of origin. A different pattern of TCR rearrangement will be seen if the transformation occurs during earlier stages of T-cell development, before the completion of the final TCRα rearrangement. Assuming that early oncogenic mutations do not arrest cell development at the early stages (as is the case in T-cell acute lymphoblastic leukemia [T-ALL]), the T-cell is able to sequentially rearrange TCR, and one would expect oligoclonal or polyclonal patterns of clonotypes, characteristic of the stage in which the transformation happened. Indeed, our previous analyses of malignant T-cell clonotypes in cutaneous TCL revealed an oligoclonal pattern of TCRα and TCRβ but monoclonality of TCRγ, which suggests that original malignant transformation occurred at the stage of a double-negative thymocyte after rearrangement of the TCRγ locus.[11,12] A similar pathogenic scenario involving early thymocytes has been proposed for anaplastic large-cell lymphoma (ALCL),[13,14] but it is unclear whether other TCLs might also be derived from immature T cells.

During analysis of our CTCL sequencing data, we were puzzled by the finding that certain malignant clonotypes were repeatedly found in samples from unrelated patients; because of the vastness of the TCR repertoire, we had expected that malignant clonotypes would be unique for each patient. We have noted a preferential usage of certain Vα and Vβ segments in cutaneous TCLs[11] and were intrigued by the finding that similar biases in V-segment usage was reported in T-ALL, which is derived from lymphoid precursors arrested at various stages of development.[15,16]

In this study, we used the previously optimized bioinformatic pipeline for de novo assembly of CDR3 regions using paired-end RNA sequencing, whole-genome DNA sequencing (WGS), and whole-exome DNA sequencing (WES)[11] to analyze malignant TCR clonotypes in 574 samples that make up 8 major subtypes of TCL. We propose that TCR rearrangement patterns across different clinical TCL entities are incompatible with the mature T-cell transformation

hypothesis and support the immature lymphoid progenitor as the cell of origin.

Methods

TCL data collection and tumor cell fraction analysis

The graphical summary of the methods is shown in Figure 1. Raw fastq files of WES, WGS, and whole-transcriptome sequencing (WTS) were collected from 13 different studies (574 patients).[17-29] Ethics clearance to access the human genome data from restricted databases such as Database of Genotype and Phenotype (dbGaP) and European Genome-phenome Archive (EGA) was provided by the Health Research Ethics Board of the Alberta Cancer Committee under ethics ID-HREBA.CC-20-0118. In addition, samples from 50 healthy controls were also obtained.[30] The accession numbers of the studies are listed in supplemental Table 1, and the diagnostic groupings are listed in supplemental Table 2. Tumor cell fraction (TCF) was determined on the basis of copy number aberrations (CNAs) using Titan CNA (v. 1.17.1)[31] and/or Ichor CNA (v. 0.2.0)[32] in the 368 samples for which DNA sequencing data were available (supplemental Table 3). Data sets from Crescenzo et al[29] (SRP044708) lacked control DNA and therefore, a panel of normal, as recommended in the Genome Analysis Toolkit (GATK) pipeline,[33] was created to analyze the TCF for samples from that study. The majority of transcriptome data were not obtained from the same samples as their exome counterparts; therefore, TCF data calculated using the DNA sequencing was not applicable to the TCRs identified from the RNA sequencing.

TCR clonotype identification and quantification

TCR clonotypes were identified and quantified using MiXCR (v. 3.0.3 or 3.1.0).[34] We used superscript letters to indicate whether the clonotype was identified by the DNA, RNA, or amino acid (aa) sequence (eg, TCRα^{DNA}). Default recommendations for analysis of exome and transcriptome data were used with the exception of eliminating the assembly of short reads to identify additional TCR clonotypes. Short- and long-read alignments were included for WTS; however, for WES data, partial reads were filtered out because they might be the captures of only V or J sequences. After clonotypes were identified, further elimination of clonotypes based on mutational positions was performed and best predictive VJ combinations were included in further analyses. When TCFs were available, the malignant clonotypes (ie, clonotypes of the malignant T cells) were determined by including the most abundant clonotypes to match the TCFs of the sample (Figure 1). Shared clonotype identification was based on TCRβ^{aa} and was used to calculate clonotype sharing (Jaccard index) with Immunarch (v. 0.6.6).[35] An estimate of the diversity was limited to malignant clonotypes. To estimate the number of malignant TCRβ clonotypes, the clonotypes of the cumulative frequency matching TCFs from every sample were counted and averaged. On the basis of the average number of clonotypes of 9.4, the 10 most frequent clonotypes from every sample were used for the inverse Simpson index. Probability of generation (P_{gen}) of CDR3 was calculated using OLGA (v. 1.2.3)[36] for the TCL samples and samples from the 50 healthy controls. Immunarch, Graphpad Prism (v. 9), and Excel (v. 16.16.27) were used for data representation.
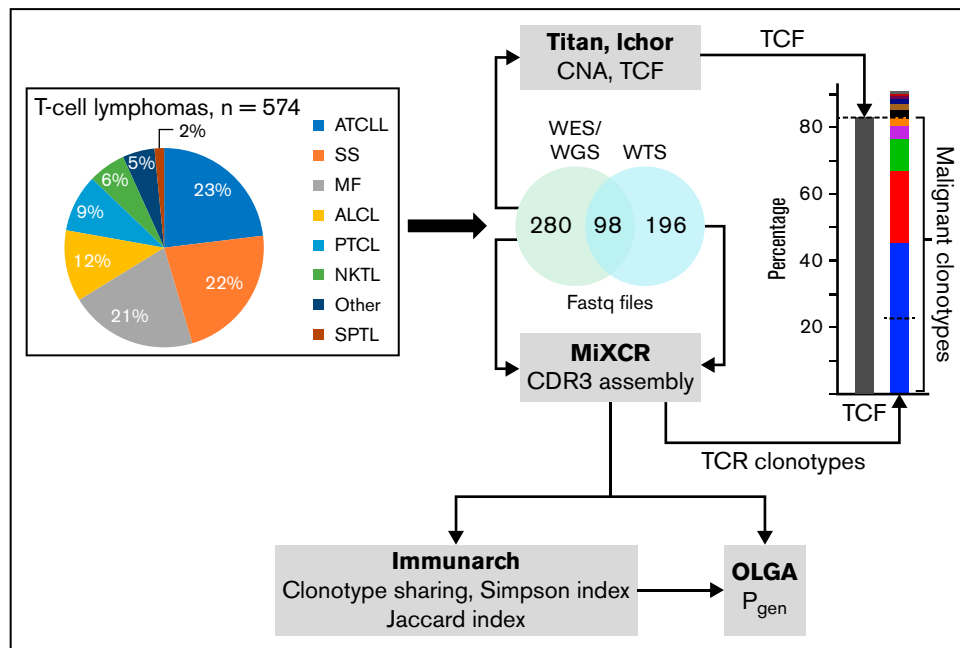
Figure 1. Summary of the methodological approach to TCR analysis in TCLs. The pie chart represents the percentage of sequencing samples collected from different subtypes of TCLs that were analyzed for TCF and the TCR clonotypes. Details for the diagnostic groupings are provided in supplemental Table 2. TCF was estimated on the basis of the chromosomal aberration identified using TitanCNA or IchorCNA. TCR clonotypes were identified using MiXCR. The diversity index (inverse Simpson), clonotype sharing (Jaccard index), and probability of clonotype generation (P_{gen}) were analyzed using Immunarch and OLGA pipelines, as described in "Methods." SPTL, subcutaneous panniculitic TCL.

Results

Clonotypic TCRβ diversity in TCLs

We analyzed data from 574 samples from the following TCL diagnoses: mycosis fungoides (MF; a primary cutaneous TCL; n = 122), SS (leukemic form of primary cutaneous TCL; n = 130), ALCL (n = 67), adult TCL or leukemia (ATCLL; n = 135), natural killer TCL (NKTL; n = 25), peripheral TCL (PTCL; n = 55), and subcutaneous panniculitis-like TCL (SPTL; n = 10). We were unable to classify 30 samples into any of the above categories, which were grouped as "other TCLs" (supplemental Table 2).

To determine the sequences of CDR3, we used the previously described bioinformatic analysis of bulk sequencing data from WES, WGS, and WTS.[11] To differentiate between malignant clonotypes (ie, clonotypes derived from malignant T cells) and those contributed from the infiltrating reactive cells, we ranked the TCR^{DNA} clonotypes (clonotypes obtained from DNA sequencing) starting with the most abundant and included only those of the cumulative frequency matching the TCF in the sample (Figure 1).[12]

Among 378 samples for which genomic data (WES or WGS) was available for calculating TCFs, the majority (75%) showed more than 1 TCRβ^{DNA} clonotype attributed to malignant T cells (matching TCF) (Figure 2). We considered 3 rearrangements as evidence of oligoclonality because although ~60% of T cells rearrange TCRβ on only 1 chromosome, the remaining 40% are biallelic rearrangements (the 60-40 rule).[37,38] Oligoclonality of malignant clonotypes was most prevalent in ALCL (43 of 47) and least common in NKTL (12 of 25) (Figure 2).

The allelic exclusion of TCRβ locus ensures that a mature T cell expresses only 1 clonotype,[37] even in the presence of biallelic rearrangements.[39,40] Thus, the single TCRβ clonotype at the level of RNA (TCRβ^{RNA}) unequivocally identifies the T-cell clone. TCFs could not be calculated for WTS samples because the matching WES or WGS data were not available; therefore, we analyzed TCRβ clonality by applying the relative frequency threshold of 25%. This threshold has been proposed by other investigators,[41,42] and we considered it reasonable to adopt that frequency threshold because the majority (79% of TCL samples) had TCFs of 25% or higher (supplemental Table 3). In 26% of the samples, the most abundant TCRβ^{RNA} clonotype had a frequency below 25% (ie, the criterion of monoclonality was not met; Figure 3). In 6 samples. the relative frequency of the second most frequent TCRβ^{RNA} clonotype was >25%, and in 38 samples, the frequencies of the first and second clonotypes were comparable (no single dominant clonotype), indicating oligoclonality of malignant clonotypes.

We further assessed the diversity of the TCRβ clonotypes by using the inverse Simpson index, which varies from 1 (1 clonotype) to the maximum value representing the number of clonotypes (in our analysis, n = 10), taking into account clonotype concentration (supplemental Figure 1). For the TCRβ^{DNA} clonotypes, the diversity was highest in SS and ALCL, and lowest in ATCLL. The diversity of TCRβ^{RNA} clonotypes had a slightly different pattern with a decrease in diversity for SS and ALCL when compared with TCRβ^{DNA} clonotypes. This can be explained by the increase in the relative contribution of the most abundant TCRβ^{RNA} clonotype (higher clonotype concentration) relative to TCRβ^{DNA} which, in turn, is likely to be caused by a higher proportion of non-productively rearranged TCRβ
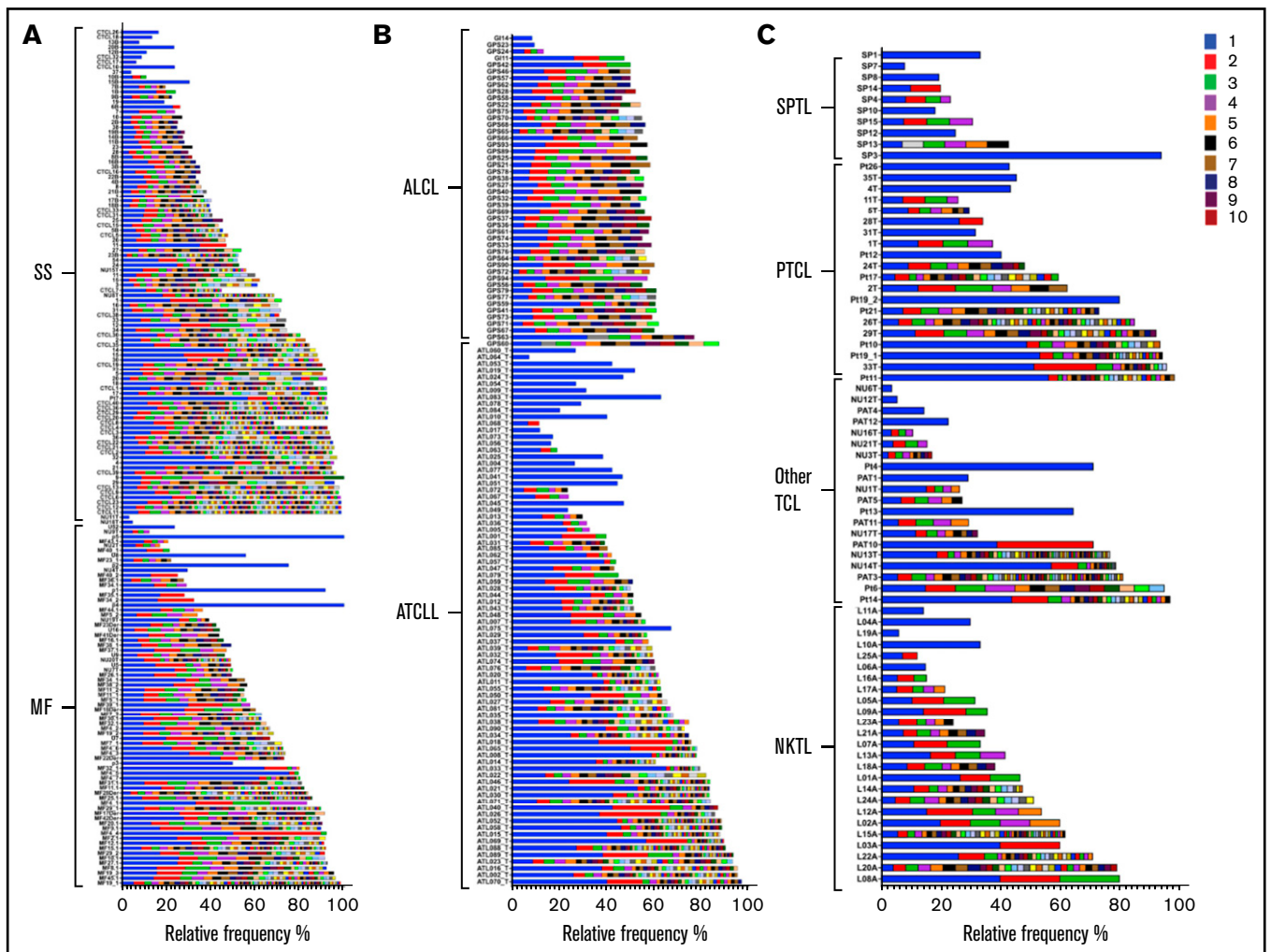
Figure 2. Frequency of TCRβ^{DNA} clonotypes in TCLs. A total of 378 DNA samples from 8 subtypes of TCLs were analyzed to identify the frequency of TCF and TCRβ^{DNA} clonotypes using WES and WGS data. TCRβ^{DNA} clonotypes corresponding to the TCF are indicated by the colored bars in the descending order of relative frequencies (the most abundant, rank 1, is the bright blue bar). The data are split by diagnosis: (A) cutaneous lymphomas: MF, and SS, (B) ALCL and ATCLL, and (C) SPTL, PTCL, Other TCL, and NKTL.

loci. The mean diversity for MF, ATCLL, and PTCL remain the same for TCRβ^{DNA} and TCRβ^{RNA}.

Patterns of TCRα, TCRβ, and TCRγ corroborate a branched evolution model of TCLs

The analyses described above suggested that TCR monoclonality is not a constant feature of TCLs, and approximately one-third of all samples display an oligoclonal pattern of malignant TCRβ, which is incompatible with the hypothesis of the mature T-cell origin hypothesis. To explore this further, we analyzed the frequencies of TCRα and TCRγ clonotypes. Unlike TCRβ, TCRα and TCRγ loci are not subjected to allelic exclusion and are usually rearranged at both chromosomes (Figure 4A). Moreover, TCRγ is rearranged before TCRβ, so the cells with the same TCRγ rearrangement may harbor several TCRβ rearrangements. The same is true for TCRα; a single TCRβ clone may comprise several (usually 2-4) TCRα clonotypes because of biallelic rearrangement and secondary rearrangements during expansion of the TCRβ clone (Figure 4A). It is thus possible

to deduce the stage of T-cell transformation from the relative frequencies of TCRα, TCRβ, and TCRγ.

Interestingly, most TCL samples showed a clearly monoclonal pattern of TCRγ dominated by 1 or 2 clonotypes (supplemental Figure 2). By plotting the frequency of the dominant (most abundant) TCRβ^{DNA} clonotype vs that of the 2 most abundant TCRγ clonotypes (Figure 4B), most of the cases fall into 2 categories. The samples located in the lower left quadrant had low-frequency TCRβ^{DNA} and low-frequency TCRγ and represented 32% of samples classified as oligoclonal for both TCRβ and TCRγ. Most ALCL cases were found in this group. The remaining (68%) of the samples showed a high proportion of TCRγ and a disproportionately low frequency of TCRβ^{DNA} (lower right quadrant), which indicates TCRγ monoclonality but oligoclonality with respect to TCRβ. Only single cases of TCL were in the right upper quadrant representing the high frequency of TCRβ and TCRγ, which would be characteristic for an expansion of a single mature T-cell clone. A similar plot of the frequencies of the dominant TCRβ^{DNA} vs the combined
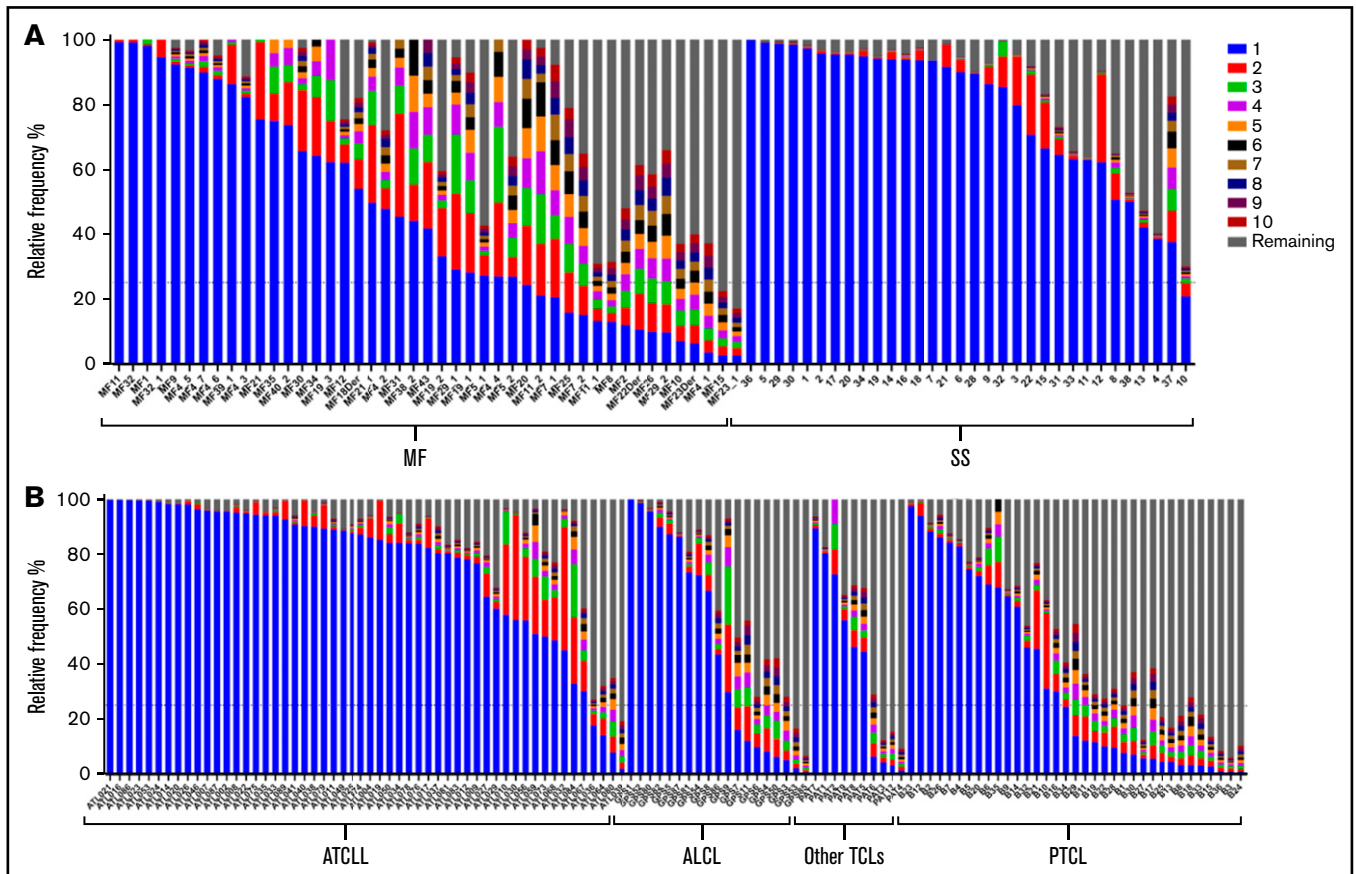
Figure 3. Frequency of TCRβ^{RNA} clonotypes in TCLs. A total of 196 RNA samples from 6 subtypes of TCLs were analyzed to identify the frequency of the TCRβ^{RNA} clonotypes. The 10 most abundant clonotypes are indicated by using the same color coding as in Figure 2; the remaining clonotypes are merged and their combined frequency is indicated by the gray bars. The data are split by diagnosis: (A) cutaneous lymphomas: MF and SS, (B) ATCLL, ALCL, Other TCLs and PTCL.

frequencies of the first and second (biallelic) TCRα^{DNA} showed a linear correlation within the lower left quadrant, which represents oligoclonality in both chains (Figure 4C). The samples clustered into 2 major groups: those with the higher frequency of TCRα and TCRβ (mostly ATCLL) and those with a lower frequency occupying the low-frequency portion of the quadrant (Figure 4C). The analysis of the TCRα^{RNA} and TCRβ^{RNA} clonotypes helped interpret the findings. As shown in Figure 4D, most ATCLL and SS samples showed monoclonality because they were located in the right upper quadrant, which represents high frequencies of the dominant clonotypes. However, ALCL, PTCL, MF, and other TCLs were scattered in the remaining quadrants, which underscores their oligoclonality not only at the DNA level but also at the level of TCR gene expression.

Although the above described findings showed significant variation in TCR clonality patterns within the same TCL diagnostic entities, general conclusions can be made regarding the putative cell of origin (Figure 4A). ALCL was the most heterogeneous with the clear oligoclonality of TCRα, TCRβ, and TCRγ, which indicated that the transformation takes place in immature thymocytes before the TCRγ recombination step. On the opposite end of the spectrum were SS and ATCLL, which in the majority of cases showed TCRγ monoclonality and high frequency of TCRβ, mapping the transformation event approximately at the TCRβ rearrangement step. However, none of the TCLs exhibited a consistent pattern of monoclonality,

defined at the DNA level by a maximum of 2 rearrangements of each TCRγ, TCRβ, and TCRα genes matching the TCF and at the RNA level by a single dominant TCRβ clonotype and a maximum of 2 TCRα clonotypes.

On the basis of the above described results, we hypothesized that TCRβ^{DNA} clonotypes may label neoplastic T-cell clones in the sample, and the distribution of those clonotypes is not random but might be a result of the clonal evolution of the disease. Systems comprising objects that grow in size and compete with each other (such as growth of cancer subclones in a tumor) often obey power laws such as Zipf-Mandelbrot law expressed as $S_{(k)} \sim S_{(1)}/k^{\gamma}$ where $S_{(k)}$ is the size of the species (in our case, the size of the TCRβ^{DNA} clonotype) of the rank k, $S_{(1)}$ is the size of the first ranked species, and γ is a scaling constant.[43,44] Indeed, the log-log plots in which mean abundance of the TCRβ^{DNA} clonotype was plotted vs its rank revealed the expected linear relationship for all TCLs, although the slope of the linear regression line varied from −0.34 for ALCL to −1.21 for ATCLL, which reflects different values of the constant γ for different types of TCLs (supplemental Figure 3). Reproducible linearity of the rank-size TCRβ plots for different TCLs were therefore consistent with zipfian dynamics[45] and supported the hypothesis that malignant T-cell clones are dynamically coherent and interactively evolve via neutral clonal evolution,[44] as postulated by us for MF.[17]
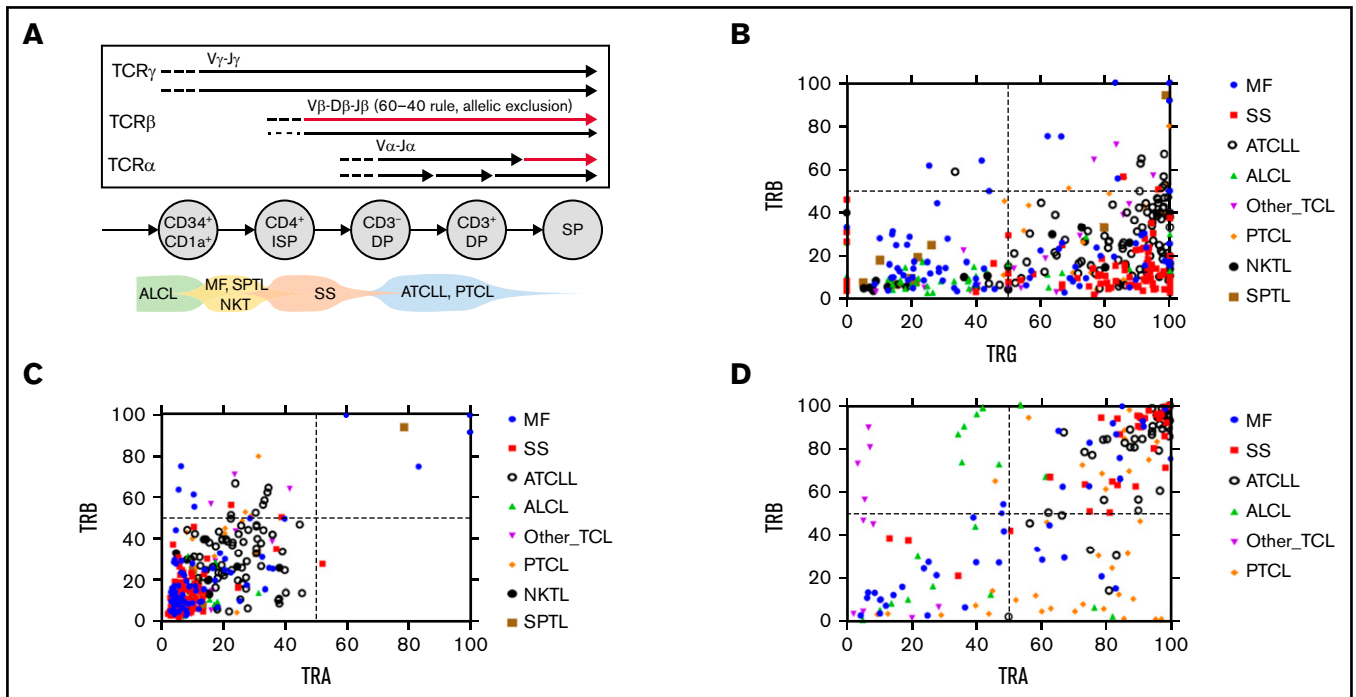
Figure 4. Oligoclonality in TCLs. (A) Schematic presentation of the TCR gene rearrangement during T-cell development. The arrows represent rearrangement on each chromosome mapped to different stages of a normal human thymocyte. The thin arrow for TCRβ symbolizes the 60-40 rule. Predicted time points of the transformation in various subtypes of TCLs are shown at the bottom. (B-C) WES and WGS samples (n = 367) were analyzed to identify the TCRα^{DNA}, TCRβ^{DNA}, and TCRγ^{DNA} clonotypes. (C) The frequency of the most abundant TCRβ^{DNA} clonotype was plotted vs the added frequency of the 2 most frequent TCRγ^{DNA} clonotypes or the 2 most abundant TCRα^{DNA} clonotypes for each sample. (D) A similar correlation plot of the frequency of dominant TCRβ^{RNA} clonotypes vs the dominant TCRα^{RNA} clonotypes (data from 196 WTS samples). DP, double positive; ISP, immature single positive; SP, single positive.

Biased V and J usage in TCLs

Previous studies have shown biased rearrangement of certain TCR V and J gene combinations,[46,47] but it is not clear whether those findings are generalizable to all TCLs. We calculated the frequencies with which V and J genes are represented in the most abundant clonotype of TCRβ (Figure 5) or the 2 most abundant clonotypes in TCRα and TCRγ (Figure 6). There was a strong preferential usage of certain Vβ and Vα segments across different TCLs. The most striking was Vβ20-1 (*TRBV20-1*), which was found in a large proportion of clonotypes in MF, SS, ATCLL, ALCL, and NKTL, at the level of both DNA and RNA. Vβ4-1 (*TRBV4-1*) was overrepresented in ALCL, NKTL, and SPTL. Other Vβ segments were more disease-specific, such as *TRBV18*, which was often found in MF and *TRBV24-1*, which was frequent in PTCL. Biased use of Vα segments was also obvious; for example, Vα16 (*TRAV16*), Vα35 (*TRAV35*), Vα41 (*TRAV41*), and Vα3 (*TRAV3*) were overrepresented in most TCLs. The Jβ and Jα segments also seemed to be unequally used (eg, *TRBJ2-1* and *TRBJ2-7* overexpression in all subtypes of TCLs) but because of a small number of J genes, the patterns are less obvious (Figures 5 and 6).

We have also analyzed V and J segment usage in the 10 most abundant TCRβ clonotypes, which represents the median number of malignant clonotypes determined in comparison with TCFs. The overall pattern was similar to that of the most abundant clonotypes, which confirms biased use of *TRBV20-1*, *TRBV4-1*, *TRBJ2-1*, and *TRBJ2-7*.

Previous studies in CTCL have indicated overrepresentation of *TRGV10-TRGJP1* in the malignant T cells.[48] We confirmed this in our analysis, and we also found that *TRGV2*, *TRGV8*, and *TRGV9* were frequently used across different TCLs, but *TRGV1*, *TRGV5*, *TRGV11* were relatively underrepresented (Figure 6C-D).

Shared malignant clonotypes represent public clonotypes and have high recombination probability

The results described above indicated that certain V-J combinations are overrepresented. This suggests that the clonotypes (as defined by the amino acid sequence of the highly variable CDR3 region) may also be shared between patients with different TCLs. Indeed, as shown in supplemental Tables 4 and 5, we have found numerous shared TCRα^{aa} and TCRβ^{aa}. Some clonotypes were shared across a large number of patients with different TCLs. For example, CARRKSSFF was detected in 100 samples (54 samples used *TRBV28-TRAJ1-1*) and CDNNNDMRF (*TRAV16-TRAJ43*) was present in 213 samples.

We estimated the fraction of the malignant clonotypes shared by 2 or more individuals using the Jaccard index (ranging from 0 [no overlap] to 1 [full overlap] of TCR repertoires). Both TCRα^{aa} and TCRβ^{aa} had a high degree of sharing (Jaccard indices of 0.4 and 0.34, respectively), predictably showing higher sharing for TCRα.

Shared malignant clonotypes resembled public clonotypes found in the blood of healthy controls. Public TCRs largely result from V-J
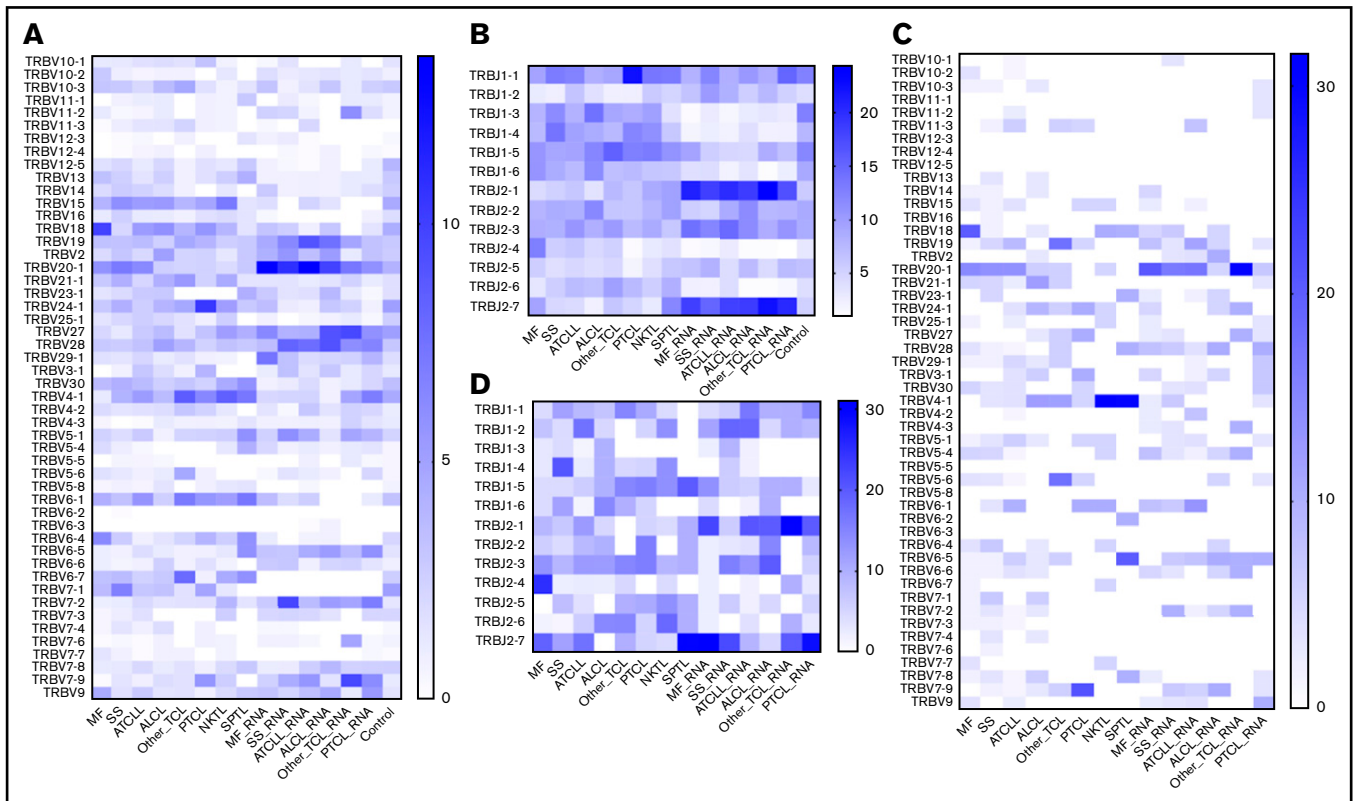
Figure 5. Vβ and Jβ (*TRBV/TRBJ*) usage in TCLs. The identified TCRβ[DNA] and TCRβ[RNA] clonotypes from WES, WGS, and WTS data for the TCL diagnoses and control blood lymphocytes to identify preferential *TRBV* and *TRBJ* segment usage. (A-B) *TRBV* and *TRBJ* frequencies, respectively, in the 10 most abundant clonotypes. (C-D) Plots similar to those in panels A and B with the inclusion of the single most abundant clonotype. The control group is absent in panels C and D because of the lack of a high-frequency clonotype in these samples.

recombination biases and convergent recombination in which rearrangement of different segments can result in the same amino acid sequence.[49] Public TCR chains can pair with a diverse repertoire of private TCR chains and therefore they are not determining antigen specificity.[49] To determine whether the detected shared malignant clonotypes might represent the result of biased, convergent V-J recombination, we calculated the probability of CDR3 sequence generation (P_{gen}) value[36,50] for the most abundant TCRβ and TCRα clonotypes. P_{gen} of those shared malignant clonotypes had 2 probability peaks at 10^{-9} and 10^{-6}, respectively (Figure 7A; supplemental Tables 4-7). Shared TCRβ[RNA] and TCRα[RNA] had even higher values with the maxima at 10^{-6} and 10^{-4}, respectively (Figure 7B), which were significantly higher than the P_{gen} of peripheral blood private (unique) clonotypes ($<10^{-9}$).

Discussion

TCLs represent a clinically diverse spectrum of T-cell cancers that vary from aggressive to relatively indolent and from nodal to organ specific (such as the skin). Nevertheless, analysis of TCR rearrangement patterns revealed striking similarities across different TCLs. Because TCR loci are rearranged sequentially and the rearrangement happens only during the thymic development of T-cell precursors, the rearrangement products retained in the genome allow us to trace the clonal history of TCL and illuminate the much-debated question of the cell of origin.

The current model of TCLs as neoplasms originating from various subpopulations of normal mature T cells[3] is not only unsupported but is directly contradicted by our findings. Malignant transformation of a single lymphocyte randomly drawn from a vast pool of 10^6 to 10^7 singular T-cell clonotypes should yield lymphomas that are monoclonal (all malignant cells share the same clonotype), unique (low probability of shared clonotypes between tumors from unrelated patients), and unconnected (random frequency distribution of clonotypes). We show here that TCLs are mostly oligoclonal, have a high degree of interindividual clonotype sharing across different disease entities, and represent connected systems of clonotypes, the frequency of which follows the Zipf-Mandelbrot power law. We propose that considering the T-cell lymphatic precursor as a cell of origin allows not only for a more accurate interpretation of our findings but also accounts for previous observations that could not be explained by a mature T-cell model.

Clonotypic diversity arises if a transformed lymphoid precursor passes through different stages of TCR rearrangements resulting in a clonotypically heterogeneous pool of early cancer cells, mimicking the differentiation of normal T cells. The model schematically shown in Figure 4A assumes that precursor cells for mature TCLs progress through the sequential rearrangements analogous to those of normal thymocytes and in a stark contrast to those of immature TCLs and leukemias (eg, T-ALL), which become arrested at a given stage of development. Thus, transformation before TCRγ rearrangement
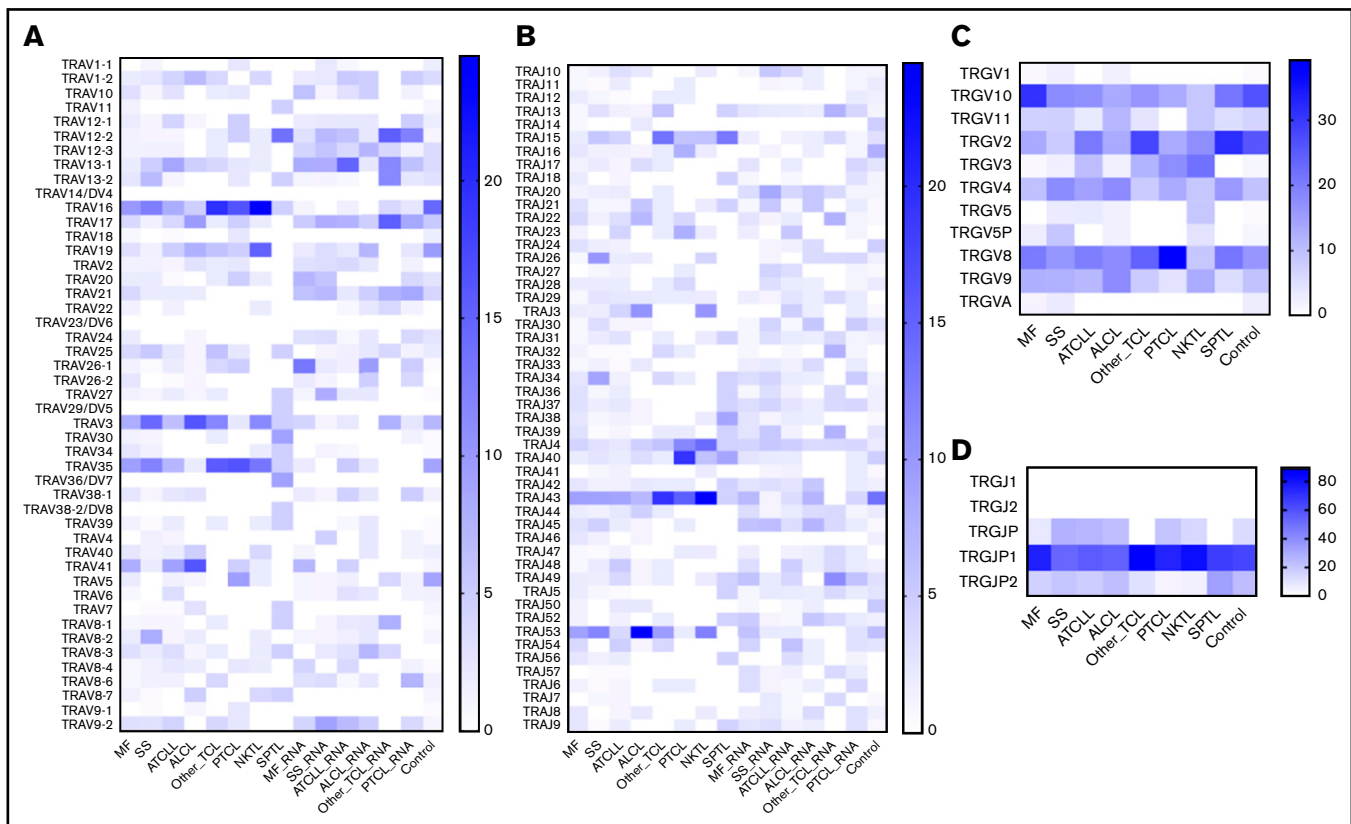
Figure 6. V and J (*TRAV*, *TRAJ*, *TRGV*, and *TRGJ*) usage in TCLs. The frequency of V and J segments of TCRα and TCRγ clonotypes are plotted as in Figure 7. The 2 most abundant TCRα^{DNA} and TCRγ^{DNA} clonotypes from TCLs and control samples were identified, and the frequency of V and J segments were plotted on the heat map. The RNA samples were used only for identification of the TCRα^{RNA} clonotypes. (A-B) TRAV and TRAJ segment usage; (C-D) *TRGV* and *TRGJ* segment usage.

would result in oligoclonality for all TCR rearrangements, whereas transformation after TCRγ rearrangement would render cancer cells monoclonal for TCRγ but oligoclonal for TCRβ and TCRα. Transformation after the TCRβ rearrangement stage would render the malignant cells monoclonal (single TCRβ^{RNA} clonotype and 1 to 2 TCRβ^{DNA} clonotypes) but oligoclonal for TCRα. Across different TCLs, ALCL showed a fingerprint for transformation at the earliest stage (before or at the TCRγ rearrangement), and ATCLL and PTCL seemed to derive from a cell in a later stage of development (TCRβ rearrangement or later). However, significant overlap in the patterns of clonality existed within each disease, indicating that the stage of the transformation does not fully define the phenotype of the disease. Because the available sequencing data did not have detailed clinical information, we cannot exclude that different subtypes within each diagnosis (eg, anaplastic lymphoma kinase [ALK^{+}], ALK^{-}, primary cutaneous ALCL, and breast implant–associated ALCL) differ with regard to the differentiation stage of the cell of origin as well.

The existence of more than 1 malignant clonotype has previously been shown in TCLs, but this finding has largely been ignored, and the clonotypes have been considered as artifacts caused by contamination of the sample with normal reactive T cells. We consider those explanations unlikely. First, in our analysis we include only the most abundant clonotypes of cumulative frequency equal to the proportion of tumor cells in the sample (TCFs). Second,

across a large number of samples, we demonstrated the power-law relationship between the frequency and the rank of the clonotype. Power-law distribution characterizes complex, evolving, out-of-equilibrium, multiplicative systems composed of interconnected components, such as cities in a country, words in a language, repertoire of protein domains, or migrating human populations.[43,45,51,52] It has been shown that growing malignant tumors demonstrate power-law distribution between tumor subclones.[44] The distribution of clonotypes in TCLs followed power-law distribution as well, which is understandable only if TCLs are considered as evolving systems of interacting and expanding T-cell clones identified by their clonotypic signature.

We considered alternative explanations of our findings because it could be argued that clonotypic oligoclonality might also be a result of secondary TCR rearrangements in the already established tumor. However, this is unlikely because the TCR rearrangement machinery is irreversibly inactivated after completion of TCRα rearrangement, and the essential enzymes RAG1 and RAG2 are not expressed in TCLs.[53] Moreover, additional supportive evidence for the lymphoid precursor model for TCLs comes from the pattern of V-J usage, which invalidates the argument of secondary mutations as the cause for multiple high-frequency malignant clonotypes. TCR rearrangements in human thymocytes are characterized by biased V and J segment usage, reflecting the probability of recombination (P_{gen}) that produces shared
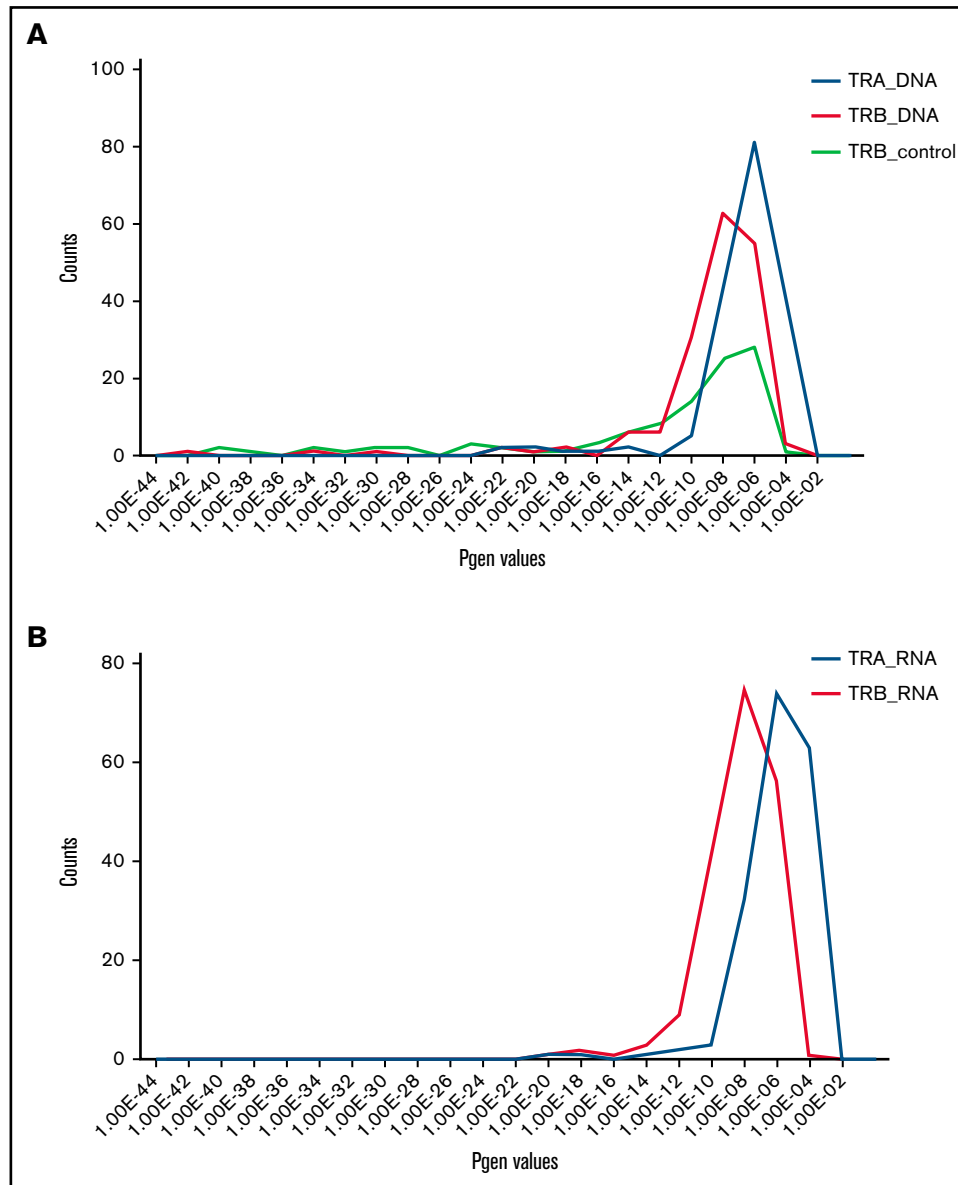
Figure 7. P_{gen} for TCRα and TCRβ clonotypes. Shared TCRα^{aa} and TCRβ^{aa} clonotypes were identified using the Immunarch pipeline and their P_{gen} was calculated using the OLGA algorithm (Figure 1) from DNA and RNA data. (A) P_{gen} histogram for the shared TCRα^{DNA} and TCRβ^{DNA} clonotypes. As a comparison, the P_{gen} values for the TCRβ^{DNA} clonotypes for the healthy control group are plotted. (B) Analogous histogram (as in panel A) showing P_{gen} for the shared clonotypes in TCRα^{RNA} and TCRβ^{RNA}.

(public) clonotypes.[54,55] A very similar situation is seen in TCLs. The V segments preferentially recombined in TCLs (*TRBV20-1*, *TRBV4-1*, *TRBV5-1*, *TRAV14*, *TRAV35*, or *TRAV41*) were also the most frequently used segments in normal human thymocytes, and the shared clonotypes had comparable high P_{gen} values ($>10^{-7}$ for TCRα and $>10^{-9}$ for TCRβ).[54,55] Differences in combinatorial probability is a probable explanation for a higher Jaccard index (degree of clonotype sharing) for TCRα more than for TCRβ.

It has been postulated that V-J bias in some TCLs is a result of selection pressure from common antigens.[56-60] For example, *TRBV20-1* has the highest frequency across all TCLs and is known to be involved in response to Epstein-Barr virus or cytomegalovirus.[46,61] However, none of the known CDR3 regions that recognize viral antigen was found among the malignant clonotypes. We postulate that the frequent finding of *TRBV20-1* in TCLs is not a result of its antigenic selection but simply its ubiquity. *TRBV20-1* is the prominent component of the TCRβ clonotypes of normal blood,[62,63] tumor-infiltrating T cells,[64,65] and autoimmune reactions.[66-69] *TRBV20-1* is the most frequently rearranged segment in double positive thymocytes[55] and is frequently found in T-ALL, which develops directly from lymphoid precursors.[15]

Although our data strongly indicate the role of the lymphoid precursor as the cell of origin in TCLs, the origin of those cells and their pathways of development cannot be inferred from this study.

Compelling experimental data suggest that the precursor T cell in the thymus might be the cell of origin for ALCL,[13,14] but we cannot exclude the possibility that extrathymic sites are involved as well. T-cell maturation and TCR recombination may happen in extrathymic sites such as bone marrow, lymph nodes, or the tonsils.[70-72] In a patient with lymphomatoid papulosis, we have found early precursor cells in the bone marrow before the clinical onset of the disease[73] and cases of transmission of TCL by bone marrow transplant from asymptomatic donors who later developed the disease have been observed,[74] which indicates that bone marrow could be a potential reservoir of progenitor malignant cells. Single-cell sequencing experiments in which TCRα can be paired with TCRβ in the context of the transcriptomic profile of the cells would probably allow further elucidation of the origin of lymphoma precursors.

In summary, we propose that mature TCLs do not originate from T cells in the periphery but rather from lymphoid precursors at different stages of their differentiation. Better understanding of the developmental trajectories of the early precursor into clinical lymphoid malignancy would improve diagnosing and staging of TCLs and help with designing targeted therapies that intervene with the disease at the initial stages.

Acknowledgments

Authorship

Contribution: A.I. collected, analyzed, and assembled data, identified TCL studies, and wrote the manuscript; D.H. analyzed the data to identify TCFs and the TCR clonotypes; and R.G. supervised the research, advised on data analysis and assembly, and wrote the manuscript.

Conflict-of-interest disclosure: R.G. received speaker and advisory board honoraria from Mallinckrodt, Eli Lilly, Sanofi, AbbVie, Sun Pharma, and Kyowa Kirin outside this study. The remaining authors declare no competing financial interests.

ORCID profiles: A.I., 0000-0001-7370-2378; D.H., 0000-0001-7983-8524; R.G., 0000-0002-2310-8300.

Correspondence: Aishwarya Iyer, University of Alberta, 260 HMRC, 114th St and 85th Ave, Edmonton, AB T6G 2R3, Canada; e-mail: aiyer2@ualberta.ca.

References

1. Jiang M, Bennani NN, Feldman AL. Lymphoma classification update: T-cell lymphomas, Hodgkin lymphomas, and histiocytic/dendritic cell neoplasms. *Expert Rev Hematol.* 2017;10(3):239-249.

2. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood.* 2016; 127(20):2375-2390.

3. Fiore D, Cappelli LV, Broccoli A, Zinzani PL, Chan WC, Inghirami G. Peripheral T cell lymphomas: from the bench to the clinic. *Nat Rev Cancer.* 2020;20(6):323-342.

4. Marchi E, O'Connor OA. The rapidly changing landscape in mature T-cell lymphoma (MTCL) biology and management. *CA Cancer J Clin.* 2020; 70(1):47-70.

5. Timmins MA, Wagner SD, Ahearne MJ. The new biology of PTCL-NOS and AITL: current status and future clinical impact. *Br J Haematol.* 2020; 189(1):54-66.

6. Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. *Nature.* 1988;334(6181):395-402.

7. Warren RL, Freeman JD, Zeng T, et al. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res.* 2011;21(5):790-797.

8. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A direct estimate of the human alphabeta T cell receptor diversity. *Science.* 1999;286(5441):958-961.

9. Qi Q, Liu Y, Cheng Y, et al. Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci USA.* 2014;111(36):13139-13144.

10. Langerak AW, Groenen PJ, Brüggemann M, et al. EuroClonality/BIOMED-2 guidelines for interpretation and reporting of Ig/TCR clonality testing in suspected lymphoproliferations. *Leukemia.* 2012;26(10):2159-2171.

11. Iyer A, Hennessey D, O'Keefe S, et al. Clonotypic heterogeneity in cutaneous T-cell lymphoma (mycosis fungoides) revealed by comprehensive whole-exome sequencing. *Blood Adv.* 2019;3(7):1175-1184.

12. Iyer A, Hennessey D, O'Keefe S, et al. Skin colonization by circulating neoplastic clones in cutaneous T-cell lymphoma. *Blood.* 2019;134(18):1517-1527.

13. Moti N, Malcolm T, Hamoudi R, et al. Anaplastic large cell lymphoma-propagating cells are detectable by side population analysis and possess an expression profile reflective of a primitive origin. *Oncogene.* 2015;34(14):1843-1852.

14. Congras A, Hoareau-Aveilla C, Caillet N, et al. ALK-transformed mature T lymphocytes restore early thymus progenitor features. *J Clin Invest.* 2020;130(12):6395-6408.

15. Brüggemann M, van der Velden VH, Raff T, et al. Rearranged T-cell receptor beta genes represent powerful targets for quantification of minimal residual disease in childhood and adult T-cell acute lymphoblastic leukemia. *Leukemia.* 2004;18(4):709-719.

16. Asnafi V, Beldjord K, Boulanger E, et al. Analysis of TCR, pT alpha, and RAG-1 in T-acute lymphoblastic leukemias improves understanding of early human T-lymphoid lineage commitment. *Blood.* 2003;101(7):2693-2703.

17. Iyer A, Hennessey D, O'Keefe S, et al. Branched evolution and genomic intratumor heterogeneity in the pathogenesis of cutaneous T-cell lymphoma. *Blood Adv.* 2020;4(11):2489-2500.

18. Choi J, Goh G, Walradt T, et al. Genomic landscape of cutaneous T cell lymphoma. *Nat Genet.* 2015;47(9):1011-1019.

19. Ungewickell A, Bhaduri A, Rios E, et al. Genomic analysis of mycosis fungoides and Sézary syndrome identifies recurrent alterations in TNFR2. *Nat Genet.* 2015;47(9):1056-1060.

20. McGirt LY, Jia P, Baerenwald DA, et al. Whole-genome sequencing reveals oncogenic mutations in mycosis fungoides. *Blood.* 2015;126(4): 508-519.

21. da Silva Almeida AC, Abate F, Khiabanian H, et al. The mutational landscape of cutaneous T cell lymphoma and Sézary syndrome. *Nat Genet.* 2015;47(12):1465-1470.

22. Li Z, Lu L, Zhou Z, et al. Recurrent mutations in epigenetic modifiers and the PI3K/AKT/mTOR pathway in subcutaneous panniculitis-like T-cell lymphoma. *Br J Haematol.* 2018;181(3):406-410.

23. Simpson HM, Khan RZ, Song C, et al. Concurrent mutations in ATM and genes associated with common γ chain signaling in peripheral T cell lymphoma. *PLoS One.* 2015;10(11):e0141906.

24. Palomero T, Couronné L, Khiabanian H, et al. Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nat Genet.* 2014;46(2):166-170.

25. Yoo HY, Sung MK, Lee SH, et al. A recurrent inactivating mutation in RHOA GTPase in angioimmunoblastic T cell lymphoma. *Nat Genet.* 2014; 46(4):371-375.

26. Jiang L, Gu ZH, Yan ZX, et al. Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. *Nat Genet.* 2015;47(9): 1061-1066.

27. Kataoka K, Nagata Y, Kitanaka A, et al. Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat Genet.* 2015;47(11):1304-1315.

28. Wang L, Ni X, Covington KR, et al. Genomic profiling of Sézary syndrome identifies alterations of key T cell signaling and differentiation genes. *Nat Genet.* 2015;47(12):1426-1434.

29. Crescenzo R, Abate F, Lasorsa E, et al; European T-Cell Lymphoma Study Group, T-Cell Project: Prospective Collection of Data in Patients with Peripheral T-Cell Lymphoma and the AIRC 5xMille Consortium "Genetics-Driven Targeted Management of Lymphoid Malignancies". Convergent mutations and kinase fusions lead to oncogenic STAT3 activation in anaplastic large cell lymphoma. *Cancer Cell.* 2015;27(4):516-532.

30. Wang Y, McKay JD, Rafnar T, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet.* 2014;46(7):736-741.

31. Ha G, Roth A, Khattra J, et al. TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res.* 2014;24(11):1881-1893.

32. Adalsteinsson VA, Ha G, Freeman SS, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun.* 2017;8(1):1324.

33. Broad Institute, Genome Analysis Toolkit (GATK). Panel of Normals (PON). https://gatk.broadinstitute.org/hc/en-us/articles/360035890631-Panel-of-Normals-PON-. Accessed 25 June 2021.

34. Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods.* 2015;12(5):380-381.

35. Nazarov V, Rumynskiy E. immunomind/immunarch: 0.6.5: Basic single-cell support. Zenodo. .

36. Sethna Z, Elhanati Y, Callan CG, Walczak AM, Mora T. OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics.* 2019;35(17):2974-2981.

37. Bonnet M, Ferrier P, Spicuglia S. Molecular genetics at the T-cell receptor beta locus: insights into the regulation of V(D)J recombination. *Adv Exp Med Biol.* 2009;650:116-132.

38. Mostoslavsky R, Alt FW, Rajewsky K. The lingering enigma of the allelic exclusion mechanism. *Cell.* 2004;118(5):539-544.

39. Hosoya T, Li H, Ku CJ, Wu Q, Guan Y, Engel JD. High-throughput single-cell sequencing of both *TCR-β* alleles. *J Immunol.* 2018;201(11): 3465-3470.

40. Wu GS, Bassing CH. Inefficient V(D)J recombination underlies monogenic T cell receptor β expression. *Proc Natl Acad Sci U S A.* 2020;117(31): 18172-18174.

41. Kirsch IR, Watanabe R, O'Malley JT, et al. TCR sequencing facilitates diagnosis and identifies mature T cells as the cell of origin in CTCL. *Sci Transl Med.* 2015;7(308):308ra158.

42. de Masson A, O'Malley JT, Elco CP, et al; Masson A de et al. High-throughput sequencing of the T cell receptor β gene identifies aggressive early-stage mycosis fungoides. *Sci Transl Med.* 2018;10(440):eaar5894.

43. Corominas-Murtra B, Seoane LF, Solé R. Zipf's Law, unbounded complexity and open-ended evolution. *J R Soc Interface.* 2018;15(149): 20180395.

44. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. Identification of neutral tumor evolution across cancer types. *Nat Genet.* 2016;48(3): 238-244.

45. De Marzo G, Gabrielli A, Zaccaria A, Pietronero L. Dynamical approach to Zipf's law. *Phys Rev Res.* 2021;3(1):013084.

46. Zhang W, Hawkins PG, He J, et al. A framework for highly multiplexed dextramer mapping and prediction of T cell receptor sequences to antigen specificity. *Sci Adv.* 2021;7(20):eabf5835.

47. Turner SJ, Doherty PC, McCluskey J, Rossjohn J. Structural determinants of T-cell receptor bias in immunity. *Nat Rev Immunol.* 2006;6(12):883-894.

48. Hara J, Benedict SH, Yumura K, Ha-Kawa K, Gelfand EW. Rearrangement of variable region T cell receptor gamma genes in acute lymphoblastic leukemia. V gamma gene usage differs in mature and immature T cells. *J Clin Invest.* 1989;83(4):1277-1283.

49. Venturi V, Kedzierska K, Price DA, et al. Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proc Natl Acad Sci USA.* 2006;103(49):18691-18696.

50. Marcou Q, Mora T, Walczak AM. High-throughput immune repertoire analysis with IGoR. *Nat Commun.* 2018;9(1):561.

51. Schläpfer M, Dong L, O'Keeffe K, et al. The universal visitation law of human mobility. *Nature.* 2021;593(7860):522-527.

52. Gabaix X. Zipf's Law for cities: an explanation. *Q J Econ.* 1999;114(3):739-767.

53. Yoneda N, Tatsumi E, Kawano S, Matsuo Y, Minowada J, Yamaguchi N. Human recombination activating gene-1 in leukemia/lymphoma cells: expression depends on stage of lymphoid differentiation defined by phenotype and genotype. *Blood.* 1993;82(1):207-216.

54. Heikkilä N, Vanhanen R, Yohannes DA, et al. Human thymic T cell repertoire is imprinted with strong convergence to shared sequences. *Mol Immunol.* 2020;127:112-123.

55. Park J-E, Botting RA, Domínguez Conde C, et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science.* 2020;367(6480):eaay3224.

56. Clemente MJ, Przychodzen B, Jerez A, et al. Deep sequencing of the T-cell receptor repertoire in CD8[+] T-large granular lymphocyte leukemia identifies signature landscapes. *Blood.* 2013;122(25):4077-4085.

57. Malcolm TIM, Hodson DJ, Macintyre EA, Turner SD. Challenging perspectives on the cellular origins of lymphoma. *Open Biol.* 2016;6(9):160232.

58. Litvinov IV, Shtreis A, Kobayashi K, et al. Investigating potential exogenous tumor initiating and promoting factors for Cutaneous T-Cell Lymphomas (CTCL), a rare skin malignancy. *OncoImmunology.* 2016;5(7):e1175799.

59. Tokura Y, Heald PW, Yan SL, Edelson RL. Stimulation of cutaneous T-cell lymphoma cells with superantigenic staphylococcal toxins. *J Invest Dermatol.* 1992;98(1):33-37.

60. Morgan SM, Hodges E, Mitchell TJ, Harris S, Whittaker SJ, Smith JL. Molecular analysis of T-cell receptor β genes in cutaneous T-cell lymphoma reveals Jbeta1 bias. *J Invest Dermatol.* 2006;126(8):1893-1899.

61. Miles JJ, Bulek AM, Cole DK, et al. Genetic and structural basis for selection of a ubiquitous T cell receptor deployed in Epstein-Barr virus infection. *PLoS Pathog.* 2010;6(11):e1001198.

62. Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.* 2009;19(10):1817-1824.

63. Hou X, Chen W, Zhang X, et al. Preselection TCR repertoire predicts CD4[+] and CD8[+] T-cell differentiation state. *Immunology.* 2020;161(4):354-363.

64. Li B, Li T, Pignon JC, et al. Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat Genet.* 2016;48(7):725-732.

65. Omland SH, Hamrouni A, Gniadecki R. High diversity of the T-cell receptor repertoire of tumor-infiltrating lymphocytes in basal cell carcinoma. *Exp Dermatol.* 2017;26(5):454-456.

66. Petersen J, Montserrat V, Mujico JR, et al. T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with celiac disease. *Nat Struct Mol Biol.* 2014;21(5):480-488.

67. Luo G, Ambati A, Lin L, et al. Autoimmunity to hypocretin and molecular mimicry to flu in type 1 narcolepsy. *Proc Natl Acad Sci USA.* 2018;115(52):E12323-E12332.

68. Rosati E, Pogorelyy MV, Dowds CM, et al. Identification of disease-associated traits and clonotypes in the T cell receptor repertoire of monozygotic twins affected by inflammatory bowel diseases. *J Crohn's Colitis.* 2020;14(6):778-790.

69. Yohannes DA, Freitag TL, de Kauwe A, et al. Deep sequencing of blood and gut T-cell receptor β-chains reveals gluten-induced immune signatures in celiac disease. *Sci Rep.* 2017;7(1):17977.

70. McClory S, Hughes T, Freud AG, et al. Evidence for a stepwise program of extrathymic T cell development within the human tonsil. *J Clin Invest.* 2012;122(4):1403-1415.

71. García-Ojeda ME, Dejbakhsh-Jones S, Chatterjea-Matthes D, et al. Stepwise development of committed progenitors in the bone marrow that generate functional T cells in the absence of the thymus. *J Immunol.* 2005;175(7):4363-4373.

72. Holland AM, Zakrzewski JL, Tsai JJ, et al. Extrathymic development of murine T cells after bone marrow transplantation. *J Clin Invest.* 2012;122(12):4716-4726.

73. Gniadecki R, Lukowsky A, Rossen K, Madsen HO, Thomsen K, Wulf HC. Bone marrow precursor of extranodal T-cell lymphoma. *Blood.* 2003;102(10):3797-3799.

74. Berg KD, Brinster NK, Huhn KM, et al. Transmission of a T-cell lymphoma by allogeneic bone marrow transplantation. *N Engl J Med.* 2001;345(20):1458-1463.