Table E1. List of candidate variants identified using WES (MAF <=0.01) and segregate within the family (variants in P1, P2 and not their unaffected sibling or parents).

Genetic model	Gene name	Exonic function	Amino acid change	Homozygosity	PolyPhen	SIFT
Recessive	NADK	insertion (6 bases)	p.G414delinsRRG:p.G446delinsRRG:p.G591delinsRRG		na	na
Recessive	FAT4	nonsynonymous	p.Q1257E		B	T
Recessive	LY6G5C	nonsynonymous	p.F53L:p.F56L:p.F54L		B	T
Recessive	AGER	nonsynonymous	p.G82S:p.G68S:p.G113S		P	T
Recessive	NOTCH4	nonsynonymous	p.G294R		D	T
Recessive	C6ORF10	nonsynonymous	p.G477V:p.G479V:p.G463V:p.G478V		D	D
Recessive	C6ORF10	nonsynonymous	p.L264W:p.L266W:p.L250W:p.L257W:p.L265W		D	D
Recessive	C6ORF10	nonsynonymous	p.G143R:p.G122R:p.G145R:p.G129R		na	D
Recessive	HLA-DRB5	stopgain	p.Q220X		na	na
Recessive	HLA-DRB1	nonsynonymous	p.V73M		D	D
Recessive	HLA-DRB1	nonsynonymous	p.V73L		D	T
Recessive	HLA-DRB1	nonsynonymous	p.D70N		D	D
Recessive	HLA-DRB1	nonsynonymous	p.D57Y		P	D
Recessive	HLA-DRB1	nonsynonymous	p.D57N		B	D
Recessive	HLA-DQA1	nonsynonymous	p.M18T		B	T
Recessive	TRAF3IP2	nonsynonymous	p.D10N:p.D19N		D	D
Recessive	TULP4	nonsynonymous	p.S522N		B	T
Recessive	GLT6D1	nonsynonymous	p.P219S		D	T
Recessive	PITRM1	nonsynonymous	p.L64F:p.L441F:p.L785F:p.L883F:p.L884F		D	D
Recessive	PITRM1	nonsynonymous	p.L113V:p.L145V		B	T
Recessive	C14ORF178	nonsynonymous	p.G31D:p.G61D		B	D
Recessive	C19ORF33	insertion (15 bases)	p.K90delinsKEGEGQ		na	na
Recessive	GGN	nonsynonymous	p.A517V:p.A434V		D	T
Recessive	RASGRP4	nonsynonymous	p.G165R		B	D
Recessive	MIA	nonsynonymous	p.P16L	14.7 MB	na	na
Recessive	PSG6	nonsynonymous	p.S312F:p.S405F	14.7 MB	B	D
Recessive	PSG6	nonsynonymous	p.I122M:p.I243M	14.7 MB	D	D
Recessive	ZNF221	nonsynonymous	p.V165M	14.7 MB	D	D
Recessive	ZNF225	nonsynonymous	p.R352H	14.7 MB	D	D
Recessive	CEACAM16	nonsynonymous	p.S32I	14.7 MB	B	D
Recessive	RSPH6A	nonsynonymous	p.Q184H:p.Q448H	14.7 MB	D	D
Recessive	POLD1	nonsynonymous	p.R1060C*	14.7 MB	D	D
Recessive	SIGLEC12	nonsynonymous	p.A77T	14.7 MB	P	D
Compound-het	LRP1B	nonsynonymous	p.E3955K		B	T
Compound-het	LRP1B	nonsynonymous	p.V2146F		B	D
Compound-het	ZFHX3	nonsynonymous	p.Y865C		D	D
Compound-het	ZFHX3	nonsynonymous	p.K520N		B	T
De novo	MUC12	nonsynonymous	p.S1610I:p.S1753I		B	D

*POLD1 R1060C is in reference to *POLD1* isoform 1 (NM_001256849.1). PolyPhen and SFIT scores: B: benign; T: tolerant; P: probably deleterious; D: deleterious

Table E2. Homozygous regions identified in autosomal chromosomes using WES and segregating within the family.

Chromosome	Start Position	End Position	Size (MB)	SNPs (Homozygous)
3	130,098,639	132,105,588	2.007	262
6	33,032,788	38,650,628	5.618	1732
13	41,567,248	45,563,464	3.996	555
19	38,040,492	38,314,767	0.274	119
19	39,219,780	53,990,002	14.770	8617

Table E3 Summary of immune repertoire analysis.

Sample	Cell Input	Total templates	Unique clonotypes	Productive templates	Unique productive clonotypes
HC1-CD8	160,000	7367	4942	5825	3964
HC2-CD8	100,000	2269	2075	1838	1687
S002a-CD8	67,000	1660	1547	1415	1318
P002-CD8	93,000	2568	243	1653	187
S002b-CD8	50,000	1401	496	910	416
P3-CD8	30,000	735	278	588	212
HC1-CD4	100,000	6550	5862	5223	4628
HC2-CD4	100,000	1341	1234	1101	1008
S002a-CD4	100,000	1475	1375	1247	1161
P002-CD4	64,000	4528	3655	3693	2936
S002b-CD4	100,000	2046	1964	1773	1703
P3-CD4	57,000	4905	3358	4085	2772
HC1-B	50,000	1550	1545	1272	719
HC2-B	100,000	2017	2012	1676	1032
S002a-B	199,000	11838	11824	9938	6557
P002-B	28,000	3148	3143	2604	1359
S002b-B	75,000	1652	1643	1294	770
P3-B	54,000	503	497	438	249

HC1, healthy control 1; HC2, healthy control 2; S002a, healthy sibling; P002, patient 1; S002b, patient 2; P3, patient 3.

Fig E1

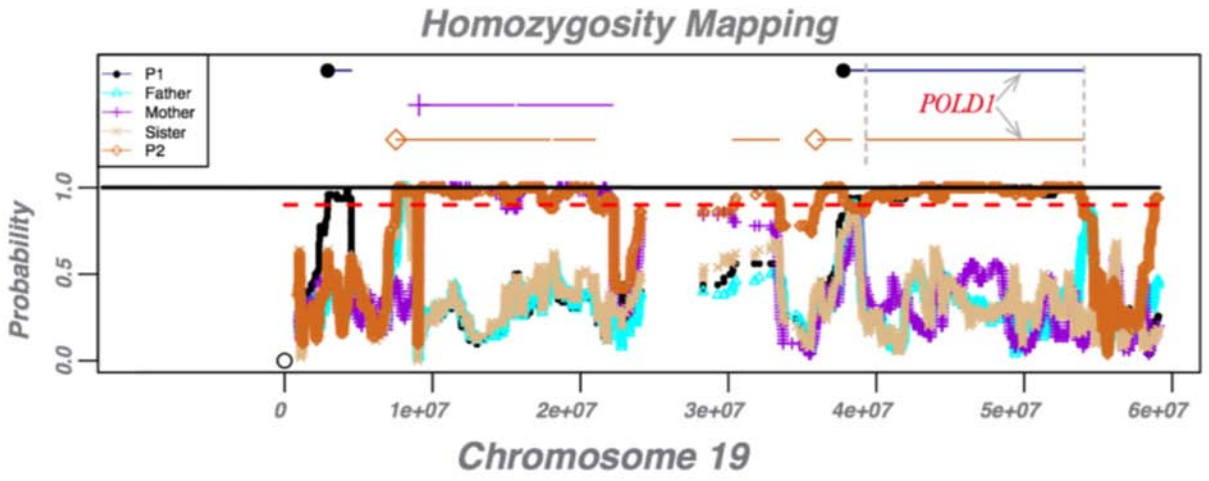


Fig E2

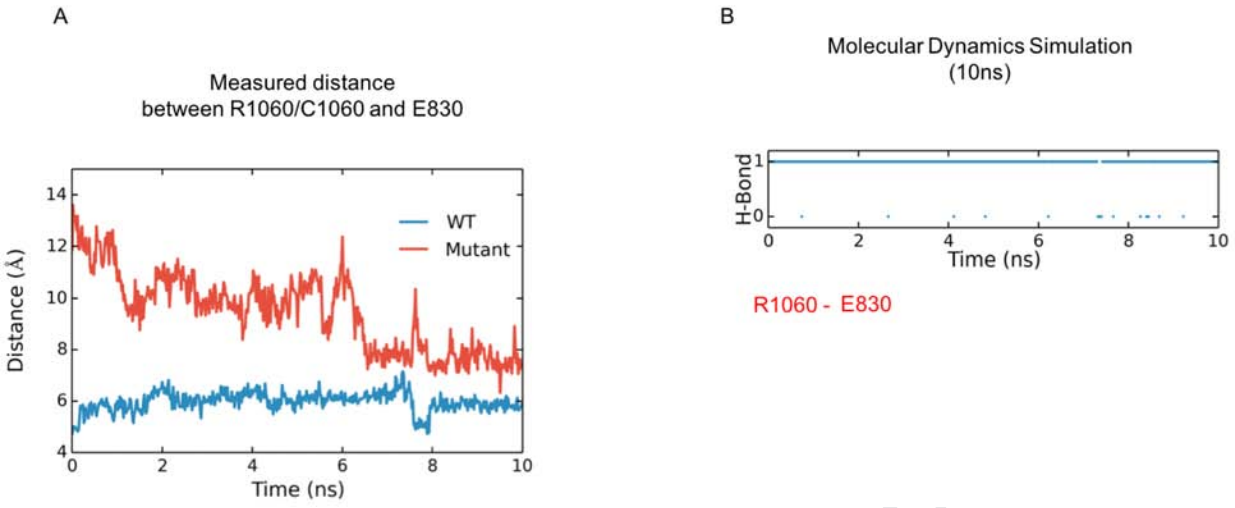


Fig E3

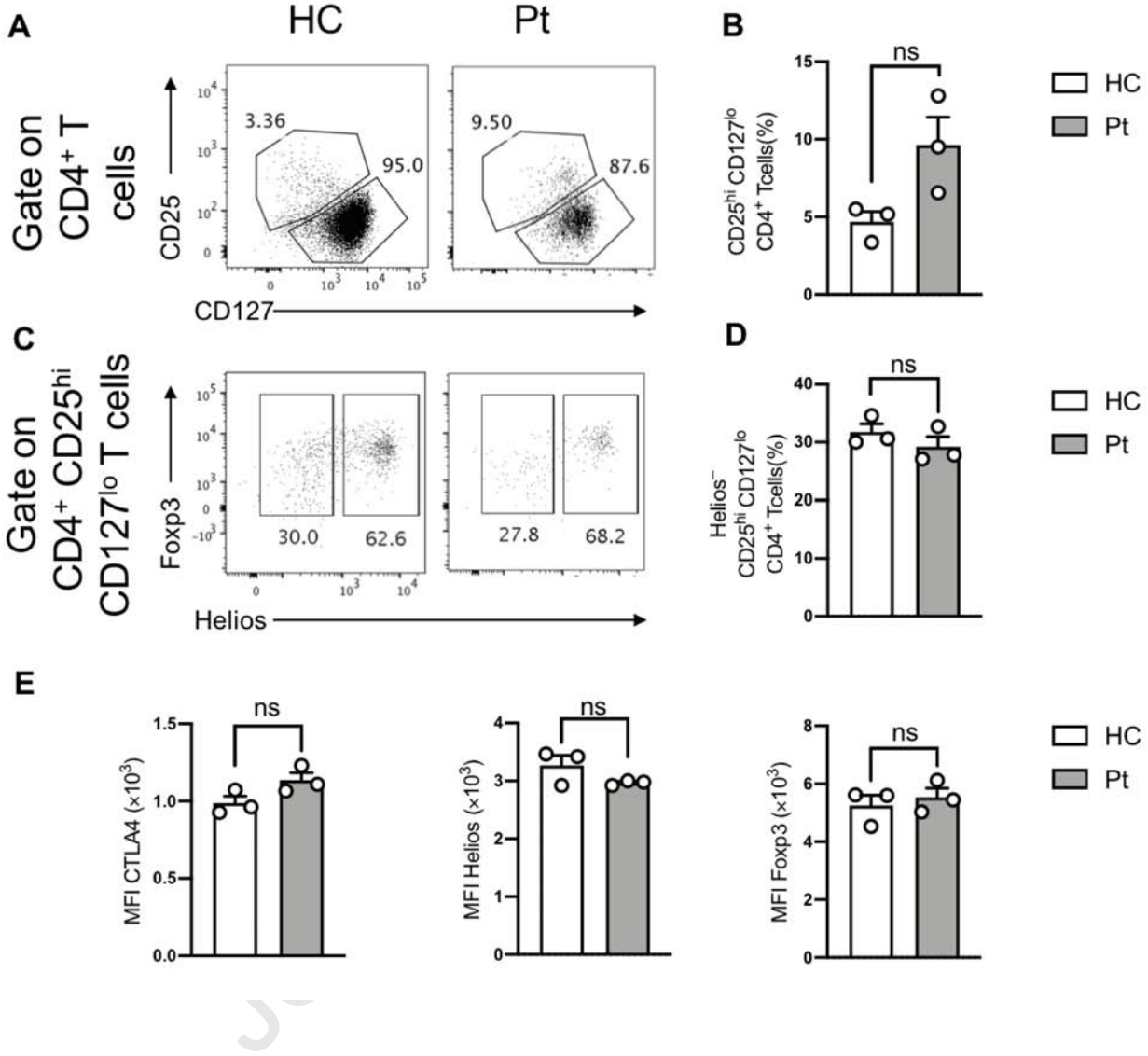
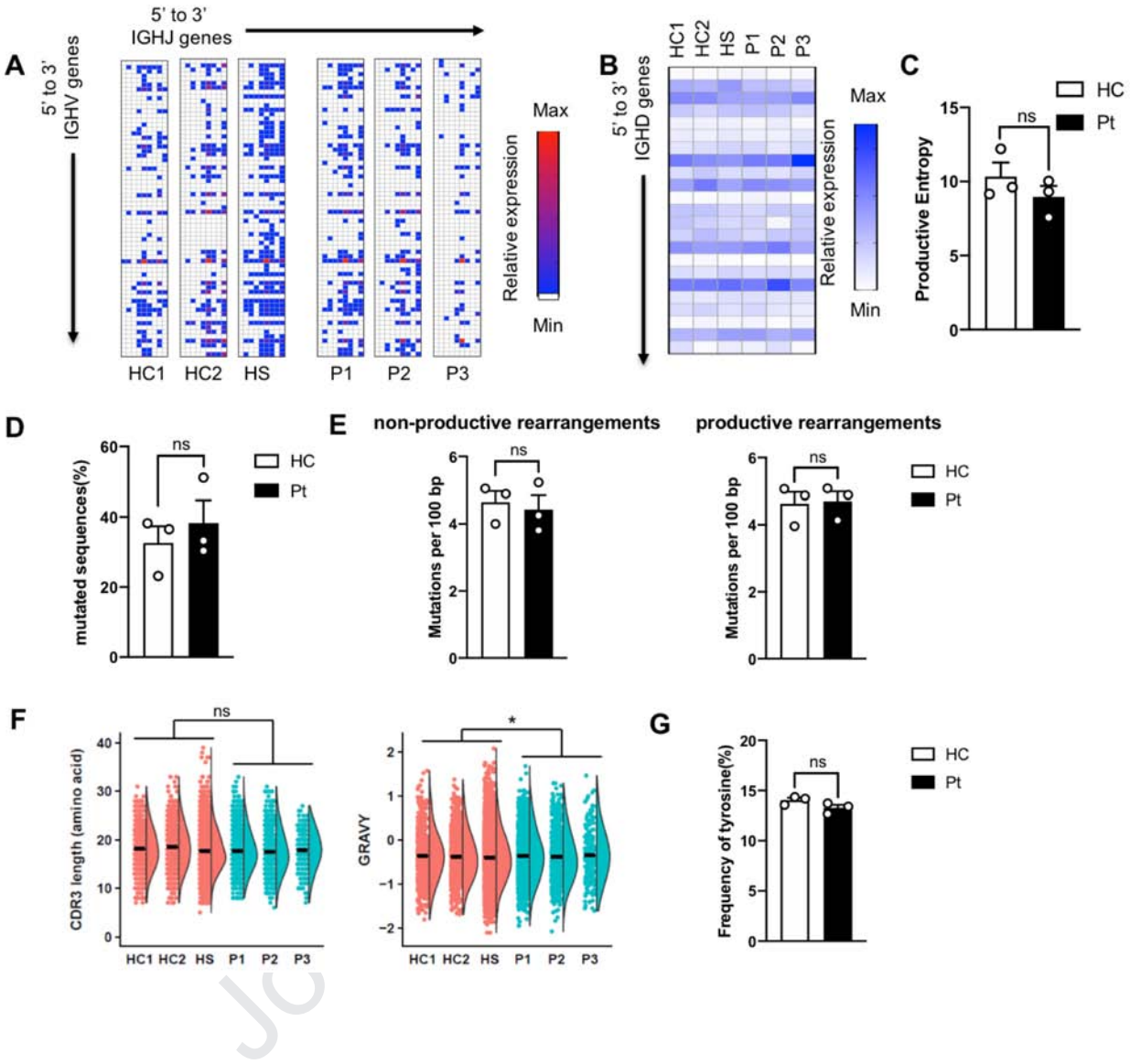

A
HC
Pt
Gate on CD4+ T cells
CD25
CD127
3.36
95.0
9.50
87.6
B
ns
CD25hi CD127lo CD4+ Tcells(%)
HC
Pt
C
Gate on CD4+ CD25hi CD127lo T cells
Foxp3
Helios
30.0
62.6
27.8
68.2
D
ns
Helios− CD25hi CD127lo CD4+ Tcells(%)
HC
Pt
E
ns
MFI CTLA4 (×10³)
ns
MFI Helios (×10³)
ns
MFI Foxp3 (×10³)
HC
Pt

Fig E4


A
5' to 3'
IGHJ genes
5' to 3'
IGHV genes
HC1
HC2
HS
P1
P2
P3
Max
Relative expression
Min
B
HC1
HC2
HS
P1
P2
P3
5' to 3'
IGHD genes
Max
Relative expression
Min
C
HC
Pt
ns
Productive Entropy
15
10
5
0
D
ns
HC
Pt
mutated sequences(%)
60
40
20
0
E
non-productive rearrangements
productive rearrangements
ns
ns
HC
Pt
Mutations per 100 bp
6
4
2
0
F
ns
*
CDR3 length (amino acid)
GRAVY
HC1 HC2 HS P1 P2 P3
G
ns
HC
Pt
Frequency of tyrosine(%)
20
15
10
5
0

Combined Immunodeficiency due to a loss of function mutation in DNA Polymerase Delta 1

Ye Cui, PhD,[1] Sevgi Keles, MD,[2] Louis-Marie Charbonnier, PhD,[1] Amélie M. Julé, PhD,[1] Lauren Henderson, MD, MSc.[1] Seyma Celikbilek Celik, BSc, [2] Ismail Reisli, MD, [2] Chen Shen, PhD,[3,4] Wen Jun Xie, PhD,[5] Klaus Schmitz-Abe, PhD,[6] Hao Wu, PhD,[1,3] Talal A. Chatila, MD, MSc[1]

[1]Division of Immunology, The Boston Children's Hospital. Department of Pediatrics, Harvard Medical School, Boston; [2]Necmettin Erbakan University, Meram Medical Faculty, Division of Pediatric Allergy and Immunology, Konya; [3]Program in Molecular and Cellular Medicine, Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston; [4]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston; [5]Department of Chemistry, Massachusetts Institute of Technology, Cambridge; [6]Division of Newborn Medicine and Neonatal Genomics Program, Boston Children's Hospital, Harvard Medical School, Boston.

Corresponding Author: Talal A. Chatila at the Division of Immunology, Boston Children's Hospital and the Department of Pediatrics, Harvard Medical School.

Address: Karp Family Building, Room 10-214. 1 Blackfan Street, Boston, MA 02115

Email: talal.chatila@childrens.harvard.edu

Supplementary Methods

Whole exome sequencing and data analysis: WES data was processed through Variant Explorer Pipeline (VExP) using BWA aligner (version 0.7.17) for mapping reads to the human genome (hg19) and PICARDtools (V2.20.2) to mark/delete duplicate reads. Single Nucleotide Variants (SNVs) and small insertions / deletions (indels) were jointly called across all samples using both GATK (multi-sample variant calling, v4.1) and SAMTools (v1.9). Further, VExP was performed to annotate 21 relevant genetic databases (from Allele frequency and Gene-phenotype consortiums) and 23 coding/non-coding variant pathogenicity predictors into the output of the system. Variant analysis was performed using different inheritance models (assuming full penetrance) based on three filtering criteria: first, include variants predicted to have a potential functional coding consequence, including stop gain or loss, splice site disruption, indel, and nonsynonymous. Second, variants were filtered based on allele frequency in control populations (gnomAD, ExAC, EVS, 1000GP and internal data from 2,114 unaffected individuals from BCH). The variants were further prioritized to include those with read depth ≥10X and deleterious prediction (2 or more of 23 softwares, including PolyPhen, SIFT, FATHMM, CADD, etc). For pedigree-consistency analysis, VExP had verified consistency within all family members.

Homozygosity mapping: We use Variant Explorer Pipeline "VExP" to determine homozygous regions using whole genome sequencing data (WES). In summary, the method uses a sliding window approach, 100 SNPs, and retained segments with a minimum of 98% homozygosity. Homozygous SNPs cannot be more than 100 Kb away from each other. Next, VExP joins all the homozygous regions using several

considerations including regions with no genes or noncoding genes. It retains only segments where observed homozygosity exceeds 3 cM and avoid the effect of residual population homozygosity that is likely innocuous and tolerated by natural selection. We use genetic, as opposed to physical distance, for all calculations. To calculate overall homozygosity for every sample, we sum all segments exceeding 3 cM. Homozygosity Mapping is applied to the results from the whole family, obtaining overlapping homozygous regions between affected individuals with no overlapping with unaffected samples (**Fig E1, Table E2**).

Reference:

1. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982; 157:105-32.

Supplementary Figure Legends

Figure E1. Homozygous regions identified in chromosome 19 using WES and segregating within the family (Homozygosity mapping).

Figure E2. A. Distance measurement between residue 1060 and 830 in wild-type and mutant POLD1. Longer distance in the mutant POLD1 reflects the interaction lost between CysB and POLBc_delta domains. **B**. Hydrogen bond (HB) tracing of the interaction between R1060 and E830 in POLD1. 1 means HB exists, 0 means HB disappears.

Figure E3. POLD1^{R1060C} patients are normal in regulatory T cell frequency and phenotype. **A.** Representative dot plot analysis of CD25^{hi} CD127^{lo} CD4^{+} T cells in patient vs. a control subject. **B.** Percentages of CD25^{hi} CD127^{lo} CD4^{+} T cells in the peripheral blood of healthy controls (n=3; open circles) and POLD1^{R1060C} patients (n=3; closed circles). **C.** Representative dot plot analysis of Helios^{-} CD25^{hi} CD127^{lo} CD4^{+} T cells in patient vs. a control subject. **D.** Mean fluorescence intensity of the respective regulatory T cell marker in patient and control subjects. ns, not significant, by unpaired two-tailed Student's t-test.

Figure E4. Profiles of IGHV-IGHJ pairing and somatic hypermutation (SHM) analysis in POLD1^{R1060C} B Cells. **A**. Frequencies of specific IGHV and IGHJ pairing in unique *IGH* clonotypes of B cells from healthy controls and patients. White represents the absence of a given IGHV and IGHJ pairing. Blue reflects a low frequency while red represents a higher frequency of usage. **B**. Frequencies of specific IGD gene usage in unique *IGH* clonotypes of B cells from healthy controls and patients. White represents the lowest gene usage. Blue reflects a higher frequency of usage. **C**. Productive entropy of B cells

in healthy controls versus patients (n=3; open circles). **D.** Percentage of sequences carrying at least one somatic hypermutation (SHM) among all unique rearrangements (productive and non-productive) with resolved V family, gene or allele. **E.** Number of mutations per 100 bp within the IGH V segment of unique rearrangements with at least one mutation in non-productive rearrangements and productive rearrangements. **F**. Amino acid properties of *in silico* translated *IGH*-CDR3 region, for unique productive rearrangements encoding a complete CDR3 region (starting and ending with consensus codons). CDR3 length: control vs. patient group, p>0.1; subject effect, p<10e-6. GRAVY, Grand Average of Hydrophobicity[1]: control vs. patient group, p= 0.0490; subject effect, p>0.4.). **G**. Proportion of tyrosine residue in the *IGH*-CDR3 of unique sequences: control vs. patient group (n=3; open circles). **C, D, E** and **G** were analyzed with Student's unpaired two tailed t test. Results represent means ± S.E.M. ns, not significant. **F** was analyzed with 2-way ANOVA, to contrast sequences patterns between patients and control while accounting for per-subject variations. ns, not significant; *, p<0.5.