

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

Aishwarya IYER¹, Dylan HENNESSEY¹ and Robert GNIADOCKI^{1,2,3}

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

Correspondence: Dr. A. Iyer, aiyer2@ualberta.ca

Abstract

Background

Mature T-cell lymphomas (TCLs) are rare, clinically heterogeneous hematologic cancers of high medical need. TCLs have inferior prognosis compared with their B-cell counterparts, which is attributed to poor understanding of their pathogenesis. Based on 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.

Methods and findings

To address the debated question of the cell of origin in TCLs we analyzed to identify the highly variable complementarity determining regions (CDR3) regions of T-cell receptor (TCR) to trace the clonal history of the T-cells. We have collected previously published whole genome -exome, and -transcriptome sequencing data from 574 TCL patients comprising five nodal lymphomas [anaplastic large cell lymphoma (n=67), peripheral T-cell lymphoma (PTCL, n=55), adult T-cell lymphoma/leukemia (n=135), natural killer T-cell lymphoma (NKCL, n=25), not specified/other (n=30)] and three extranodal, cutaneous T-cell lymphomas [mycosis fungoides (n=122), Sezary syndrome (n=130), and subcutaneous panniculitis-like T-cell lymphoma (n=10)]. TCR clonotypes contained in the tumor cell fraction, representing the clonotypes of malignant cells, were identified by de novo assembly of CDR3 regions of TCR γ , β and α . We have found that the vast majority of TCLs are clonotypically oligoclonal, although the pattern oligoclonality varied. Anaplastic large cell lymphoma was most diverse comprising multiple clonotypes of TCR γ , β and α whereas adult T-cell lymphoma/leukemia and peripheral T-cell lymphomas often showed monoclonality for TCR γ and β but had diverse TCR α clonotypes. These patterns of rearrangements were not compatible with the current mature T-cell precursor model and 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 showing biased usage of V and J segments of high combinatorial probability resulting in recurrent, “public” CDR3 sequences shared between unrelated patients and across different clinical TCL entities. Frequencies of malignant clonotypes followed Zipf-Mandelbrot scaling law suggesting that TCLs comprise an interconnected system of expanding tumor clones.

The major limitation of this study is that it is based on the analysis of the TCR clonotypes and does not directly inform about developmental trajectories of cellular clones.

Conclusions

Lymphoid precursors are the likely cells of origin for mature T-cell lymphomas. Anaplastic large cell lymphoma seems to be derived from the most immature precursors with germline TCR whereas peripheral T-cell lymphoma and adult T-cell lymphoma/leukemia map to the later stages after TCR β rearrangement stage. Clonotypically diverse initiating cells may seed target tissues being responsible for disease relapses after therapy.

Introduction

T-cell lymphomas (TCLs) is a heterogenous group of 29 malignancies comprising approximately 10% of non-Hodgkin's lymphomas and the incidence of 6000 cases per million [1,2]. Most TCLs are classified as mature T-cell lymphomas because malignant cells phenotypically resemble mature T-cells, harbor rearranged T-cell receptor (TCR) V(D)J genes and many express TCR $\alpha\beta$ heterodimer [3,4]. Striking similarities between subsets of normal T-cells and the malignant cells in TCLs (e.g. to T-regulatory cells in the adult T-cell lymphoma/leukemia or central memory T-cells to Sezary syndrome) led to the 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 TCR γ , TCR β , and TCR α loci are sequentially rearranged during different stages of intrathymic maturation of T cells from diverse pool of variable (V), diversity (D), and joining (J) gene. The TCR β locus contains 47 V (*TRBV*), 2 D (*TRBD*) and 13 J (*TRBJ*) segments whereas the TCR α locus comprises 42 *TRAV* and 61 *TRAJ* segments, which recombine yielding unique DNA sequences that are retained in genomes of all daughter cells [6]. Analysis of the complementarity determining region 3 (CDR3), the most diverse fragment of TCR chain coded by the V-(D)-J joining region and involved in antigen recognition, has long been used in the molecular diagnosis of T-cell lymphomas and leukemias. The repertoire of normal human TCR is in the range of 10^6 - 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 (i.e. malignant clonotypes) [10].

Clonotypic analyzes of T-cell lymphomas 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 TCR α and TCR β clonotypes would be inherited from the cell of origin. Transformation during earlier stages of T-cell development, before the completion of the final TCR α rearrangement, would produce oligoclonal or polyclonal patterns of clonotypes, characteristic for the stage in which the transformation happened. Indeed, our previous analyses of malignant T-cell clonotypes in cutaneous T-cell lymphoma revealed an oligoclonal pattern of TCR β and TCR α , but monoclonality of TCR γ suggesting that original malignant transformation occurred at the stage of a double-negative thymocyte after rearrangement of the

TCR γ locus [11,12]. A similar pathogenic scenario involving early thymocytes has been proposed for the anaplastic large cell lymphoma [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 also puzzled by the finding that certain malignant clonotypes were repeatedly found in samples from unrelated patients whereas due to the vastness of TCR repertoire we had expected that malignant clonotypes would have been unique for each patient. We have noted a preferential usage of certain V β and V α segments in cutaneous T-cell lymphomas [11] and were intrigued by the finding that similar biases in V segment usage was reported in T-cell acute lymphoblastic leukemia (T-ALL) which is derived from lymphoid precursors arrested at various stages of development [15,16].

In this study we employed the previously optimized bioinformatic pipeline for de novo assembly of CDR3 regions using paired-end RNA sequencing, whole genome- and whole exome DNA sequencing [11] to analyze malignant TCR clonotypes in 574 samples comprising 8 major subtypes of TCL. We propose that TCR rearrangement pattern across different clinical TCL entities is incompatible with the mature T-cell transformation hypothesis and supports the immature lymphoid progenitor as the cell of origin.

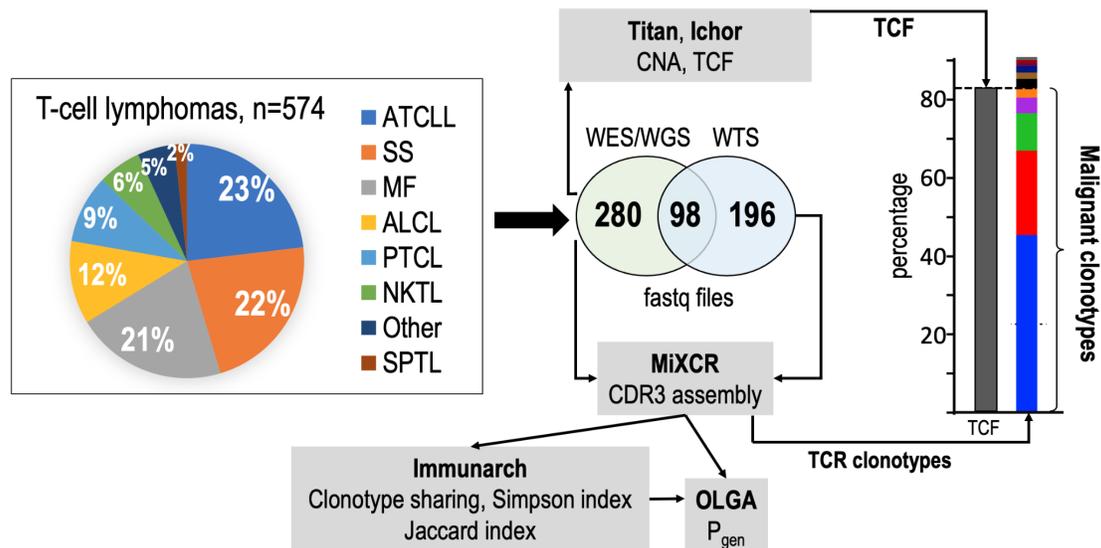


Figure 1. Summary of the methodological approach to TCR analysis in TCLs. Pie-chart representing the percentage of sequencing samples collected from different subtypes of TCLs were analyzed for tumor cell fraction (TCF) and the TCR clonotypes. The diagnostic groupings are detailed in supplementary Table S2. TCF was estimated based on the chromosomal aberration identified using TitanCNA or IchorCNA. TCR clonotypes were identified using MiXCR and the diversity index (inverse Simpson), clonotype sharing (Jaccard index), and probability of clonotype generation (P_{gen}) were analyzed using Immunarch and OLGA pipelines, as detailed in the Methods. WES, whole exome sequencing; WTS, whole transcriptome sequencing; WGS, whole genome sequencing; ATCLL, adult T-cell lymphoma/leukemia; SS, Sezary syndrome; MF, mycosis fungoides; ALCL, anaplastic large cell lymphoma; PTCL, peripheral T-cell lymphoma; NKTL, natural killer/T-cell lymphoma; SPTL, subcutaneous panniculitic T-cell lymphoma.

Methods

TCL data collection and TCF analysis

The graphical summary of the methods is shown in **Fig 1**. Raw fastq files of whole exome, whole genome and whole transcriptome sequencing and whole transcriptome sequencing were collected from 13 different studies (574 patients)[17–29]. To access the human genome data from restricted databases such as dbGaP and EGA was provided by Health Research Ethics Board of Alberta Cancer Committee under ethics ID- HREBA.CC-20-0118. Additionally 50 healthy controls samples were also obtained[30]. The accession number of the studies are listed in **supplementary Table S1** and the diagnostic grouping in the **supplementary Table S2**. Tumor cell fraction (TCF) was determined based on copy number aberrations (CNA) 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 (**supplementary Table S3**). Dataset of Crescenzo et al. (SRP044708) lacked control DNA and therefore, a panel of normal as recommended in the GATK pipeline [33] was created to analyze the TCF for samples from that study. Majority of transcriptome data was not obtained from the same samples as their exome counterparts and therefore, TCF data calculated using the DNA sequencing was not applicable to the TCR identified from the RNA sequencing.

TCR clonotype identification and quantification

TCR clonotypes were identified and quantified using MiXCR version 3.0.3 or 3.1.0 [34]. We used the superscript to indicate whether the clonotype was identified by the DNA, RNA, or amino acid (aa) sequence (e.g. TCR α^{DNA}). Default recommendations for analysis of exome and transcriptome data were used with the exception of eliminating assembly of short reads to identify additional TCR clonotypes. Short and long read alignments were included for whole-transcriptome sequencing (WTS); however, for WES data, partial reads were filtered out because they might be the captures of only V or J sequences. After identification of clonotypes further elimination of clonotypes based on mutational positions were performed and best predictive VJ combinations were included in further analyses. For the cases where TCF was available, the malignant clonotypes (i.e. clonotypes of the malignant T-cells) were determined by including the most abundant clonotypes to match TCF of the sample (**Fig 1**). Shared clonotype identification was based on TCR β^{aa} and was used to calculate diversity (inverse Simpson index) and clonotype sharing (Jaccard index) with Immunarch version 0.6.6 [35]. Probability of generation of CDR3 (P_{gen}) was calculated using OLGA version 1.2.3[36] for the TCL samples and the 50 healthy controls. Immunarch, Graphpad Prism (version 9) and Excel (version 16.16.27) were used for data representation.

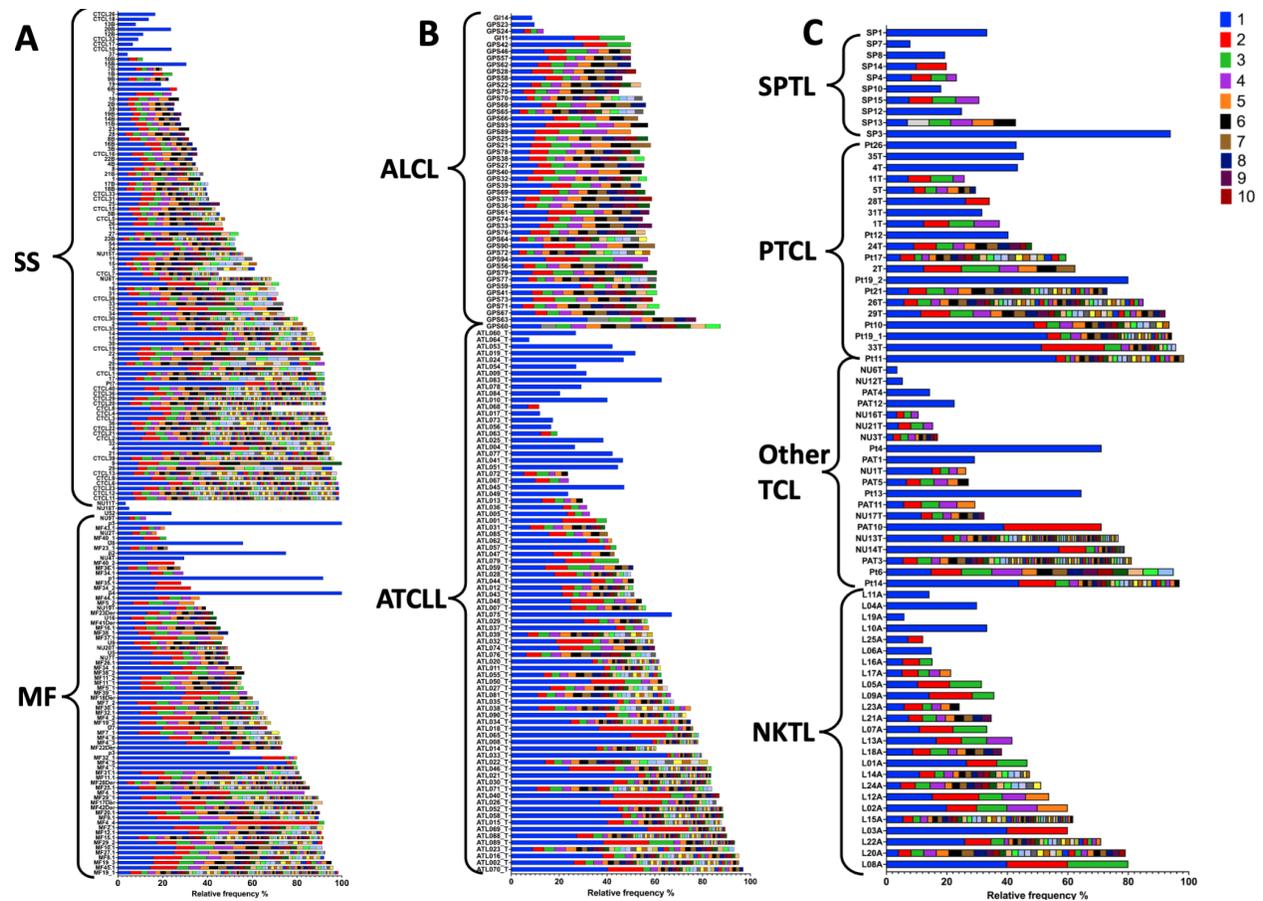


Figure 2. Frequency of TCR β ^{DNA} clonotypes in TCLs. Total of 378 samples from 8 subtypes of T-cell lymphomas (TCLs) were analyzed for identifying tumor cell fraction (TCF) and TCR β ^{DNA} clonotypes using the whole exome/genome sequencing data. TCR β ^{DNA} clonotypes corresponding to the TCF are indicated by the coloured bars in the descending order of relative frequencies (the most abundant, rank 1, is the dark blue bar). The data are split by diagnosis: **A** - cutaneous lymphomas, MF and SS; **B** - ATCLL and ALCL, **C** - NKTL, PTCL, SPTL and Other. Abbreviations of disease entities as in Fig 1.

Results

Clonotypic TCR β diversity in T-cell lymphomas

We analyzed data from 574 samples comprising the following TCL diagnoses: mycosis fungoides (MF, a primary cutaneous T-cell lymphoma, n=122), Sezary syndrome (SS, leukemic form of primary cutaneous T-cell lymphoma, n=130), anaplastic large cell lymphoma (ALCL, n=67), adult T-cell lymphoma/leukemia (ATCLL, n=135), natural killer T-cell lymphoma (NKTL, n=25), peripheral T-cell lymphoma (PTCL, n=55), subcutaneous panniculitis-like T-cell lymphoma (SPCL, n=10). We were unable to classify 30 samples into one of the above categories, which were grouped as “other TCLs” (supplementary Table S2).

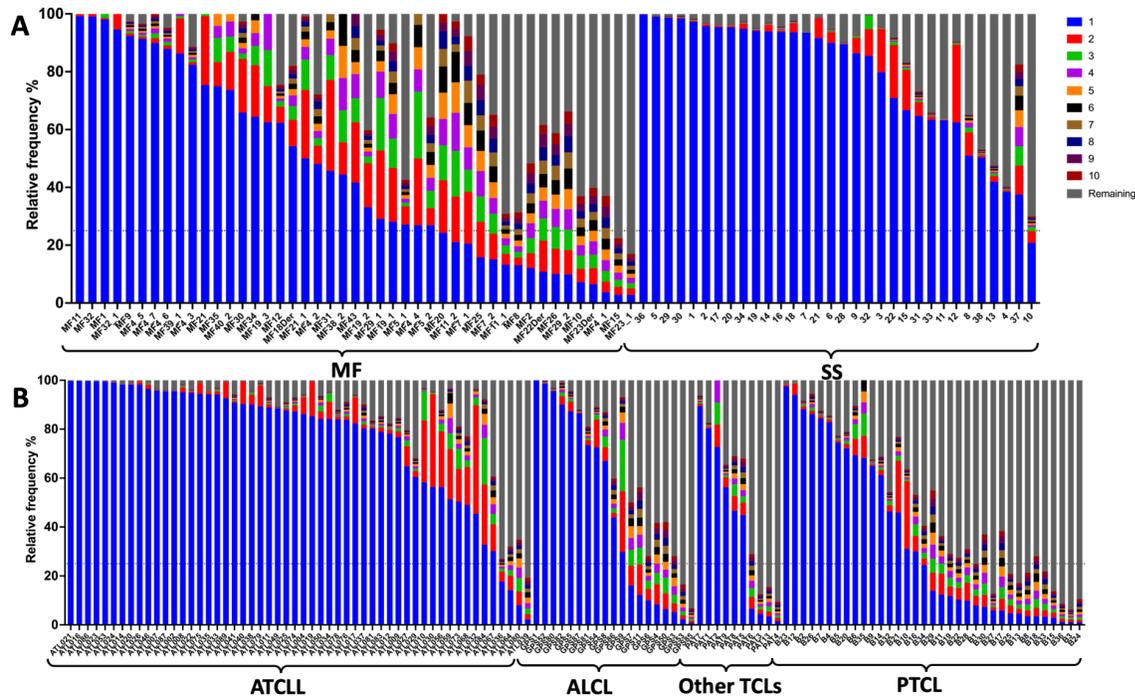


Figure 3. Frequency of TCR β ^{RNA} clonotypes in TCLs. Total of 196 RNA samples from 6 subtypes of TCLs were analyzed to identify the frequency of the TCR β ^{RNA} clonotypes. Ten most abundant clonotypes are indicated in colour using the same coding as in Fig 2; the remaining clonotypes are merged and their combined frequency indicated by the gray bar. The data are split by diagnosis: **A** - cutaneous lymphomas, MF and SS; **B** - ATCLL, ALCL, PTCL and Other. Abbreviations of disease entities as in Fig 1.

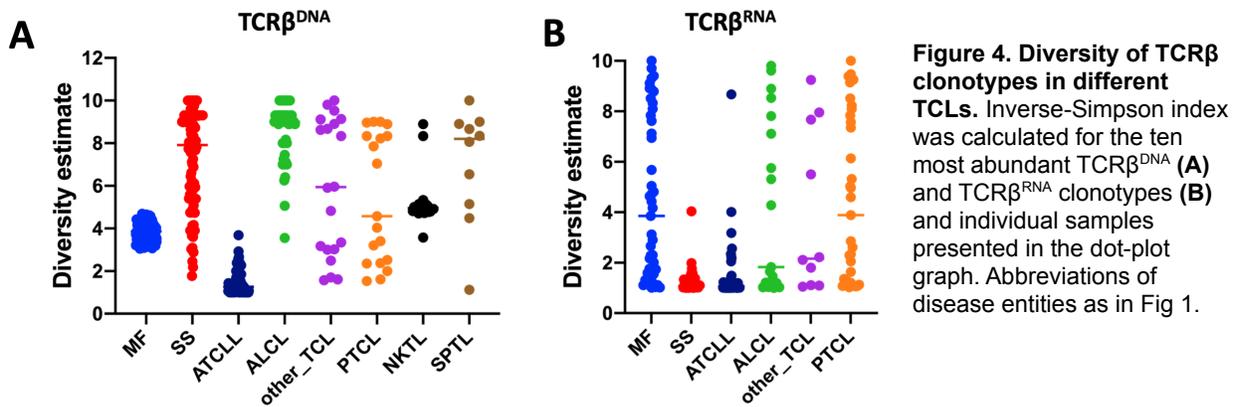


Figure 4. Diversity of TCR β clonotypes in different TCLs. Inverse-Simpson index was calculated for the ten most abundant TCR β ^{DNA} (**A**) and TCR β ^{RNA} clonotypes (**B**) and individual samples presented in the dot-plot graph. Abbreviations of disease entities as in Fig 1.

To determine the sequences of CDR3, we used the previously described bioinformatic analysis of bulk sequencing data from whole exome sequencing (WES), whole genome sequencing (WGS), and whole transcriptome sequencing (WTS) [11]. To differentiate between malignant clonotypes (i.e. 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 tumor cell fraction (TCF) in the sample (**Fig 1**) [12].

Among 378 samples for which genomic data (WES/WGS) was available to calculate TCF, the majority (75%) showed more than one TCR β^{DNA} clonotype attributed to malignant T-cells (matching TCF) (**Fig 2**). We considered three rearrangements as evidence of oligoclonality because although approximately 60% of T-cells rearrange TCR β on only one chromosome, the remaining 40% are bi-allelic rearrangements (the 60/40 rule) [37,38]. Oligoclonality of malignant clonotypes was most prevalent in ALCL (43/47) and least common in NKTL (12/25) (**Fig 2**).

The allelic exclusion of TCR β locus assures that a mature T-cell only expresses one clonotype [37], even in the presence of biallelic rearrangements [39,40]. Thus, the single TCR β clonotype at the level of RNA (TCR β^{RNA}) unequivocally identifies the T-cell clone. Since TCF could not be calculated for WTS samples because the matching WES/WGS data were not available, we analyzed TCR β 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 because the majority (79% of TCL samples) had TCF of 25% or higher (**supplementary Table S3**). In 26% of samples the most abundant TCR β^{RNA} clonotype had a frequency below 25%, i.e. the criterion of monoclonality was not met (**Fig 3**). In 6 samples the relative frequency of the second most frequent TCR β^{RNA} clonotype was >25% and in 38 cases the frequencies of the first and the second clonotypes were comparable (no single dominant clonotype), indicating oligoclonality of malignant clonotypes.

To further assess the diversity of the TCR β clonotypes using the inverse Simpson index which varies from 1 (one clonotype) to the maximum value representing the number of clonotypes (in our analysis n=10), taking into account clonotype concentration (**Fig 4**). For the TCR β^{DNA} clonotypes, the diversity was highest in SS and ALCL and lowest in ATCLL. The diversity of TCR β^{RNA} clonotypes had a slightly different pattern with a decrease in diversity for SS and ALCL when compared to TCR β^{DNA} clonotypes. This can be explained by the increase in the relative contribution of the most abundant TCR β^{RNA} clonotype (higher clonotype concentration) relative to TCR β^{DNA} which, in turn, is likely to be caused by a higher proportion of non-productively rearranged TCR β loci. The mean diversity for MF, ATCLL, and PTCL remain the same for TCR β^{DNA} and TCR β^{RNA} .

Patterns of TCR γ , - β and - α corroborate a branched evolution model of TCLs

The analyses described above suggested that TCR monoclonality is not a constant feature of TCLs and approximately 1/3 of all samples display an oligoclonal pattern of malignant TCR β , which is incompatible with the hypothesis of the mature T-cell origin hypothesis. To explore this further, we analyzed the frequencies of TCR γ and - α clonotypes. Unlike TCR β , TCR- γ and - α loci are not subjected to allelic exclusion and are usually rearranged at both chromosomes (**Fig 5 A**). Moreover, TCR γ is rearranged before TCR β so the cells with the same TCR γ rearrangement may harbor several TCR β rearrangements. The same is true for TCR α ; a single TCR β clone may comprise several (usually 2-4) TCR α clonotypes because of biallelic rearrangement and secondary rearrangements during expansion of the TCR β clone (**Fig 5 A**). It is thus possible to deduce the stage of T-cell transformation from the relative frequencies of TCR γ , β and α .

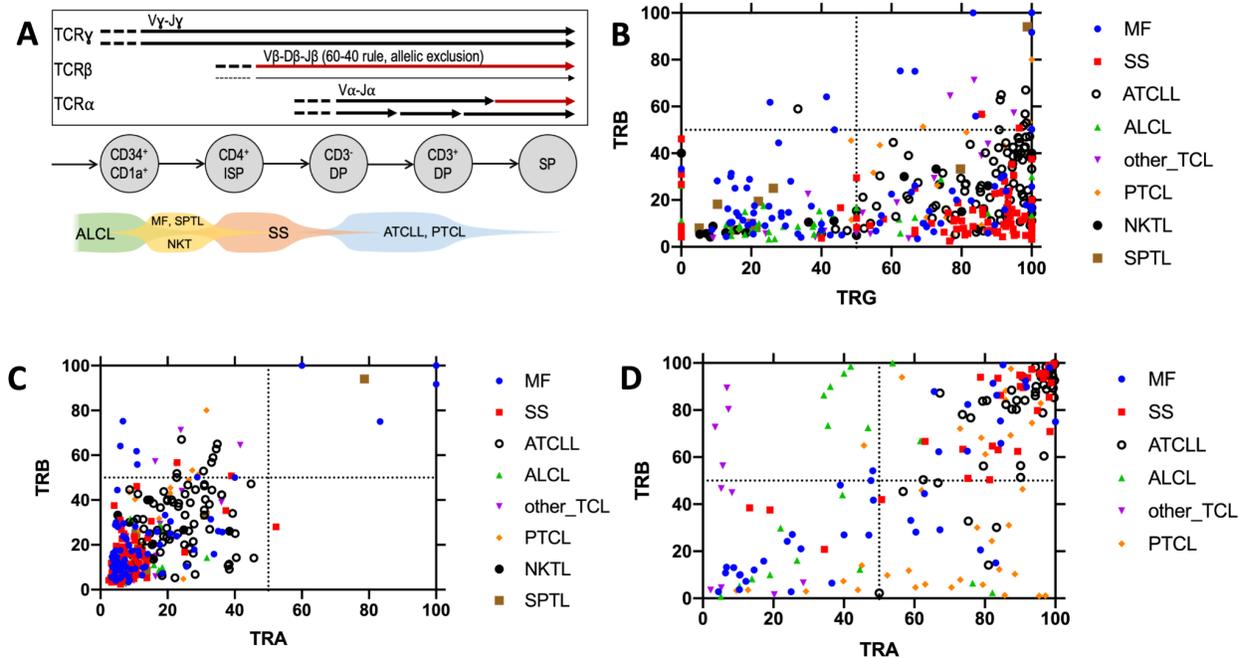


Figure 5. Oligoclonality in T-cell lymphomas. **A** - Schematic presentation of the TCR gene rearrangement during T-cell development. The arrows represent rearrangement on each chromosome mapped to different stages of normal human thymocyte. The thin arrowline for TCR β symbolizes the 60-40 rule. Predicted time points of the transformation in various subtypes of TCLs is shown in the bottom row. **B**, **C** - WES/WGS samples (n=367) were analyzed to identify the TCR α^{DNA} , TCR β^{DNA} and TCR γ^{DNA} clonotypes. The frequency of the most abundant TCR β^{DNA} clonotype was plotted versus the added frequency of two most frequent TCR γ^{DNA} clonotypes or the two most abundant TCR α^{DNA} clonotypes (C) for each sample. **D** - A similar correlation plot of the frequency of dominant TCR β^{RNA} clonotype versus the dominant TCR α^{RNA} clonotype (data from 196 WTS samples). Abbreviations of disease entities as in Fig 1.

Interestingly, most TCL samples showed a clearly monoclonal pattern of TCR γ dominated by one or two clonotypes (**Supplementary Fig S1**). By plotting the frequency of the dominant (most abundant) TCR β^{DNA} clonotype versus that of the two most abundant TCR γ clonotypes (**Figure 5 B**) most of the cases fall into two categories. The samples located in the lower left quadrant had low-frequency TCR β^{DNA} and low-frequency TCR γ and represented 32% of samples classified as oligoclonal for both TCR γ and TCR β . Most ALCL cases were found in this group. The remaining (68%) of the samples showed a high proportion of TCR γ and a disproportionately low frequency of TCR β^{DNA} (lower right quadrant) indicating TCR γ monoclonality but oligoclonality with respect to TCR β . Only single cases of TCL were in the right upper quadrant representing high frequency of TCR β and TCR γ which would be characteristic for an expansion of a single, mature T-cell clone. A similar plot of the frequencies of the dominant TCR β^{DNA} vs the combined frequencies of the first and second (biallelic) TCR α^{DNA} showed a linear correlation within the lower left quadrant representing oligoclonality in both chains (**Fig 5 C**). The samples clustered in two major groups: those with the higher frequency of TCR β and α (mostly ATCL) and the remaining occupying the low-frequency portion of the quadrant (**Fig 5 C**). The analysis of the TCR α^{RNA} and TCR β^{RNA} clonotypes helped in the interpretation of the findings. As shown in **Fig 5 D**, most ATCLL and SS samples showed monoclonality being located in the right upper quadrant representing high frequencies of the dominant clonotypes. However, ALCL, PTCL, MF and other TCL were scattered in the remaining

quadrants underscoring their oligoclonality not only at the DNA level but also at the level of TCR gene expression.

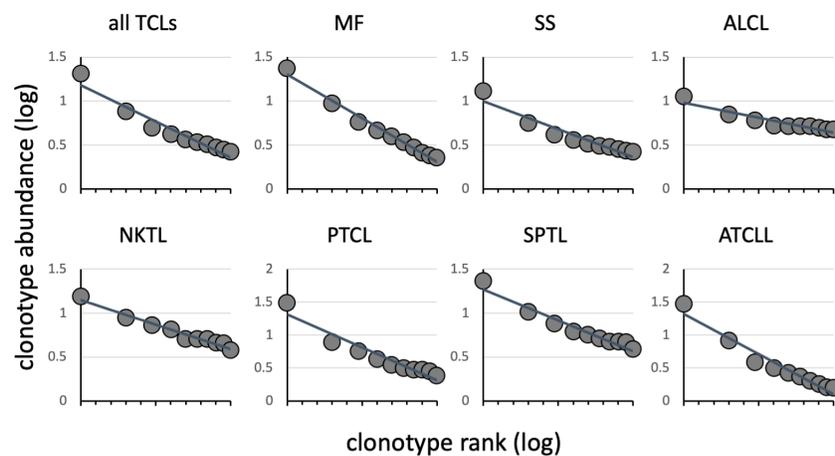


Figure 6. Rank-frequency plots showing power-law distribution of clonotypes in TCLs. TCR β^{DNA} clonotypes identified from 378 TCL samples were analyzed to estimate the frequency of each clonotype in the individual samples. The frequency of each clonotype was plotted versus its abundance rank (log-log plot). **A** - TCR β^{DNA} clonotypes from all TCLs subtypes. **B-H** - TCR β^{DNA} clonotypes from individual TCL entities. The line is the linear regression. Abbreviations of disease entities as in Fig 1.

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 (**Fig 5 A**). ALCL was the most heterogeneous with the clear oligoclonality of TCR γ , $-\beta$, and $-\alpha$ which indicated that the transformation takes place in immature thymocytes before TCR γ recombination step. On the opposite pole of the spectrum were SS and ATCLL which in the majority of cases showed TCR γ monoclonality and high frequency of TCR β , mapping the transformation event approximately at the TCR β rearrangement step. However, none of the TCLs exhibited a consistent pattern of monoclonality, defined at the DNA level by maximum two rearrangements of each TCR γ , $-\beta$ and $-\alpha$ genes matching the TCF and at the RNA level by a single dominant TCR β clonotype and maximum two TCR α clonotypes.

Based on the above described results we hypothesized that TCR β^{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 which 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 TCR β^{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 TCR β^{DNA} clonotype was plotted versus 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 reflecting different values of the constant γ for different types of TCLs (**Fig 6**). Reproducible linearity of the rank-size TCR β 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].

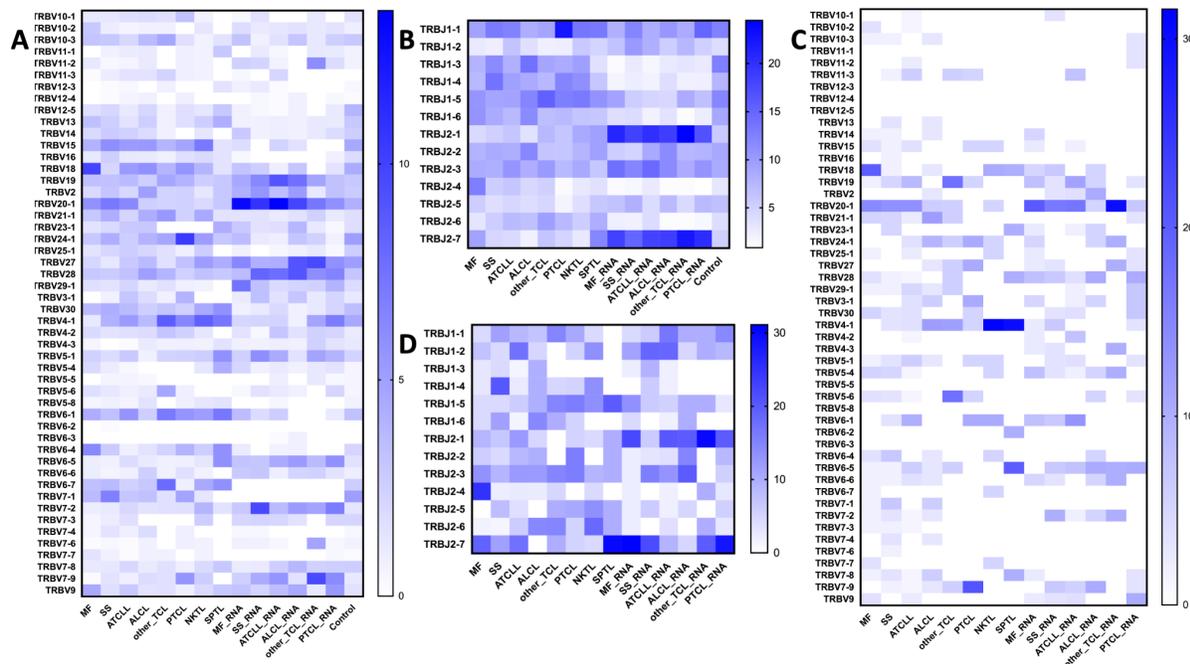


Figure 7. V β and J β (*TRBV/TRBJ*) usage in TCLs. The identified TCR β^{DNA} and TCR β^{RNA} clonotypes from WES/WGS and WTS data for the TCL diagnoses and control blood lymphocytes to identify preferential *TRBV* and *TRBJ* segment usage. **A, B** - represent *TRBV* and *TRBJ* frequencies, respectively, in the 10 most abundant clonotypes. **C, D** - similar plot as in A, B but including the single, most abundant clonotype. Control group is absent in C and D due to the lack of a high frequency clonotype in these samples. Abbreviations of disease entities as in Fig 1.

Biased V and J usage in TCLs

Previous studies have shown biased rearrangement of certain TCR V and J genes 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 TCR β (**Fig 7**) or the two most abundant clonotypes in case of TCR α and TCR γ (**Fig 8**). There was a strong preferential usage of certain V β and V α segments across different TCLs. The most striking was V β 20-1 (*TRBV20-1*) which was found in a large proportion of clonotypes in MF, SS, ATCLL, ALCL and NKTL, both at the level of DNA and RNA. V β 4-1 (*TRBV4-1*) was overrepresented in ALCL, NKTL, and SPTL. Other V β segments were more disease-specific, e.g. *TRBV18* was often found in MF and *TRBV24-1* was frequent in PTCL. Biased use of V α segments was also obvious, for example V α 16 (*TRAV16*), V α 35 (*TRAV35*), V α 41 (*TRAV41*) and V α 3 (*TRAV3*) were overrepresented in most TCLs. The J β and J α segments also seemed to be unequally used (e.g. *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 (**Fig 7 and 8**).

We have also analyzed V and J segment usage in the 10 most abundant TCR β clonotypes, representing the median number of malignant clonotypes, determined in comparison to TCF. The overall pattern was similar to that of the most abundant clonotypes confirming biased use of *TRBV20-1*, *TRBV4-1*, *TRBJ2-1* and *TRBJ2-7*.

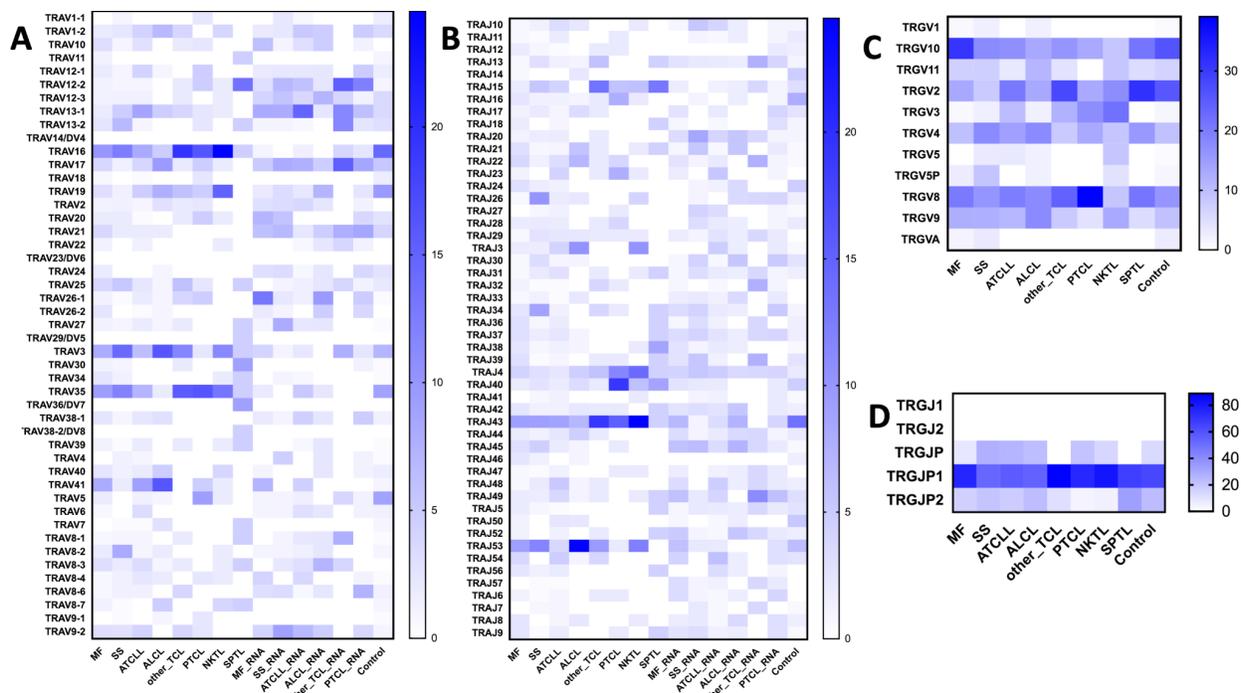


Figure 8. V and J (TRAV, TRAJ, TRGV, TRGJ) usage in TCLs. The frequency of V and J segments of TCR α and TCR γ clonotypes are plotted as in Fig 7. Two most abundant TCR α^{DNA} TCR γ^{DNA} clonotypes from TCLs and control samples were identified and the frequency of V and J segments plotted on the heat map. The RNA samples were used only for identification of the TCR α^{RNA} clonotypes. **A, B** - TRAV and TRAJ segment usage; **C, D** - TRGV and TRGJ segment usage.

Lastly, previous studies in CTCL have indicated over-representation of the *TRGV10-TRGJP1* in the malignant T-cells [48]. We could confirm this in our analysis and also found that *TRGV2*, *TRGV8* and *TRGV9* were frequently used across different TCLs, but *TRGV1*, *5*, *11* were relatively underrepresented (**Fig 8 C,D**).

Shared malignant clonotypes represent public clonotypes and have high recombination probability

The results described above indicated that certain V-J combinations are overrepresented suggesting that the clonotypes (defined as by the aminoacid sequence of the highly variable CDR3 region) may also be shared between patients with different TCLs. Indeed, as shown in **supplementary Tables S4 and S5** we have found numerous shared TCR α^{aa} and TCR β^{aa} . Some clonotypes were shared across a large number of patients with different TCLs. For example, CARRKSSFF was detected in 100 samples (54 samples using *TRBV28-TRAJI-1*) and CDNNNDMRF (*TRAV16-TRAJ43*) was present in 213 samples.

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

Shared malignant clonotypes resembled public clonotypes found in the blood of healthy subjects. Public TCRs largely result from V-J recombination biases and convergent recombination where rearrangement of

different segments can result in the same aminoacid sequence [49]. Public TCR chains can pair with a diverse repertory 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 TCR β and TCR α clonotypes. P_{gen} of those shared malignant clonotypes had two probability peaks at 10^{-9} and 10^{-6} respectively (**Fig 9 A** and **supplementary Tables S4-S7**). Shared TCR β^{RNA} and TCR α^{RNA} had even higher values with the maxima at 10^{-6} and 10^{-4} , respectively (**Fig 9 B**) which were significantly higher than the P_{gen} of peripheral blood private (unique) clonotypes ($<10^{-9}$).

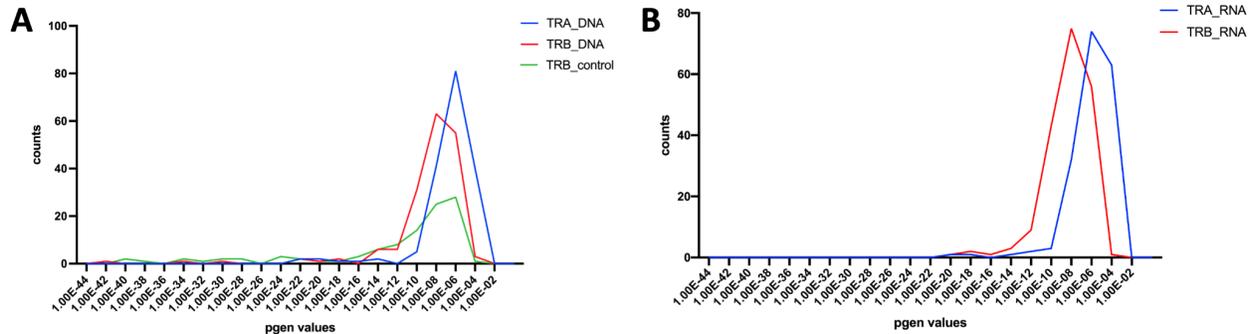


Figure 9. Probability of generation (P_{gen}) for TCR α and TCR β clonotypes. Shared TCR α^{aa} and TCR β^{aa} clonotypes were identified using the Immunarch pipeline and their P_{gen} was calculated using OLGA algorithm (Fig 1) from DNA and RNA data. **A** - P_{gen} histogram for the shared TCR α^{DNA} and TCR β^{DNA} clonotypes. As a comparison, the P_{gen} values for the TCR β^{DNA} clonotypes for the healthy control group are plotted. **B** - analogous histogram as in (A) showing P_{gen} for the shared clonotypes in TCR α^{RNA} and TCR β^{RNA} .

Discussion

TCLs represent a clinically diverse spectrum of T-cells cancers, varying 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 to trace the clonal history of T-cell lymphoma 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 directly contradicted by our findings. Malignant transformation of a single lymphocyte randomly drawn from a vast pool of 10^6 - 10^7 singular T-cell clonotypes should yield lymphomas which 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 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 which 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 4 A**, assumes that precursor cell for mature TCLs progress through the sequential rearrangements, analogously to normal thymocytes and in a stark contrast to immature T-cell lymphomas and leukemias (e.g. T-ALL) which become ‘arrested’ at a given stage of development. Thus, transformation before TCR γ rearrangement would result in oligoclonality for all TCR rearrangements whereas transformation after TCR γ rearrangement would render cancer cells monoclonal for TCR γ but oligoclonal for TCR β and TCR α . Transformation after the TCR β rearrangement stage would render the malignant cells monoclonal (single TCR β^{RNA} clonotype and 1-2 TCR β^{DNA} clonotypes) but oligoclonal for TCR α . Across different TCLs, ALCL showed a fingerprint of transformation at the earliest stage (before or at the TCR γ rearrangement) and ATCLL and PTCL seemed to derive from a cell in a later stage of development (TCR β 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.

Existence of more than one malignant clonotype has previously been shown in TCLs, but this finding has largely been ignored 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 (TCF). 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 the 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 only understandable 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 as it could be argued that clonotypic oligoclonality might also be a result of secondary TCR rearrangements in the already established tumor. This is however unlikely as the TCR rearrangement machinery is irreversibly inactivated after completion of TCR α 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 and clonotype sharing. TCR rearrangements in human thymocytes are characterized by biased V and J segment usage reflecting the probability of recombination (P_{gen}) producing shared (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* or *TRAV35*, *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 TCR α and $>10^{-9}$ for TCR β) [54,55]. Differences in combinatorial probability is a probable explanation for a higher Jaccard index (degree of clonotype sharing) for TCR α than TCR β .

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 or cytomegalovirus [46,61]. However, none of the known CDR3 regions recognizing viral antigen was found among the malignant clonotypes. We postulate that the frequent finding of *TRBV20-1* in TCLs is not due to its antigenic selection but simply its ubiquity. *TRBV20-1* is the prominent component of the TCR β clonotypes of normal blood [62,63], tumor-infiltrating T-cells [64,65], or autoimmune reactions [66–69]. *TRBV20-1* is the most frequently rearranged segment in DP thymocytes [55] and is frequently found in T-cell acute lymphoblastic leukemia (T-ALL) which directly develops 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]. We have found early MF precursor cells in the bone marrow before the clinical onset of the disease [73] and cases of transmission of T-cell lymphoma by bone marrow transplant from asymptomatic donors who later developed the disease have been observed [74] indicating that bone marrow could be a potential reservoir or progenitor malignant cells.

In summary, we propose that mature TCLs do not originate from the 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 to design targeted therapies intervening with the disease at the initial stages.

Acknowledgements

The authors would like to thank the NIH Data access committee and EGA European Genome-Phenome Archive database for their help with getting access to certain restricted datasets.

This research was funded by the unrestricted grants from the University of Alberta (Canada) and Bispebjerg Hospital (Copenhagen, Denmark), Danish Cancer Society (Kræftens Bekæmpelse; R124-A7592 Rp12350), Canadian Dermatology Foundation (CDF) and University Hospital Foundation (University of Alberta). Additionally Dr. Aishwarya Iyer was funded under the accelerate postdoctoral fellowship under a joint grant from the Mitacs and Sun Pharma Global FZE.

Competing interests: All authors have completed the ICMJE uniform disclosure form and declare: RG - no support from any organisation for the submitted work; speaker honoraria and advisory board honoraria from the following organisations that might have an interest in the submitted work - Mallinckrodt, Lilly, Sanofi, Abbvie, Sunpharma, Kyowa Kirin; no other relationships or activities that could appear to have influenced the submitted work. Other authors - no competing interests.

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: 239–249.
2. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016;127: 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: 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: 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: 54–66.
6. Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. *Nature*. 1988;334: 395–402.
7. Warren RL, Freeman JD, Zeng T, Choe G, Munro S, Moore R, 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: 790–797.
8. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A direct estimate of the human alpha T cell receptor diversity. *Science*. 1999;286: 958–961.
9. Qi Q, Liu Y, Cheng Y, Glanville J, Zhang D, Lee J-Y, et al. Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci U S A*. 2014;111: 13139–13144.
10. Langerak AW, Groenen PJTA, Brüggemann M, Beldjord K, Bellan C, Bonello L, et al. EuroClonality/BIOMED-2 guidelines for interpretation and reporting of Ig/TCR clonality testing in suspected lymphoproliferations. *Leukemia*. 2012;26: 2159–2171.
11. Iyer A, Hennessey D, O'Keefe S, Patterson J, Wang W, Salopek T, et al. Clonotypic heterogeneity in cutaneous T-cell lymphoma (mycosis fungoides) revealed by comprehensive whole-exome sequencing. *Blood Adv*. 2019;3: 1175–1184.
12. Iyer A, Hennessey D, O'Keefe S, Patterson J, Wang W, Wong GK-S, et al. Skin colonization by circulating neoplastic clones in cutaneous T-cell lymphoma. *Blood*. 2019;134: 1517–1527.
13. Moti N, Malcolm T, Hamoudi R, Mian S, Garland G, Hook CE, 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: 1843–1852.
14. Congras A, Hoareau-Aveilla C, Caillet N, Tosolini M, Villarese P, Cieslak A, et al. ALK-transformed mature T lymphocytes restore early thymus progenitor features. *J Clin Invest*. 2020;130: 6395–6408.
15. Brüggemann M, van der Velden VHJ, Raff T, Droese J, Ritgen M, Pott C, 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: 709–719.
16. Asnafi V, Beldjord K, Boulanger E, Comba B, Le Tutour P, Estienne M-H, 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: 2693–2703.
 17. Iyer A, Hennessey D, O’Keefe S, Patterson J, Wang W, Wong GK-S, et al. Branched evolution and genomic intratumor heterogeneity in the pathogenesis of cutaneous T-cell lymphoma. *Blood Adv*. 2020;4: 2489–2500.
 18. Choi J, Goh G, Walradt T, Hong BS, Bunick CG, Chen K, et al. Genomic landscape of cutaneous T cell lymphoma. *Nat Genet*. 2015;47: 1011–1019.
 19. Ungewickell A, Bhaduri A, Rios E, Reuter J, Lee CS, Mah A, et al. Genomic analysis of mycosis fungoides and Sézary syndrome identifies recurrent alterations in TNFR2. *Nat Genet*. 2015;47: 1056–1060.
 20. McGirt LY, Jia P, Baerenwald DA, Duszynski RJ, Dahlman KB, Zic JA, et al. Whole-genome sequencing reveals oncogenic mutations in mycosis fungoides. *Blood*. 2015;126: 508–519.
 21. da Silva Almeida AC, Abate F, Khiabani H, Martinez-Escala E, Guitart J, Tensen CP, et al. The mutational landscape of cutaneous T cell lymphoma and Sezary syndrome. *Nat Genet*. 2015;47: 1465–1470.
 22. Li Z, Lu L, Zhou Z, Xue W, Wang Y, Jin M, 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: 406–410.
 23. Simpson HM, Khan RZ, Song C, Sharma D, Sadashivaiah K, Furusawa A, et al. Concurrent Mutations in ATM and Genes Associated with Common γ Chain Signaling in Peripheral T Cell Lymphoma. *PLoS One*. 2015;10: e0141906.
 24. Palomero T, Couronné L, Khiabani H, Kim M-Y, Ambesi-Impiombato A, Perez-Garcia A, et al. Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nat Genet*. 2014;46: 166–170.
 25. Yoo HY, Sung MK, Lee SH, Kim S, Lee H, Park S, et al. A recurrent inactivating mutation in RHOA GTPase in angioimmunoblastic T cell lymphoma. *Nat Genet*. 2014;46: 371–375.
 26. Jiang L, Gu Z-H, Yan Z-X, Zhao X, Xie Y-Y, Zhang Z-G, et al. Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. *Nat Genet*. 2015;47: 1061–1066.
 27. Kataoka K, Nagata Y, Kitanaka A, Shiraishi Y, Shimamura T, Yasunaga J-I, et al. Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat Genet*. 2015;47: 1304–1315.
 28. Wang L, Ni X, Covington KR, Yang BY, Shiu J, Zhang X, et al. Genomic profiling of Sézary syndrome identifies alterations of key T cell signaling and differentiation genes. *Nature Genetics*. 2015. pp. 1426–1434. doi:10.1038/ng.3444
 29. Crescenzo R, Abate F, Lasorsa E, Tabbo’ F, Gaudiano M, Chiesa N, et al. Convergent mutations and kinase fusions lead to oncogenic STAT3 activation in anaplastic large cell lymphoma. *Cancer Cell*.

- 2015;27: 516–532.
30. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet.* 2014;46: 736–741.
 31. Ha G, Roth A, Khattra J, Ho J, Yap D, Prentice LM, et al. TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res.* 2014;24: 1881–1893.
 32. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun.* 2017;8: 1324.
 33. Website. [cited 25 Jun 2021]. Available: <https://gatk.broadinstitute.org/hc/en-us/articles/360035890631-Panel-of-Normals-PON->
 34. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods.* 2015;12: 380.
 35. Nazarov V, Rumynskiy E. immunomind/immunarch: 0.6.5: Basic single-cell support. 2020 [cited 20 Jun 2021]. doi:10.5281/zenodo.3893991
 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: 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: 539–544.
 39. Hosoya T, Li H, Ku C-J, Wu Q, Guan Y, Engel JD. High-Throughput Single-Cell Sequencing of both Alleles. *J Immunol.* 2018;201: 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: 18172–18174.
 41. Kirsch IR, Watanabe R, O'Malley JT, Williamson DW, Scott L-L, Elco CP, et al. TCR sequencing facilitates diagnosis and identifies mature T cells as the cell of origin in CTCL. *Sci Transl Med.* 2015;7: 308ra158.
 42. Masson A de, de Masson A, O'Malley JT, Elco CP, Garcia SS, Divito SJ, et al. High-throughput sequencing of the T cell receptor β gene identifies aggressive early-stage mycosis fungoides. *Science Translational Medicine.* 2018. p. eaar5894. doi:10.1126/scitranslmed.aar5894
 43. Corominas-Murtra B, Seoane LF, Solé R. Zipf's Law, unbounded complexity and open-ended evolution. *J R Soc Interface.* 2018;15: 20180395.
 44. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. Identification of neutral tumor

- evolution across cancer types. *Nat Genet.* 2016;48: 238–244.
45. Marzo GD, De Marzo G, Gabrielli A, Zaccaria A, Pietronero L. Dynamical approach to Zipf’s law. *Physical Review Research.* 2021. doi:10.1103/physrevresearch.3.013084
 46. Zhang W, Hawkins PG, He J, Gupta NT, Liu J, Choonoo G, et al. A framework for highly multiplexed dextramer mapping and prediction of T cell receptor sequences to antigen specificity. *Sci Adv.* 2021;7. doi:10.1126/sciadv.abf5835
 47. Turner SJ, Doherty PC, McCluskey J, Rossjohn J. Structural determinants of T-cell receptor bias in immunity. *Nat Rev Immunol.* 2006;6: 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: 1277–1283.
 49. Venturi V, Kedzierska K, Price DA, Doherty PC, Douek DC, Turner SJ, et al. Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proc Natl Acad Sci U S A.* 2006;103: 18691–18696.
 50. Marcou Q, Mora T, Walczak AM. High-throughput immune repertoire analysis with IGoR. *Nat Commun.* 2018;9: 561.
 51. Schlöpfer M, Dong L, O’Keeffe K, Santi P, Szell M, Salat H, et al. The universal visitation law of human mobility. *Nature.* 2021;593: 522–527.
 52. Gabaix X. Zipf’s Law for Cities: An Explanation. *The Quarterly Journal of Economics.* 1999. pp. 739–767. doi:10.1162/003355399556133
 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: 207–216.
 54. Heikkilä N, Vanhanen R, Yohannes DA, Kleino I, Mattila IP, Saramäki J, 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, Popescu D-M, Lavaert M, Kunz DJ, et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science.* 2020;367. doi:10.1126/science.aay3224
 56. Clemente MJ, Przychodzen B, Jerez A, Dienes BE, Afable MG, Husseinzadeh H, et al. Deep sequencing of the T-cell receptor repertoire in CD8+ T-large granular lymphocyte leukemia identifies signature landscapes. *Blood.* 2013;122: 4077–4085.
 57. Malcolm TIM, Hodson DJ, Macintyre EA, Turner SD. Challenging perspectives on the cellular origins of lymphoma. *Open Biol.* 2016;6. doi:10.1098/rsob.160232
 58. Litvinov IV, Shtreis A, Kobayashi K, Glassman S, Tsang M, Woetmann A, et al. Investigating potential exogenous tumor initiating and promoting factors for Cutaneous T-Cell Lymphomas

- (CTCL), a rare skin malignancy. *Oncoimmunology*. 2016;5: 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: 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 J β 1 Bias. *Journal of Investigative Dermatology*. 2006. pp. 1893–1899. doi:10.1038/sj.jid.5700304
 61. Miles JJ, Bulek AM, Cole DK, Gostick E, Schauenburg AJA, Dolton G, et al. Genetic and Structural Basis for Selection of a Ubiquitous T Cell Receptor Deployed in Epstein-Barr Virus Infection. *PLoS Pathogens*. 2010. p. e1001198. doi:10.1371/journal.ppat.1001198
 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: 1817–1824.
 63. Hou X, Chen W, Zhang X, Wang G, Chen J, Zeng P, et al. Preselection TCR repertoire predicts CD4 and CD8 T-cell differentiation state. *Immunology*. 2020;161: 354–363.
 64. Li B, Li T, Pignon J-C, Wang B, Wang J, Shukla SA, et al. Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat Genet*. 2016;48: 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: 454–456.
 66. Petersen J, Montserrat V, Mujico JR, Loh KL, Beringer DX, van Lummel M, et al. T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with celiac disease. *Nat Struct Mol Biol*. 2014;21: 480–488.
 67. Luo G, Ambati A, Lin L, Bonvalet M, Partinen M, Ji X, et al. Autoimmunity to hypocretin and molecular mimicry to flu in type 1 narcolepsy. *Proc Natl Acad Sci U S A*. 2018;115: E12323–E12332.
 68. Rosati E, Pogorelyy MV, Dowds CM, Moller FT, Sorensen SB, Lebedev YB, et al. Identification of Disease-associated Traits and Clonotypes in the T Cell Receptor Repertoire of Monozygotic Twins Affected by Inflammatory Bowel Diseases. *J Crohns Colitis*. 2020;14: 778–790.
 69. Yohannes DA, Freitag TL, de Kauwe A, Kaukinen K, Kurppa K, Wacklin P, 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: 17977.
 70. McClory S, Hughes T, Freud AG, Briercheck EL, Martin C, Trimboli AJ, et al. Evidence for a stepwise program of extrathymic T cell development within the human tonsil. *J Clin Invest*. 2012;122: 1403–1415.
 71. García-Ojeda ME, Dejbakhsh-Jones S, Chatterjea-Matthes D, Mukhopadhyay A, BitMansour A, Weissman IL, 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: 4363–4373.
 72. Holland AM, Zakrzewski JL, Tsai JJ, Hanash AM, Dudakov JA, Smith OM, et al. Extrathymic development of murine T cells after bone marrow transplantation. *J Clin Invest*. 2012;122:

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: 3797–3799.
74. Berg KD, Brinster NK, Huhn KM, Goggins MG, Jones RJ, Makary A, et al. Transmission of a T-cell lymphoma by allogeneic bone marrow transplantation. *N Engl J Med*. 2001;345: 1458–1463.