Supplementary Materials

fig. S1. Diversity indices of *IGH* and *TRB* repertoire.

fig. S2. Differential usage of *V, D* and *J* genes in the *IGH* and *TRB* repertoires of patients with RAG deficiency.

fig. S3. Characteristics of the CDR3 region of *IGH* and *TRB* total sequences in peripheral blood lymphocytes.

fig. S4. P and N nucleotide addition in the CDR3 of unique *IGH* and *TRB* sequences from controls and patients with RAG mutations and various clinical phenotypes.

fig. S5. Germline index and frequency of sequences without P and N nucleotides in the *TRB* repertoire of patients with RAG deficiency and controls.

fig. S6. Distribution of immunoglobulin heavy chain isotypes among total *IGH* sequences from controls and patients with RAG deficiency.

fig. S7. Inference of antigen-mediated selection in combined and individual IgM, IgG and IgE isotype transcripts from healthy controls and patients with RAG deficiency.

Table S1. Clinical and laboratory features of patients with RAG deficiency

Table S2. Summary of *RAG* mutations, recombination activity, and results of *IGH* and *TRB* sequencing in healthy donors and patients with RAG deficiency

Table S3. Raw data for all the figures (named “data for Fig 1-6” and “data for Suppl Fig 2-6) and sequence of *IGH* and *TRB* transcripts from healthy controls and patients with RAG deficiency.

Supplementary Materials

Table S1. Clinical and laboratory features of patients with RAG deficiency

	Patients grouped by clinical phenotype											
	CID with autoimmunity and/or granulomas				***Leaky SCID***			***Omenn syndrome***				
	CID1	CID2	CID3	CID4	LS1	LS2	LS3	OS1	OS2	OS3	OS4	OS5
Gene defect (mutation)	*RAG1* H612R;K86fs	*RAG2* F62L;F62L	*RAG1* F974L;R841Q	*RAG1* W522C;R973C	*RAG1* K86fs;K86fs	*RAG1* K86fs;R404Q	*RAG1* R108X;R108X	*RAG1* R561H;R624C	*RAG1* K391E;K391E	*RAG1* R396C;R737H	*RAG1* W959X;W959X	*RAG1* C469fs;C469fs
Age at diagnosis	16 y	27 y	24 mo	15 mo	13 mo	22 mo	4 mo	2 mo	6 mo	8 mo	5 weeks	3 mo
ALC (cells/μL)	1,820	687	9,630	4,100	3,400	830	4,300	n.a.	1,700	1,230	8,230	12,500
Eosinophils/μL	260	n.a.	200	100	590	980	300	n.a.	150	580	13,200	1,780
$CD3^+$ cells/μL	581	391	160	3,973	2,300	423	2,550	11140	629	1,119	3,424	4,442
$CD4^+$ cells/μL	530	225	108	3,296	80	191	212	6204	497	726	3,368	3,109
$CD8^+$ cells/μL	80	162	10	668	840	41	1,731	4931	101	394	0	1,036
$CD45RA^+$ cells (% of $CD4^+$ cells)	6	11	n.d.	3	0	2	9.2	1	3.8	1	0	n.d.
γδT cells (% of $CD3^+$ cells)	n.d.	n.d.	n.d.	2.1	69	7.2	23.8	n.d.	n.d.	0.1	0	n.d.
$CD19^+$ cells/μL	460	78	359	$0^{\#}$	620	100	193	0	2	0	0	30
$CD16/56^+$ cells/μL	450	215	340	66	460	290	1,121	4,564	1,015	98	5,003	8,000
Proliferative response to PHA	↓	normal	↓	↓	↓	↓	n.d.	↓↓	absent	absent	absent	↓↓
IgG (mg/dL)	411	1,000^	1,570	153	2,420	1104	2,080	10	900^	1,003^	39	2,340
IgA (mg/dL)	<5	<5	77	15	194	153	98	6	110	323	22	72
IgM (mg/dL)	38	25	230	<10	328	108	275	302	35	7	16	27
IgE (IU/mL)	<1	<2	227	126	5.6	15	213	>5000	12	>50,000	<2	5

Skin rash	no	Freckling (café au lait) in axilla and chest	yes (vasculitis)	yes, but not consistent with Omenn	no	yes	yes, diffuse	yes, diffuse	yes, diffuse	yes, diffuse	yes, diffuse	yes, diffuse
Lymphadenopathy	yes	no	no	no	no	no	yes	yes	yes	no	yes	yes
Hepatosplenomegaly	no	no	no	yes	no	no	no	no	yes	yes	no	yes
Infections	VZV (zoster), rec. sinusitis	disseminated VZV, rec. sinusitis, *Cryptococcus* meningitis, *Pseudomonas* pneumonia	rec. URTI	Norovirus, CVL sepsis	chronic CMV, pyodermitis, pneumonia	CMV	CMV, candida	CMV	diarrhea, pneumonia	PJP, adenovirus	pneumonia, skin infection (MRSA)	diarrhea, lip abscess
Autoimmunity	ITP, Coombs +, vitiligo	suspected ITP	vasculitis, AIHA, ITP, myositis	AIHA	AIHA	AIHA	AIHA	no	no	autoimmune hepatitis?	no	no
Granulomas	lungs	lungs	no	no	no	no	no	no	no	no	no	no
Reference	(*48*)	CID-9 in (*49*)	(*6*)	unpubl.	LS6 in (*49*)	unpubl.	LS5 in (*49*)	unpubl.	unpubl.	unpubl.	unpubl.	(*50*)

[#]after rituximab (previous count: 538 $CD19^+$ cells/µl)

^on intravenous immunoglobulins

AIHA: autoimmune hemolytic anemia; CID: combined immune deficiency; CMV: cytomegalovirus; CVL: central venous line; γδT: leaky SCID with expansion of $TCR\gamma\delta^+$ T cells; ITP: immune thrombocytopenic purpura; LS: leaky SCID; MRSA: methicillin-resistant *S. aureus*; PJP: *Pneumocystis jiroveci* pneumonia; rec: recurrent; TCR: T cell receptor; unpubl.: unpublished; URTI: upper respiratory tract infections; VZV: varicella zoster virus

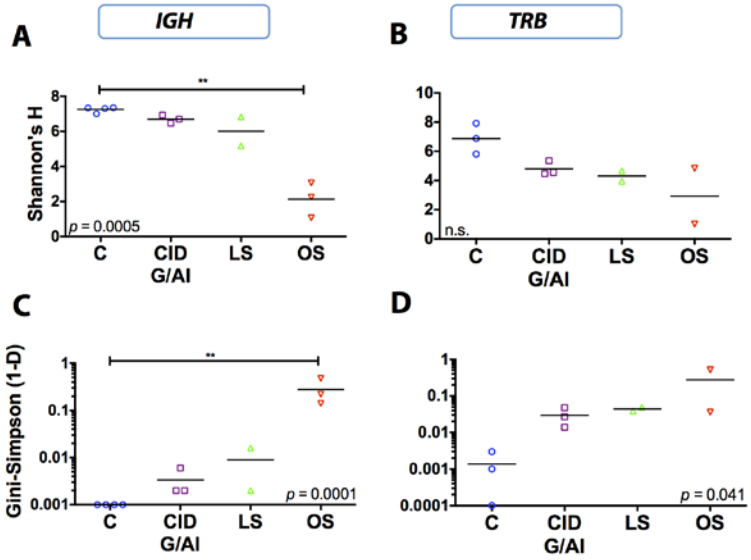
↓, ≥15 and <30% of healthy control value; ↓↓, <15% of healthy control value; absent, <5% of healthy control value

Table S2. Summary of *RAG* mutations, recombination activity, and results of *IGH* and *TRB* sequencing in healthy donors and patients with RAG deficiency

Patient		Gene	Allele 1		Allele 2		*IGH* Sequences			*TRB* Sequences		
Code	***Diagnosis***		***Mutation***	***Activity (% of WT)***	***Mutation***	***Activity (% of WT)***	***Total***	***Unique***	***Unique/Total***	***Total***	***Unique***	***Unique/Total***
C1	HD						3899	3587	92.00%	1007	363	36.05%
C2	HD						5894	4943	83.86%	852	363	42.61%
C3	HD						2006	1888	94.12%	4915	3140	63.89%
C4	HD						4389	4223	96.22%	NA	NA	
CID1	CID	*RAG1*	H612R	100%	K86Vfs*33	2.65%	5667	4516	79.69%	ND	ND	
CID2	CID	*RAG2*	F62L	19.60%	F62L	19.60%	2494	1975	79.19%	1086	231	21.27%
CID3	CID	*RAG1*	F974L	56.50%	R841Q	0.00%	3538	1903	53.79%	12399	545	4.40%
CID4	CID	*RAG1*	W522C	41.65%	R973C	0.00%	NA	NA		2092	385	18.40%
LS1	γδT	*RAG1*	K86Vfs*33	2.65%	K86Vfs*33	2.65%	4038	2752	68.15%	ND	ND	
LS2	LS	*RAG1*	K86Vfs*33	2.65%	R404Q	1.18%	NA	NA		3821	490	12.82%
LS3	γδT	*RAG1*	R108X	1.80%	R108X	1.80%	1122	440	39.22%	9370	177	1.89%
OS1	OS	*RAG1*	R561H	1.98%	R624C	0.00%	1521	30	1.97%	NA	NA	
OS2	OS	*RAG1*	K391E	6.54%	K391E	6.54%	ND	ND		6301	464	7.36%
OS3	OS	*RAG1*	R396C	0.56%	R737H	0.23%	801	55	6.87%	NA	NA	
OS4	OS	*RAG1*	W959X	0.01%	W959X	0.01%	ND	ND		17884	43	0.24%
OS5	OS	*RAG1*	C469Wfs*20	0.00%	C469Wfs*20	0.00%	6706	216	3.22%	ND	ND	

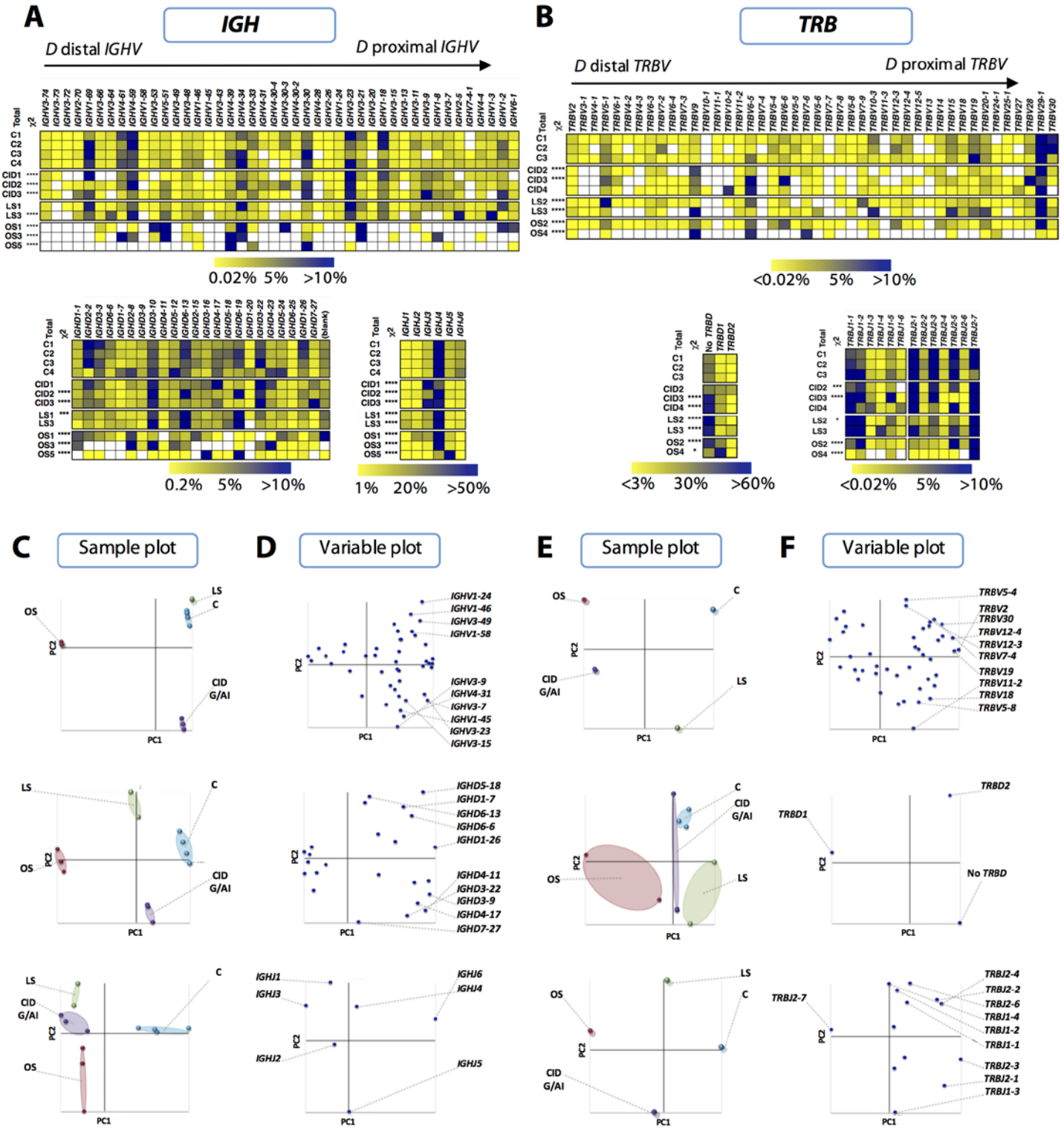
C: control; CID: combined immune deficiency; γδT: “leaky” severe combined immune deficiency with expansion of T-cell receptor γδ+ T cells; HD: healthy donor; *IGH*: Immunoglobulin Heavy Chain locus; LS: leaky SCID; ND: not detectable; NA: not available; OS: Omenn syndrome; SCID: severe combined immune deficiency; *TRB*: T cell receptor β locus; WT: wild-type

fig. S1. Diversity indices of *IGH* and *TRB* repertoire.



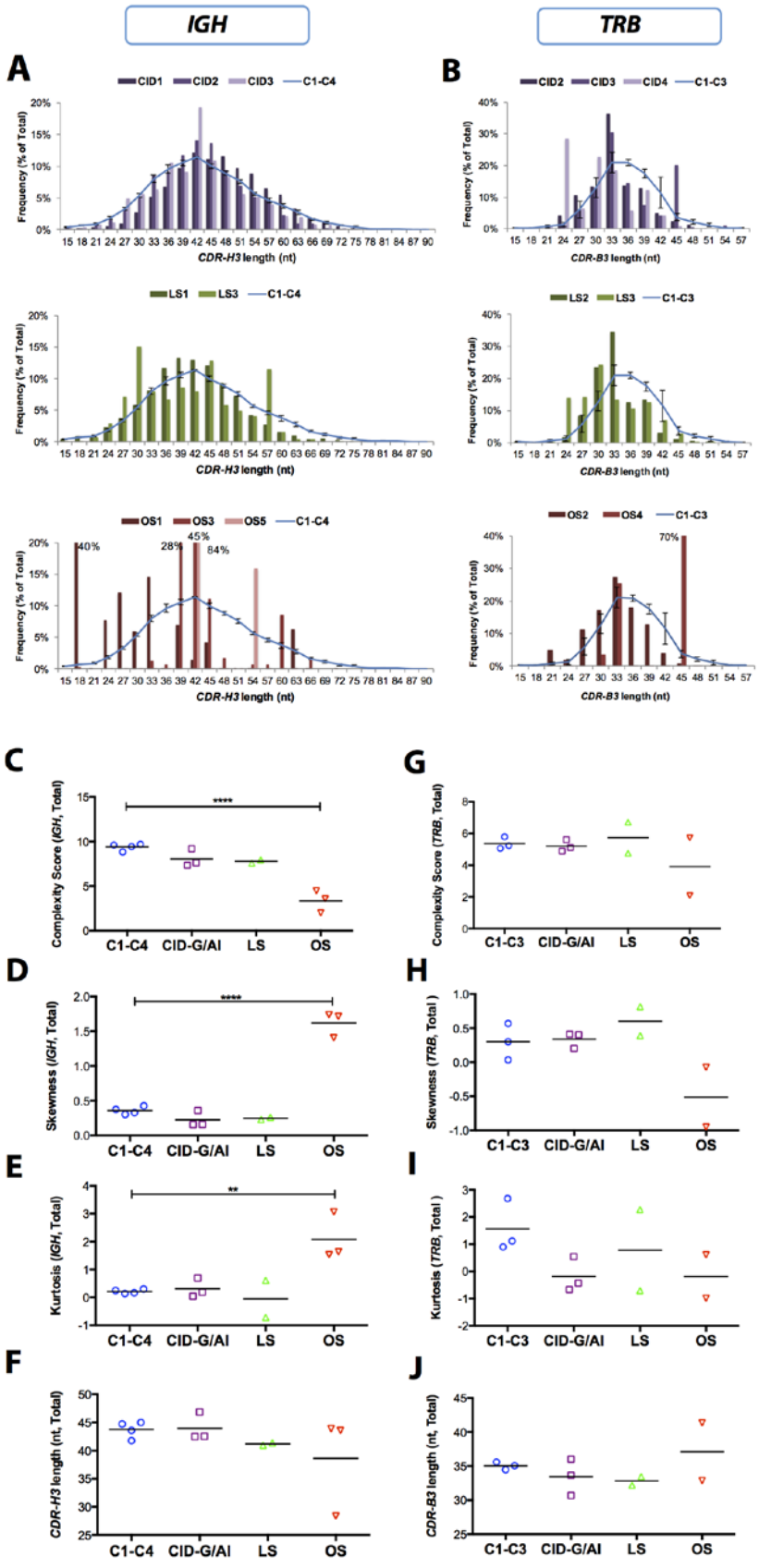
Shannon's H index of diversity for *IGH* (**A**) and *TRB* (**B**) repertoires for patients with RAG deficiency and healthy controls. Gini-Simpson's index of unevenness for *IGH* (**C**) and *TRB* (**D**) repertoires for patients with RAG deficiency and healthy controls. (Mean; ANOVA with Dunnett's correction for multiple comparison; p>0.05 (n.s.) not significant). Data shown are on total sequences.

fig. S2. Differential usage of *V*, *D* and *J* genes in the *IGH* and *TRB* repertoires of patients with RAG deficiency.



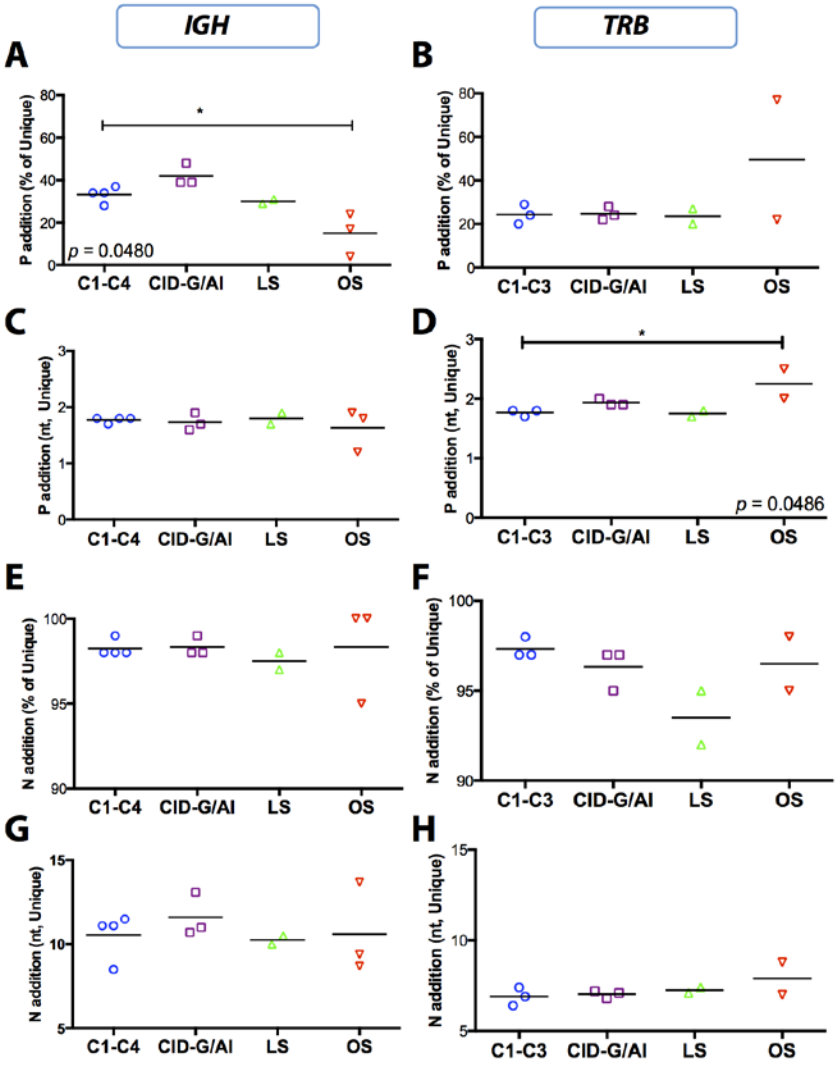
Heat map representing the frequency of *V*, *D* and *J* gene usage among total *IGH* (**A**) and *TRB* (**B**) sequences from healthy controls and patients with *RAG* mutations. Sample (**C**) and variable (**D**) plots for the differential usage of *IGHV*, *IGHD* and *IGHJ* genes, segregating control and patient samples, and the various genes according to Primary Component (PC)1 and PC2. Sample (**E**) and variable (**F**) plots for the differential usage of *TRBV*, *TRBD* and *TRBJ* genes, segregating control and patient samples and the various genes according to PC1 and PC2.

fig. S3. Characterization of the CDR3 region of *IGH* and *TRB* total sequences in peripheral blood lymphocytes.



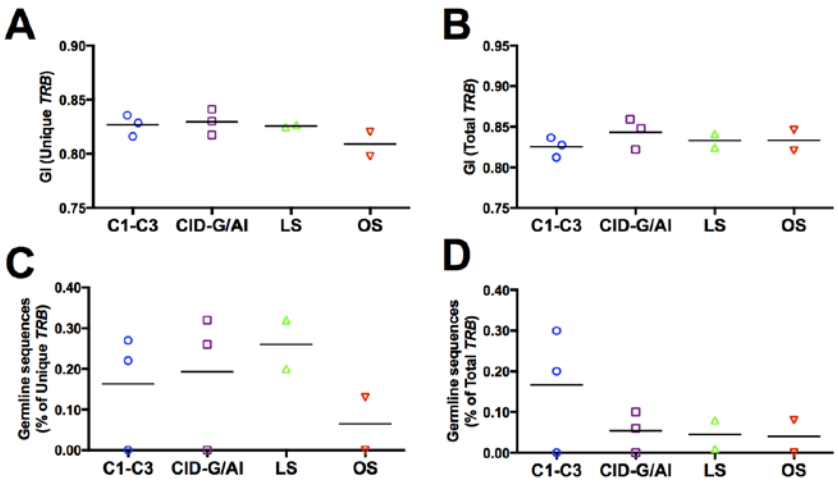
Distribution of the length of the CDR3 region of *IGH* (CDR-H3) (**A**) and *TRB* (CDR-B3) (**B**) total sequences from peripheral blood of patients with RAG deficiency and healthy controls (C1-C4; C6-C8). The distribution of the CDR3 length in healthy controls is depicted as a blue line (representing mean values ± SE). (mean ± SE). Complexity scores (**C, G**), skewness (**D, H**), kurtosis (**E, I**) and average length in nucleotides (nt) (**F, J**) of the *IGH* (**C-F**) and *TRB* (**G-J**) CDR3 unique sequences in patients with RAG deficiency and controls. In panels C-J, for each group, mean values are shown, and statistical significance was assessed by ANOVA.

fig. S4. P and N nucleotide addition in the CDR3 region of unique *IGH* and *TRB* sequences in healthy controls and patients with RAG deficiency.



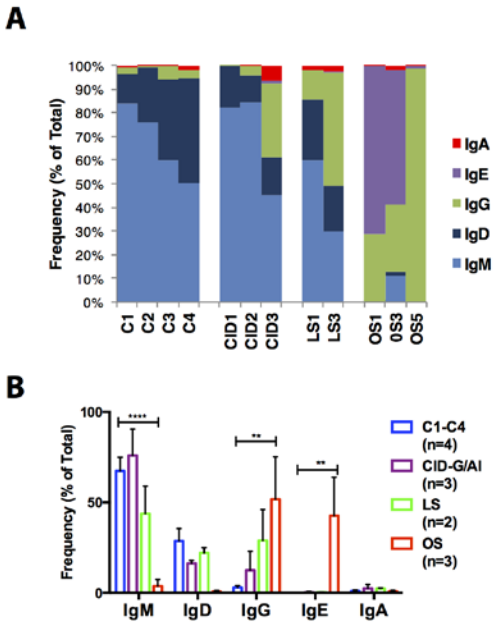
Percentage of unique *IGH* (**A**) and *TRB* (**B**) sequences with P nucleotide addition in healthy controls (C1-C4; C6-C8) and patients with *RAG* mutations and various clinical phenotypes. Average number of P nucleotide addition in *IGH* (**C**) and *TRB* (**D**) unique sequences from healthy controls and patients with *RAG* mutations. Percentage of unique *IGH* (**E**) and *TRB* (**F**) sequences with N nucleotide addition in healthy controls and patients with *RAG* mutations. Average number of N nucleotide addition in *IGH* (**C**) and *TRB* (**D**) unique sequences from healthy controls and patients with RAG mutations. Horizontal bars represent mean values. Statistical significance was assessed by ANOVA.

fig. S5. Germline index and frequency of sequences without P and N nucleotides in the *TRB* repertoire of patients with RAG deficiency and controls.



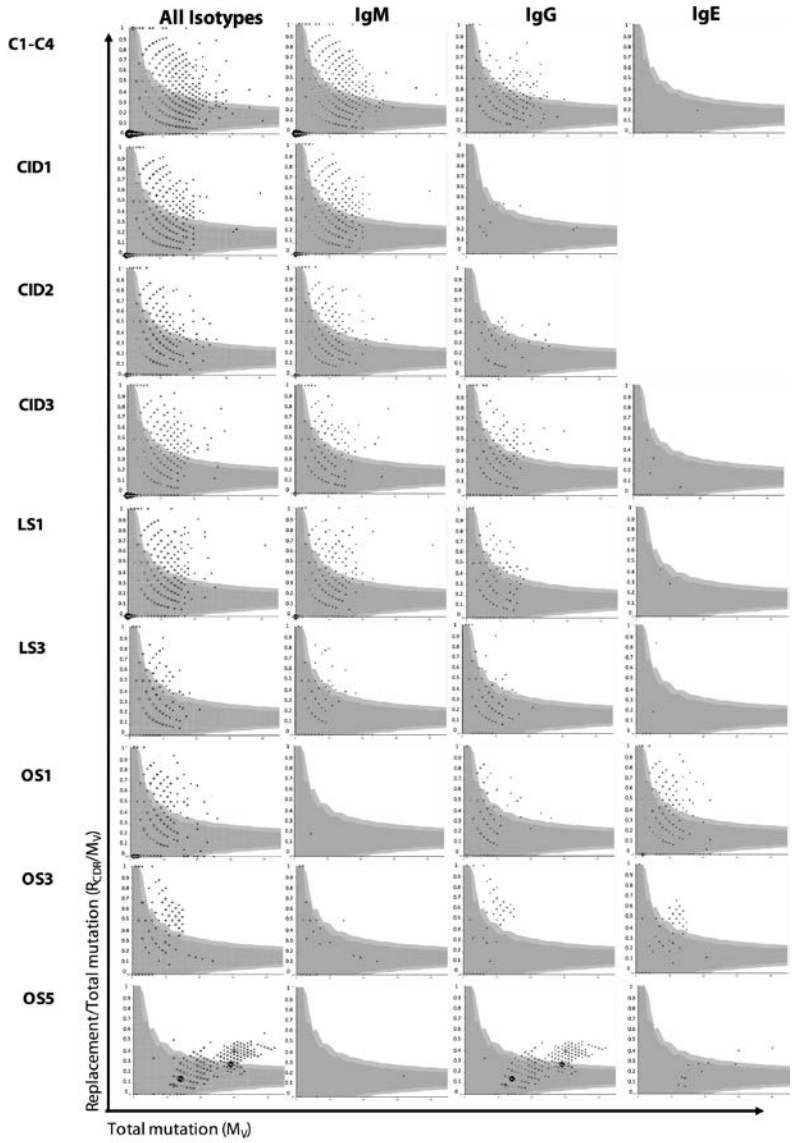
Germline index (GI) of unique (**A**) and total (**B**) *TRB* sequences from healthy controls (C1-C4; C6-C8) and *RAG*-mutated patients. Proportion of unique (**C**) and total (**D**) *TRB* transcripts without P and N nucleotide addition.

fig. S6. Distribution of immunoglobulin heavy chain isotypes among total *IGH* sequences from controls and patients with RAG deficiency.



Frequency of immunoglobulin heavy chain constant gene usage among total *IGH* sequences from peripheral blood lymphocytes of RAG-deficient patients and healthy controls (**A, B**). In panel **B**, mean values ± SE are shown. Statistical analysis was performed with unpaired t-test.

fig. S7. Inference of antigen-mediated selection in combined and individual IgM, IgG and IgE transcripts from healthy controls (C1-C4) and patients with RAG deficiency.



Graphical representation of antigen-mediated selection in all isotypes, and in *IGHM*, *IGHG* and *IGHE* transcripts, with the ratio of replacement mutations in CDR-H1 and CDR-H2 (R_{CDR}) to the total number of mutations in the V region (M_V), plotted against M_V. The dark shaded area represents 90% confidence limit for the probability of random mutations, and the light grey shaded area represents 95% confidence limits for the probability of random mutations. Data points falling outside the light grey shaded area represent the portion for antigen-mediated selection, with probability of random mutation being $p = 0.1$ (dark grey area) and $p = 0.05$ (light grey area).