Bhattacharya, Dipabarna; Teramo, Antonella; Gasparini, Vanessa Rebecca; Huuhtanen, Jani; Kim, Daehong; Theodoropoulos, Jason; Schiavoni, Gianluca; Barilà, Gregorio; Vicenzetto, Cristina; Calabretto, Giulia; Facco, Monica; Kawakami, Toru; Nakazawa, Hideyuki; Falini, Brunangelo; Tiacci, Enrico; Ishida, Fumihiro; Semenzato, Gianpietro; Kelkka, Tiina; Zambello, Renato; Mustjoki, Satu

Identification of novel STAT5B mutations and characterization of TCRβ signatures in CD4+ T-cell large granular lymphocyte leukemia

Published in:
Blood Cancer Journal

DOI:
10.1038/s41408-022-00630-8

Published: 24/02/2022

Please cite the original version:
Bhattacharya, D., Teramo, A., Gasparini, V. R., Huuhtanen, J., Kim, D., Theodoropoulos, J., Schiavoni, G., Barilà, G., Vicenzetto, C., Calabretto, G., Facco, M., Kawakami, T., Nakazawa, H., Falini, B., Tiacci, E., Ishida, F., Semenzato, G., Kelkka, T., Zambello, R., & Mustjoki, S. (2022). Identification of novel STAT5B mutations and characterization of TCRβ signatures in CD4+ T-cell large granular lymphocyte leukemia. Blood Cancer Journal, 12(2), 1-12. [31]. https://doi.org/10.1038/s41408-022-00630-8
CD4+ T-cell large granular lymphocyte leukemia (T-LGLL) is a rare subtype of T-LGLL with unknown etiology. In this study, we molecularly characterized a cohort of patients (n = 35) by studying their T-cell receptor (TCR) repertoire and the presence of somatic STAT5B mutations. In addition to the previously described gain-of-function mutations (N642H, Y665F, Q706L, S715F), we discovered six novel STAT5B mutations (Q220H, E433K, T628S, P658R, P702A, and V712E). Multiple STAT5B mutations were present in 22% (5/23) of STAT5B mutated CD4+ T-LGLL cases, either coexisting in one clone or in distinct clones. Patients with STAT5B mutations had increased lymphocyte and LGL counts when compared to STAT5B wild-type patients. TCRβ sequencing showed that, in addition to large LGL expansions, non-leukemic T cell repertoires were more clonal in CD4+ T-LGLL compared to healthy. Interestingly, 25% (15/59) of CD4+ T-LGLL clones were found, albeit in much lower frequencies, in the non-leukemic CD4+ T cell repertoires of the CD4+ T-LGLL patients. Additionally, we further confirmed the previously reported clonal dominance of TRB86-expressing clones in CD4+ T-LGLL. In conclusion, CD4+ T-LGLL patients have a typical TCR and mutation profile suggestive of aberrant antigen response underlying the disease.

**INTRODUCTION**

T-cell large granular lymphocyte leukemia (T-LGLL) is a rare lymphoproliferative disease characterized by chronic expansion of clonal, mature cytotoxic T cells in the peripheral blood and bone marrow [1]. Two subtypes of T-LGLL are commonly recognized: the most common CD8+ T-LGLL (70% of cases) and the less frequent CD4+ T-LGLL (30% of cases) [1]. Clinically, CD4+ T-LGLL is usually an indolent disease and unlike in its CD8+ counterpart, these patients rarely have cytopenias or autoimmune symptoms. However, CD4+ T-LGLL has been reported more frequently associated with secondary neoplasms such as monoclonal B-cell lymphocytosis and plasma cell disorders [2, 3]. In CD4+ T-LGLL, the leukemic T-LGLs express CD4 (either alone or together with CD8) and the α/β T-cell receptor (TCR) together with a typical mature cytotoxic (Granzyme B+, CD56+, CD57+) and activated/memory T-cell (CD2+dim, CD7bright, CD11a+bright, CD28−, CD62L−) phenotype [3].

Up to 55% of CD4+ T-LGLL patients have been shown to harbor STAT5B mutations [4, 5]. In CD8+ T-LGLL, the most common mutated gene is STAT3 [6], whereas STAT5B mutations are rare and often associated with an aggressive disease form [7–9]. All reported STAT5B mutations in CD4+ T-LGLL are point mutations within the SH2 or transactivation domains of STAT5B. N642H and Y665F are the most common STAT5B mutations, and they both have been shown to increase STAT5B protein activity [4, 10, 11].

The etiology of CD4+ T-LGLL remains unknown. An initial antigen-driven expansion of CD4+ T cells, followed by the occurrence of oncogenic events (i.e., somatic mutations), has been suggested to lead to the persistence of abnormal T-cell clones [11]. Non-self-antigen(s) instead of autoantigens [5] are proposed targets of CD4+ T-LGLL clones as CD4+ T-LGLL are not associated with autoimmune diseases. In some earlier reports, CD4+ T-LGLL clones have been implied to recognize cytomegalovirus (CMV) antigens [12, 13]. In CD4+ T-LGLL [3], the enrichment of the Vß13.1 gene usage has also been reported, differentiating it from
CD8+ T-LGLL clones is not a stochastic process, but rather a result of an antigen-driven immune response [18–22].

As CD4+ T-LGLL is a rare disease, previous studies evaluating the clinical impact of STAT5B mutations have been limited in size, and no deep TCR profiling has been performed with modern sequencing and bioinformatic tools [23–25]. Therefore, we aimed to collect a large cohort of CD4+ T-LGLL patients (n = 35) and examine by deep amplicon sequencing the STAT5B mutation status and correlate the genotype information with the clinical data. Additionally, by deep TCRβ sequencing, we studied the landscape of both the leukemic and non-leukemic T-cell repertoires in CD4+ T-LGLL and compared that to the landscapes from the healthy controls and patients with CD8+ T-LGLL.

**Materials and Methods**

**Patient cohort**

CD4+ T-LGLL patients (n = 35) were recruited from the hematology units of 3 academic institutions, namely Padova (Italy; n = 22 patients), Helsinki (Finland; n = 6 patients) and Shinshu (Japan; n = 7 patients). Buffy coats from healthy blood donors (n = 37) were provided by the Finnish Red Cross Blood service (Helsinki, Finland). The study was conducted in accordance with the Ethics Committees (Padova University Hospital ethics committee, approval number 4213/AO/17; Helsinki University Hospital ethics committee, approval number 303/12/03/01/2011; Shinshu University Hospital ethics committee, approval number 581). All patients gave written informed consent prior to their inclusion in the study in accordance with the Declaration of Helsinki.

**Sample preparation**

Peripheral blood mononuclear cells (PBMC) were isolated with Ficoll-Hypaque (Sigma Aldrich) gradient centrifugation. Leukemic clones were purified using magnetic micro-beads coated with monoclonal anti-human CD4, CD57, or CD56 antibodies (Miltenyi Biotech). When available, cryopreserved cells (n = 15) were first subjected to depletion of CD8+ cells, after which the CD8-negative population (or CD8+ in case of CD4+ CD8+ double-positive LGL clone) was further purified with anti-human CD4-conjugated microbeads. Alternatively, CD57 (n = 8) or CD56 (n = 3) purified LGLs or PBMC (n = 9) were used depending on sample availability. A similar approach was used to purify CD4+ cells from buffy coats obtained from healthy controls (n = 37). Flow cytometry analysis was conducted to control the purity of the obtained cell fractions (>95% among lymphocytes) with FACS Canto II in Padova and FACS Verse in Helsinki; data were analyzed with FACS Diva and FACS Suite software, respectively (all Becton Dickinson). Surface staining was performed according to the purification: CD4 FITC CD8 PE or CD4 FITC CD57 PE or CD4 FITC CD56 APC (all Becton Dickinson). DNA was extracted using Gentra Puregene Cell and Tissue kit (Qiagen) and quantified using the Qubit 2.0 fluorometer with DNA high sensitivity kit (Thermo Fisher Scientific).

**Amplonc sequencing of STAT5B**

STAT5B sequencing was performed from LGL or PBMC fraction (Fig. 1) to detect somatic mutations. Firstly, targeted sequencing was performed starting from 50 ng of DNA. Libraries, generated according to the manufacturer’s instructions, were sequenced on an Illumina MiSeq instrument using MiSeq Reagent kit v2 or v3 and following the TruSeq Custom Amplicon Assay (illumina) pipeline. Bioinformatic variant calling was performed using MiSeq Reporter software (version 2.5.1) with default settings while for variant annotation Illumina Variant Studio 2.0 was used. Next, the observed mutations were confirmed using amplicon sequencing as previously described in Kim et al. [26]. Amplonc primers used for both techniques are reported in Supplementary Table 1. Germline variants were excluded by examination of a non-leukemic cell population from each patient. All identified variants were visually confirmed using the Integrative Genome Viewer (IGV) [27].

**Western blot assay and STAT5B luciferase reporter assay**

Please refer to Supplementary methods and Supplementary Table 2.

**TCRβ sequencing and data analysis**

TCRβ sequencing from genomic DNA (the same DNA as used for amplicon sequencing, Fig. 1 and Table 1) of CD4+ T-LGLL patients (n = 27) was conducted with Adaptive Biotechnologies’ ImmunoSEQ assay with “Survey” resolution [28]. Only productive, i.e., complete, in-frame, TCRβ rearrangements were included in the analyses. Sample clonality was assessed using the Simpson clonality as provided by the Adaptive ImmunoSeq Analyzer (ver 3.0). Healthy controls (n = 785) from peripheral blood [29] were also used after downsampling to 40,000 reads. Clones, defined as unique rearranged receptor sequences in the ImmunoSEQ Analyzer, were classified as very rare (<0.01%), small (0.01–0.1%), medium (0.1–1%), large (1–20%), and hyperexpanded (>20%). The lowest detectable productive frequency in the TCR sequencing data was 2.59 × 10−5. OLGA [25] (ver 1.2.3) was used to calculate the generation probabilities of the TCR rearrangements with default parameters.

VDjtools was used to generate circus plots for V gene usage from TCRβ-sequencing data [30]. The query against the VDJ database (VDJdb; ver 2021-01-10) [24] was done for all available TCRs from CD4+ cells. To evaluate possible amino acid level similarities of TCRs, GLIPH (ver 1.0.0 and 2.0.0) [23, 31] was used on a remote server with default parameters.

**Statistical methods**

Statistical analysis for the evaluation of differential STAT5B transcriptional activities through luciferase reporter assay was performed with a two-way ANOVA followed by Dunn’s multiple correction test. Fisher’s exact chi-square, or Bonferroni corrected Mann-Whitney tests were used to analyze the relationship between STAT5B mutational status, TCRβ repertoire covariates, and relative clinical features of the cohort of patients profiled (see Table 2 for more details). Fisher’s exact test with Bonferroni multiple correction was used to calculate V genes enrichment for CD4+ T-LGLL clones. All the analyses were performed.
Table 1. STAT5B mutations and TCRβ clonotypes in CD4+ T-LGLL patients.

| PATIENT | CELLS | Mutation 1 VAF | Other mutations VAF | Clone 1 | % | Clone 2 | % | Other clones | % | aa sequence |
|---------|-------|----------------|---------------------|---------|---|---------|---|-------------|---|-------------|
| 1       | CD4+  | 23%            |                     | TRBV10-03*01 | 39%|                     |   |             |    | CAVRIKGTQDQYF |
| 2       | CD4+CD8+ | Q706L 45% |                     | TRBV06-05*01 | 64%| TRBV03 12% |   | TRBV27-01*01 | 10%| CASSAAGTEAFF |
| 3       | CD4+  | 36%            |                     | TRBV12   78%|   |         |   |             |    | CASSLVSQGTVYF |
| 4       | CD4+  | 43%            |                     | TRBV05-04*01 | 74%|                     |   |             |    | CASSWSNQFEYF |
| 5       | CD4+  | 30%            |                     | TRBV12   63%|   |         |   |             |    | CASSLVSQGTVYF |
| 6       | CD4+  | 15%            |                     | TRBV03   32%| TRBV07-03*01 | 18%| TRBV12   7%|   | CASSQDDGSGNQFEYF |
| 7       | CD57+ | 47%            |                     | TRBV06-06 | 97%|                     |   |             |    | CASSSSTTSSTQYF |
| 8       | CD56+ | 32%            |                     | TRBV06-05*01 | 90%|                     |   |             |    | CASSSTGTANTAFF |
| 9       | CD4+CD8+ | T628S (G>C) |                     | TRBV06-05*01 | 46%| TRBV27-01*01 | 7%| TRBV24   6%|   | CASSSGQKTEAFF |
| 10      | CD4+CD8+ | T628S (G>C) | T628S (T>A) 2% | TRBV18-01*01 | 84%|                     |   |             |    | CASSFSQASTEYF |
| 11      | PBMC  | 33%            |                     | TRBV18-01*01 | 84%|                     |   |             |    | CASSFSQASTEYF |
| 12      | PBMC  | 49%            |                     | TRBV06-05*01 | 85%|                     |   |             |    | CASSFSQASTEYF |
| 13      | CD4+  | 23%            |                     | TRBV21-01*01 | 24%| TRBV06-01*01 | 22%| TRBV10-02*01 | 15%| CASSSSREKQTEAFF |
| 14      | PBMC  | 34%            |                     | TRBV06-05*01 | 59%| TRBV09-01 | 7%| TRBV12   2%|   | CASSGFGTANTEAFF |
| 15      | PBMC  | 31%            |                     | TRBV06-05*01 | 59%| TRBV09-01 | 7%| TRBV12   2%|   | CASSGFGTANTEAFF |
| 16      | PBMC  | 40%            |                     | TRBV09-01 | 71%|                     |   |             |    | CASSSSTTSSTQYF |
| 17      | CD57+ | 1%             |                     | TRBV06-01*01 | 61%| TRBV28-01*01 | 17%| TRBV10-03*01 | 11.74%| CASSRNYQOQF |
| 18      | CD4+  | 49%            |                     | TRBV09-01 | 71%|                     |   |             |    | CASSSSTTSSTQYF |
| 19      | CD4+CD8+ | T628S 42% |                     | TRBV06-05*01 | 85%|                     |   |             |    | CASSFFQGNTANTAFF |
| 20      | CD4+CD8+ | T628S 49% |                     | TRBV09-01 | 71%|                     |   |             |    | CASSFFQGNTANTAFF |
| 21      | CD56+ | 23%            |                     | TRBV09-01 | 71%|                     |   |             |    | CASSFFQGNTANTAFF |
| 22      | PBMC  | 49%            |                     | TRBV06-05*01 | 59%| TRBV09-01 | 7%| TRBV12   2%|   | CASSFFQGNTANTAFF |
| 23      | PBMC  | 30%            |                     | TRBV06-05*01 | 59%| TRBV09-01 | 7%| TRBV12   2%|   | CASSFFQGNTANTAFF |
| 24      | PBMC  | 31%            |                     | TRBV06-05*01 | 59%| TRBV09-01 | 7%| TRBV12   2%|   | CASSFFQGNTANTAFF |
| 25      | CD4+  | 6%             |                     | TRBV06-05*01 | 59%| TRBV09-01 | 7%| TRBV12   2%|   | CASSFFQGNTANTAFF |
| 26      | CD4+CD8+ | Q220H 2% |                     | TRBV09-01 | 71%|                     |   |             |    | CASSFFQGNTANTAFF |
| 27      | CD4+  | 6%             |                     | TRBV06-05*01 | 59%| TRBV09-01 | 7%| TRBV12   2%|   | CASSFFQGNTANTAFF |
| 28      | CD4+  | 56%            |                     | TRBV05-08*01 | 26%| TRBV25-01*01 | 21%| TRBV04-01*01 | 6%| CASSQGERNVGRTQYF |
| 29      | CD57+ | 34%            |                     | TRBV05-08*01 | 26%| TRBV25-01*01 | 21%| TRBV04-01*01 | 6%| CASSQGERNVGRTQYF |
| 30      | PBMC  | wt             |                     | TRBV05-08*01 | 26%| TRBV25-01*01 | 21%| TRBV04-01*01 | 6%| CASSQGERNVGRTQYF |
| 31      | PBMC  | wt             |                     | TRBV05-08*01 | 26%| TRBV25-01*01 | 21%| TRBV04-01*01 | 6%| CASSQGERNVGRTQYF |

STAT5B mutation status, type of the STAT5B mutation and TCR Vβ chain sequences (>5%) with their relative percentage (%) are given for each patient. For patients #17 and #24, two additional clones <5% are reported considering the putative match with the low STAT5B mutation VAF (in italics). The CDR3 amino acid sequences are listed according to the % of the Vβ expansions (from the highest to the lowest). Color codes the cell of origin used in the sequencing: green, CD4+ sorted cells; orange, CD4+CD8+ sorted cells; blue, CD57+ sorted LGLs; yellow, CD56+ sorted LGLs; pink, PBMC. VAF variant allele frequency, wt wild-type, PBMC peripheral blood mononuclear cells.
with GraphPad Prism 6 (ver 6.07) or R (ver 4.0.2) and p < 0.05 was considered significant.

RESULTS

STAT5B mutation status is associated with clinical features in CD4+ T-LGLL

Clinical characteristics of the entire CD4+ T-LGLL cohort (n = 35) are shown in Table 2a. The median age at diagnosis was 72 years (ranging from 41 to 86 years) and no gender bias was observed (ratio Males/Females = 1:0.6). White blood cell (WBC) and LGL counts were overall high (9.4 ± 4 x 10⁹/l and 3.45 ± 2.40 x 10⁶/l, respectively). Absolute neutrophil count (ANC), hemoglobin (Hb), and platelets (PLT) were mostly within the normal range. Half of the patients (49%), hemoglobin (Hb), g/l (range) 140 (123–163) (p = 0.06, respectively; Fig. 2d, e).

|      | CD4 + LGLL n = 35 | CD4+CD8+ | CD4+CD8- | CD56+CD16+ | CD56+CD16- | CD56-CD16+ | CD56-CD16- | ANCs, x10⁹/l | PLTs, x10⁹/l | Hbs, g/l | Platelets, x10⁹/l |
|------|-------------------|----------|----------|------------|------------|------------|------------|--------------|--------------|----------|------------------|
| Females (%) | 17 (49%) | 8 (35%) | 9 (75%) | 0.035 |
| Males (%) | 18 (51%) | 15 (65%) | 3 (25%) |
| Median age (range) | 72 (41–86) | 74 (42–86) | 69 (41–86) | 0.23 |
| Median WBC, x10⁹/l (range) | 9.4 (1.4–20.9) | 10.5 (1.4–20.9) | 6.9 (3.4–11.4) | 0.006 |
| LGL count, x10⁶/l (range) | 3.45 (0.43–8.86) | 4.31 (0.43–8.86) | 2.03 (0.43-6.10) |

Clinical information was collected at diagnosis. Data (a) from the entire cohort of CD4+ T-LGLL patients and data from the same patients (b) split by the presence of somatic STAT5B mutations. For two STAT5B-mutated patients no CD57, CD56, and CD16 expression data was available. Two STAT5B mutated and one wild-type patients presented a double leukemic expansion (CD57+CD56±CD16-) and were associated to CD57+CD56+CD16- group since it was the prevalent clone. Statistically significant p-values (p < 0.05) between the STAT5B mutated and wild-type patients are highlighted in bold. CD4/CD8 and CD57/CD16/CD56/CD16 immunophenotype distributions between the STAT5B mutated and wild-type patients were analyzed using the Chi-square test, whereas sex prevalence was analyzed with Fisher’s exact test. All other clinical features were analyzed with Mann-Whitney U-test.

The patients present a double expansion, in six cases the main clone is CD4+CD8- and the minor one is CD4+CD8+ whereas only one case presents a major clone CD4+CD8- and a minor one CD4–CD8+.

Among the mutated cases (n = 23), 78% of the patients (n = 18) harbored STAT5B mutations with a variant allele frequency (VAF) over 15%, whereas the remaining 22% of cases (n = 5) presented lower VAFs ranging from 1% to 6% (Fig. 1 and Table 1).

Multiple STAT5B mutations were detected in five patients (#9, #11, #12, #18, #29) (Fig. 1), corresponding to 22% of the mutated CD4+ T-LGLL cases. Careful examination of the sequencing reads revealed that in two cases (#12 and #18) the closely located genetic lesions were present in different alleles. For patient #9, amplicon sequencing was performed independently in the two genetic lesions were present in different alleles. For patient #29, the E433K mutation (VAF 34%). Patient #9 harbored P685R, N642H, and E433G mutations with almost identical VAF (47%), thus suggesting that these heterozygous mutations co-occurred in the same patients (Fig. 2d, e). Furthermore, all patients (n = 4) having CD57+CD56+CD16+ phenotype had STAT5B mutation, whereas in patients with CD57+CD56+CD16- phenotype mutation prevalence was 62% (16 of 26 patients).

Identification of multiple and novel STAT5B mutations

Among the mutated cases (n = 23), 78% of the patients (n = 18) harbored STAT5B mutations with a variant allele frequency (VAF) over 15%, whereas the remaining 22% of cases (n = 5) presented lower VAFs ranging from 1% to 6% (Fig. 1 and Table 1).

Multiple STAT5B mutations were detected in five patients (#9, #11, #12, #18, #29) (Fig. 1), corresponding to 22% of the mutated CD4+ T-LGLL cases. Careful examination of the sequencing reads revealed that in two cases (#12 and #18) the closely located genetic lesions were present in different alleles. For patient #9, amplicon sequencing was performed independently in the two genetic lesions were present in different alleles. For patient #29, the E433K mutation (VAF 34%). Patient #9 harbored P685R, N642H, and E433G mutations with almost identical VAF (47%), thus suggesting that these heterozygous mutations co-occurred in the same patients (Fig. 1 and Table 1).

By targeted amplicon sequencing, we discovered that 23 out of 35 patients (66%) harbored somatic STAT5B point mutations (Fig. 1). No differences were detected in the median age between the two groups, whereas STAT5B mutations were more frequent in males (Fisher’s exact test, p = 0.035; Fig. 2a and Table 2b). Patients with STAT5B mutations displayed increased LGL counts and lymphocytosis when compared to the STAT5B wild-type patients (Mann-Whitney test, p = 0.006 for both variables; Fig. 2b, c and Table 1b). There was no statistically significant difference in the
transactivation domain (TAD), both of which are mutational hotspot regions of the \textit{STAT5B} gene. In addition, we identified four additional mutations that are located in the coiled-coil domain (CCD; Q220H, \( n = 1 \)), DNA binding domain (DBD; E433G/K, \( n = 2 \)), and in the inter-domain region (P702A, \( n = 1 \)).

The most frequent mutation was N642H (9/23, 39\%), followed by T628S (5/23, 22\%), S715F (4/23, 17\%) and Y665F (3/23, 13\%). Two different base substitutions led to the same missense T628S mutation: the wild-type nucleotide triplet ACC was found to mutate to either AGC or TCC, both inducing the conversion from tyrosine (T) to serine (S). The first nucleotide alteration has already been listed in COSMIC [32], whereas the latter is novel. Interestingly, the two possible substitutions leading to the same amino acid conversion were both identified in different alleles (i.e., in different reads) of patient #12.

Fig. 2 Most relevant clinical features associated with \textit{STAT5B} mutations. Box and bar plots of the clinical features in \textit{STAT5B} mutated and wild-type patients: \( \text{a} \) Sex, \( \text{b} \) LGL count, \( \text{c} \) WBC, \( \text{d} \) ANC, \( \text{e} \) Hb All clinical features were analyzed with Mann–Whitney \( U \)-test except for sex prevalence which was analyzed with Fisher’s exact test. \textit{WBC} white blood cells, \textit{LGL} large granular lymphocytes, \textit{ANC} absolute neutrophil count, \textit{Hb} hemoglobin, \textit{wt} wild-type.
E433K and V712E are novel activating STAT5B mutations associated with CD4+ T-LGLL

Functional characterization of E433K, V712E, and P685R mutations was performed to determine their impact on the transcriptional activity of the STAT5B gene; known activating variant N642H and the wild-type STAT5B constructs were used as positive and negative controls, respectively (Fig. 3a). Q220H and P702A mutations were excluded from the analysis because of their low VAF (2% and 3%, respectively). T628S mutation with gain-of-function property has already been described before in T-prolymphocytic leukemia (T-PLL) [33].

As shown in Fig. 3a, both E433K and V712E were found to be highly activating when compared to wild-type STAT5B (p < 0.001, two-way ANOVA followed by Dunn’s multiple correction test). There was at least a 10-fold increase in the luminescent signal for the E433K mutation and a 15-fold increase for the V712E mutation when compared to the wild-type STAT5B. The signal from the P685R mutated plasmid did not differ from the wild-type construct. This mutation was detected in the patient with co-existing N642H and E433K mutations. In addition, HeLa cells transiently carrying the STAT5B GOF mutants exhibited upregulated phosphorylation of STAT5B protein compared to the cells carrying WT STAT5B (Fig. 3b). Furthermore, STAT5B mutated cells had increased protein expression levels of c-MYC, Bcl-2, and Pim-1 which are downstream targets of STAT5B [34].

Both the leukemic and non-leukemic TCRβ repertoires are highly clonal in CD4+ T-LGLL

TCRβ sequencing (TCRβ-seq) was performed on samples from CD4+ T-LGLL patients (n = 27) and healthy controls (n = 37). From T-LGLL patients, genomic DNA was isolated from bead-separated cells (CD4+, CD4+CD8+, CD56+ or CD57+) or PMBC depending on sample availability (Table 1). In healthy controls, CD4+ selected fractions were used. As expected, CD4+ T-LGLL patients harbored larger T-cell clones than in healthy controls (Fig. 4a). Both monoclonal and oligoclonal expansion patterns were observed. By focusing on clones with clone size exceeding 5% of the repertoire, we identified a total of 59 possible T-LGLL clones from the 27 patients (Supplementary Table 4).

For the CD4+ sorted samples, the overall TCRβ clonality was higher in CD4+ T-LGLL (n = 6) as compared to healthy controls (n = 37) (Fig. 4b, padj < 0.05, Bonferroni corrected Mann–Whitney test). Even after removing the leukemic T-LGLL clones from the patient samples, the observed difference in the clonality persisted (Fig. 4c, padj < 0.05, Bonferroni corrected Mann–Whitney test). Thus, the non-leukemic TCR repertoires in CD4+ T-LGLL patients are less diverse than the CD4+ TCR repertoires in the healthy controls. By repeating the analysis with varying thresholds of 3–7% for selecting the putative T-LGLL clones, we showed that this was invariant to our chosen threshold of 5% (Supplementary Fig. 1).

We also compared a similarly profiled CD8+ T-LGLL cohort [35] (n = 9) with the CD4+ T-LGLL cohort. Surprisingly, the overall clonality in CD4+ T-LGLL samples was higher than in the CD8+ T-LGLL samples (Fig. 4d, padj < 0.05, Bonferroni corrected Mann–Whitney test), although it is known that CD8+ TCR repertoires are more clonal than CD4+ repertoires [36]. This was however not seen in the non-leukemic repertoires, where the non-leukemic compartments from CD8+ T-LGLL samples had higher clonality than the non-leukemic compartments from CD4+ T-LGLL samples (Fig. 4e, padj < 0.05, Bonferroni corrected Mann–Whitney test).

Most CD4+ T-LGLL-associated STAT5B mutations are clonally restricted

For most of the patients (71%) presenting with a STAT5B mutated clone over 15%, the VAF of the STAT5B mutation corresponded to the size of the main T-LGLL clone: the mutation VAFs were half of the size of the TCR clones suggesting heterozygous mutations (Table 1). Only in patients #2 and #15, the mutation VAFs were slightly higher, suggesting that there was more than one T-cell clone harboring the same mutation. In patient #21 the mutation VAF (23%) matched with the combined size of the two main T-LGLL clones (24% and 22%), suggesting that the mutation is present either in one clone in homozygous/hemizygous fashion or in both clones as heterozygous (Table 1). Interestingly, patients with low VAF (<5%) STAT5B mutations (#8, #13, #17, and #24) had nevertheless large T-LGLL clones (37–71%) suggesting that they had either subclonal STAT5B mutations or the mutations were located outside the main clone (Table 1).

In patients with multiple STAT5B mutations, the clonal structure was evaluated based on the VAF and TCR clone size. In patient #9, the three mutations co-existed in the same clone based on their similar VAFs (47%) matching with the size of the main clone (97%) (Table 1). In patient #18, the Y665F (VAF 34%) and N642H (VAF 5%) mutations were detected in different subclones, and accordingly, also two expanded TCR clonotypes were detected (Table 1). Patient #11 had one large main TCR clone (90%) but two different
CD4+ CD8+

PBMC

n

clonality index from CD4

CD8

types (Supplementary Table 4) could be a result of convergent

37).

c

purified

sum of the rest (clones 4–10) is shown in one color. The cell isolation method (CD4, CD4 CD8, CD57, CD56, or non-selected PBMC) of the analyzed
cells from healthy controls (Fig.5b). We further looked for the presence of the CD4

repertoire between the CD4

β

repertoires between the CD4

+ T−LGLL clonotypes with a size < 5%) as normalized Shannon

Wiener clonality index from CD4+ T-LGLL (patient samples purified for CD4, n = 6) and CD4+ cells from healthy controls (n = 37). c Non-leukemic CD4+ repertoire clonality (selecting the CD4+ T-LGLL clones with a size < 5%) as normalized Shannon–Wiener clonality index from CD4+ T-LGLL (patient samples purified for CD4, n = 6) and CD4+ cells from healthy controls (n = 37). d Total repertoire clonality as normalized

Shannon–Wiener clonality index from CD4+ T-LGLL (patient samples purified for CD4, n = 6) and CD8+ cells from CD8+ T-LGLL (patient samples purified for CD8, n = 9). e Non-leukemic repertoire clonality (selecting the CD4+ T-LGLL clones with a size < 5%) as normalized Shannon–Wiener clonality index from CD4+ T-LGLL (patient samples purified for CD4, n = 6) and CD8+ cells from CD8+ T-LGLL (patient samples purified for CD8, n = 9). All p-values were calculated with a Mann–Whitney test and corrected with Bonferroni.

STAT5B mutations with different VAFs (Y665F: 32% and V712E: 11%) (Fig. 1 and Table 1). This suggests possible clonal evolution and the presence of different genetic lesions in individual cells bearing the same TCR.

**CD4+ T-LGLL clonotypes are predominantly restricted to individual patients**

To understand whether the 59 identified CD4+ T-LGLL clono-
types (Supplementary Table 4) could be a result of convergent

evolution, we examined if these clones were shared between patients. In concordance with what has been reported before in

CD8+ T-LGLL [14], none of the patients had the same TCR

rearrangement in their main leukemic CD4+ T-LGLL clonotypes (Fig. 5a). However, 25% (15/59) of T-LGLL clonotypes were found

in the non-leukemic TCR repertoires of other patients, albeit in smaller frequencies (ranging between frequencies of 1.41 × 10−5–1.27 × 10−2, Supplementary Table 5, Fig. 5a). Of these 59 clones, 27% (16/59) were found in healthy CD4+ sorted TCR repertoire (n = 37), again in much smaller frequencies (ranging between 8.24 × 10−5–2.59 × 10−5) (Supplementary Table 5, Fig. 5a). We further looked for the presence of the CD4+ T-LGLL clonotypes in a larger publicly available cohort (n = 786) of healthy donors’ unsorted TCR repertoire from Emerson et al. [29]. We observed that 36/59 of the patient clones (around 61%) were found at least once in the healthy donors with the most common clone being found in 163 healthy donors (Fig. 5b). However, most of these clones were found in extremely rare frequencies (ranging between 8.2 × 10−5 and 1.05 × 10−4).
Fig. 5  Publicness of the CD4+ T-LGLL clones. a CD4+ T-LGLL clonotypes are shared between CD4+ T-LGLL patients (n = 27) and healthy cohort (n = 37) in extremely rare frequencies. Shared clonotypes are at the bottommost rows and the colors refer to the size of the clonotype. b CD4+ T-LGLL clonotypes found in smaller frequencies in publicly available, unsorted, healthy cohort data (n = 786) [29] where the colors refer to the size of the clonotype.
We performed a similar analysis with our CD8+ T-LGLL cohort. None of the CD8+ leukemic clones (defined similarly as clones above 5%) was shared between the patients (Supplementary Fig. 2a). 50% of these clones were found in the Emerson et al. [29] cohort in low frequencies (Supplementary Fig. 2b).

Next, we calculated the generation probabilities of the leukemic TCRs with OLGA [25] as high generation probabilities could indicate a bias in the recombination and low generation probabilities towards convergent evolution. The TCR generation probabilities were higher in clones that were shared between healthy individuals. Conversely, the generation probabilities in clones that were exclusive to CD4+ T-LGLL were low (Supplementary Fig. 3).

Skewed Vβ gene usage in CD4+ T-LGLL clones

Previous studies using Vβ flow cytometry analysis have suggested a biased usage of the Vβ13.1 family (consisting of TRBV06-05, TRBV06-06, and TRBV06-09 genes) in CD4+ T-LGLL [16]. In our data, the predominant V-gene of CD4+ T-LGLL clonotypes belonged to the TRBV06 family in 20% (12/59) of CD4+ T-LGLL clonotypes (Supplementary Table 6, Supplementary Fig. 4). By selecting the single most expanded clones, we found that 8 out of 27 T-LGLL clones belonged to the TRBV06 family, in comparison to only 2 out of the 35 clones belonging to the TRBV06 family in healthy (Fig. 6a, p = 0.01, Fisher’s two-sided exact test). Representative examples from CD4+ T-LGLL, with both monoclonal and polyclonal expansions, and healthy are shown in Fig. 6b.
To test whether the CD4+ T-LGLL clones target a known viral antigen (from e.g., CMV, EBV, or Influenza A), we queried the 59 T-LGLL clones against the largest database containing known antigen specificities (VDJdb [24]) but no exact matches for TCRs recognizing epitope bound to class II HLAs were found. The antigen specificities of the non-leukemic TCRs returned multiple matches, with Influenza A as the most common target (Supplementary Fig. 5a).

Since structurally similar TCRs are known to recognize similar antigens [23, 31, 37], we used GLIPH to identify significantly enriched motifs within CD4+ T-LGLL clonotypes in comparison to healthy naive repertoires. Interestingly, we identified two conserved motifs within the 59 T-LGLL clonotypes. The first one, ”SDP”, in three T-LGLL clones all of which shared the TRBV24 gene, and the second one, ”SLRG”, in three other T-LGLL clones with different V genes (Supplementary Table 7). No correlation was observed between motifs and the mutational statuses of the patients. Next, we analyzed whether the T-LGLL clonotypes shared motifs within their repertoire, i.e., whether the T-LGLL responses could be polyclonal antigen-specific responses. In 33% (2/6) of CD4+ sorted patients’ samples, we could identify a motif that was shared between leukemic and non-leukemic repertoires. Structurally similar TCRs from patient #18 are shown in Supplementary Fig. 5b, where the leukemic clone can be seen to share similarities with the multiple non-leukemic TCRs.

Although GLIPH does not consider motifs less than three amino acids long, Garrido et al. have previously reported the presence of the “QG” motif in the middle of the CDR3 region in patients with monoclonal Vβ13.1 expansion and HLA-DRB1*07:01 genotype [16]. We found this “QG” motif in 18.6% (11/59) of T-LGLL clones (of which only two expressed Vβ13.1), but also in 13.4% (20/144) of top 5 expanded clonotypes in healthy CD4+ repertoire [Fig. 6c, Supplementary Table 4, p = 0.39, Fisher’s two-sided exact test]. To address whether “QG” is specific to CD4+ T-LGLL or specific to HLA-DRB1*07:01 in general, we analyzed whether “QG” was more common in CD4+ clones of HLA-DRB1*07:01 positive (n = 5) than negative (n = 32) healthy donors. We found that “QG” was enriched in HLA-DRB1*07:01 positive healthy donors (3.5% of CD4+ clones) in comparison to negative donors (0.54% of CD4+ clones; p = 7.53 × 10^-122, Fisher’s one-sided exact test), further suggesting its association with CD4+ cells and HLA-DRB1*07:01 in general and not specifically to CD4+ T-LGLL.

DISCUSSION
Somatic STAT5B mutations were initially described in CD8+ T-LGLL [9], but follow-up analyses have revealed that they more often occur in CD4+ and immature T-cell malignancies [4, 33, 38–42]. Here, we performed a detailed molecular characterization of 35 CD4+ T-LGLL patients, which is the largest cohort described so far. We identified novel gain of function STAT5B mutations and found that leukemic clonotypes are predominantly private to the patients.

At least one STAT5B mutation was detected in 66% of the 35 CD4+ T-LGLL patients. This is the highest frequency of STAT5B mutations thus far reported in the literature. While the previously published studies have mostly focused on the mutational hotspot regions of STAT5B [5, 9], our sequencing assay covered the whole protein-coding sequence of STAT5B. Moreover, the use of a highly sensitive method allowed us to detect low-frequency mutations down to 1% VAF.

Functional analyses of the newly discovered STAT5B mutations confirmed STAT5 pathway activation, as the downstream targets of STAT5 (c-MYC, Bcl-2, and Pim-1 proteins) [34] were over-expressed. Similarly, a previous study has shown that STAT5B mutant-transduced cell lines upregulate Bcl-2 [43]. c-MYC is an essential regulator of cell proliferation and survival [34], and Pim-1, an anti-apoptotic gene, is known to synergize with c-MYC in leukemogenesis [44]. Altogether, this data suggests that the STAT5B GOF mutations are likely to be involved in clonal dominance and maintenance, probably by conferring proliferation advantage for the mutated cells. Accordingly, STAT5B mutated CD4+ T-LGLL patients had higher lymphocyte and LGL counts compared to wild-type patients.

In addition to the earlier described N642H, S715F, and Y665F mutations, we discovered the T628S mutation to be common in CD4+ T-LGLL (22% of our cohort). This mutation has been previously reported as a highly activating mutation in T-PLL [33, 45] and hepatosplenic T-cell lymphoma (HSTCL) [42]. Although SH2 and TAD domains are the known mutation hotspot area in the STAT5B gene, we also identified novel variants in the CCD (Q220H) and DBD (E433K/G) of STAT5B. Similarly, the majority of the STAT3 mutations are in the SH2 domain in CD8+ T-LGLL but additional mutations are also found in other domains [4] underlining the importance of covering all mutation hotspots in the diagnostic tests of TLL.

Among the STAT5B mutated CD4+ T-LGLL patients, 22% harbored multiple STAT5B mutations either in the main T-LGLL clone or in separate smaller clones. Interestingly, mutation VAFs did not always correlate with the T-cell clone size based on TCRβ sequencing, suggesting sub-clonal heterogeneity. Similarly, 17% of CD8+ T-LGLL patients have been reported to harbor multiple STAT3 mutations [46]. Our results are consistent with previous data, suggesting that in T-LGLL additional somatic mutations can arise in the pre-expanded clonotypes [47].

In CD8+ T-LGLL, STAT3 mutations have been associated with autoimmune manifestations such as rheumatoid arthritis and neutropenia [2, 5, 8, 48, 49]. In CD4+ T-LGLL, neutropenia is uncommon, and similarly, in our cohort STAT5B mutated patients had normal ANC levels but increased lymphocyte and LGL counts. Irrespective of the mutation status, the phenotype of the CD4+ T-LGLL clone was in most (75%) cases CD16–CD56+. Although CD16+CD56+ phenotype was detected in only four cases, all these had mutated STAT5B. CD56+ immunophenotype has also been associated with STAT5B mutations in CD8+ T-LGLL [9] in which the LGLs are usually CD56-.

TCRB-seq has provided new insights into the nature of clonal expansions in CD8+ T-LGLL [14] but has not been performed in CD4+ T-LGLL earlier. In accordance with the previous flow cytometry-based data, in CD4+ T-LGLL the TRBV06 family was enriched in the most hyper-expanded clones. Interestingly, 25% of T-LGLL clonotypes were also found among the non-expanded TCR receptor in CD4+ T-LGLL patients and 27% were found in healthy individuals. This was not observed in the CD8+ T-LGLL disease, consistent with previous results [14]. Although our data was not HLA-matched, this could hint that the eliciting antigen in CD4+-T-LGLL could be commonly encountered. Further, in CD4+ T-LGLL also the non-leukemic T-cell compartment was significantly more clonal than in healthy controls. It has been shown previously, that CD4+ T cells in healthy controls are much richer in their diversity as compared to CD8+ T cells [50]. Hence, it was interesting to note that the overall clonality was higher in CD4+ T-LGLL compared to CD8+ T-LGLL, but this could also be due to the relatively small sample size. However, the eliciting antigen(s) remain unknown, as no exact matches were found for the known antigen-specific TCRs, which could be due to the small amount of known class II pMHC–TCR pairs and potential bias in the HLA genotypes of our samples. Additionally, since previous literature has indicated that a significant proportion of CD4+ T-LGLL patients has an underlying disease or malignancy [3], it cannot be excluded that the triggering event is abnormal antigen stimulation such as tumor antigen.

More research is warranted to clarify whether CD4+ T-LGLL is a true hematological malignancy or a reactive disorder of the immune system as clonal T cell proliferations and somatic mutations are also observed in healthy controls and non-
malignant diseases [26, 51]. Following the initial event, activating STAT5B mutations may boost the aberrant proliferation and clonal persistence. Despite the STAT5B hyperactivation and the presence of multiple STAT5B mutations, the clinical course of CD4+ T-LGLL is indolent, and patients rarely have symptoms, differentiating it from CD8+ T-LGLL and other mature and immature CD4+ T-cell leukemias and lymphomas.

DATA AVAILABILITY
TCRB sequencing data is freely available at https://clients.adaptivebiotech.com/pub/bhattacharya-2022-BJCI (https://doi.org/10.21417/D82022B8C).

REFERENCES
1. Matutes E. The 2017 WHO update on mature T- and natural killer (NK) cell neoplasms. Int J Lab Hematol. 2018;40:97–103.
2. Kuwahara N, Kodaka T, Zushi Y, Sasaki M, Goka T, Maruoka H, et al. T-cell large granular lymphocytic (LGL) leukemia consists of CD4/CD8 and CD4/CD8 LGL populations in association with immune thrombocytopenia, autoimmune neuropaenia, and monoclonal B-cell lymphocytosis. J Clin Exp Hematop. 2019;59:202–6.
3. Lima M, Almeida J, dos Anjos Teixeira M, del Carmen Alguero M, Santos AH, Balanzategui A, et al. TCRβTCRBβ+CD4-positive T cells. Blood. 2016;128:2465–8.
4. Barilà G, Teramo A, Calabretto G, Vicenzetto C, Gasparini VR, Pavan L, et al. Role of STAT3 and STAT5B mutations in the genomics of T-cell LGL leukemia. Blood. 2009;114:4999–107.
5. Barilà G, Teramo A, Calabretto G, Vicenzetto C, Gasparini VR, Semenzato G, et al. Structural and functional consequences of the STAT5B driver mutation in chronic T-cell lymphocytic leukemia. J Exp Med. 2007;206:1729–43.
6. Glavine C, Juan H, Nau A, Hatton O, Wage L, Rubelt F, et al. Identifying specific groups in the T cell receptor repertoire. Nature 2017;547:94–98.
7. López C, Bergmann AK, Paul U, Murga Penas EM, Nagel I, Betts MJ, et al. Genes encoding members of the JAK-STAT pathway or epigenetic regulators are recurrently mutated in T-cell prolymphocytic leukemia. Br J Haematol. 2016;173:265–73.
8. De Araujo ED, Erdogan F, Neubauer HA, Meneksedag-Erol D, Manaswiyoungkul P, Erber ER, et al. Lack of common TCRA and TCRB clonotypes in CD8(+) T-LGL. Biol Haematol. 2007;109:4890–904.
9. D'Avanzo D, Renard B, De Vos P, van der Ploeg L, van de Waterbeemd H, et al. The T-cell receptor repertoire in non-Hodgkin lymphoma: a genome-wide antigen screening. Nat Biotechnol. 2020;38:1194–202.
10. Barbieri DJ, Vroomans R, Huh W, Svoboda U, Rius C, Dolton G, et al. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. Nucleic Acids Res. 2020;48:D1057–D1062.
11. Setyna Z, Elhanati Y, Callan CG, Walczak AM, Mora T. OLGA: fast computation of lymphocyte repertoire motifs. Cytom B Clin Cytom. 2020;87:145–158.
12. Kim D, Park G, Huhiwan J, Lundigren S, Khajuria RK, Hurtado AM, et al. Somatic mTOR mutation in clonally expanded T lymphocytes associated with chronic graft versus host disease. Nat Commun. 2020;11:2246.
13. Rodriguez-Caballero A, García-Montero AC, Bárcena P, Almeida J, Ruiz-Cabello F, et al. Discovery of somatic STAT5b mutations in large granular lymphocytic leukemia. Blood. 2012;120:971–8.
14. Clemente MJ, Przychodzen B, Jerez A, Dienes BE, Afable MG, Husseinzadeh H, et al. Structural and functional consequences of the STAT5B driver mutation impacts biological and clinical features of T-LGL leukemia. Oncotarget. 2015;6:46616–24.
15. Kontro M, Kuusanmäki H, Eldors H, Burmeister T, Andersson EI, Bruserud O, et al. Somatic STAT5B mutations are significantly associated with c-myc and N-myc in murine leukemia virus-induced tumors. Cell. 1989;56:673–8.
16. Byrne KJ, Dalgleish AG. Chronic immune activation and inflammation as the cause of malignancy. Br J Cancer. 2001;85:473–83.
17. Cesarman E, Mesri EA. Kaposi sarcoma-associated herpesvirus and other viruses in human lymphomagenesis. Curr Top Microbiol Immunol. 2007;312:263–87.
18. Reza SK, Weiss LM. Epstein-Barr virus-associated lymphoproliferative disorders. Hum Pathol. 2007;38:1293–304.
19. D’Avanzo D, Renard B, De Vos P, van der Ploeg L, van de Waterbeemd H, et al. The T-cell receptor repertoire in non-Hodgkin lymphoma: a genome-wide antigen screening. Nat Biotechnol. 2020;38:1194–202.
20. Barbieri DJ, Vroomans R, Huh W, Svoboda U, Rius C, Dolton G, et al. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. Nucleic Acids Res. 2020;48:D1057–D1062.
21. D. Bhattacharya et al.
46. Rajala HLM, Olson T, Clemente MJ, Lagström S, Ellonen P, Lundan T, et al. The analysis of clonal diversity and therapy responses using STAT3 mutations as a molecular marker in large granular lymphocytic leukemia. Haematologica. 2015;100:91–99.

47. Kerr CM, Clemente MJ, Chomczynski PW, Przychodzen B, Nagata Y, Adema V, et al. Subclonal STAT3 mutations solidify clonal dominance. Blood Adv. 2019;3:917–21.

48. Mariotti B, Calabretto G, Rossato M, Teramo A, Castellucci M, Barilà G, et al. Identification of a -Fas ligand axis in the development of neutropenia in T large granular lymphocyte leukemia. Haematologica. 2020;105:1351–60.

49. Muñoz-García N, Jara-Acevedo M, Caldas C, Bárcena P, López A, Puig N, et al. Mutations in T/NK-cell chronic lymphoproliferative disorders of large granular lymphocytes (LGL) association with disease features. Cancers. 2020;12:3508.

50. Li HM, Hiroi T, Zhang Y, Shi A, Chen G, De S, et al. TCRβ repertoire of CD4+ and CD8+ T cells is distinct in richness, distribution, and CDR3 amino acid composition. J Leukoc Biol. 2016;99:505–13.

51. Mustjoki S, Young NS. Somatic mutations in ‘benign’ disease. N Engl J Med. 2021;384:2039–52.

ACKNOWLEDGEMENTS

The research was funded by the European Research Council (Project: M-IMM 647355), Academy of Finland Heal-Art consortium (314442) and ERA PerMed (JAKSTAT-TARGET consortium), Sigrid Juselius Foundation, Signe and Ane Gyllenberg Foundation, Helsinki Institute for Life Science, and Cancer Foundation Finland to SM; Italian Association for Cancer Research (AIRC #20216) to GS; Società Italiana Ematologia Sperimentale (SIES) fellowship to AT. Amplicon sequencing was performed at the Institute for Molecular Medicine Finland FIMM Technology Centre, which is supported by Biocenter Finland. We acknowledge IT Center for Science Ltd and the Aalto Science-IT project for computational resources and Finnish Red Cross Blood Service for providing samples from healthy donors. We acknowledge Emmi Rehn for providing a modified script of GLIPH that was used for the previous version of the findings. We thank Jay Klievink for technical expertise in mutagenesis assays.

AUTHOR CONTRIBUTIONS

DB, AT, VRG, TKe, FI, RZ, and SM conceived the study; DB, AT, VRG, GSc, CV, GC, TKa, HN, BF, and ET prepared the samples and performed sequencings; DB, DK and VRG performed functional assays; DB, JH and JT performed bioinformatic analyses; GB, MF, FI, GSe, RZ and SM provided the patient samples; DB, JH, AT, and VRG wrote the paper; GSe and SM provided funding; TKe, GSe, RZ and SM critically revised the paper. All authors revised the paper.

COMPETING INTERESTS

SM has received honoraria and research funding from Novartis, Bristol-Myers Squibb, Incyte, and Pfizer. GSe has received honoraria from Novartis and Roche. Others report no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41408-022-00630-8.

Correspondence and requests for materials should be addressed to Satu Mustjoki.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022