Inborn errors of immunity underlie clonal T cell expansions in large granular lymphocyte leukemia

Carlos Bravo-Perez,[1,2] Carmelo Gurnari,[1,3] Jani Huuhtanen,[4,5,6,7] Naomi Kawashima,[1] Luca Guarnera,[1,3] Aashray Mandala,[1] Nakisha D. Williams,[1] Christopher Haddad,[1] Michaela Witt,[1] Serhan Unlu,[1] Zachary Brady,[1] Olisaemeka Ogbue,[1] Mark Orland,[1] Arooj Ahmed,[1] Yasuo Kubota,[1] Simona Pagliuca,[1,8] Arda Durmaz,[1] Satu Mustjoki,[4,5,6] Valeria Visconte,[1] and Jaroslaw P. Maciejewski[1]

[1]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio, USA. [2]Department of Hematology and Medical Oncology, Hospital Universitario Morales Meseguer, University of Murcia, IMIB-Pascual Parrilla, CIBERER–Instituto de Salud Carlos III, Murcia, Spain. [3]Department of Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy. [4]Hematology Research Unit Helsinki, Department of Hematology, University of Helsinki and Helsinki University Hospital Comprehensive Cancer Center, Helsinki, Finland. [5]Translational Immunology Research Program, University of Helsinki, Helsinki, Finland. [6]ICAN Digital Precision Cancer Medicine Flagship, University of Helsinki and Helsinki University Hospital Comprehensive Cancer Center, Helsinki, Finland. [7]Department of Computer Science, Aalto University, Espoo, Finland. [8]Hematology Department, Nancy University Hospital and UMR 7365 CNRS University of Lorraine, France.

BACKGROUND. T cell large granular lymphocyte leukemia (T-LGLL) is a lymphoproliferative disorder of cytotoxic T lymphocytes (CTLs), often with gain-of-function *STAT3* mutations. T-LGLL represents a unique model for the study of persistent CTL expansions. Albeit autoimmunity is implied, various paradoxical observations led us to investigate whether immunodeficiency traits underpin T-LGLL.

METHODS. This is a comprehensive immunogenomic study of 92 consecutive patients from a large T-LGLL cohort with full laboratory-clinical characterization (n = 271). Whole-exome profiling of variants associated with inborn errors of immunity (IEI) and somatic mutations in T cell lymphoid drivers was analyzed. Single-cell RNA-Seq and TCR-Seq in T-LGLL samples and RNA-Seq in T cell cancer cell lines were utilized to establish biological correlations.

RESULTS. Lymphocytopenia and/or hypogammaglobulinemia were identified in 186 of 241 (77%) T-LGLL patients. Genetic screening for IEI revealed 43 rare heterozygous variants in 38 different immune genes in 34 of 92 (36%) patients (vs. 167/63,026 [0.26%] in controls). High-confidence deleterious variants associated with dominant, adult-onset IEIs were detected in 15 of 92 (16%) patients. Carriers showed atypical features otherwise tied to the cryptic IEI, such as earlier onset, lower lymphocyte counts, lower *STAT3* mutational rate, and higher proportions of hypogammaglobulinemia and immune cytopenia/bone marrow failure than noncarriers. Somatic mutational landscape, RNA-Seq, and TCR-Seq analyses supported immune imbalance caused by the IEI variants and interactions with somatic mutations in T cell lymphoid drivers.

CONCLUSIONS. Our findings in T-LGLL reveal that maladaptive CTL expansions may stem from cryptic immunodeficiency traits and open the horizon of IEIs to clonal hematopoiesis and bone marrow failure.

FUNDING. NIH; Aplastic Anemia and MDS International Foundation; VeloSano; Edward P. Evans Foundation; Instituto de Salud Carlos III; European Research Council; European Research Area Network on Personalised Medicine; Academy Finland; Cancer Foundation Finland.

Introduction

T cell large granular lymphocyte leukemia (T-LGLL) is a chronic lymphoproliferative disorder of cytotoxic T lymphocytes (CTLs) (1). Common in the elderly, the disease spans a continuum from reactive poly/oligoclonal responses to clonally skewed CTLs. During its natural course, patients eventually develop severe cytopenias or paraneoplastic autoimmune complications requiring therapy (2). Gain-of-function (GOF) *STAT3* mutations (*STAT3mt*) occur in 40%–50% of T-LGLL patients and normally imply a more severe clinical phenotype. However, T cell large granular lymphocyte (T-LGL) leukemic clones normally show constant JAK/STAT activation irrespective of the *STAT3*mt status. Clinical and laboratory advances have also revealed the dysregulation of alternative survival pathways in patients with *STAT3* wild-type (*STAT3*wt) T-LGLL (e.g., MCL1, NF-κB, PI3K/AKT) (3–6). Mutations in driver genes other than *STAT3*, recurrent in mature T cell neoplasms of other subtypes, have been reported, such as in *STAT5B*, *TNFAIP3*, and the epigenetic regulators *KMT2D* or *TET2* (7).

The precise pathogenesis of T-LGLL remains ill defined. However, its cooccurrence with different autoimmune, neoplastic, or infectious conditions is well known (8). Chronic antigenic stim-

Conflict of interest: The authors have declared that no conflict of interest exists.
Copyright: © 2025, Bravo-Perez et al. This is an open access article published under the terms of the Creative Commons Attribution 4.0 International License.
Submitted: July 1, 2024; **Accepted:** February 27, 2025; **Published:** March 4, 2025.
Reference information: *J Clin Invest*. 2025;135(9):e184431.
https://doi.org/10.1172/JCI184431.

Table 1. Clinical and laboratory features of the overall and genomic cohorts

	Overall cohort	Genomic subcohort
Clinical features	*n* = 241	*n* = 92
Sex (female), *n* (%)	107 (44.4)	41 (44.6)
Sex (male), *n* (%)	134 (55.6)	51 (55.4)
Age (years), median (IQR)	63 (55–72)	62 (54–71)
Autoimmunity, *n* (%)	106 (44.0)	51 (55.4)
Rheumatoid arthritis, *n* (%)	41 (17.0)	21 (22.8)
B cell dyscrasia, *n* (%)[A]	83 (34.4)	32 (34.7)
B cell lymphoma/CLPD, *n* (%)	32 (13.3)	16 (17.4)
Plasma cell dyscrasia, *n* (%)	64 (29.8)	23 (25.0)
Splenomegaly, *n* (%)	52 (21.6)	29 (31.5)
Mutational status	*n* = 233	*n* = 92
STAT3 mutation, *n* (%)	82 (35.2)	42 (45.6)
STAT5B mutation, *n* (%)	3 (1.2)	1 (1.1)
Treatment	*n* = 241	*n* = 92
Need of T-LGLL treatment, *n* (%)	141 (58.5)	54 (65.2)
Overall response rate, *n* (%)	47/91 (54.0)	23/45 (51.1)
Follow-up (mo), median (IQR)	64 (28–125)	98 (35–153)
Blood counts and cytopenias	*n* = 241	*n* = 92
HGB (g/dL), median (IQR)	12.4 (9.9–13.5)	12.2 (10.1–13.6)
WBC ($\times 10^9$/L), median (IQR)	5.3 (3.0–10.0)	4.1 (3.1–8.9)
ANC ($\times 10^9$/L), median (IQR)	1.5 (0.6–2.9)	1.2 (0.4–2.4)
ALC ($\times 10^9$/L), median (IQR)	2.9 (1.5–5.4)	2.9 (1.7–5.5)
LGL count ($\times 10^9$/L), median (IQR)	1.3 (0.7–2.7)	1.4 (0.6–2.7)
Lymphocytosis, *n* (%)	83 (34.4)	33 (35.9)
Cytopenias, *n* (%)	192 (79.7)	74 (80.4)
Anemia, *n* (%)	135 (57.5)	51 (55.4)
Leukopenia, *n* (%)	87 (37.3)	41 (44.6)
Neutropenia, *n* (%)	128 (54.9)	52 (56.5)
Immune cytopenia, *n* (%)[B]	58 (24.1)	23 (25.0)
BMF	28 (11.6)	11 (12.0)
Antibody-mediated peripheral cytopenia, *n* (%)	33 (13.7)	13 (14.1)
Blood lymphocyte subsets	*n* = 216	*n* = 78
T cells (%), median (IQR)	91.0 (84.0–96.0)	90.5 (83.8–96.0)
CD4:CD8 ratio, median (IQR)	0.4 (0.2–0.8)	0.4 (0.2–0.8)
NK cells (%), median (IQR)	3.0 (1.0–6.5)	3.0 (1.0–6.5)
B cells (%), median (IQR)	3.0 (1.0–7.0)	3.0 (1.0–7.0)
Serum Ig levels	*n* = 201	*n* = 78
IgG (mg/dL), median (IQR)	1170 (885–1550)	1200 (968–1520)
IgA (mg/dL), median (IQR)	212 (114–374)	220 (142–382)
IgM (mg/dL), median (IQR)	113 (65–117)	133 (79–216)

[A]B cell lymphoma, monoclonal B cell lymphocytosis/chronic lymphocytic leukemia, plasma-cell dyscrasias. [B]Immune cytopenias: BMF (i.e., AA or PRCA) and antibody-mediated peripheral cytopenia (AIHA, AIN, ITP). AA, aplastic anemia; AIHA, autoimmune hemolytic anemia; AIN, autoimmune neutropenia; CLPD, chronic lymphoproliferative disease; HGB, hemoglobin; ITP, immune thrombocytopenic purpura; LGL, large granular lymphocytes.

ulation has been hypothesized to initiate/drive CTL outgrowth in T-LGLL mostly based on these supporting clinical associations. More recently, laboratory evidence suggests the role of the immune microenvironment and its crosstalk with the T-LGL clone via elevated levels of proinflammatory cytokines (e.g., IL-15, TNF-α, IL-6, IFN-γ) (9, 10). For instance, high levels of IL-15 induce T-LGL lymphoproliferative disease in transgenic mice (10, 11), and blockade with anti–IL-15 in clinical trials has shown partial restoration of the resistance of T-LGL cells to apoptosis (12). Accordingly, T-LGLL appears to correspond with an overshooting immune reaction in which immunodominant clones, cumulatively fueled by chronic (auto)antigen exposure, and environmental and/or genetic growth signals may directly act as effector CTLs or cause cytokine-mediated manifestations (13, 14). Thus, despite being extremely rare, T-LGLL represents a unique disease model, at the intersection of a physiological immune response, autoimmunity, and malignancy.

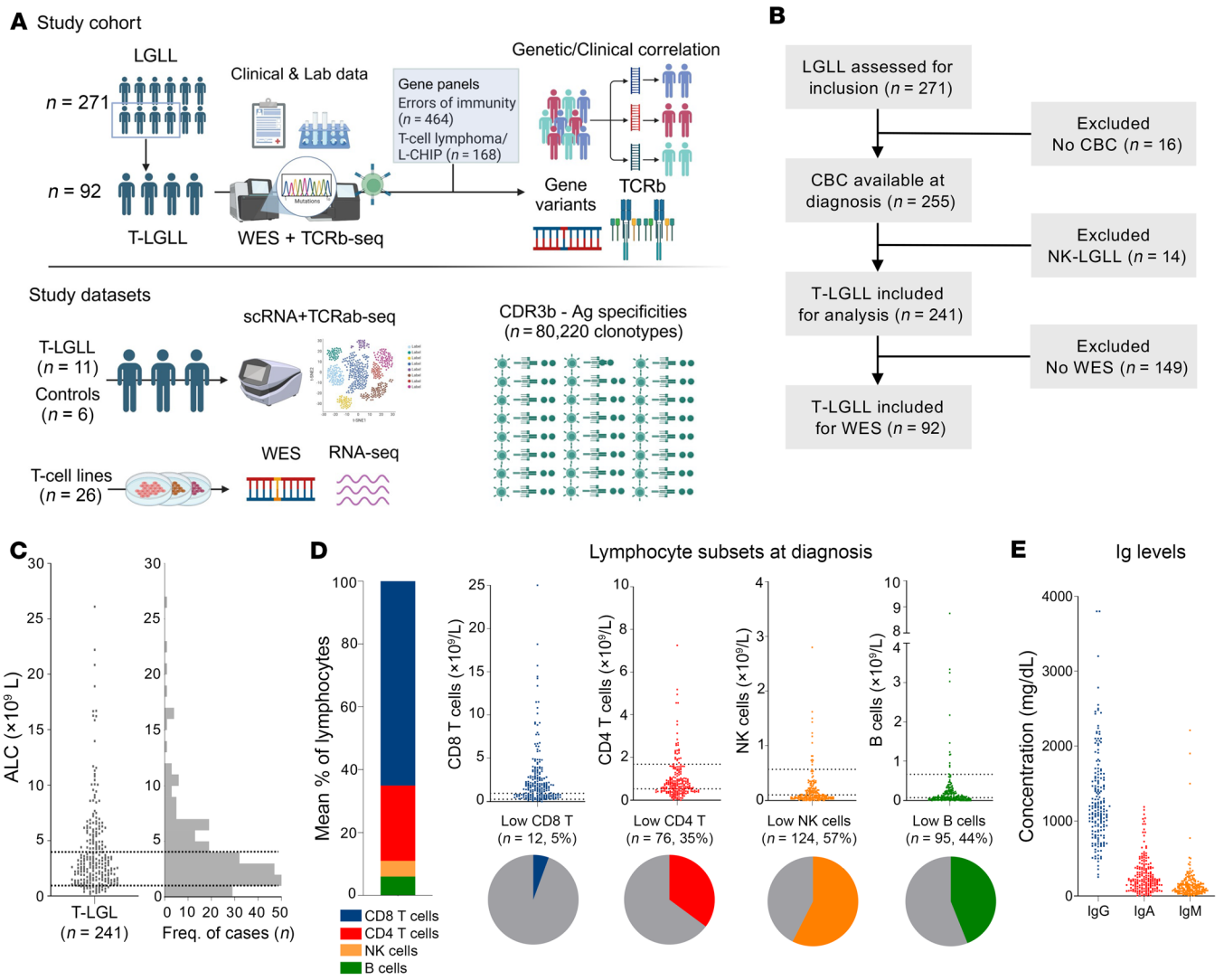
Our ongoing study of clinical features of T-LGLL suggests that, despite indications of a hyperreactive immune response, a paradoxically high frequency of immunodeficiency signs might be present at diagnosis. For instance, despite fulfilling the diagnostic criteria, the disease is not always characterized by lymphocytosis as expected from a frank leukemia, but may be associated with relatively low/normal lymphocyte counts. Moreover, hypogammaglobulinemia is also frequently encountered. We previously described hypogammaglobulinemia in T-LGLL among patients with a history of B cell dyscrasia or transplant (15, 16). We also demonstrated obvious or occult T-LGLL as a novel feature of Good syndrome (17). In patients with common variable immunodeficiency (CVID), a high rate of somatic mutations and clonality in T cells has been reported, suggesting a propensity to clonal outgrowth (18).

While admittedly development of T-LGLL is a complex process, the initial overshooting immune response might be specifically triggered in genetically predisposed subjects. We hypothesized that certain errors of immunity may induce maladaptive CTL responses, ultimately leading to the clonal shift. In this context, we aimed to systematically assess the acquired and genetic immunodeficiency underpinnings of T-LGLL.

Results

Clinical dissection of immune defects in a large T-LGLL cohort. The study was conducted on a cohort of 271 consecutive T-LGLL patients with full annotations diagnosed at Cleveland Clinic (1998-2023; Figure 1A and Table 1). Because of our focus on T cell clonal expansions, cases with NK-cell LGLL subtype were excluded. The initial analysis of clinical features confirmed a high frequency of different immune defects at T-LGLL diagnosis (186 [77%] of 241 patients). Of 241 cases with blood counts at presentation, 169 (66%) had low/normal absolute lymphocyte counts (≤4.0×10^9/L; Figure 1, B and C). Blood lymphocyte subset immunophenotyping in 216 patients revealed reduced levels of $CD4^+$ T cells in 76 (35%), low NK cells in 124 (58%), and low B cells in 95 (44%) cases (Figure 1D). In addition, 64 (32%) of 201 patients with Ig levels had hypogammaglobulinemia, including 29 (14%), 32 (16%), and 36 (18%) of 201 patients with low IgG, IgA, and IgM, respectively (Figure 1, E and F). Clinical history indicated that 39 (61%) of 64 cases with hypogammaglobulinemia could have been acquired, given the prior diagnosis of hemato-lymphoid neoplasms, anti–B cell or other immunosuppressive treatments, and Good syndrome or related CVID-like conditions. However, in 25 (39%) of 64 hypogammaglobulinemic patients, no apparent factors for secondary deficiency were identified (Figure 1G).

Immunogenomic profiling of rare variants predisposing to errors of immunity. Focusing on the investigation of genetic immunodefi-



F

Correlation Ig levels, age, and lymphocyte counts

	Age	WBC	ALC	LGL	T cells	CD4 T	CD8 T	NK cells	B cells	IgG	IgA
Age											
WBC	0.012 0.90										
ALC	0.051 0.59	0.836 <0.001									
LGL	0.032 0.73	0.525 <0.001	0.634 <0.001								
T cells	-0.038 0.69	0.742 <0.001	0.808 <0.001	0.668 <0.001							
CD4 T	-0.132 0.16	0.598 <0.001	0.569 <0.001	0.390 <0.001	0.721 <0.001						
CD8 T	0.003 0.98	0.689 <0.001	0.783 <0.001	0.623 <0.001	0.917 <0.001	0.468 <0.001					
NK cells	-0.127 0.17	0.340 <0.001	0.221 0.02	0.149 0.11	0.252 0.007	0.579 <0.001	0.072 0.44				
B cells	-0.23 0.01	0.518 <0.001	0.506 <0.001	0.344 <0.001	0.527 <0.001	0.681 <0.001	0.348 <0.001	0.510 <0.001			
IgG	0.096 0.31	-0.087 0.36	-0.036 0.71	0.048 0.61	0.076 0.42	0.163 0.08	0.037 0.70	0.077 0.41	0.059 0.53		
IgA	0.188 0.04	-0.168 0.07	-0.143 0.13	-0.020 0.83	-0.070 0.46	0.023 0.81	-0.117 0.21	-0.138 0.14	-0.057 0.54	0.486 <0.001	
IgM	-0.105 0.26	-0.057 0.55	-0.008 0.94	0.082 0.39	0.127 0.18	0.081 0.39	0.114 0.22	-0.034 0.72	0.048 0.61	0.427 <0.001	0.396 <0.001

Rho: 1.0, 0.5, 0, -0.5, -1.0

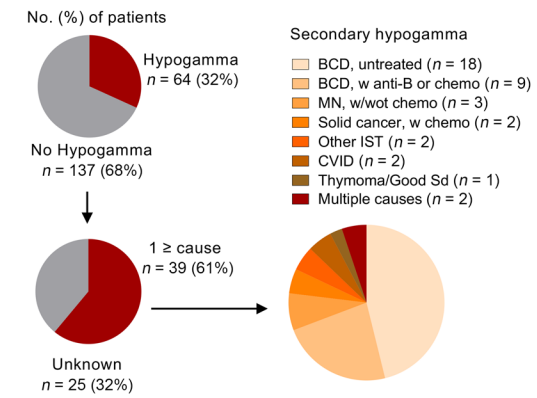


Figure 1. Study summary and systematic characterization of immunodeficiency underpinnings of T-LGLL. (**A**) Graphic summary of study design, study cohort, and study datasets (created with Biorender.com). (**B**) Flow chart of the initial immunologic laboratory characterization. (**C**) Distribution dot plot of cohort absolute lymphocyte count (ALC) (available in *n* = 241 patients). Dashed lines represent the lower and upper limits of the normal range. The histogram represents the number of cases within ALC ranges. (**D**) Lymphocyte populations in peripheral blood at diagnosis (available in *n* = 216 patients). The bar chart represents the mean percentages of lymphocyte populations ($CD8^+$ T cells, $CD4^+$ T cells, NK and B cells). Distribution dot plots of each lymphocyte subset are also shown. Dashed lines represent the lower and upper limits of the normal range. The pie charts highlight the cases with low $CD8^+$ T cells, low $CD4^+$ T cells, low NK cells, and low B cells. (**E**) Concentration of Ig isotypes (IgG, IgA, IgM) at diagnosis (available in *n* = 201 patients). (**F**) Correlogram showing the association of Ig levels with age, blood counts, and lymphocytes. Spearman's rho correlation coefficients and *P* values are shown for each pair of variables. Color scale represents the magnitude of the correlation coefficient. (**G**) Screening for potential causes of Ig hypogammaglobulinemia, detected in *n* = 64 patients at diagnosis. Anti-B, anti–B cell therapy; BCD, B cell dyscrasia; CBC, complete blood count; Chemo, chemotherapy; IST, immunosuppressive therapy; LGL, large granular lymphocyte count; WBC, white blood cell count; MN, myeloid neoplasms.

ciency traits as predisposing factors for T-LGLL, we subjected a representative subcohort of 92 consecutive patients (38% of the whole cohort, no clinically biased criteria applied) to whole-exome sequencing (WES) (Table 1). The median age at diagnosis was 62 years (IQR: 54–71) with a male:female ratio of 0.5. Mutations in *STAT3* and *STAT5B* were present in 42 (46%) and 1 (1%) of 92 patients, respectively. We then analyzed the open reading frames of 464 genes associated with inborn errors of immunity (IEI) included in the 2022 classification of the International Union of Immunological Societies (IUIS; Supplemental Table 1; supplemental material available online with this article; https://doi.org/10.1172/JCI184431DS1) (19). Rare variants were defined by minor allele frequencies (MAF) in the Genome Aggregation Database (gnomAD) of less than 1%. Each variant was assessed according to American College of Medical Genetics (ACMG) criteria using ClinVar (20) and VarSome (21). For analysis, we selected variants classified as pathogenic/likely pathogenic (P/LP), or variants of uncertain significance (VUS) overrepresented in our cohort, i.e., those with a significant *P* value for observed versus expected frequencies after Benjamini-Hochberg multiple-testing correction (FDR < 0.05).

We found 183 suspicious rare germline variants in IEI-related genes (Figure 2A). Of those, we selected 43 variants of potential clinical relevance (40 VUS; 3 P/LP) in 38 different genes that were overrepresented in our cohort after comparing observed versus expected allele frequencies in the general population. These variants were detected in 34 (37%) of 92 patients, a combined genetic burden significantly elevated when compared with the *All of US* (https://allofus.nih.gov/) healthy control population with genomic data, of whom only 167 (0.26%) of 63,026 individuals carried any of the variants identified ($P < 0.001$; Figure 2B and Supplemental Table 2). The median number of variants in carriers was 1 (range: 1–3 mutations/patient). All variants were heterozygous. Twenty-nine and 11 variants involved IEI-related genes associated with recessive and dominant traits, respectively; the remaining 3 affected recessive genes have also been previously linked to atypical IEI in heterozygous status (Figure 2, C and D). Gene-phenotypic analysis clustered the variants in combined T/B cell defects (*n* = 12), immune dysregulation/autoinflammation (*n* = 10), bone marrow failure (BMF, *n* = 6), phagocyte defects (*n* = 6), B cell defects (*n* = 5), and innate immunodeficiency (*n* = 4) (Figure 2E). The gene variants linked to immune dysregulation and innate or humoral immunodeficiency syndromes were more commonly associated with dominant and adult-onset IEI traits (Figure 2F) (19, 22).

Characterization of T-LGLL patients with IEI high-confidence deleterious variants. To further establish biological-clinical correlations, we defined carriers of high-confidence deleterious (hcD) variants, more likely to predispose to immune misbalance, as those harboring either P/LP defects, or heterozygous VUS for dominant traits. Fifteen (16%) of 92 patients carried these high-risk variants involving 13 genes (Figure 3, Supplemental Figures 1–3, and Table 2). Among them, the only case with a genetic diagnosis of an IEI was included: a 61-year-old man with a history of autoimmune hemolytic anemia prior to T-LGLL who had a heterozygous P/LP variant in *AIRE* (p.R297N). Biallelic mutations in this master regulator of central tolerance are associated with autoimmune polyendocrine syndrome type 1 (APS-1, OMIM#607358). However, heterozygous mutations involving the *plant homeodomain* (PHD) region, where R297 is located, can lead to atypical APS-1, characterized by late-onset, organ-specific autoimmunity due to a dominant-negative effect of the altered allele (23). A case of biallelic, classic APS-1 with T-LGLL was previously reported in a 34-year-old woman with concurrent pure red cell aplasia (PRCA) (24).

We also found a heterozygous P/LP variant in *TCIRG1* (c.1767-2A>G) in a 66-year-old man with chronic neutropenia and personal and family history (mother) of rheumatoid arthritis (Figure 3, Supplemental Figures 1–3, and Table 2). Biallelic mutations involving this phagocytic H^+-ATPase cause recessive malignant osteopetrosis (OMIM#604592), while heterozygous variants have been associated with dominant congenital neutropenia. This splice site variant, involving the 3′ (acceptor) AG sequence of *TCIRG1* intron 14, is predicted to have a significant, negative effect by different algorithms, including the MaxEntScan (25) model for splice site variants (score 1.91, mutant-WT Δ=7.95 –predicted as a high potential to disrupt native splicing). Notably, the same gene, through usage of an alternative initiation codon in exon 7, generates TIRC7, a different protein expressed by T cells with immune checkpoint functions (26). The location of this variant is predicted to impair the splicing of the 2 proteins, thus potentially explaining both the neutropenia and the autoimmune manifestations in this case.

The patient with the earliest diagnosis of T-LGLL within our cohort was also found to carry a hcD defect: a 17-year-old man with a prior history of CVID, who had a complicated debut mimicking severe immune-mediated BMF, and undergoing allogeneic hematopoietic stem cell transplant (Figure 3, Supplemental Figures 1–3, and Table 2). He carried a heterozygous P/LP variant in *STK4* (p.E324*), linked to combined T/B cell defects, lymphoproliferation, and cytopenias in biallelic configuration (OMIM# 614868). The patient was also a heterozygous carrier of a hypomorphic, low-frequency variant in *PRF1* (p.A91V, MAF = 0.0293), which has been described as a risk factor for late-onset hemophagocytic lymphohistiocytosis, acquired aplastic anemia, and different T cell lymphoid neoplasms (27, 28). Peripheral blood flow cytome-

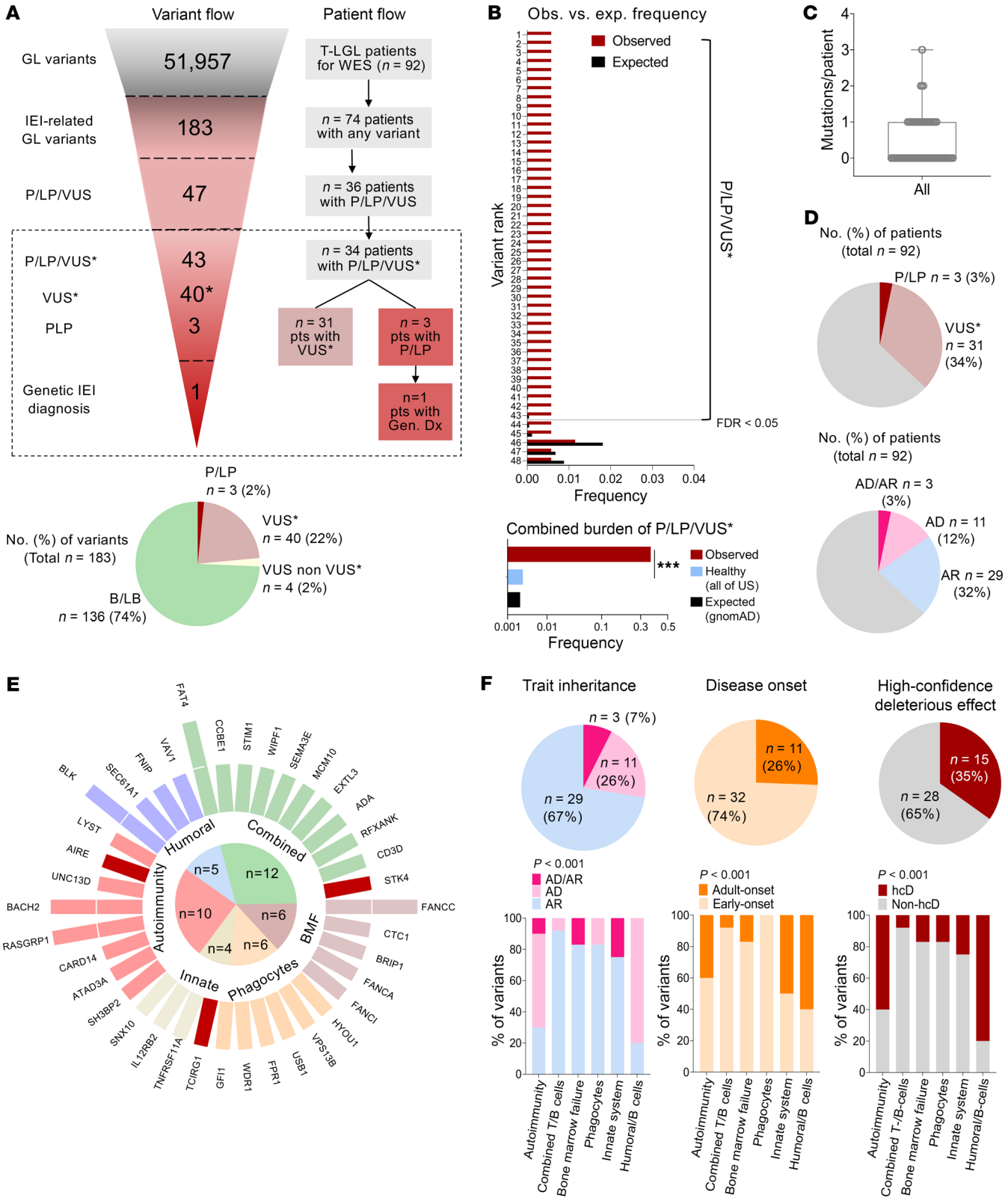

A
Variant flow
Patient flow
GL variants
51,957
IEI-related GL variants
183
P/LP/VUS
47
P/LP/VUS*
43
VUS*
40*
PLP
3
Genetic IEI diagnosis
1
T-LGL patients for WES (n = 92)
n = 74 patients with any variant
n = 36 patients with P/LP/VUS
n = 34 patients with P/LP/VUS*
n = 31 pts with VUS*
n = 3 pts with P/LP
n=1 pts with Gen. Dx
No. (%) of variants (Total n = 183)
P/LP n = 3 (2%)
VUS* n = 40 (22%)
VUS non VUS* n = 4 (2%)
B/LB n = 136 (74%)
B
Obs. vs. exp. frequency
Observed
Expected
Variant rank
P/LP/VUS*
FDR < 0.05
Frequency
Combined burden of P/LP/VUS*
***
Observed
Healthy (all of US)
Expected (gnomAD)
Frequency
C
Mutations/patient
All
D
No. (%) of patients (total n = 92)
P/LP n = 3 (3%)
VUS* n = 31 (34%)
No. (%) of patients (total n = 92)
AD/AR n = 3 (3%)
AD n = 11 (12%)
AR n = 29 (32%)
E
Humoral
Combined
Autoimmunity
BMF
Innate
Phagocytes
n=5
n=12
n=10
n=6
n=4
n=6
FAT4
CCBE1
STIM1
WIPF1
SEMA3E
MCM10
EXTL3
ADA
RFXANK
CD3D
STK4
FANCC
CTC1
BRIP1
FANCA
FANCI
HYOU1
VPS13B
USB1
FPR1
WDR1
GFI1
TCIRG1
TNFRSF11A
IL12RB2
SNX10
SH3BP2
ATAD3A
CARD14
RASGRP1
BACH2
UNC13D
AIRE
LYST
BLK
SEC61A1
FNIP
VAV1
F
Trait inheritance
n = 3 (7%)
n = 11 (26%)
n = 29 (67%)
P < 0.001
AD/AR
AD
AR
% of variants
Autoimmunity
Combined T/B cells
Bone marrow failure
Phagocytes
Innate system
Humoral/B cells
Disease onset
n = 11 (26%)
n = 32 (74%)
P < 0.001
Adult-onset
Early-onset
% of variants
Autoimmunity
Combined T/B cells
Bone marrow failure
Phagocytes
Innate system
Humoral/B cells
High-confidence deleterious effect
n = 15 (35%)
n = 28 (65%)
P < 0.001
hcD
Non-hcD
% of variants
Autoimmunity
Combined T-/B-cells
Bone marrow failure
Phagocytes
Innate system
Humoral/B-cells

Figure 2. Immunogenomic landscape of variants predisposing to IEI in patients with T-LGLL. (**A**) Variant and patient flow. Variants of potential clinical relevance included in the analysis (P/LP/VUS*) are highlighted with a dashed box. The pie chart at the bottom indicates the distribution of the variants by pathogenicity (ACMG criteria). (**B**) Comparison of the observed (Obs.) versus expected (exp.) (gnomAD) allele frequencies for the selection of variants overrepresented in our cohort (P/LP/VUS*). Multiple testing correction was done with Benjamini-Hodgberg method (FDR < 0.05). The bar chart at the bottom represents the combined mutational burden of all P/LP/VUS* observed in our cohort (39% vs. 0.26%, as compared with *All of US* healthy controls with genomic data, n = 63,026). χ^2 P values are shown. The expected probability of finding any of the variants in the general population according to gnomAD is also shown (0.22%). (**C**) Individual mutation burden among carriers. All variants were found in the heterozygous state. (**D**) Number of patients carrying the variants according to pathogenicity (upper chart) and pattern of inheritance (lower chart). (**E**) Circle plot representing the number of variants by gene according to functional clustering and pathogenicity. Maroon bars indicate P/LP and genetic diagnoses of IEI, respectively. (**F**) Proportion of variants found according to functionality and pathogenicity, trait inheritance, and disease onset. χ^2 P values are shown. ***P < 0.001. AD, autosomal dominant; AR, autosomal recessive; B/LB, benign/likely benign; Gen. Dx, genetic diagnosis; GL, germline; VUS*, VUS overrepresented in our cohort.

try performed at the referring institution before transplant revealed decreased proportions of NK and NK/T cells expressing perforin as described for hemophagocytic lymphohistiocytosis (51% [normal:86%–98%] and 2% [normal:30%–78%], respectively), suggesting combinatory effects of the variants identified in this case.

Among the other cases with hcD variants, we point out 2 patients with heterozygous gene defects in *BACH2* (Figure 3, Supplemental Figures 1–3, and Table 2). A dominant, adult-onset IEI has been recently linked to this gene, characterized by lymphocyte maturation defects, intestinal inflammation, and hypogammaglobulinemia (29). Both cases from our cohort were diagnosed early with T-LGLL, in the fourth decade of life; they had personal and family history of autoimmune disease and presented $CD4^+$ T cell defects. One of them exhibited all the features described for BACH2 deficiency syndrome: a 36-year-woman with a history of type 1 diabetes, inflammatory bowel disease, PRCA, and IgA deficiency prior to T-LGLL diagnosis. This patient was found to carry a heterozygous missense variant involving the dimerization domain (p.M11L), close to previously described P/LP variants (29).

Distinct clinical features identified in carriers of IEI hcD variants. Next, we investigated the clinical and laboratory features of carriers of hcD variants (Figure 4A and Supplemental Figure 4). As compared with noncarriers, these patients were diagnosed with T-LGLL earlier (median: 55 versus 64 years, P = 0.01) and more frequently had a family history of immune or hemato-lymphoid conditions (40% versus 4%, P < 0.001). Lower lymphocyte and large granular lymphocyte (LGL) counts were appreciated, resulting in higher rates of lymphocytopenia (27% versus 3%, P = 0.006). Carriers had notably higher frequencies of hypogammaglobulinemia (40% versus 20%) and cytopenias of any type (93% versus 75%), although the significance level was not reached, probably due to relatively small sample size (statistical power = 60%). However, they had significantly higher rates of IgA deficiency (27% versus 7%, P = 0.03) and autoimmune cytopenias or BMF (60% versus 19%, P = 0.001) than noncarriers.

There were no differences in the need of LGLL therapy, nor in the overall response rate as severity indicators. In addition, survival curves suggested an apparent favorable trajectory for hcD carriers that, however, was not confirmed in multivariate analysis (Supplemental Figure 5 and Supplemental Table 3).

Mutational landscape in T-cell lymphoid drivers. Analysis of single-nucleotide variants (SNVs) and copy-number variants (CNVs) in T cell lymphoid drivers (169 genes, Supplemental Table 4) revealed 91 variants of clinical relevance in 30 different genes in 62 (67%) of 92 patients (Supplemental Table 5). Mean mutational burden was 1 (range: 0 – 4 mutations/patient, Figure 4B). Mutations in genes other than *STAT3* were detected in 45 (49%) of 92 patients. Mutations in *STAT5B* and *STAT6* were identified in 3 (3%) patients. Mutations in epigenetic modifiers *KMT2C*, *KMT2D*, *KDM6A*, *KMT5C*, *TET2*, *DNMT3A* and *CREBBP* were detected in 16 (17%) patients, in 9 of 16 (56%) of them in cooccurrence with *STAT3* mutations. We also identified variants in other recurrent mature T cell neoplasm genes, such as the RNA helicase *DDX3X* and the tumor suppressor *TP53* — one case each. Other oncogenes/tumor suppressors harboring damaging mutations in our cohort were *CSMD3*, *POT1*, *NOTCH1*, *NRAS*, *STAG2*, *CACNA1E*, *SEPTIN4*, *CHD2*, *LRP1B*, *TTN*, *PCLO*, *VCAN*, *XIRP2*, *UBA1*, *KRAS*, and *KDR*. As expected for our focus on T-LGLL, we did not find mutations in NK-cell lymphoid neoplasm genes *CCL2* or *CCL22* (Figure 4, C–E and Supplemental Table 5).

Gene-level CNV analysis conducted in WES data did not detect the presence of copy number (CN) gains in *STAT3*, *STAT5A*, or *STAT5B* (located in chromosome [chr] 17q21.2), nor CN losses in *TP53* (chr 17p13.1, Supplemental Figure 6). We first focused on the screening for CNVs in these genes because of their being common CNV hotspots in other T cell neoplasms (30–33), as confirmed by the interrogation of genomic data from 26 human T cell cancer cell lines gathered from DepMap (Supplemental Table 6 and Supplemental Figures 7 and 8) (34). In contrast with T-LGLL, assessment of SNVs and CNVs in T cell cancer cell lines showed alterations in *STAT3* via CN gain in 3 (12%) of 26 cell lines and *TP53* SNV and CN loss in 15 (58%) and 5 (19%) of 26 cell lines, respectively. Next, we extended our analysis of CNVs to other T cell lymphoid drivers, which yielded alterations in 4 (4%) of 92 T-LGLL patients (Supplemental Table 7), including the amplification of *CACNA1E* (chr 1q25), *NEB* (chr 2q23) and *KMT5C* (chr 19q13), and the loss of *SEPTIN4* (chr 17q22, this latter case not including the *STAT3* locus).

Comparison of IEI hcD carriers versus noncarriers revealed significantly different mutational configurations (P = 0.014): 4 of 15 (27%) versus 38 of 77 (49%) cases were *STAT3*mt; conversely 9 of 15 (60%) versus 26 of 77 (34%) had mutations in genes other than *STAT3*; 1 of 15 (7%) versus 14 of 77 (18%) patients had coexistence of *STAT3* and other gene mutations. Among the genes other than *STAT3* mutated in *STAT3*wt hcD carriers, we noted hotspot mutations in *TNFAIP3* (n = 2), *NOTCH1* (n = 1), and *NRAS* (n = 1) as well as CNVs in *CACNA1E* and *SEPTIN4* in the 2 patients with mutations in *TNFAIP3* (Figure 4, C–E).

Comutation analysis of IEI and T cell lymphoid gene variants. Next, we sought to explore genetic-biological correlations. Gene ontology (GO) term pathway analysis revealed that the hub formed by the IEI-related genes having hcD variants had 28 enriched commonalities with the network formed by a list of known T-LGLL dysregulated genes (Supplemental Figure 9A and Supplemental Table

Table 2. Clinical and laboratory features of carriers of hcD variants in IEI-linked genes

UPN	IEI hcD variants (cDNA, protein)[A]	MAF	Age Dx (years)	Sex	Medical history	Family history	Cytopenias	Splenomegaly	STAT3mt	Immunodeficiency	Treatment	Response
24	AIRE (c.G889, p.D297N)	NF	61	M	CAD, IgM, MGUS	No	Anemia	Yes (splenectomy)	*STAT3* p.Y640F (VAF = 31.9%), p.D661Y (VAF = 1.0%)	Low CD4$^+$ T cells, low NK cells, low B cells, MGUS without immunoparesis	No	NA
49	STK4 (c.G970T, p.E324*)	NF	17	M	Pancytopenia and hypoplastic BM resembling severe AA (16 years)	IBD (paternal grandfather), leukemia (maternal grandfather)	Pancytopenia	Yes (mild)	No	Lymphocytopenia, low CD4$^+$ T cells, low NK cells, absent B cells, hypogamma (diagnosis of CVID). Low perforin$^+$ NK and NK/T cells (51% [86%–98%] and 2% [30%–78%], respectively).	Allo-HSCT	No
236	TCIRG1 (c.1767-2A>G, X804_splice)	NF	66	M	Rheumatoid arthritis, chronic neutropenia	Rheumatoid arthritis (mother)	Anemia, Neutropenia	Yes	No	Low NK cells, high polyclonal IgA levels	MTX	No
297	BACH2 (c.A31C, p.M11L)	NF	36	F	T1DM, IBD, PRCA	T1DM (sibling)	Anemia, neutropenia	No	No	Low CD4$^+$ T cells, low NK cells, low B cells, IgA deficiency	CSPA	Yes
232	BACH2 (c.C509T, p.T170M)	NF	36	M	Glomerulonephritis	Sarcoidosis (mother)	No	No	No	Low/borderline CD4$^+$ T cells, low NK cells, transient episodes of leukopenia and intermittent fevers	No	NA
46	BLK (c.C682A, p.L228I)	NF	58	M	Pancytopenia and hypoplastic BM resembling nonsevere AA (53 years)	No	Neutropenia	No	*STAT3* p.Y640F (VAF = 7.8%)	Low CD4$^+$ T cells, low NK cells, low B cells	No	NA
98	BLK (c.T230C, p.F77S)	NF	55	M	Rheumatoid arthritis, AIN	No	Pancytopenia	No	No	Lymphocytopenia, low CD4$^+$ T cells, low NK cells, low B cells	CY	No
62	CARD14 (c.C659T, p.S220L)	1.1×10^{-4}	69	M	Hx of ITP	Asthma (father)	Leukopenia	No	*STAT3* p.Y640F (VAF = 14.2%)	Low CD4$^+$ T cells, low B cells	No	NA
20	SH3BP2 (c.G343A, p.E115K)	4.0×10^{-6}	34	F	Positive ANA, AIN	SLE (mother)	Anemia, neutropenia	No	No	Hypogamma (diagnosis of CVID)	No	NA
52	ATAD3A (c.T1247G, p.F416C)	8.0×10^{-6}	76	M	Rheumatoid arthritis	No	Anemia, neutropenia	No	*STAT3* p.D661Y (VAF = 9.9%)	Lymphocytopenia, low CD4$^+$ T cells, low NK cells, low B cells, high polyclonal IgA levels	CSPA	Yes
96	CTC1 (c.T926G, p.L309R)	NF	58	F	Hx of CAD, MBL and IgM MGUS	No	Neutropenia	Yes (splenectomy)	No	Low CD4$^+$ T cells, MBL, IgM, MGUS with no immunoparesis	CY	No
23	GFI1 (c.G440T, p.R147L)	NF	67	F	None	No	Neutropenia	No	No	IgM deficiency	No	NA
82	SEMA3E (c.C299T, p.S100L)	7.3×10^{-5}	47	M	IgG MGUS	No	Anemia, leukopenia	Yes (splenectomy)	No	Lymphocytopenia, low CD4$^+$ T cells, low NK cells, low B cells, IgG, MGUS with immunoparesis (IgA deficiency)	CY	Yes
37	SEC61A1 (c.C569T, p.T190I)	NF	31	F	None	No	Pancytopenia	Yes	No	Normal lymphocyte count. Normal gamma fraction (Ig dosification not available)	CSPA	No
16	VAV1 (c.T1628G, p.M543R)	NF	54	F	AIHA, Light chain MGUS	No	Anemia	No	No	IgM deficiency	No	NA

[A]All variants were found in heterozygous configuration. AA, aplastic anemia; ANA, antinuclear antibodies; AIHA, autoimmune hemolytic anemia; AIN, autoimmune neutropenia; CAD, cold aglutinin disease; CSPA, cyclosporine; CY, cyclophosphamide; HSCT, hematopoietic stem cell transplant; IBD, inflammatory bowel disease; ITP, immune thrombocytopenic purpura; MBL, monoclonal B cell lymphocytosis; MGUS, monoclonal gammopathy of uncertain significance; MTX, methotrexate; NF, not found; T1DM, type 1 diabetes mellitus; UPN, unique patient number; Hx, history.

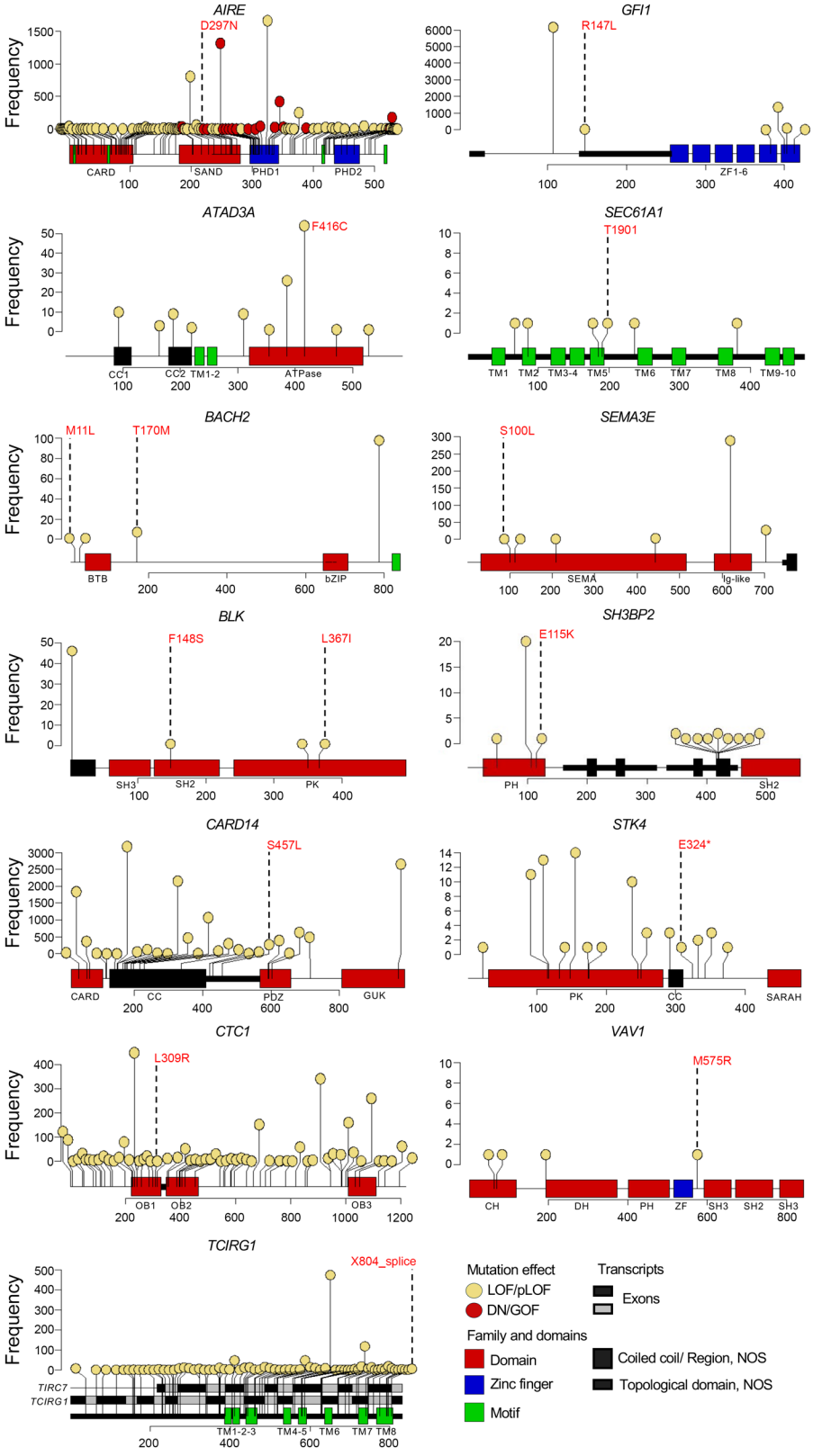


Figure 3. Mutations in 13 IEI genes with hcD variants identified in T-LGLL patients. Annotation of the domains of the proteins coded by the canonical transcripts was extracted from Ensembl and UniProt.json files. For TCIRG1, exon-protein correlations for both the canonical and alternative transcripts are shown. The mutations labeled in red with the amino acid change are the ones found in our study. The plot additionally displays rare (MAF < 1%) deleterious variants previously reported in these genes using gnomAD genomic browser, version 4.1.0, integrating pathogenicity predictors and variant frequency (number of variants reported in gnomAD). LOF/pLOF, loss of function/predicted LOF; NOS, region/domain not otherwise specified.

8) (35). Among them, cellular response to biotic stimulus, pathogen recognition, production/regulation of IL-12 and TNF-α, cellular response to IL-6, or positive regulation of T cells was encountered, supporting the pathogenic role of hcD variants in promoting CTL expansion via aberrant/defective signals involving these pathways.

The IEI hcD variants can be phenotypically dichotomized according to the type of the abnormal response as immune dysregulation (hyperactive response) or innate/adaptive immunodeficiency (defective response) defects. To further investigate the interplay between IEI hcD and T cell driver genes, we performed variant correlation and HALLMARK (https://www.gsea-msigdb.org/) pathway enrichment analysis. We observed an enrichment of mutations in elements of the IL-6/JAK/STAT3, NOTCH, and TP53 pathways in cases with IEI variants associated with immune dysregulation. In contrast, there was enrichment of mutations associated with autoimmunity and lymphoproliferation via TNF-α/NF-κB and the RAS pathways in cases with IEI variants associated with immunodeficiency (Supplemental Figure 9, B and C).

RNA-Seq analysis of TCR signaling pathway in STAT3mt versus STAT3wt expanded T cells. Chronic antigen stimulation has been suggested to initially drive the outgrowth of T-LGL clones. However, there are limited data supporting the role of T cell receptor (TCR) signaling in the disease. Based on a lower *STAT3*mt rate in IEI hcD carriers, we hypothesized that the clonal expansion might rely on TCR signaling more strongly in *STAT3*wt versus mutant cells, as described for other T cell lymphoid neoplasms having *STAT3*mt (36–39). To this end, 2 transcriptomic datasets were interrogated: (a) single-cell RNA+TCRαβ-Seq (scRNA-Seq + TCRαβ-Seq) of T-LGLL (n = 11) and healthy control (n = 6) samples, independently repurposed from the study by Huuhtanen and Bhattacharya et al. (Supplemental Table 9) (13) and (b) bulk RNA-Seq of mature T cell cancer cell lines from DepMap (n = 22, Supplemental Table 6, Supplemental Figure 8) (34).

Gene-expression levels and differential expression analyses were performed in *STAT3*mt versus *STAT3*wt hyperexpanded T cells (>10 TCR templates) from scRNA-

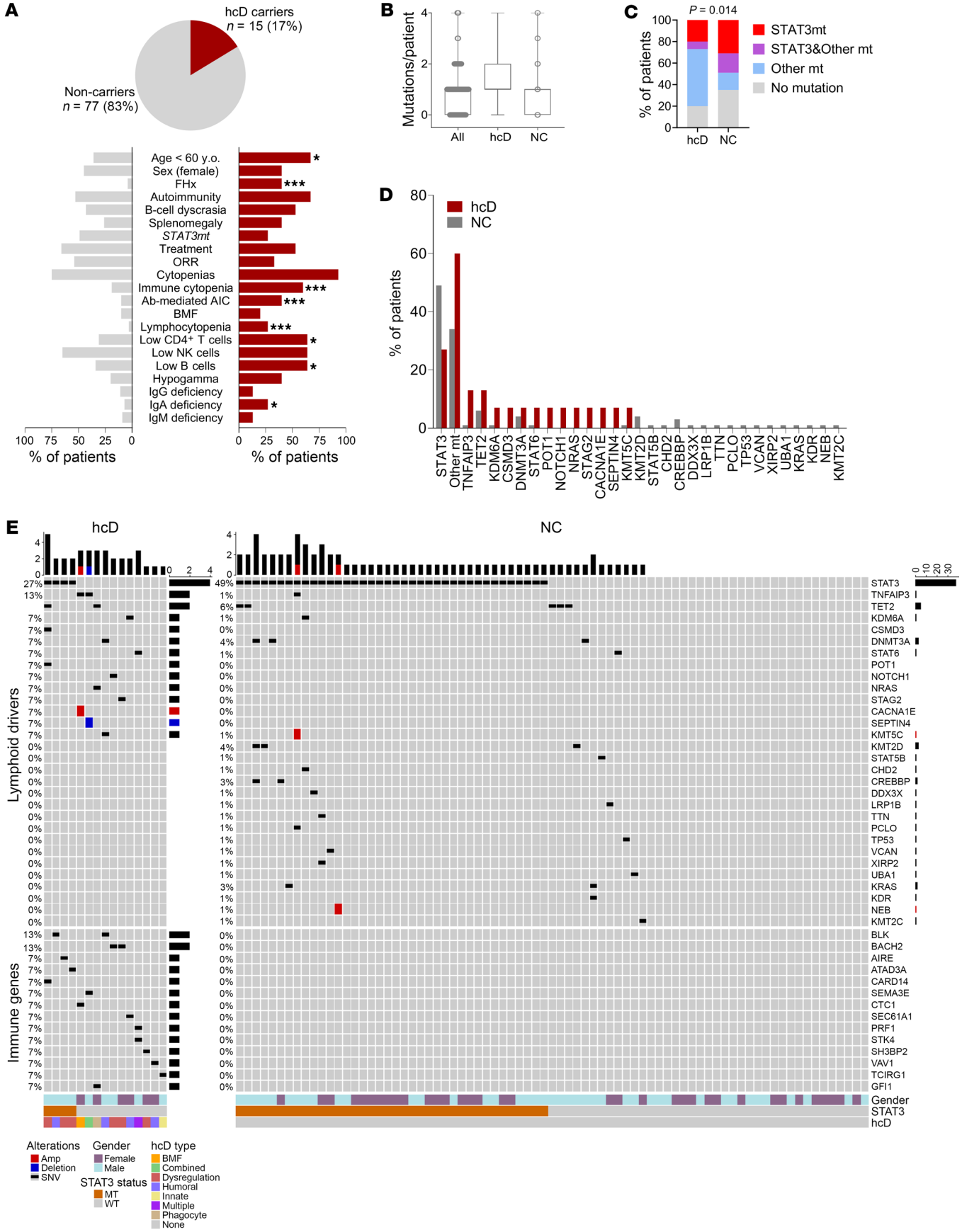

A
hcD carriers
n = 15 (17%)
Non-carriers
n = 77 (83%)
Age < 60 y.o.
Sex (female)
FHx
Autoimmunity
B-cell dyscrasia
Splenomegaly
STAT3mt
Treatment
ORR
Cytopenias
Immune cytopenia
Ab-mediated AIC
BMF
Lymphocytopenia
Low CD4+ T cells
Low NK cells
Low B cells
Hypogamma
IgG deficiency
IgA deficiency
IgM deficiency
% of patients
% of patients
B
Mutations/patient
All
hcD
NC
C
P = 0.014
STAT3mt
STAT3&Other mt
Other mt
No mutation
% of patients
hcD
NC
D
hcD
NC
% of patients
STAT3
Other mt
TNFAIP3
TET2
KDM6A
CSMD3
DNMT3A
STAT6
POT1
NOTCH1
NRAS
STAG2
CACNA1E
SEPTIN4
KMT5C
KMT2D
STAT5B
CHD2
CREBBP
DDX3X
LRP1B
TTN
PCLO
TP53
VCAN
XIRP2
UBA1
KRAS
KDR
NEB
KMT2C
E
hcD
NC
Lymphoid drivers
Immune genes
STAT3
TNFAIP3
TET2
KDM6A
CSMD3
DNMT3A
STAT6
POT1
NOTCH1
NRAS
STAG2
CACNA1E
SEPTIN4
KMT5C
KMT2D
STAT5B
CHD2
CREBBP
DDX3X
LRP1B
TTN
PCLO
TP53
VCAN
XIRP2
UBA1
KRAS
KDR
NEB
KMT2C
BLK
BACH2
AIRE
ATAD3A
CARD14
SEMA3E
CTC1
SEC61A1
PRF1
STK4
SH3BP2
VAV1
TCIRG1
GFI1
Gender
STAT3
hcD
Alterations
Amp
Deletion
SNV
Gender
Female
Male
STAT3 status
MT
WT
hcD type
BMF
Combined
Dysregulation
Humoral
Innate
Multiple
Phagocyte
None

Figure 4. Clinical, laboratory, and genetic characterization of patients with hcD IEI variants. (**A**) Clinical and laboratory features of the carriers of hcD variants (hcD, red) versus noncarriers (NC, gray). Family history (FHx) designates positive family history for immune or hemato-lymphoid conditions. The category immune cytopenia includes both BMF (i.e., aplastic anemia or PRCA) and antibody-mediated peripheral autoimmune cytopenias (Ab-mediated autoimmune cytopenia, i.e., autoimmune hemolytic anemia, autoimmune neutropenia, or immune thrombocytopenia). (**B**) Mutational burden (mutations/patient). (**C**) Mutational configuration in *STAT3* and other T cell lymphoid driver genes. (**D**) Frequency (%) of mutations in individual T cell lymphoid driver genes. (**E**) Mutational profile in carriers and noncarriers of hcD variants. The upper plot represents somatic mutations in T cell lymphoid driver genes. The lower plot shows the germline hcD variants in IEI-linked genes, clustered by immune pathways. The specific mutated IEI gene is detailed for each case. The low-frequency, hypomorphic PRF1 variant identified in P49 is also shown. Main clinical features and color legends are indicated below. χ^2 *P* values are shown. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. Ab-mediated AIC, antibody-mediated peripheral autoimmune cytopenia; ANC, absolute neutrophil count; FHx, family history of immune/hemato-lymphoid conditions; ORR, overall response rate; y.o., years old.

Seq+ TCRαβ-Seq samples (Supplemental Figure 10). We observed that several genes proximal to the TCR complex were differentially expressed between *STAT3*mt and *STAT3*wt clones. Specifically, *TRBC* and *TRBV* genes, *CD2*, *CD3E*, *CD5*, *CD8A*, *LAT*, *LCK*, *SYK*, or *ZAP70* were upregulated in *STAT3*wt cells (i.e. downregulated in *STAT3*mt clones), which showed expression levels like those of hyperexpanded T cells from healthy controls (Supplemental Table 10 and Supplemental Figures 11 and 12). Previous differential gene-pathway analysis of this experimental dataset also suggested upregulation of the TCR signaling pathway in *STAT3*wt versus *STAT3*mt cells (13). Next, to validate these findings, we analyzed the expression of TCR-related genes in 22 T cell cancer cell lines according to the *STAT3* amplification status previously assessed (*STAT3*-amplified [*STAT3*amp] $n = 3$; *STAT3*wt $n = 19$, Supplemental Figures 7 and 8). We confirmed that *STAT3*wt versus *STAT3*amp cells had generally lower expression of TCR-related genes, including *CD3Z*/*CD247*, *ZAP70*, *LCK*, and *LAT* (Supplemental Figure 13). Differential expression and pathway analysis (40) of *STAT3*amp versus fusion-matched *STAT3*wt lines (Supplemental Methods and Supplemental Tables 6 and 11) also showed significant upregulation of immune receptor–based signaling pathways in *STAT3*wt versus *STAT3*amp cells (Supplemental Figure 14).

To further investigate the interplay between *STAT3*mt status and TCR signaling, we calculated a TCR signaling expression score using the Seurat AddModuleScore function (41). To this end, we used 15 TCR-related genes, including *TRAC* and *TRDC* as constant elements of the TCR complex (GO:0042101) and components of the TCR signalosome (GO:0036398). A score with the same set of genes was also calculated in bulk RNA-Seq from the T cell cancer cell lines as geometric means as previously described (13). In T-LGLL samples, *STAT3*mt clones had significantly lower TCR scores than *STAT3*wt and healthy control cells. Similarly, lower TCR scores were found in *STAT3*amp versus *STAT3*wt T cell cancer cell lines (Figure 5, A and B, and Supplemental Figure 15).

T-LGLs correspond to mature T cells with generally high expression of TCR. Therefore, to assess the relationship between strong TCR and STAT3 signaling in a single-cell basis, we categorized the TCR score as high (above the 90th percentile) or low (below the 90th percentile) and compared the distribution of these TCR-high/low cells depending on the *STAT3*mt status. In addition, using the lists of genes upregulated in *STAT3*mt versus *STAT3*wt cells generated here and gathered from MSigDB (42), we calculated a 9-gene STAT3 signaling score as an indicator of STAT3 activation (Supplemental Methods). Grouping of the cells as STAT3 score high/low was also done using the 90th percentile threshold. Coexpression analysis showed that STAT3-high scores were overrepresented in the *STAT3*mt clones (14% versus 8% and 2% in *STAT3*wt and healthy control cells, respectively, $P < 0.0001$, Figure 5, C and D), while TCR-high scores were more frequent in *STAT3*wt and healthy cells (22% and 17% versus 2% in *STAT3*mt clones, respectively, $P < 0.0001$). Overall, cells with double STAT3-high/TCR-high scores were infrequent (0.71%). Altogether, our results suggest that the TCR signaling might have a pathogenic contribution on CTL outgrowth, particularly prior to the acquisition of GOF *STAT3* mutations during T-LGLL evolution.

TCR repertoires and target specificities in T-LGLL samples. Next, to further investigate the diversity and the nature of the T cell responses promoted, we studied the TCR repertoire using deep variable β (VB) complementarity determining regions (CDR3) sequencing. We analyzed productive rearrangements within normalized repertoires in a subset of 18 T-LGLL patients, 3 of whom carried hcD variants associated with innate/adaptive immunodeficiency (Figure 6A). As compared with the repertoires of 145 healthy controls, normalized in a similar fashion (43, 44), all metrics supported a lesser level of TCR diversity. Consistent with the diagnosis, we also found a higher number and size of pathologically expanded clonotypes in T-LGLL patients. However, no substantial differences in these parameters were appreciated between carriers and noncarriers of hcD variants (Figure 6, B and C). In an effort to track the possible targets of immune responses, we then assessed clonotype specificities according to our previously published metanalytic collection of CDR3b sequences (Supplemental Table 12) (45). Initial analysis of the whole repertoires revealed exact matches with the reference in 4% of the sample clonotypes, with similar proportions between groups associated with response to infection, autoimmunity, and tumor surveillance. Analysis focused on pathologically expanded clonotypes showed exact matches in 6% clones; most of these specificities (185/227, 81%), and all found in T-LGLL patients with hcD variants (7/7, 100%), were targeted against pathogens, supporting the notion that microbial agents might be the prevailing antigenic triggers in these cases (Figure 6D).

A graphical summary of the proposed model on the pathogenic role of underlying IEI in CTL proliferations and clonal shift is shown in Figure 6E.

Discussion

Several recent advances have been made toward a better understanding of the pathogenesis of T-LGLL (7), for instance, the identification of mutations in genes other than *STAT3* (e.g., *STAT5B*, *TNFAIP3*, epigenetic regulators) or the role of the immune microenvironment besides the leukemic clone (13). Chronic antigen stimulation has been classically hypothesized to initiate/drive CTL outgrowth, based on both clinical and translational observations

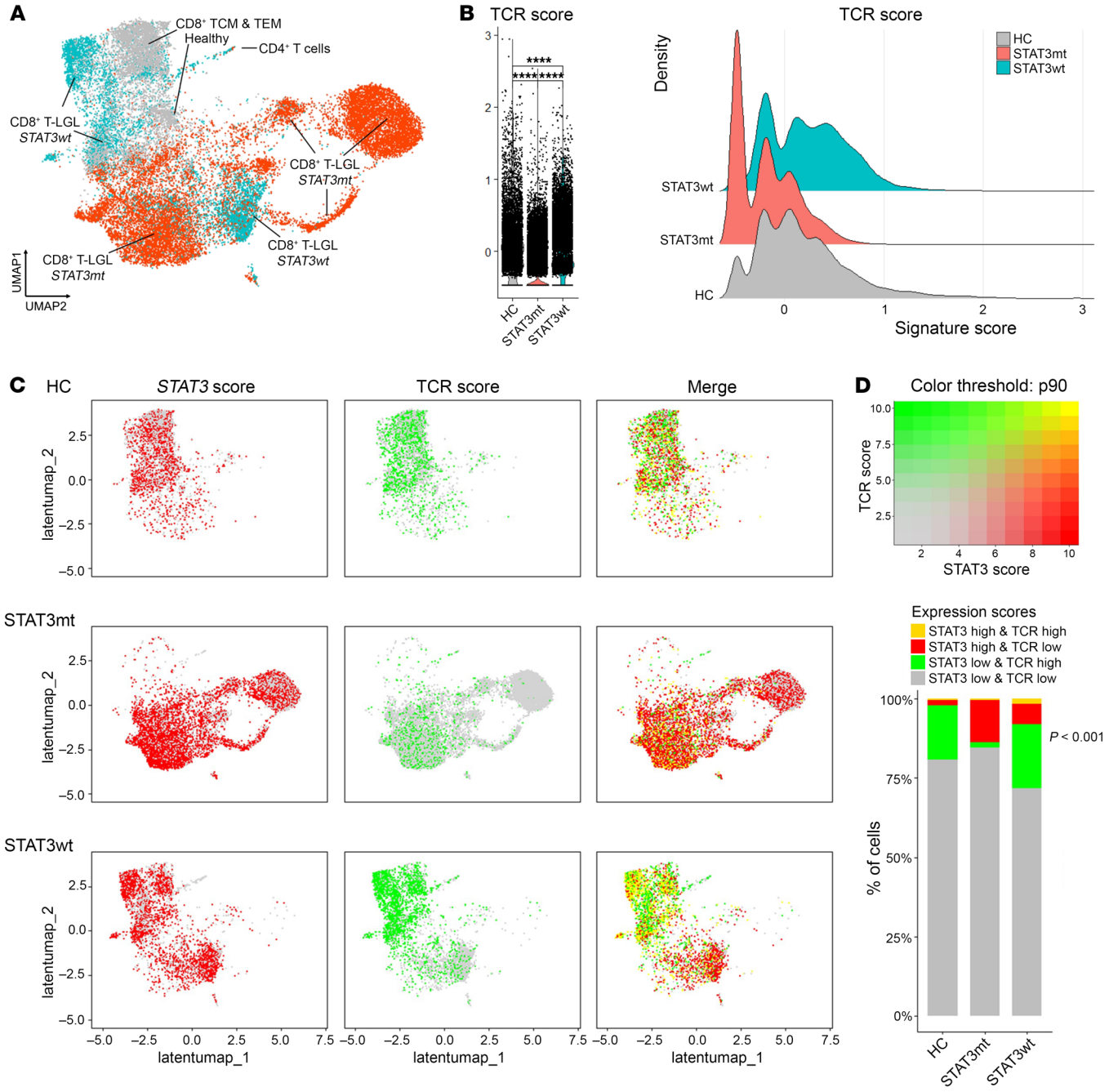


Figure 5. Single-cell RNA+TCRαβ expression analysis of TCR signaling genes in *STAT3*mt and *STAT3*wt cells from T-LGLL samples and healthy controls. (**A**) UMAP representation of the reclustered hyperexpanded T cells (>10 templates) annotated by cell phenotype and STAT3mt status. (**B**) TCR signalosome score in *STAT3*mt versus *STAT3*wt cells. The violin and ridge plots show the TCR score in *STAT3*mt and *STAT3*wt T-LGLL clones and in healthy control hyperexpanded T cells. Wilcoxon's test $P < 0.10$ are shown. (**C**) Coexpression analysis of TCR and *STAT3* signaling scores split by the type of cell. STAT3mt status and scaled expression of *STAT3* activation (red) and TCR signalosome scores (green) are highlighted in the same UMAP representation. Color scale corresponds to the signature score, and the color thresholds correspond to the 90th percentile of the scores. The cells are divided as *STAT3*mt, *STAT3*wt, and healthy controls. Gray dots correspond to *STAT3* and TCR expression below the 90th percentile; red and green signals correspond to strong *STAT3* and TCR signaling, respectively; yellow signals correspond to double STAT3/TCR-high signaling. (**D**) Proportion of cells with high values of *STAT3* and TCR signaling scores according to the STAT3mt status. χ^2 *P* values are shown. ****$P < 0.0001$. HC, healthy controls.

(8). There is, however, limited direct evidence on the role of antigen recognition in T-LGLL or on the identity of the culprit(s) epitope(s). Indeed, no common TCR sequences have been identified across T-LGL clones of different patients (13). However, this does not contradict that T-LGLL can be driven by an abnormal response to an antigen. For instance, it has been shown that, in a significant fraction of T-LGLL patients, the leukemic clones have been found to share CDR3 motifs with their nonleukemic counterpart (13, 46). An increased costimulatory cell-cell and cytokine cross-talk with nonleukemic lymphocytes and antigen-presenting cells

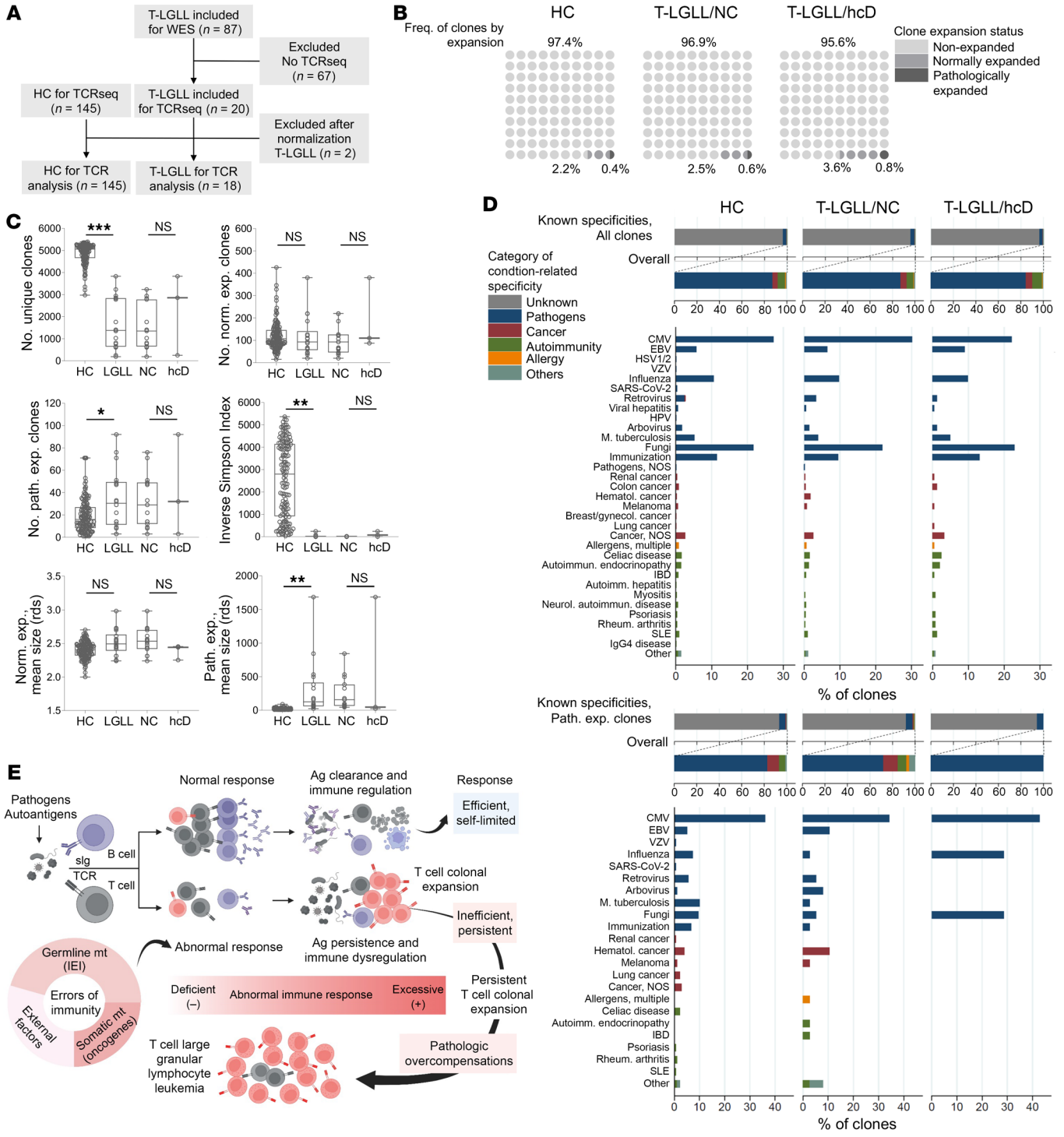


Figure 6. TCR repertoire analyses of the implications of the hcD variants associated with dominant IEI. (**A**) Patient flow for TCR immunosequencing. Of 20 T-LGLL patients with WES and deep TCR sequencing, 18 patients were downstream analyzed after resampling to 5,420 clones. Data from 145 healthy controls from Pagliuca et al. were also used (45). (**B**) Pooled distribution of clones according to the expansion status. Nonexpanded, normally expanded, and pathologically expanded clonotypes are defined by 1, 2–5, and >5 templates, respectively. (**C**) Number of unique, normally expanded, and pathologically expanded clonotypes, inverse Simpson index, and mean clone size between T-LGLL versus healthy controls and between hcD carriers and noncarriers (NC) within the T-LGLL cases. One dot per sample. Mann-Whitney *U* test *P* values are shown. (**D**) Bar plot illustrating the condition-related known specificities of the whole repertoire (upper plot) and only the pathologically expanded clonotypes (lower plot) in healthy controls and T-LGLL patients with/without hcD variants. (**E**) Proposed model on the pathogenic role of underlying IEI in CTL proliferations and clonal shift (created with Biorender.com). Aberrant immune responses caused by genetic or acquired factors, both deficient or hyperreactive abnormal responses, may lead to antigen persistence and/or immune dysregulation and eventually result in T-LGLL as a pathologic overcompensation. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Ag: antigen; CMV, cytomegalovirus; IBD, inflammatory bowel disease; HPV, human papillomavirus; NOS, not otherwise specified; sIg, surface Ig; SLE, systemic lupus erythematosus; VZV, varicella zoster virus.

has also been reported (13). Altogether, current evidence suggests that, despite the fact that the triggering event might be caused by different epitopes, common mechanisms independent of the inciting antigen may be at the root of the disorder.

In this line, we systematically explored in one of the largest genomic T-LGLL cohorts the possibility that the initial overshooting immune response in T-LGLL might be specifically induced in individuals genetically predisposed to errors of immunity. Overall, our results suggest that IEI variants are found in a substantial fraction of T-LGLL cases. Prior to this work, 6 cases of T-LGLL in severe IEI were reported, including 4 cases with recessive SCID (18, 47–49), a patient with X-linked agammaglobulinemia (50), and a case with APS-1 (24). Among them, we highlight the report of 2 identical twins affected by ADA2 deficiency who developed *STAT3*wt T-LGLL as a consistent manifestation with a staggered time course (49). Importantly, T-LGLL is a typical disease of the elderly, making the diagnosis of such severe syndromes during adulthood very unlikely. Conversely, the defects we found here were all heterozygous and mostly associated with dominant, adult-onset IEI, supporting their pathogenic role (22). As compared with classic IEI, and as expected for the growing family of adult-onset IEI, less-deleterious variants for dominant traits and monoallelic variants for recessive diseases were found in our study, probably explaining incomplete penetrance, variable expressivity, and delayed or atypical manifestations.

Despite IEI constituting a growing and highly heterogeneous family of disease, these syndromes can be phenotypically dichotomized as per an excessive or defective immune response (i.e., immune dysregulation and innate/adaptive immunodeficiency syndromes, respectively) (19, 22). Interestingly, both IEI categories were represented by the hcD variants found in the T-LGLL WES cohort, suggesting 2 routes by which IEI may promote CTL expansions: (a) immune dysregulation, intrinsically overshooting autoimmunity, and (b) innate/adaptive immune defects, probably due to increased risk of infection or inability to efficiently clear pathogens. Therefore, both deficient and excessive abnormal immune responses might fuel inefficient, persistent CTL outgrowth. A baseline compromised T cell repertoire may additionally predispose to the oligoclonal/clonal shift. In this study, we found that carriers of IEI hcD variants might exhibit distinct features, such as early T-LGLL diagnosis, low lymphocyte counts, low rates of somatic *STAT3* mutations, and high proportions of hypogammaglobulinemia, autoimmune cytopenias, or BMF. These atypical features reinforce an alternative pathogenesis of modifying genetic factors present in a fraction of cases with T-LGLL. Importantly, in the clinical setting, they may serve as red flags tied to the underlying/cryptic IEI. A better characterization of the immune defects in a patient may open the possibility of using supportive measures (e.g., antimicrobial prophylaxis, Ig replacement) or tailored therapies as well as avoiding multiple lines of ineffective treatment. Ultimately, the accurate diagnosis of a cryptic IEI in young adults with refractory severe cytopenias may lead to evaluation of hematopoietic stem cell transplant as a definitive treatment. Multidisciplinary care is recommended in this complex patient population (51).

In this study, we comprehensively characterized the mutational landscape of T-LGLL. We identified SNVs and CNVs in T cell lymphoid drivers in 67% patients. Mutations in genes other than *STAT3* were detected in 49% of patients, including validated targets such as *STAT5B*, *TNFAIP3*, and epigenetic regulators. We also identified SNVs and CNVs in recurrently mutated genes in T cell neoplasms and lymphoid clonal hemopoiesis. Amplification of *STAT3* and *STAT5A/B* and deletion of *TP53* are frequent events in T cell leukemia/lymphomas, particularly after relapse/transformation as well as in cell lines due to increased genomic instability (30–32, 52). We have confirmed this in a comprehensive analysis of 26 T cell cancer cell lines from DepMap. In contrast, analysis of our WES cohort did not identify such CNVs in T-LGLL samples. We identified, however, CNVs in other T cell lymphoid drivers in 4% of cases. This is consistent with a recent study on 105 T-LGLL patients using WES, which evidenced a similar rate of CNVs (5). The use of whole-genome sequencing, third-generation sequencing based in long-reads, or optical genome mapping would be necessary in future studies to better assess the full CNV spectrum of T-LGLL (53, 54).

Comparison of IEI hcD carriers versus noncarriers revealed different mutational configurations. We highlight a lower proportion of *STAT3*mt cases in carriers than noncarriers. Based on this finding, we investigated the interplay between TCR signaling and *STAT3*mt status by using scRNA+TCR-Seq of T-LGLL samples and RNA-Seq of T cell cancer cell lines. Our genomic T-LGLL cohort did not have coupled RNA-Seq for integrated expression analysis. However, both RNA-Seq datasets utilized instead here allowed the possibility to accurately separate and compare *STAT3*mt versus *STAT3*wt cells. Gene-expression levels, differential expression, and pathway analyses consistently suggested downregulation of TCR-related genes in *STAT3*mt T-LGLs, which were upregulated in *STAT3*wt clones. An analogous bypass of TCR complex signaling by enhanced STAT3 activation has been described in T cell lymphoid neoplasms of other subtypes and was confirmed in our analysis of RNA-Seq of T cell cancer cell lines (36–39). This finding provides more indirect evidence supporting the role of TCR signaling in T-LGLL pathogenesis. Notably, *STAT3*mt is considered a subclonal, late event occurring in the context of an initial oligoclonal/clonal T cell expansion (6, 7, 55). Therefore, TCR signaling might not only be important in *STAT3*wt patients, but also in *STAT3*mt cases prior to the acquisition of the GOF mutation and even after that to support the continuous growth of pathologically expanded *STAT3*wt clones.

Apart from a lower *STAT3*mt rate, IEI hcD carriers had a greater proportion of mutations in genes other than *STAT3* (60%), including an enrichment in hotspot mutations in *TNFAIP3*, *NOTCH1*, and the RAS pathway. Somatic mutations in these genes have been associated with acquired immune dysregulation and lymphoproliferation syndromes. Similar results have been published in a cohort of 17 patients with primary immunodeficiency interrogating somatic mutations in T cells (18, 56). Illustratively, in this study, one patient with SCID due to ADA2 deficiency and *STAT3*wt T-LGL expansion was found to carry a somatic mutation in *KRAS*. Altogether, these findings point toward different molecular configurations driving the T cell expansion resulting in T-LGLL. For patients with variants predisposing to errors of immunity, this might constitute an example of a maladaptive somatic gene rescue, a phenomenon described for patients with SCID or with inherited BMF syndromes (57–60).

Our study has several limitations. The assessment of the functional impact of the numerous variants identified through genomics is an important step in searching for disease-causing mutations.

Given the heterogeneity and diverse mechanism across IEI, we approached this issue by (a) selecting variants overrepresented in our cohort (significant *P* value after multiple-testing correction); (b) using the ACMG criteria for unified analysis; (c) testing the variants by using different in silico prediction tools; and (d) stratifying the analysis of the variants based on pathogenicity, validated phenotypic classifications, and disease inheritance and/or age of onset. Establishment of gene-phenotype correlations can also be challenging, in particular when applied to rare diseases. Therefore, we screened a panel of immune genes only included in the IUIS consensus of IEI (19), an updated classification whereby genotypic data are accompanied with clinical and laboratory phenotypic features. However, given that the selection of genes is empiric, we acknowledge that the list may not be complete. In addition, and most importantly, acquired immunodeficiency due to comorbidities or age-related decline in immune function may also be causative. Assessment of the TCR clonotyping was also limited by the relatively low sample size and the lack of in vitro validations. Specificities were assessed both in the whole repertoire and pathologically expanded clonotypes. Despite the relatively high number of clonotypes included in the reference (>80,000), the proportion of exact sample-reference matches was low (≈5%), a finding that is, however, expected for this approach. Among the matches, most of them were associated with infectious agents, indirectly supporting the notion that microbes might be the prevailing antigenic drivers in some of these cases.

Our study was focused on T-LGLL. Cases with NK cell proliferations were excluded because, although unified by some features, such as *STAT3* mutations, NK-LGLL constitutes a less frequent (<10% of LGLL cases), and less clinically/molecularly defined variant of the disease (7). The diagnostic boundaries between clonal and nonclonal NK cell expansions are blurred, and mutations in *CCL22*, characteristic of NK-LGLL (52, 61), have not been reported in T-LGLL cases, as confirmed by our genomic assessment. Future studies conducted in multicenter, collaborative cohorts should be designed to characterize the immunogenetic landscape of NK-LGLL.

In conclusion, our study proposes an alternative pathogenesis for T-LGLL paradoxically related to underlying seemingly indolent errors of immunity. These, under certain circumstances, might lead to aberrant CTL responses. Altogether, our findings constitute an illustrative lesson of nature on the understanding of clonal CTL expansions and open the horizon of IEI to other clonal hematological disorders and immune-mediated BMF.

Methods

Full description of methods is available in the Supplemental Methods.

Sex as a biological variable

This study examined male and female participants, as both men and women were eligible, and findings were similar for both sexes.

Clinical cohort

This cross-sectional, genetic association study was performed in a cohort of consecutive patients with T-LGLL, diagnosed and managed at Taussig Cancer Center, Cleveland Clinic Foundation from 1998 to 2023, described elsewhere (Table 1) (2). Because of our focus on T cell clonal expansions, cases with NK-LGLL subtype were excluded. Clinical data collected comprised patient demographics, comorbidities, personal and family history of immune disorders or hemato-lymphoid neoplasms, splenomegaly, cytopenias, and BMF (62). Conditions potentially leading to hypogammaglobulinemia (63) were screened in patient's medical charts prior to T-LGLL diagnosis. Clinical outcomes analyzed were treatment initiation, transfusion dependency, splenectomy, hematopoietic stem cell transplant, transformation to high-grade lymphoma, or death.

Laboratory data at diagnosis comprised complete blood count, LGL count, Ig levels, and M protein assessment. Lymphocyte subset characterization by flow cytometry (in $CD45^{Ly}$ gated cells) routinely performed in the clinic in peripheral blood at LGL diagnosis was also collected for the study of the following populations: T cells ($CD3^+$, normal range [NR]: 0.96–2.39 × 10^9/L), $CD4^+$ T helper cells ($CD3^+CD4^+$, NR: 0.53–1.67 × 10^9/L), $CD8^+$ cytotoxic T lymphocytes (CTLs, $CD3^+CD8^+$, NR: 0.28–0.96 × 10^9/L), NK cells ($CD3^-CD16/CD56^+$, NR: 0.10–0.57 × 10^9/L), and B cells ($CD19^+$, NR: 0.08–0.66 × 10^9/L). The Ig levels analyzed included the quantification of IgG (NR: 717–1411 mg/dL), IgA (NR: 78–391 mg/dL), and IgM (NR: 53–334 mg/dL).

Targeted sequencing was performed as previously described using a custom panel for detection of hematological neoplasm gene variants from TruSeq or Nextera platforms (Illumina) (64, 65). *STAT3* mutational status was assessed by deep DNA sequencing (66).

WES

WES was performed by Novogene in genomic DNA (gDNA) extracted from peripheral blood mononuclear cells. Raw read files were first converted to FASTQ format, then aligned to human genome hg38 using the Burrows-Wheeler aligner (BWA) (67). Aligned reads were processed using Genome Analysis Toolkit (GATK), which also extracted candidate variants/ polymorphisms to reduce sequencing errors (68). Variant annotation was performed by using ANNOVAR (69). A stringent categorization algorithm to avoid false positives was devised, removing: (a) variants with minimum depth of less than 10 or less than 4 reads supporting the alternate allele; (b) synonymous SNVs; (c) variants in repetitive genomic regions. The variant coordinates were crosschecked with the list of somatic mutations in the same patients, and any commonalities were omitted from the germline list.

Errors of immunity-linked genes. We screened this genomic cohort for the presence of rare germline variants associated with primary immunodeficiency in a panel comprised of 464 immune genes defined by the 2022 Updated Classification of Human Inborn Errors of Immunity (IEI) of the IUIS Expert Committee (Supplemental Table 1) (19). Rare variants were defined as those with population allele frequencies of less than 1% obtained from gnomAD. Only variants annotated as missense, nonsense, indel, or splice site were considered for downstream analyses. Each variant was assessed according to the ACMG criteria, using ClinVar (20) and VarSome tools (21). We selected pathogenic/likely pathogenic (P/LP) and VUS overrepresented in our cohort, i.e., with a significant corrected *P* value for the comparison of observed versus expected frequencies according to gnomAD. Multiple-testing correction of *P* values by using Benjamini-Hochberg with a FDR level of 0.05 was applied. Exclusion criteria were: (a) variants with a variant allele frequency (VAF) of less than 40%, (b) variants estimated by the ACMG pathogenicity criteria to be benign or likely benign, and (c) nonoverrepresented VUS, corresponding to a Benjamini-Hochberg FDR above 0.05.

T cell lymphoid drivers. We screened this genomic cohort for the presence of somatic variants in a list of 168 recurrent T cell lymphoid

drivers (Supplemental Table 3). The selection of the genes was based on 2 criteria: (a) previously described in T-LGLL according to 2 seminal publications (5, 70); (b) alternatively, not described in T-LGLL but identified as recurrent genes in either mature T cell neoplasms and/or lymphoid clonal hemopoiesis (L-CHIP) (30–33). Missense, nonsense, frameshift, and indel variants were further filtered by pathogenicity criteria according to COSMIC (https://cancer.sanger.ac.uk/cosmic), ClinVar, and VarSome somatic filters. Only P/LP variants were selected to increase stringency in terms of clinical consequences.

Gene-level somatic CNVs were primarily called using CNVkit (71). Values were calculated by mapping genes onto the segment level calls and computing a weighted average along the genomic coordinates. Normalized read depths ($\log_2$), b-allele frequency (BAF), and CN estimates for ref/alt alleles given the VAF data were extracted. CNVs in hypervariable chromosomic regions (Supplemental Table 13) or CNVs observed in general population datasets (DECIPHER, Database of Genomic Variants [DGV]) (72, 73) were excluded.

Variant analysis plan

To determine the burden of the rare variants of potential clinical interest in IEI-linked genes, we estimated (a) the individual IEI mutational burden (number of IEI variants per patient); (b) the combined IEI mutational burden in the cohort (proportion of subjects with at least 1 IEI variant); and (c) the simplified expected probability of finding any of the IEI variants in our cohort (sum of individual IEI variant allelic frequencies according to gnomAD). As a control population for statistical comparisons, we estimated the combined mutational burden of the IEI variants found in a cohort of healthy subjects in *All of US* (74). This is a NIH research program aiming to enroll more than one million of US residents aged 18 years or older to create a nationwide population study cohort. Demographics, surveys, clinical information, and biospecimens are donated. To date, short-read whole genome sequencing is available from 245,368 individuals.

The IEI-linked variants included in this work were further clustered and analyzed according to (a) pathogenicity, (b) immune-functional phenotypic implications, and/or (c) pattern of inheritance according to 2022 IUIS Classification of IEI (Supplemental Table 1) (19), and (d) the presumed age period of onset of the associated IEI (i.e., early vs. adult-onset disease) (19, 22). To establish correlations between genomic and biological or clinical data, we further defined a category of hcD variants, considered more likely to predispose to immune misbalance in a carrier, as those being either: (a) P/LP variants, (b) heterozygous VUS for dominant traits, or (c) homozygous/compound heterozygous P/LP/VUS for recessive diseases. Clinical variables, survival outcomes, and laboratory and biological parameters of carriers versus noncarriers of these high-risk variants were compared.

TCR immunosequencing and analysis

Sequencing of the complementarity determining regions (CDR3) regions of the human TCR β gene was performed using the ImmunoSEQ Assay (Adaptive Biotechnologies), as previously described (45, 75). Deep TCR sequencing data of 145 healthy controls originated from Emerson and DeWitt (original publication and ImmuneACCESS) (43, 44). Downstream analysis of the TCR repertoire was performed exclusively in productive rearrangements (i.e., reads that were in-frame and did not contain a stop codon in their sequence). To overcome intersample differences in depth, downsampling of the TCR repertoires to the optimal threshold of 5,420 clones was performed as a normalization procedure. The diversity metrics calculated per sample included the number of unique clonotypes, unique clone size, and the inverse Simpson index (45) (the lowest value for this index is 1 and the highest value is equal to the number of species). The expansion status of the clones within a repertoire was defined according to the number of templates as (a) nonexpanded (1 template), (b) normally expanded (2–5 templates), (c) pathologically expanded (>5 templates), and (d) hyperexpanded (>10 templates). Condition-related known specificities of the identified clonotypes were annotated according to the dataset from Pagliuca et al. (n = 80,220 reference clonotypes, Supplemental Table 12) (45).

Single-cell RNA+TCRαβ-Seq from T-LGLL and healthy control samples

Preprocessed Seurat objects of scRNA-Seq+ TCRαβ-Seq of flow cytometry–sorted CD45$^+$ blood mononuclear cells from T-LGLL samples (n = 11) and healthy controls (n = 6) from Huuhtanen and Bhattacharya et al., available at https://zenodo.org/records/4739231 (13), were utilized. Clinical characteristics of the samples used are summarized in Supplemental Table 9. Extensive methodological description of this study dataset is available elsewhere (13). We focused our analysis on hyperexpanded T cell clonotypes (>10 TCR templates). Batch-corrected latent embeddings from scVI (version 0.5.0) (76) were used for graph-based clustering and uniform manifold approximation and projection (UMAP) dimensionality reduction implemented in Seurat (v. 3.0.0) with RunUMAP function, and scaled with 3,000 most highly variable genes with the FindVariable function and ScaleData functions with default parameters (77.

Genomics and transcriptomics from mature T cell cancer cell lines

Genomic data from 26 mature T cell neoplasm cell lines was gathered from The Cancer Dependency Map Project (https://depmap.org, DepMap, Broad Institute) (34). Briefly, DepMap Data Release is a publicly available comprehensive omics resource for understanding cancer biology and identifying potential therapeutic targets. We selected all cell lines matching with the context "Mature T NK cell neoplasms." Gene-level damaging-supporting SNVs and CN normalized read datasets were analyzed. Bulk RNA-Seq was available for 22 of the 26 cell lines. Read count data from RNA-Seq by expectation-maximization (RSEM) (unstranded mode) was normalized with the Trimmed Mean of M-values (TMM) method in edgeR default option (78). A summary of main biological characteristics and RNA-Seq data used in this study are provided in Supplemental Table 6.

Gene-expression levels and differential gene-expression analysis

Single-cell and bulk RNA-Seq mean gene-expression levels were compared using t test and Wilcoxon's tests, respectively (79). Differential expression analyses were performed using DESeq2 with default parameters, based on the Wald test with Bonferroni's correction of P values (80). In scRNA-Seq T-LGLL samples, we compared *STAT3*mt versus *STAT3*wt cells. In cell lines with RNA-Seq, we compared *STAT3*mt versus fusion-matched *STAT3*wt cell lines, based on the oncogene fusion present in the *STAT3*mt cells. Enrichment analysis was performed with the list of dysregulated genes (abs $\log_2$ fold change [$\log_2$FC] > 0.2; adjusted P value [Padj] < 0.05) with hypergeometric testing implemented in ClusterProfiler (40) with GO terms gathered from MSigDB (42). Gene-expression scores in scRNA-Seq were calculated with the Seurat AddModuleScore function (41). The TCR

score was calculated with 15 genes, including components of the TCR complex (GO:0042101) and the TCR signalosome (GO:0036398). A score with the same set of genes was also calculated in bulk RNA-Seq from the T cell cancer cell lines as geometric means (13). A *STAT3* score, indicative of *STAT3* activation, was also calculated for scRNA-Seq with 9 genes identified by interrogating the set of genes upregulated in *STAT3*mt versus *STAT3*wt T-LGL clones and selecting those genes matching with either a list of upregulated genes in human cells expressing *STAT3* gathered from MSigDB (42) or with the lists of upregulated genes in *STAT3*mt versus *STAT3*wt T cell neoplasm lines from DepMap generated here. To visualize gene expression in scRNA-Seq, scaled expressions were used with the Seurat FeaturePlot function (77). Gene-expression scores were visualized in scRNA-Seq in a similar fashion. Thresholds corresponding to the 90th percentiles of the gene expression scores were set to identify and quantify the proportion of cells with high TCR/STAT3 signaling scores.

Statistics

Categorical variables were presented as percentages and compared using Pearson's χ^2 and Fisher's exact tests. Continuous variables were presented as mean and SD if normally distributed and as median and IQR if nonnormally distributed.

Differential analysis of categorical variables included Pearson's χ^2 or Fisher's exact tests; comparison of continuous variables included 2-tailed Student's *t* test or nonparametric Mann-Whitney *U* test. Differences with a 2-tailed *P* values less than 0.05 were considered statistically significant. Survival analysis between groups was done with the log-rank test. Associations between clinical data and survival outcomes were assessed with unadjusted (univariable) and adjusted (multivariable) Cox's regression. Overall survival estimations were presented with 95% CI.

Gene pathway analysis was performed with using GeneMANIA (University of Toronto, Canada) (81) and Cytoscape (NIH, Bethesda, MD) (82).

Statistical analysis and graphic representation was performed using GraphPad Prism, version 9.4 (GraphPad Software Inc.), and STATA, version 16 (StataCorp LLC) or R (R Core Team, Vienna, Austria) (83). The R packages and functions used for SNVs, CNVs, and scRNA-Seq and bulk RNA-Seq analyses are indicated as per their mention throughout this section.

Study approval

This research was approved by the institutional review board of Cleveland Clinic. Patient enrollment into clinical registry and specimen collection was performed in accordance with IRB 15-1278 and IRB 5024, respectively. The study was performed in accordance with the Declaration of Helsinki. Written, informed consent was obtained from all participants.

Data availability

Processed genomic data from patient samples used in this study is available at Zenodo (https://doi.org/10.5281/zenodo.14888890). Processed scRNA+TCR from T-LGLL and healthy control samples is available at Zenodo (https://zenodo.org/records/4739231). Genomic and transcriptomic data from T cell cancer cell lines is available at DepMap (https://depmap.org). All other information, including values of all data points in graphs, is provided in the Supplemental Material and Supporting Data Values files. Additional information not included in the Supplemental Material and Supporting Data Values files is available from the authors upon reasonable request.

Author contributions

CBP collected and analyzed clinical and molecular data, interpreted results, and wrote the manuscript. MW and ZB performed sample processing. CG, NK, AM, LG, CH, NDW, SU, MO, AA, and OO collected data and edited the manuscript. JH and SM generated and provided scRNA-Seq data for analysis and edited the manuscript. JH, YK, SP, and AD analyzed molecular data and edited the manuscript. CG, VV, and JPM provided invaluable help to the manuscript preparation, generated and conceived the study design, designed figures and tables, and wrote the manuscript. All authors participated in the critical review of the final paper and submission.

Acknowledgments

We thank our sources of funding: the NIH (R35HL135795to JPM), the Aplastic Anemia and MDS International Foundation (AA&MDSIF to CBP, VV, and JPM), the VeloSano 9 Pilot Award, the Vera and Joseph Dresner Foundation–MDS (to VV), the European Research Council (M-IMM 647355 to SM), the Academy of Finland Heal-Art consortium (314442 to SM), ERA PerMed (JAK-STAT-TARGET consortium to SM), the Sigrid Juselius Foundation (to SM), and the Cancer Foundation Finland (to SM). CBP has a postdoctoral fellowship from Instituto de Salud Carlos III (JR22/00041). CG was supported by a grant from the Edward P. Evans Foundation. We are deeply grateful to all patients who participated in the study, as well as to *All of Us* participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's *All of Us* Research Program for making available the participant data examined in this study.

Address correspondence to: Jaroslaw P. Maciejewski, Department of Translational Hematology and Oncology Research, Taussig Cancer Institute. 9620 Carnegie Ave n building, Building, NE6-314, Cleveland, Ohio 44106, USA. Email: maciejj@ccf.org.

1. Zawit M, et al. Large granular lymphocytic leukemia: from immunopathogenesis to treatment of refractory disease. *Cancers (Basel)*. 2021;13(17):4418.
2. Sanikommu SR, et al. Clinical features and treatment outcomes in larg e granular lymphocytic leukemia (LGLL). *Leuk Lymphoma*. 2018;59(2):416–422.
3. Yang J, et al. TRAIL mediates and sustains constitutive NF-κB activation in LGL leukemia. *Blood*. 2018;131(25):2803–2815.
4. Zhang R, et al. Network model of survival signaling in large granular lymphocyte leukemia. *Proc Natl Acad Sci U S A*. 2008;105(42):16308–16313.
5. Cheon H, et al. Genomic landscape of TCRαβ and TCRγδ T-large granular lymphocyte leukemia. *Blood*. 2022;139(20):3058–3072.
6. Koskela HLM, et al. Somatic STAT3 mutations in large granular lymphocytic leukemia. *N Engl J Med*. 2012;366(20):1905–1913.
7. Marchand T, et al. A modern view of LGL leukemia. *Blood*. 2024;144(18):1910–1923.
8. Bravo-Pérez C, et al. Persistent large granular lymphocyte clonal expansions: "the root of many evils"-and of some goodness. *Cancers (Basel)*. 2022;14(5):1340.
9. Kothapalli R, et al. Constitutive production of proinflammatory cytokines RANTES, MIP-1beta and IL-18 characterizes LGL leukemia. *Int J Oncol*. 2005;26(2):529–535.
10. Chen J, et al. Increased serum soluble IL-15Rα levels in T-cell large granular lymphocyte leukemia. *Blood*. 2012;119(1):137–143.

11. Teramo A, et al. Intrinsic and extrinsic mechanisms contribute to maintain the JAK/STAT pathway aberrantly activated in T-type large granular lymphocyte leukemia. *Blood*. 2013;121(19):3843–54, S1.
12. Brammer JE, et al. Effective treatment with the selective cytokine inhibitor BNZ-1 reveals the cytokine dependency of T-LGL leukemia. *Blood*. 2023;142(15):1271–1280.
13. Huuhtanen J, et al. Single-cell characterization of leukemic and non-leukemic immune repertoires in CD8+ T-cell large granular lymphocytic leukemia. *Nat Commun*. 2022;13(1):1981.
14. Calabretto G, et al. LGL clonal expansion and unexplained cytopenia: two clues don't make an evidence. *Cancers (Basel)*. 2022;14(21):5236.
15. Viny AD, et al. Chronic B-cell dyscrasias are an important clinical feature of T-LGL leukemia. *Leuk Lymphoma*. 2008;49(5):932–938.
16. Awada H, et al. Large granular lymphocytic leukaemia after solid organ and haematopoietic stem cell transplantation. *Br J Haematol*. 2020;189(2):318–322.
17. Gurnari C, et al. Novel invariant features of Good syndrome. *Leukemia*. 2021;35(6):1792–1796.
18. Savola P, et al. Somatic mutations and T-cell clonality in patients with immunodeficiency. *Haematologica*. 2020;105(12):2757–2768.
19. Bousfiha A, et al. The 2022 Update of IUIS phenotypical classification for human inborn errors of immunity. *J Clin Immunol*. 2022;42(7):1508–1520.
20. Landrum MJ, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(d1):D1062–D1067.
21. Kopanos C, et al. VarSome: the human genomic variant search engine. *Bioinformatics*. 2019;35(11):1978–1980.
22. Staels F, et al. Monogenic adult-onset inborn errors of immunity. *Front Immunol*. 2021;12:753978.
23. Oftedal BE, et al. Dominant mutations in the autoimmune regulator AIRE are associated with common organ-specific autoimmune diseases. *Immunity*. 2015;42(6):1185–1196.
24. Dhodapkar MV, et al. T-cell large granular lymphocytic leukemia and pure red cell aplasia in a patient with type I autoimmune polyendocrinopathy: response to immunosuppressive therapy. *Mayo Clin Proc*. 1994;69(11):1085–1088.
25. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*. 2004;11(2–3):377–394.
26. Makaryan V, et al. TCIRG1-associated congenital neutropenia. *Hum Mutat*. 2014;35(7):824–827.
27. Voskoboinik I, et al. Perforin and granzymes: function, dysfunction and human pathology. *Nat Rev Immunol*. 2015;15(6):388–400.
28. Solomou EE, et al. Perforin gene mutations in patients with acquired aplastic anemia. *Blood*. 2007;109(12):5234–5237.
29. Afzali B, et al. BACH2 immunodeficiency illustrates an association between super-enhancers and haploinsufficiency. *Nat Immunol*. 2017;18(7):813–823.
30. Sandell RF, et al. Genetic landscape and classification of peripheral T cell lymphomas. *Curr Oncol Rep*. 2017;19(4):28.
31. da Silva Almeida AC, et al. The mutational landscape of cutaneous T cell lymphoma and Sézary syndrome. *Nat Genet*. 2015;47(12):1465–1470.
32. Pizzi M, et al. Pathogenesis of peripheral T cell lymphoma. *Annu Rev Pathol*. 2018;13:293–320.
33. Niroula A, et al. Distinction of lymphoid and myeloid clonal hematopoiesis. *Nat Med*. 2021;27(11):1921–1927.
34. Tsherniak A, et al. Defining a cancer dependency map. *Cell*. 2017;170(3):564–576.
35. Leblanc F, et al. Large granular lymphocyte leukemia: from dysregulated pathways to therapeutic targets. *Future Oncol*. 2012;8(7):787–801.
36. Qin Z, et al. TCR signaling induces STAT3 phosphorylation to promote TH17 cell differentiation. *J Exp Med*. 2024;221(3):e20230683.
37. Seffens A, et al. STAT3 dysregulation in mature T and NK cell lymphomas. *Cancers (Basel)*. 2019;11(11):1711.
38. Moerman-Herzog AM, et al. Transcriptome analysis of Sézary syndrome and lymphocytic-variant hypereosinophilic syndrome T cells reveals common and divergent genes. *Oncotarget*. 2019;10(49):5052–5069.
39. Gaud G, et al. Regulatory mechanisms in T cell receptor signalling. *Nat Rev Immunol*. 2018;18(8):485–497.
40. Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–287.
41. Tirosh I, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016;352(6282):189–196.
42. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.
43. Emerson RO, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet*. 2017;49(5):659–665.
44. Dean J, et al. Annotation of pseudogenic gene segments by massively parallel sequencing of rearranged lymphocyte receptor loci. *Genome Med*. 2015;7:123.
45. Pagliuca S, et al. Clinical and basic implications of dynamic T cell receptor clonotyping in hematopoietic cell transplantation. *JCI Insight*. 2021;6(13):e149080.
46. Clemente MJ, et al. Deep sequencing of the T-cell receptor repertoire in CD8+ T-large granular lymphocyte leukemia identifies signature landscapes. *Blood*. 2013;122(25):4077–4085.
47. Savaşan S, et al. Different clonal T-large granular lymphocyte proliferations in SCID. *J Clin Immunol*. 2019;39(3):245–248.
48. Bajin IY, et al. Atypical combined immunodeficiency due to Artemis defect: a case presenting as hyperimmunoglobulin M syndrome and with LGLL. *Mol Immunol*. 2013;56(4):354–357.
49. Barzaghi F, et al. Case Report: Consistent disease manifestations with a staggered time course in two identical twins affected by adenosine deaminase 2 deficiency. *Front Immunol*. 2022;13:910021.
50. Daou H, et al. CD8-positive cutaneous lymphoproliferation associated with large granular lymphocyte leukemia in a patient with X-linked agammaglobulinemia. *J Cutan Pathol*. 2021;48(4):567–571.
51. Seidel MG. Treatment of immune-mediated cytopenias in patients with primary immunodeficiencies and immune regulatory disorders (PIRDs). *Hematology Am Soc Hematol Educ Program*. 2020;2020(1):673–679.
52. Chen Z, et al. Full-spectral genome analysis of natural killer/T cell lymphoma highlights impacts of genome instability in driving its progression. *Genome Med*. 2024;16(1):48.
53. Coccaro N, et al. Feasibility of optical genome mapping in cytogenetic diagnostics of hematological neoplasms: a new way to look at DNA. *Diagnostics (Basel)*. 2023;13(11):1841.
54. Bravo-Perez C, et al. The whole is greater than the sum of its parts: long-read sequencing for solving clinical problems in haematology. *J Cell Mol Med*. 2024;28(3):e17961.
55. Kerr CM, et al. Subclonal STAT3 mutations solidify clonal dominance. *Blood Adv*. 2019;3(6):917–921.
56. Lundgren S, et al. Somatic mutations associate with clonal expansion of CD8+ T cells. *Sci Adv*. 2024;10(23):eadj0787.
57. Aluri J, Cooper MA. Somatic mosaicism in inborn errors of immunity: Current knowledge, challenges, and future perspectives. *Semin Immunol*. 2023;67:101761.
58. Gutierrez-Rodrigues F, et al. Somatic mosaicism in inherited bone marrow failure syndromes. *Best Pract Res Clin Haematol*. 2021;34(2):101279.
59. Lasho T, Patnaik MM. Adaptive and maladaptive clonal hematopoiesis in telomere biology disorders. *Curr Hematol Malig Rep*. 2024;19(1):35–44.
60. Gordon J, et al. Non-canonical FLT3 alterations reveal novel germline FLT3 variants leading to somatic gene rescue mutations. *Blood Cancer J*. 2024;14(1):125.
61. Baer C, et al. CCL22 mutations drive natural killer cell lymphoproliferative disease by deregulating microenvironmental crosstalk. *Nat Genet*. 2022;54(5):637–648.
62. Alaggio R, et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: lymphoid neoplasms. *Leukemia*. 2022;36(7):1720–1748.
63. Compagno N, et al. Immunoglobulin replacement therapy in secondary hypogammaglobulinemia. *Front Immunol*. 2014;5:626.
64. Gurnari C, et al. A study of telomerase reverse transcriptase rare variants in myeloid neoplasia. *Hematol Oncol*. 2022;40(4):812–817.
65. Adema V, et al. Pathophysiologic and clinical implications of molecular profiles resultant from deletion 5q. *EBioMedicine*. 2022;80:104059.
66. Andersson E, et al. Activating somatic mutations outside the SH2-domain of STAT3 in LGL leukemia. *Leukemia*. 2016;30(5):1204–1208.
67. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–595.
68. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303.
69. Wang K, et al. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
70. Coppe A, et al. Genomic landscape characterization of large granular lymphocyte leukemia

with a systems genetics approach. *Leukemia*. 2017;31(5):1243–1246.
71. Talevich E, et al. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol*. 2016;12(4):e1004873.
72. MacDonald JR, et al. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res*. 2014;42(d1):D986–D992.
73. Firth HV, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am J Hum Genet*. 2009;84(4):524–533.
74. Denny JC, et al. The "all of us" research program. *N Engl J Med*. 2019;381(7):668–676.
75. Robins HS, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood*. 2009;114(19):4099–4107.
76. Lopez R, et al. Deep generative modeling for single-cell transcriptomics. *Nat Methods*. 2018;15(12):1053–1058.
77. Stuart T, et al. Comprehensive integration of single-cell data. *Cell*. 2019;177(7):1888–1902.
78. Robinson MD, et al. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–140.
79. Soneson C, Robinson MD. Bias, robustness and scalability in single-cell differential expression analysis. *Nat Methods*. 2018;15(4):255–261.
80. Love MI, et al. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
81. Warde-Farley D, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010;38(suppl_2):214–220.
82. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–2504.
83. *R: A language and environment for statistical computing*. Version 4.4.3. R Foundation for Statistical Computing; 2025. https://www.R-project.org/.