Protein functionality as a potential bottleneck for somatic revertant variants

Kaiser, F.M.P.; Reisli, I.; Pico-Knijnenburg, I.; Langerak, A.W.; Kavelaars, F.G.; Artac, H.; ... ; Burg, M. van der

Protein functionality as a potential bottleneck for somatic revertant variants

To the Editor:

Somatic genome stability has been considered the rule, rather than the exception; yet, somatic mutations are driving factors in many diseases with a genetic background. Conversely, somatic reversion of germline mutations has been shown to mitigate pathogenicity by partially or completely restoring gene function. Such somatic revertant variants have been observed across a broad spectrum of immunodeficiencies and often result in atypical phenotypes that present a diagnostic and therapeutic challenge.[1] In the majority of cases, one or a few reverting somatic variants have been identified per patient and the limitations of this "natural form of gene therapy" are unknown.[1]

We describe a 1-year-10-month-old patient with a germline-encoded, homozygous 2 bp deletion in the *CD247* gene (*CD247* c.43_44delCA; p.Gln15ValfsTer72) (Fig 1, *A* and *B*), which encodes for the CD3ζ chain, and dramatically reduced T-cell numbers (Fig 1, *C* and *D*; see Fig E1 and Table E1 in this article's Online Repository at www.jacionline.org) combined with a relatively mild clinical phenotype (see the Case Report in this article's Online Repository at www.jacionline.org). The mutation maps to the signal peptide of the protein and results in a premature stop codon at amino acid position 72. CD3ζ is required for assembly of T-cell receptor (TCR) complexes and the subsequent shuttling to the cell membrane.[2] Given the essential role in TCR expression, CD3ζ deficiency presents as a classic $T^-/B^+/NK^+$ severe combined immunodeficiency with reduced peripheral T lymphocytes devoid of TCR/CD3 surface coexpression.[3,4] Unexpectedly, we identified a small population of T cells with TCRαβ/CD3 coexpression in the patient (Fig 1, *E* and *F*). We hypothesized that the mild clinical phenotype was caused by a somatic reversion of the mutation in this population and therefore performed deep sequencing of the first exon of the *CD247* gene in $CD3^+$, $TCR\alpha\beta^+$ T cells to identify potential revertant variants. We discovered 52 unique somatic variants of which the majority of 49 restored the reading

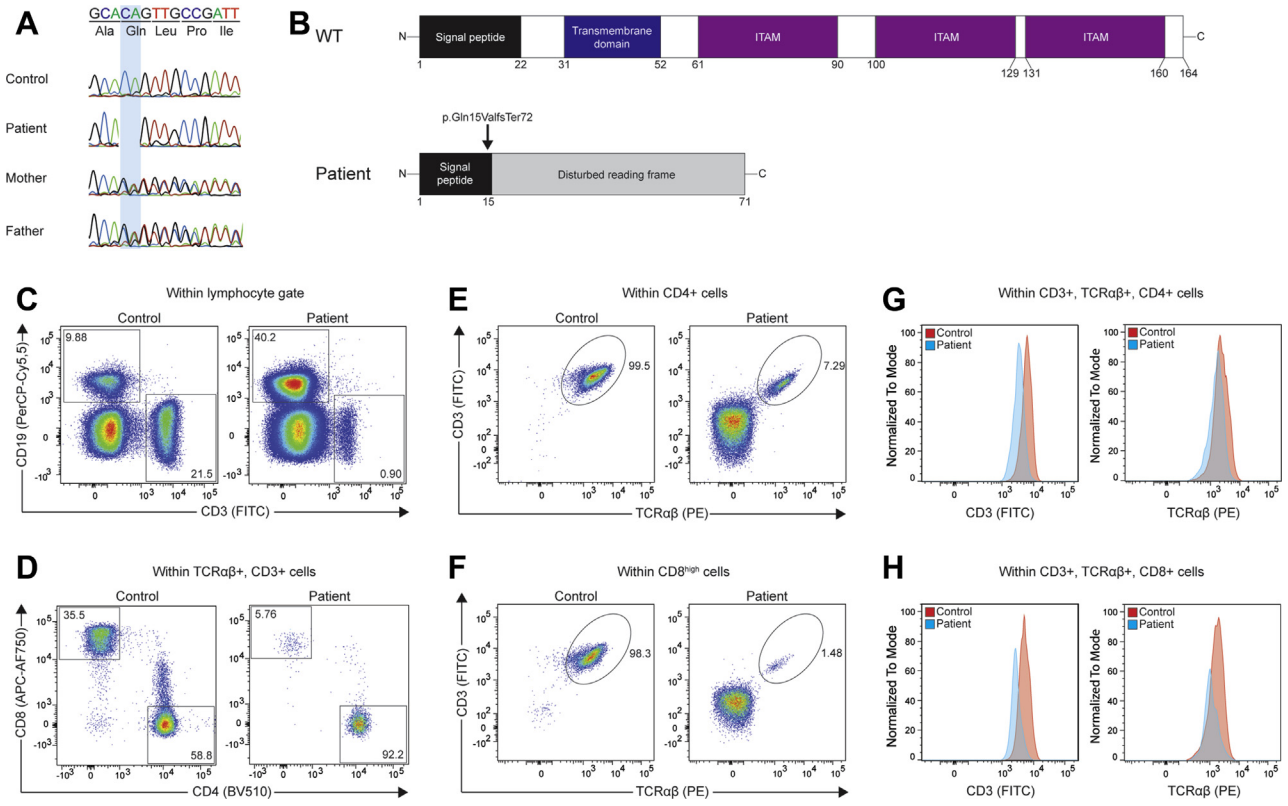

FIG 1. Genetic diagnosis and immunophenotyping of patient. **A,** Sequencing chromatogram showing homozygous and heterozygous germline mutation in patient and parents, respectively. **B,** Schematic representation of wild-type (*WT*) and patient CD3ζ chain with germline mutation indicated by *black arrow*. **C-H,** Flow cytometric immunophenotyping of PBMCs. **C,** $CD3^+$ T and $CD19^+$ B cells within lymphogate. **D,** Proportion of $CD4^+$ and $CD8^+$ T cells after pregating on $CD3^+$, $TCR\alpha\beta^+$ cells. **E** and **F,** Expression of CD3 and TCRαβ after exclusion of B and natural killer cells and pregating on $CD4^+$ and $CD8^{high}$ T lymphocytes, respectively. **G** and **H,** Histogram plots showing expression of CD3 and TCRαβ within $CD4^+$ and $CD8^{high}$ populations, respectively. *ITAM,* Immunoreceptor tyrosine-based activation motif.

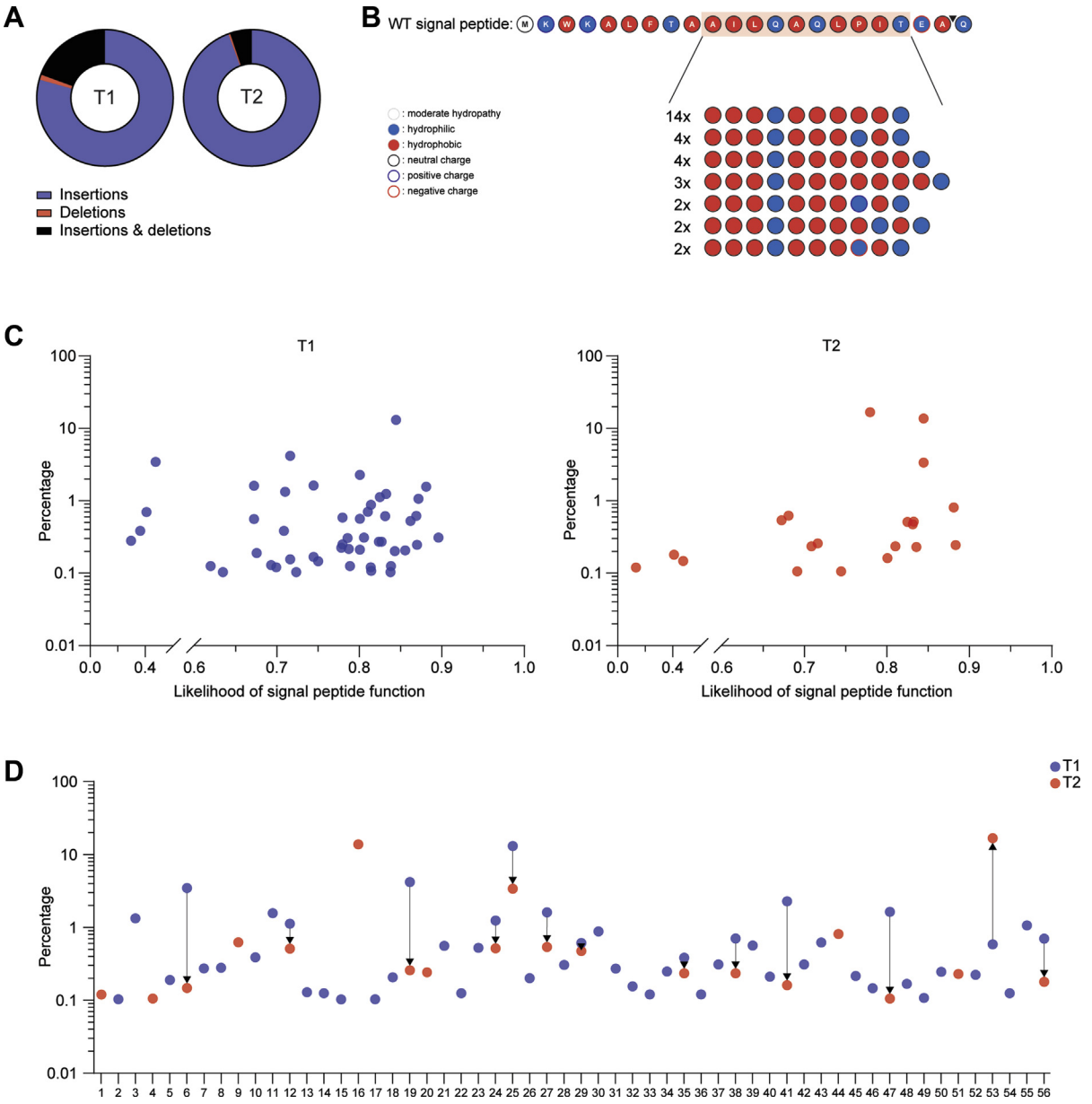

FIG 2. Complex spectrum of somatic reverting variants in CD3$^+$, TCR$\alpha\beta^+$ T cells. **A,** Distribution of insertions and deletions in CD3$^+$, TCR$\alpha\beta^+$ T cells at both time points, referred to as T1 (n = 52) and T2 (n = 23). **B,** Identical somatic revertant variants based on biochemical characteristics of amino acids with indicated frequency of recurrence for both time points. Amino acids are symbolized as indicated in the legend. **C,** Somatic variants that restore the reading frame of the protein at both time points with likelihood of signal peptide function and relative frequency at given time point. **D,** Percentage of somatic variants that restore the reading frame of the protein and relative distribution at both time points. Upward and downward pointing *arrows* indicate relative increase or decrease between both time points, respectively. The number of each variant is based on Fig E2.

frame of the gene and thus prevented the premature stop during translation (see Fig E2 and Table E2 in this article's Online Repository at www.jacionline.org). The variants were mostly located downstream of the *CD247* c.43_44delCA mutation within the region that encodes the signal peptide and mainly encompassed in-frame additions of 2 or 5 nucleotides (Fig 2, *A*, Fig E2). To follow up, we repeated the analysis with a sample that had been obtained 1 year later at the age of 2 years and 10 months. At this time point, we detected 23 somatic variants, including 9 novel variants that had not been present earlier (Fig 2, *A*, Fig E2, Table E2). Twenty variants restored the disrupted reading frame of the protein. The disease-causing germline mutation was found in approximately one half of the reads of the CD3$^+$ population at both time points (Table E2),

suggesting that most CD3-expressing T cells were compound heterozygous for the germline mutation and a somatic variant, which is also reflected by the reduced CD3 and TCR expression (Fig 1, *G* and *H*). Of note, we only detected a few rare somatic variants in the $CD3^-$ T-cell population and none reverted the germline mutation (see Table E3 in this article's Online Repository at www.jacionline.org). Furthermore, we could only identify the germline mutation in the natural killer cells of the patient and no variants in T cells of healthy controls (see Table E4 in this article's Online Repository at www.jacionline.org).

Signal peptides serve to shuttle proteins to the secretory pathway, membranous cellular compartments, or the plasma membrane prior to their cleavage from the mature protein.[5,6] Interestingly, signal peptides usually do not have a fixed primary sequence, but begin with a few positively charged amino acids followed by a long stretch of hydrophobic residues. Most somatic variants retained such a characteristic distribution of amino acids despite a difference in the primary structure and several variants were found recurrently (Fig 2, *B*, Fig E2). Bioinformatic analysis of each somatic variant confirmed that the majority of variants conferred adequate signal peptide function, based on which we consider them somatic reverting variants (Fig 2, *C*, Fig E2).[7]

These observations raise the question at which stage the revertant variants had arisen. Given that the spectrum of variants decreased after 1 year (Fig 2, *D*, Fig E2), one could speculate that dominant clones with the potential for self-renewal had been selected. However, the spectrum of variants alternated between the 2 time points and all except for 1 were found at a lower frequency in the second sample (Fig 2, *D*, Fig E2), suggesting that the generation of T cells with revertant variants was an ongoing process that occurred at later stages of development. In addition, varying antigen exposure could have shaped the distribution of functional T-cell clones with somatic revertants.

Somatic mosaicism caused by revertant variants has been observed in many cases of severe combined immunodeficiency—including several reports on *CD247* deficiency—but never to this extreme extent, which may be explained by the lack of a deep-sequencing–based approach.[3,8,9] We believe that the unique location of the disease-causing mutation in the signal peptide sequence is integral to understanding the pathogenicity of this case. Disease-causing germline mutations are mostly located in regions that are integral to protein function. In order to revert the deleterious effects of such a mutation, somatic reverting variants must therefore either directly restore the primary sequence of the protein, encode for an amino acid with similar properties, or remove an adjacent nonsense or frameshift mutation without affecting the functionality of the domain. Because signal peptides are cleaved from most mature proteins and lack a highly conserved motif,[5] they may tolerate a certain degree of mutations as long as the reading frame of the mature protein remains unaffected, which we have observed in our case for most revertant variants. We therefore postulate that the unique location of the germline mutation is key to unmasking the striking spectrum of revertant mutations within the T-cell pool.

The true degree of mosaicism may be difficult to assess as the majority of variants in coding regions could potentially alter protein function and would therefore subject the mutated cell to positive or negative selection, thereby affecting its frequency within the cell pool. However, if the affected domain can buffer a certain degree of genetic variability—such as signal peptides—the true spectrum of mosaicism may be unmasked. We hypothesize that somatic revertant variants in highly proliferative tissues—such as the hematopoietic system—are common and suggest that somatic revertant variants occur frequently in the slipstream of pathogenic germline mutations, but are limited by the necessity for protein functionality, which acts as the bottleneck.

We thank the patient and the family for their participation in the study. We are indebted to Barbara Barendregt and Peter Valk for technical support. The research for this manuscript was performed within the framework of the Erasmus Postgraduate School Molecular Medicine.

Fabian M. P. Kaiser, MD, MSc[a,b]
Ismail Reisli, MD[c]
Ingrid Pico-Knijnenburg, BSc[d]
Anton W. Langerak, PhD[a]
François G. Kavelaars, BSc[e]
Hasibe Artac, MD[f]
Hanna IJspeert, PhD[a]*
Mirjam van der Burg, PhD[a,d]*

From [a]the Department of Immunology, [b]the Department of Pediatrics, and [c]the Department of Hematology, University Medical Center Rotterdam, Rotterdam, The Netherlands; [c]the Department of Pediatric Immunology and Allergy, Meram Medical Faculty, Necmettin Erbakan University, Konya, Turkey; [d]the Laboratory for Pediatric Immunology, Department of Pediatrics, Leiden University Medical Center, Leiden, The Netherlands; and [f]the Department of Pediatric Immunology and Allergy, Selçuk University Medical Faculty, Konya, Turkey. E-mail: m.van_der_burg@lumc.nl.

*These authors contributed equally to this work.

This work has been funded by the Stichting Sophia Kinderziekenhuis Fonds (grant no. S15-07 Genes and Immunity in SCID) (M.v.d.B.).

Disclosure of potential conflict of interest: The authors declare no competing financial interests.

REFERENCES

1. Mensa-Vilaro A, Bravo Garcia-Morato M, de la Calle-Martin O, Franco-Jarava C, Martinez-Saavedra MT, Gonzalez-Granado LI, et al. Unexpected relevant role of gene mosaicism in patients with primary immunodeficiency diseases. J Allergy Clin Immunol 2019;143:359-68.
2. Call ME, Wucherpfennig KW. The T cell receptor: critical role of the membrane environment in receptor assembly and function. Annu Rev Immunol 2005;23:101-25.
3. Rieux-Laucat F, Hivroz C, Lim A, Mateo V, Pellier I, Selz F, et al. Inherited and somatic CD3zeta mutations in a patient with T-cell deficiency. N Engl J Med 2006;354:1913-21.
4. Roberts JL, Lauritsen JP, Cooney M, Parrott RE, Sajaroff EO, Win CM, et al. T-B+NK+ severe combined immunodeficiency caused by complete deficiency of the CD3zeta subunit of the T-cell antigen receptor complex. Blood 2007;109:3198-206.
5. von Heijne G. The signal peptide. J Membr Biol 1990;115:195-201.
6. Harter C, Wieland F. The secretory pathway: mechanisms of protein sorting and transport. Biochim Biophys Acta 1996;1286:75-93.
7. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 2019;37:420-3.
8. Blazquez-Moreno A, Perez-Portilla A, Agundez-Llaca M, Dukovska D, Vales-Gomez M, Aydogmus C, et al. Analysis of the recovery of CD247 expression in a PID patient: insights into the spontaneous repair of defective genes. Blood 2017;130:1205-8.
9. Marin AV, Jimenez-Reinoso A, Briones AC, Munoz-Ruiz M, Aydogmus C, Pasick LJ, et al. Primary T-cell immunodeficiency with functional revertant somatic mosaicism in CD247. J Allergy Clin Immunol 2017;139:347-9.e8.

Available online May 15, 2020.
https://doi.org/10.1016/j.jaci.2020.04.045

STUDY DESIGN

Peripheral blood and clinical data were obtained according to the guidelines of the Medical Ethics Committees of Meram Medical Faculty at Selçuk University in Konya and the Erasmus Medical Center in Rotterdam. The family gave written informed consent to the study. Detailed information regarding the methodology follows.

METHODS

Immunophenotyping and FACS

Flow cytometric analysis and fluorescence-activated cell sorting (FACS) was performed with antibodies against CD3 (FITC; clone SK7; BD Biosciences, San Jose, Calif), CD19 (PerCP-Cy5,5; clone SJ25C1; BD Biosciences); CD16 (PE; clone 3G8; BD Biosciences); CD56 (APC; N901 clone NKH1; Beckman Coulter Immunotech; Marseille, France); CD4 (BV510; clone OKT4; BioLegend, San Diego, Calif); CD8 (APC-AF750; clone B9.11; Beckman Coulter, Brea, Calif); TCRαβ (PE; clone IP26A; Beckman Coulter). Flow cytometric analysis and FACS were performed on a BD FACSCanto II cell analyzer and a BD FACSAria III sorter (BD, Franklin Lakes, NJ), respectively. Data analysis was performed with FlowJo (version 10; BD).

Sanger sequencing

Candidate genes (*CD3D*, *CD3E*, *CD3G,* and *CD247*) were amplified by PCR with AmpliTaq Gold DNA Polymerase (Thermo Fisher Scientific, Waltham, Mass) and subjected to Sanger sequencing as part of routine diagnostics.

Deep sequencing of *CD247*

Sorted $CD3^+$, $TCR\alpha\beta^+$ T cells were centrifuged in a microcentrifuge tube and the cell pellet was resuspended in 20 μL lysis buffer (10 mmol/L Tris-HCl [pH 7.6], 50 mmol/L NaCl, 6.25 mmol/L $MgCl_2$, 0.045% NP40, 0.45% Tween-20). A total volume of 1 μL proteinase K (20 mg/mL) was added and samples were incubated for 1 hour at 56°C before heat-inactivation for 15 minutes at 95°C. Exon 1 of *CD247* was amplified by PCR and subjected to deep sequencing on either a 454 GS junior instrument (Roche, Branford, Conn) or the MiSeq System (Illumina, San Diego, Calif).

For Roche 454 sequencing, exon 1 and flanking parts of the intron were amplified using primers that were adapted with Roche Lib-A adapters and sample-specific multiplex identifier (MID) tags:

Forward: 5′-CAGACAGATACATACACACACCCCAA-3′

Reverse: 5′-AAGGAGACCCCAGCCCCTCAC-3′

PCR products were purified by means of gel extraction using the QIAgen Gel Extraction Kit (Qiagen, Hilden, Germany) and Agencourt AMPure XP beads (Beckman Coulter). Subsequently, DNA concentrations of the libraries were measured using the Quant-iT Picogreen dsDNA Assay Kit (Invitrogen, Thermo Fisher Scientific).

For the MiSeq System, the coding region of *CD247* was amplified by PCR over 30 cycles with the following primers:

Forward: 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGC CTCTTTCTGAGGGAAA-3′

Reverse: 5′-TCGCGAGTTAATGCAACGATCGTCGAAATTCGCTCAC TTGCCCATTGATTTGA-3′

Subsequently, PCR products were purified using Ampure XP beads (Beckman Coulter) followed by a nested PCR reaction (10 cycles) to include the sample-specific indices and Illumina sequencing adapters, using primers from the Truseq Custom Amplicon Index Kit (Illumina). The final concentrations of the libraries were measured using the Quant-it Picogreen dsDNA Assay Kit (Invitrogen). Libraries were paired-end sequenced (2 × 221 bps) on a MiSeq System with use of a MiSeq Reagent Kit v3, according to the manufacturer's protocol (Illumina). Paired sequences were merged using paired-end read merger (PEAR) to create a FASTQ file of each sample.[E1] Sequences were filtered and analysed using Microsoft Excel 2016. Only sequences with an exact match of the first and last 8 nucleotides of the coding region of exon 1 were included for analysis. Per variant, the frequency of variant reads and the average quality score per base were calculated. Only variant reads that were present in >0.1% of reads with a minimum average quality score above 20 were included for analysis. All detected variants were compared to the reference sequence and analyzed by 2 persons individually.

Bioinformatic prediction of signal-peptide function

Prediction of signal peptide function was performed with the software SignalP 5.0 according to the authors' instructions (http://www.cbs.dtu.dk/services/SignalP/).[E2]

CASE REPORT

We describe a girl of consanguineous parents with a history of recurrent infections and lymphopenia that was admitted to our hospital at the age of 1 year and 10 months with severe tachypnea, productive coughing, and fever. Prior to admission, the patient had suffered from recurring bouts of oral moniliasis, sinusitis, otitis media, and pneumonia. She regularly required intravenous antibiotic therapy for the control of her infections. After BCG vaccination, she developed suppurative lymphadenitis. Her mother and father had not shown any signs of immunodeficiency. Her paternal aunt had passed away as an infant due to unknown reasons.

On physical examination, she was severely tachypneic (64 breaths/min) and showed diffuse bilateral crackles on pulmonary auscultation. Her tonsils and lymph nodes did not show any gross alterations in size. Her height was 84 cm (25%-50%), her weight was 11 kg (10%-25%), and her head circumference was 46.5 cm (50%-75%). The white blood cell count was 9.9×10^9/L [$6.0\text{-}17.0 \times 10^9$/L] with 62% neutrophils (15%-46%), 32% lymphocytes (45%-76%), 6% monocytes (3%-6%), and a platelet count of 380×10^9/L [$150\text{-}450 \times 10^9$/L]. T-cell counts were low, but B-cell and NK-cell counts were normal (Table E1). Her serum immunoglobulins were 10 g/L IgG (6.05-14.3 g/L), 1.56 g/L IgA (0.3-10.7 g/L), and 2.32 g/L IgM (0.6-2.2 g/L). Lymphoblastic transformation in response to phytohemagglutinin was normal, but she showed no reaction in the tuberculin skin test despite BCG vaccination. TCRβ spectratyping unveiled a restricted, yet polyclonal pattern (Fig E1).[E3] A chest computed tomography scan showed bilateral pneumonic consolidation and a small thymus. On admission, she received intravenous antibiotic treatment and was supplemented with intravenous immunoglobulins, which substantially improved her condition. Afterward, she received prophylactic treatment with trimethoprim/sulfamethoxazole and did not develop additional severe infections.

Whereas the majority of $CD4^+$ and $CD8^+$ T cells showed neither CD3 nor TCR expression, a small subset expressed reduced levels of the CD3 complex and TCRαβ. Given the essential role of the CD3 complex in TCR surface expression and signal transduction, we performed targeted DNA sequencing of potential candidate genes and identified a homozygous 2 bp deletion in *CD247* (*CD247* c.43_44delCA; p.Gln15ValfsTer72) that is located in the N-terminal signal peptide of the protein and results in a premature stop codon at amino acid position 72.

Six months after admission, she developed a generalized maculopapular rash, neutropenia, and bilateral arthritis in the ankles. Her clinical presentation was accompanied by an elevated erythrocyte sedimentation rate (74 mm/h) and increased CRP levels (19.1 mg/dL). Antinuclear antibodies were not detected.

Serology unveiled *Borrelia burgdorferi*–specific IgM. In the bone marrow, we detected a reduction of mature neutrophils. She was treated with penicillin for 2 weeks. Given that the symptoms did not improve after 6 weeks, she was diagnosed with juvenile rheumatoid arthritis and treated with ibuprofen. Three months after initiation of treatment, her symptoms waned. Shortly thereafter, she developed arthritis in the ankles, wrists, and fingers. She was treated with ibuprofen, corticosteroids, and cyclosporine, which substantially improved her condition.

At the age of 7 years, the patient developed autoimmune hemolytic anemia and thrombocytopenia. She was diagnosed with Evans syndrome and successfully treated with corticosteroids, intravenous immunoglobulins, and cyclosporine.

At the age of 9 years, she developed cervical lymphadenopathy and was diagnosed with stage 4 non-Hodgkin lymphoma. She was treated according to the LMB89 Group B protocol. Four months after therapy, she developed a cervical mass that was confirmed as a relapse. She received treatment with rituximab, ifosfamide, carboplatin, and etoposide (ICE protocol) for 2 months and underwent bone marrow transplantation at the age of 10 years, using a nonrelated, partially HLA-mismatched (9/10 matched) donor and carmustine, etoposide, and cytarabine as the conditioning regime. She received cyclosporine and methotrexate for graft-versus-host disease prophylaxis. Two weeks after transplantation, she developed cutaneous and intestinal graft-versus-host disease and was treated with tacrolimus and corticosteroids for 6 months. Two years post bone marrow transplantation, she has 100% donor chimerism and received all appropriate vaccinations. She has remained well and did not show any signs of relapse.

REFERENCES

E1. Zhang J, Kobert K, Flouri T. Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 2014;30:614-20.

E2. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 2019;37:420-3.

E3. van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 2003;17:2257-317.

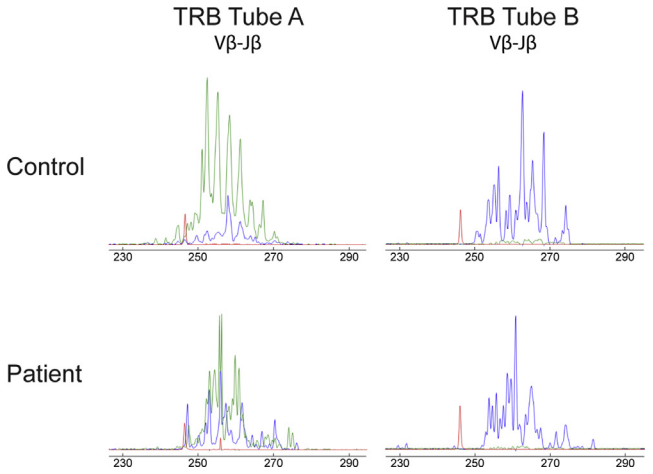

FIG E1. TCRβ spectratyping shows a restricted, yet polyclonal pattern. TCRβ spectratyping for patient and age-matched control using 23 Vβ primers and 9 Jβ primers (tube A) and 23 Vβ primers and 4 Jβ primers (tube B), respectively. *Red* indicates size standard. *Green* and *blue* graphs represent distinct T-cell receptor β joining (*TRBJ*) primer sets.

				Amino acid sequence													Signal peptide analysis		
Number	Sequence	T1 (%)	T2 (%)	1	2	3	4	5	6	7	8	9	10	11	12	13	Likelihood	Cleavage site	Probability of cleavage
1	c.29_30delCCinsT, c.43_44delCA		0.12			V	S	C		R		Q	L	P	I	T	0.1322		
2	c.36_37insTG, c.43_44delCA	0.10		A	I	L		C		R		Q	L	P	I	T	0.6346	21 and 22: TEA-QS	0.5741
3	c.37_42delCAGGCAinsGGGTCCGGC	1.33		A	I	L	G	S		G		Q	L	P	I	T	0.71	22 and 23: TEA-QS	0.5316
4	c.37_51delCAGGCACAGTTGCCG		0.11	A	I	L									I	T	0.6913	17 and 18: TEA-QS	0.4897
5	c.38_49delAGGCACAGTTGC	0.19		A	I								L	P	I	T	0.6757	17 and 18: TEA-QS	0.5340
6	c.39_55delGGCACAGTTGCCGATTAinsCCCAACTGCGT	3.46	0.15	A	I	L	H	P	T	A						S	0.4762		
7	c.40_45delGCACAG	0.27		A	I	L						Q	L	P	I	T	0.8238	19 and 20: TEA-QS	0.7459
8	c.41C>A	0.28		A	I	L	Q			E		Q	L	P	I	T	0.2974		
9	c.42delAincCCAC		0.63	A	I	L	Q	A		H		Q	L	P	I	T	0.6809	22 and 23: TEA-QS	0.5186
10	c.43_44delCA, c.30_31insTT	0.38		A		L	S	C		R		Q	L	P	I	T	0.3643		
11	c.43_44delCA, c.46_47insAC	1.56		A	I	L	Q	A	V				L	P	I	T	0.8812	21 and 22: TEA-QS	0.7827
12	c.43_44delCA, c.47_48delTGinsACCC	1.12	0.51	A	I	L	Q	A	V				P	P	I	T	0.8249	21 and 22: TEA-QS	0.7157
13	c.43_44delCA, c.48_49insATATG	0.13		A	I	L	Q	A	V	D			M	P	I	T	0.6932	22 and 23: TEA-QS	0.5484
14	c.43_44delCA, c.48_49insGG	0.12		A	I	L	Q	A	V	G				P	I	T	0.8384	21 and 22: TEA-QS	0.7296
15	c.43_44delCA, c.48_49insTG	0.10		A	I	L	Q	A	V				V	P	I	T	0.8378	21 and 22: TEA-QS	0.7418
16	c.43_44delCA, c.49_50insGC		13.77	A	I	L	Q	A	V	A				P	I	T	0.8445	21 and 22: TEA-QS	0.6897
17	c.43_44delCA, c.49_51delCCGinsAGGCC	0.10		A	I	L	Q	A	V	E				A	I	T	0.7236	17 and 18: VEA-IT	0.4544
18	c.43_44delCA, c.49_53delCCGATinsTGTG	0.21		A	I	L	Q	A	V	V					C	T	0.8556	20 and 21: TEA-QS	0.5736
19	c.43_44delCA, c.50_51insAA	4.20	0.26	A	I	L	Q	A	V	A				K	I	T	0.7162	21 and 22: TEA-QS	0.4662
20	c.43_44delCA, c.50_51insACCCC		0.24	A	I	L	Q	A	V	A			T	P	I	T	0.8836	22 and 23: TEA-QS	0.5153
21	c.43_44delCA, c.50_51insAG	0.56		A	I	L	Q	A	V	A				R	I	T	0.6723	21 and 22: TEA-QS	0.4649
22	c.43_44delCA, c.50_51insAT	0.12		A	I	L	Q	A	V	A				M	I	T	0.7886	21 and 22: TEA-QS	0.5849
23	c.43_44delCA, c.50_51insATTCC	0.53		A	I	L	Q	A	V	A			I	P	I	T	0.8622	22 and 23: TEA-QS	0.5762
24	c.43_44delCA, c.50_51insCA	1.24	0.52	A	I	L	Q	A	V	A				Q	I	T	0.8326	21 and 22: TEA-QS	0.5660
25	c.43_44delCA, c.50_51insCC	13.11	3.39	A	I	L	Q	A	V	A				P	I	T	0.8445	21 and 22: TEA-QS	0.6897
26	c.43_44delCA, c.50_51insCCTCT	0.20		A	I	L	Q	A	V	A	P			S	I	T	0.8431	22 and 23: TEA-QS	0.5677
27	c.43_44delCA, c.50_51insCG	1.61	0.54	A	I	L	Q	A	V	A				R	I	T	0.6723	21 and 22: TEA-QS	0.4649
28	c.43_44delCA, c.50_51insCT	0.31		A	I	L	Q	A	V	A	L				I	T	0.7859	21 and 22: TEA-QS	0.5577
29	c.43_44delCA, c.50_51insTC	0.61	0.47	A	I	L	Q	A	V	A				S	I	T	0.8313	21 and 22: TEA-QS	0.5700
30	c.43_44delCA, c.50_51insTGCAG	0.88		A	I	L	Q	A	V	A			C	R	I	T	0.8143	22 and 23: TEA-QS	0.4475
31	c.43_44delCA, c.50_51insTGGTGTCC	0.27		A	I	L	Q	A	V	A		W	C	P	I	T	0.8271	23 and 24: TEA-QS	0.6003
32	c.43_44delCA, c.50delCinsAAA	0.16		A	I	L	Q	A	V	A				K	I	T	0.7162	21 and 22: TEA-QS	0.4662
33	c.43_44delCA, c.50delCinsGAACGA	0.12		A	I	L	Q	A	V	A	N			E	I	T	0.8138	16 and 17: AVA-NE	0.4951
34	c.43_44delCA, c.50delCinsGGA	0.25		A	I	L	Q	A	V	A				E	I	T	0.7796	21 and 22: TEA-QS	0.4208
35	c.43_44delCA, c.50delCinsGTG	0.38	0.23	A	I	L	Q	A	V	A	W				I	T	0.7085	21 and 22: TEA-QS	0.5315
36	c.43_44delCA, c.50delCinsTTTCGTCGA	0.12		A	I	L	Q	A	V	A	F		V	E	I	T	0.6994	23 and 24: TEA-QS	0.4476
37	c.43_44delCA, c.51_52insCCCCG	0.31		A	I	L	Q	A	V	A		A		P	I	T	0.8961	16 and 17: AVA-AP	0.3903
38	c.43_44delCA, c.51_52insCG	0.71	0.23	A	I	L	Q	A	V	A				A	I	T	0.8103	21 and 22: TEA-QS	0.4258
39	c.43_44delCA, c.51_52insGA	0.56		A	I	L	Q	A	V	A		G			I	T	0.8008	21 and 22: TEA-QS	0.5489
40	c.43_44delCA, c.51_52insGC	0.21		A	I	L	Q	A	V	A		G			I	T	0.8008	21 and 22: TEA-QS	0.5489
41	c.43_44delCA, c.51_52insGG	2.28	0.16	A	I	L	Q	A	V	A		G			I	T	0.8008	21 and 22: TEA-QS	0.5489
42	c.43_44delCA, c.51_52insGGACA	0.31		A	I	L	Q	A	V	A	G			T	I	T	0.8057	22 and 23: TEA-QS	0.4015
43	c.43_44delCA, c.51_52insGGCCCCCG	0.62		A	I	L	Q	A	V	A		G	P	P	I	T	0.8691	23 and 24: TEA-QS	0.4015
44	c.43_44delCA, c.51_52insGGCCG		0.81	A	I	L	Q	A	V	A		G		P	I	T	0.8811	22 and 23: TEA-QS	0.4740
45	c.43_44delCA, c.51_52insGGGGT	0.22		A	I	L	Q	A	V	A		G		G	I	T	0.7874	22 and 23: TEA-QS	0.3840
46	c.43_44delCA, c.51_52insGGGGTGCG	0.15		A	I	L	Q	A	V	A		G	G	A	I	T	0.7501	18 and 19: AGG-AI	0.2193
47	c.43_44delCA, c.51_52insTA	1.63	0.11	A	I	L	Q	A	V	A	V				I	T	0.7445	21 and 22: TEA-QS	0.5215
48	c.43_44delCA, c.51_52insTG	0.17		A	I	L	Q	A	V	A	V				I	T	0.7445	21 and 22: TEA-QS	0.5215
49	c.43_44delCA, c.51_52insTTGGAAATCAG	0.11		A	I	L	Q	A	V	A	V	G	N	Q	I	T	0.8145	18 and 19: AVG-NQ	0.4399
50	c.43_44delCA, c.51_54delGATTinsTCGACA	0.25		A	I	L	Q	A	V	A	S			T		T	0.8699	21 and 22: TEA-QS	0.5630
51	c.43_44delCA, c.51delGinsAGTACC		0.23	A	I	L	Q	A	V	A	S			T	I	T	0.8359	22 and 23: TEA-QS	0.4420
52	c.43_44delCA, c.51delGinsTTACCCTTA	0.22		A	I	L	Q	A	V	A	L	P	L		I	T	0.7782	23 and 24: TEA-QS	0.4937
53	c.43_44delCA, c.52_53insGA	0.59	16.76	A	I	L	Q	A	V	A				E	I	T	0.7796	21 and 22: TEA-QS	0.4208
54	c.43_44delCA, c.52delAinsGGGGAAGGT	0.12		A	I	L	Q	A	V	A		G	G	R	F	T	0.6198	17 and 18: VAG-GR	0.1844
55	c.43_44delCA, c.53_54insGC	1.06		A	I	L	Q	A	V	A				D	A	T	0.8717	18 and 19: ADA-TE	0.5144
56	c.43_44delCAinsAGGCA	0.70	0.18	A	I	L	Q	A		R		Q	L	P	I	T	0.4096		

FIG E2. Somatic variants that restore the reading frame in $CD3^+$, $TCR\alpha\beta^+$ T cells. Spectrum of functional revertant variants that restore the reading frame of the protein with indicated frequency at each time point, amino acid sequence that differs from WT protein, and signal peptide analysis. Amino acids are symbolized as *white circles* for moderate hydropathy, *blue circles* for hydrophilic amino acids, and *red circles* for hydrophobic amino acids. *Black outline* denotes neutral charge, whereas *blue* and *red outlines* refer to positive and negative charges, respectively. Signal peptide analysis was performed as recommended for SignalP 5.0, including likelihood of signal peptide function, cleavage site, and probability of cleavage. Variants with signal peptide score <0.5 are not predicted to be signal peptides.

TABLE E1. WBC count and lymphocyte subsets at admission

Variable	Patient (× 10^9/L)	Normal range (× 10^9/L)
WBCs	7.9	6.0-17.0
Neutrophils	6.1 (77%)	1.5-8.5 (15%-46%)
Lymphocytes	1.2 (15%)	1.5-9.5 (45%-76%)
Monocytes	0.6 (7.6%)	0.2-1.2 (3%-6%)
Platelets	380	150-450
T cells ($CD3^+$)	0.3	1.4-8.0
T helper cells ($CD3^+$, $CD4^+$)	0.25	0.9-5.5
Cytotoxic T cells ($CD3^+$, $CD8^+$)	0.01	0.4-2.3
B cells ($CD19^+$)	0.5	0.6-3.1
NK cells ($CD16^+$, $CD56^+$)	0.3	0.1-1.4

NK, Natural killer; *WBC,* white blood cell.

TABLE E2. Somatic variants that do not restore the reading frame in $CD3^+$, $TCR\alpha\beta^+$ T cells

Number	Sequence	T1 (%)	T2 (%)	Position of stop codon
1	c.43_44delCA (germline mutation)	54.16	59.58	72
2	c.43_44delCA, c.50_51insC	1.15	0.43	51
3	c.43_44delCA, c.51_52insG	0.16	0.00	51
4	c.5A>G, c.43_44delCA	0.18	0.00	72
5	c.43_44delCA, c.52_53insCGA	0.00	0.21	73
6	c.43_44delCA, c.49_50insG	0.00	0.15	51

TABLE E3. Somatic variants in $CD3^-$, $TCR\alpha\beta^-$ T cells

Number	Sequence	Frequency (%)
1	c.43_44delCA (germline mutation)	98.27
2	c.5A>G, c.43_44delCA	0.20
3	c.11A>G, c.43_44delCA	0.19
4	c.14C>A, c.43_44delCA	0.14
5	c.16C>T, c.43_44delCA	0.24
6	c.22A>G, c.43_44delCA	0.10
7	c.26C>T, c.43_44delCA	0.32
8	c.30C>T, c.43_44delCA	0.31
9	c.32T>C, c.43_44delCA	0.13
10	c.38A>G, c.43_44delCA	0.10

TABLE E4. Overview of next-generation sequencing data

Sample	Cell population	Sequencing platform	Total reads	Read count after filtering	Number of variants	Reads with c.43_44delCA variant (%)	Reference sequence (%)
Patient	TCRαβ$^-$, CD3$^-$ T cells	Roche 454	22,417	20,414	10	98.27	0
Patient	TCRαβ$^-$, CD3$^-$ T cells at T1	Roche 454	24,808	23,189	53	54.20	0
Patient	TCRαβ$^-$, CD3$^-$ T cells at T2	Roche 454	23,232	21,687	24	59.72	0
Patient	NK cells	Illumina	102,118	98,239	0	100	0
Control 1	T cells	Illumina	32,842	31,935	0	0	100
Control 2	T cells	Illumina	45,421	44,177	0	0	100
Control 3	T cells	Illumina	33,734	32,774	0	0	100
Control 4	T cells	Illumina	33,312	32,380	0	0	100
Control 5	T cells	Illumina	34,720	33,736	0	0	100
Control 6	T cells	Illumina	31,688	30,830	0	0	100
Control 7	T cells	Illumina	26,587	25,851	0	0	100