Published in final edited form as:

Sci Immunol. 2016 December 16; 1(6): . doi:10.1126/sciimmunol.aah6109.

Characterization of T and B cell repertoire diversity in patients with RAG deficiency*

Yu Nee Lee, Ph.D.[1,2], **Francesco Frugoni, Ph.D.**[1], **Kerry Dobbs, B.S.**[3], **Irit Tirosh, M.D.**[1], **Likun Du, Ph.D.**[1], **Francesca A. Ververs, M.D.**[1], **Heng Ru, Ph.D.**[4], **Lisa Ott de Bruin, M.D.**[1], **Mehdi Adeli, M.D.**[5], **Jacob H. Bleesing, M.D., Ph.D.**[6], **David Buchbinder, M.D.**[7], **Manish J. Butte, M.D., Ph.D.**[8], **Caterina Cancrini, M.D.**[9], **Karin Chen, M.D.**[10], **Sharon Choo, M.D.**[11], **Reem A. Elfeky, M.D.**[12], **Andrea Finocchi, M.D.**[9], **Ramsay L. Fuleihan, M.D.**[13], **Andrew R. Gennery, M.D.**[14], **Dalia H. El-Ghoneimy, M.D.**[12], **Lauren A. Henderson, M.D.**[1], **Waleed Al-Herz, M.D.**[15], **Elham Hossny, M.D. Ph.D.**[12], **Robert P. Nelson, M.D**[16], **Sung-Yun Pai, M.D.**[17], **Niraj C. Patel, M.D.**[18], **Shereen M. Reda, M.D. Ph.D.**[12], **Pere Soler-Palacin, M.D.**[19], **Raz Somech, M.D., Ph.D.**[2], **aolo Palma, M.D.**[9], **Hao Wu, Ph.D.**[4], **Silvia Giliani, Ph.D.**[20], **Jolan E. Walter, M.D., Ph.D.**[1,21], and **Luigi D. Notarangelo, M.D.**[3,*]

[1]Division of Immunology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA.

[2]Pediatric Department A and the Immunology Service, "Edmond and Lily Safra" Children's Hospital, Jeffrey Modell Foundation Center, Sheba Medical Center, Tel Hashomer, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

[3]Laboratory of Host Defenses, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA.

[4]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA; Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA 02115, USA.

[5]Pediatrics Department, Weill Cornell Medical College, Hamad Medical Corporation, Doha, Qatar.

[6]Division of Hematology/Oncology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA.

[7]Division of Hematology, Children's Hospital Orange County, Orange County, CA, USA.

*Address correspondence to: L.D.N. luigi.notarangelo2@nih.gov, Laboratory of Host Defenses, NIAID, NIH, 10 Center Drive, MSC 1456, Building 10 CRC, Room 5-3950, Bethesda, MD 20892.

Author contributions: YNL and LDN designed the study, supervised analysis of the data and wrote the manuscript; YNL, FF, KD, IT, LD, FV, HR, and HW performed the experiments and analyzed the data; YNL performed statistical analysis; MA, JHB, DB, MB, CC, KC, SC, RAEF, AF, RLF, ARG, DHEG, LAH, WAH, EH, RPN, SYP, NCP, SMR, PSP, RS, PP, JEW, SG and LDN provided valuable samples and gathered the data. HR and HW analyzed the potential effects of the mutations on the RAG1/RAG2 heterotetrameric structure. LODB helped with the manuscript writing.

Competing interests: The authors declare no competing interests.

Data and material availability: Please see Supplementary Excel file for the *IGH* and *TRB* sequences.

[8]Department of Pediatrics, Division of Allergy, Immunology, Rheumatology, Stanford University, Stanford, California, USA.

[9]DPUO, University Department of Pediatrics, Bambino Gesù Children's Hospital and University of Tor Vergata School of Medicine, Rome, Italy.

[10]Division of Allergy and Immunology, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, USA.

[11]Department of Immunology, The Royal Children's Hospital, Melbourne, Australia.

[12]Department of Pediatric Allergy and Immunology, Children's Hospital, Faculty of Medicine, Ain Shams University, Cairo, Egypt.

[13]Division of Allergy and Immunology, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, IL, USA.

[14]Department of Paediatric Immunology, Great North Children's Hospital, United Kingdom, and Institute of Cellular Medicine, Newcastle University, Newcastle Upon Tyne, UK.

[15]Department of Pediatrics, Faculty of Medicine, Kuwait University, Kuwait.

[16]Division of Hematology and Oncology, Indiana University School of Medicine, Indianapolis, IN, USA.

[17]Division of Hematology-Oncology, Boston Children's Hospital, Boston, MA, USA.

[18]Department of Pediatrics, Division of Infectious Disease and Immunology, Levine Children's Hospital, Carolinas Medical Center, Charlotte, NC, USA.

[19]Paediatric Infectious Diseases and Immunodeficiencies Unit, Vall d'Hebron University Hospital, Barcelona, Spain.

[20]A. Nocivelli Institute for Molecular Medicine, Department of Molecular and Translational Medicine, University of Brescia, and the Section of Medical Genetics, Department of Pathology Spedali Civili, Brescia, Italy.

[21]Division of Pediatric Allergy/Immunology, University of South Florida, and Johns Hopkins All Children's Hospital, St. Petersburg, FL, USA.

Abstract

Recombination Activating Genes 1 and 2 (RAG1 and RAG2) play a critical role in T and B cell development by initiating the recombination process that controls expression of T cell receptor (TCR) and immunoglobulin genes. Mutations in the RAG1 and RAG2 genes in humans cause a broad spectrum of phenotypes, including severe combined immune deficiency (SCID) with lack of T and B cells, Omenn syndrome, leaky SCID, and combined immune deficiency with granulomas or autoimmunity (CID-G/AI). Using next generation sequencing, we analyzed the T and B cell receptor (TCR, BCR) repertoire in 12 patients with RAG mutations presenting with Omenn syndrome (n=5), leaky SCID (n=3), or CID-G/AI (n=4). Restriction of repertoire diversity skewed usage of Variable (V), Diversity (D), and Joining (J) segment genes, and abnormalities of CDR3 length distribution were progressively more prominent in patients with a more severe phenotype. Skewed usage of V,D and J segment genes was present also within unique sequences, indicating a

primary restriction of repertoire. Patients with Omenn syndrome had a high proportion of class-switched immunoglobulin heavy chain transcripts and increased somatic hypermutation rate, suggesting in vivo activation of these B cells. These data provide a framework for better understanding the phenotypic heterogeneity of RAG deficiency.

Introduction

The RAG1 and RAG2 proteins are expressed in developing lymphocytes and play a critical role in the assembly of interspersed Variable (V), Diversity (D), and Joining (J) gene elements at the immunoglobulin (Ig) and T cell receptor (TCR) loci, thereby initiating the VDJ recombination process that allows development of B and T cells and the establishment of adaptive immunity. Patients with null mutations in the *RAG1* or *RAG2* genes manifest a block in the development of B and T cells, resulting in T^- B^- severe combined immune deficiency (T^- B^- SCID). However, hypomorphic mutations in the *RAG* genes may allow development of a variable number of B and T cells, associated with various distinct clinical and immunological phenotypes. In particular, Omenn syndrome (OS) is characterized by generalized skin rash, lymphadenopathy, hepatosplenomegaly, eosinophilia, hypogammaglobulinemia but elevated serum IgE, lack of circulating B cells and the presence of oligoclonal, activated, autologous T cells (1). Atypical or leaky SCID (LS) is characterized by the presence of T (and in some cases, B) cells, with variably affected T cell function and without clinical features of OS (2). Another form of LS with expansion of T cells expressing the γδ form of the TCR occurs especially in patients with cytomegalovirus infection (3, 4). More recently, hypomorphic *RAG* mutations were identified in patients with delayed-onset combined immunodeficiency associated with granulomas and/or autoimmunity (CID-G/AI) (5, 6), or in other, more rare, milder and atypical presentations, including CD4 lymphopenia (7), common variable immune deficiency (8), selective deficiency of anti-polysaccharide antibody responses (9), and pyoderma gangrenosum (10). These heterogeneous clinical phenotypes are associated with a broad spectrum of nonsense, frameshift, in-frame deletion or insertion, and missense mutations of the *RAG1* and *RAG2* genes that affect various domains of the respective proteins (11).

By individually introducing a large number of human *RAG1* and *RAG2* genetic variants into Abelson virus-transformed *Rag1*$^{-/-}$ (or *Rag2*$^{-/-}$) pro-B cells carrying an inverted GFP cassette flanked by recombination signal sequences (RSS), we previously demonstrated that the severity of the clinical presentation correlates with the level of residual recombination activity supported by the mutant RAG1 protein (12, 13). In this assay, mutations with low levels of recombination activity generated fewer rearrangements at the endogenous immunoglobulin heavy chain (*Ighc*) locus, as compared to mutations with higher residual activity (12), suggesting that individual *RAG* mutations may exert different effects on immune repertoire diversity and composition. Here, we report the results of next generation sequencing (NGS) of T and B cell repertoire composition and diversity in 12 patients with *RAG* mutations, representative of the extended phenotypic spectrum of the disease. Our results demonstrate that abnormalities of T and B cell repertoires correlate with the severity of the clinical and immunological phenotype, thus further supporting genotype-phenotype correlation in this disease. Distinctive signatures of individual *V*, *D*, and *J* gene usage, and of

CDR3 composition and length distribution have been identified in patients with different phenotypes, and may contribute to the generation of an immune repertoire enriched in self-reactive specificities.

Results

Patient Characteristics

The 12 patients included in this study were assigned to three distinct groups, based on the clinical and immunological phenotype (2) (Table S1). Five patients (OS1-OS5) presented with clinical and laboratory features of OS. Three patients (LS1-LS3) presented with LS and severe CMV infection, and two of them had an increased proportion of TCR$\gamma\delta^+$ T cells ($\gamma\delta$T). Four patients (CID1-CID4) were included in the CID-G/AI group based on a clinical history of autoimmunity and/or presence of granulomas. Eleven patients carried *RAG1* and one patient carried *RAG2* bi-allelic mutations, for a total of fifteen *RAG1* and one *RAG2* distinct mutant alleles. Recombination activity of the mutant alleles was tested using Abelson virusimmortalized *Rag1*$^{-/-}$ or *Rag2*$^{-/-}$ pro-B cells as previously described (12, 13). Patients in the OS and LS subgroups carried *RAG* mutant alleles that supported only modest levels of recombination activity (<7% of wild-type, with a mean of 2.29%). By contrast, patients in the CID-G/AI group carried at least on one *RAG* mutant allele that conferred higher levels of recombination activity (Table S2).

Progressive restriction of the immune repertoire correlates with the severity of the clinical phenotype

To analyze and compare *IGH* and *TRB* repertoire diversity in patients with various clinical phenotypes associated with *RAG* mutations (Fig. 1A and Table S1), we have performed NGS of the *IGH* and *TRB* transcripts expressed by circulating B and T cells, respectively. The number of total and unique sequences of rearranged *IGH* and *TRB* products for each of the RAG-deficient patients and healthy infant controls is reported in Table S2. Of note, productive *IGH* rearrangements were detected in 3 out of five OS patients, despite virtual lack of circulating B cells. A lower ratio of unique/total *IGH* and *TRB* sequences was detected in *RAG*-mutated patients vs. healthy controls (Table S2), and this difference reached statistical significance for the *IGH* repertoire ($p<0.05$).

A graphical representation of repertoire diversity is conveyed by tree maps of the *IGH* and *TRB* (Fig. 1B) repertoires, where each dot represents a unique V-J pair and the size of each dot corresponds to the frequency of that rearrangement in the total population of sequences obtained. Marked reduction of both *TRB* and *IGH* repertoire diversity, associated with clonotypic expansions, was detected in samples from OS patients, and to a lesser extent in patients with LS. By contrast, a more diversified *IGH* repertoire was present in samples from the CID-G/AI group. However, the *TRB* repertoire of CID-G/AI patients was characterized by restrictions and clonotypic expansions.

To provide more quantitative measures of repertoire diversity and complexity, we took advantage of commonly used ecological parameters. In particular, the Shannon's H index measures repertoire diversity, taking into account both the number of total sequences and

clonal size distribution in the overall repertoire. As compared to healthy donors, RAG-deficient patients had a lower Shannon's H index for both *IGH* (Fig. 1C) and *TRB* (Fig. 1D) repertoires. When the same analysis was applied to each of the three subgroups of patients with RAG deficiency (CID-G/AI, LS, and OS), a significant reduction of the Shannon's H index was observed only for the *IGH* repertoire in OS patients (fig. S1, A and B). Next, to assess more precisely clonal size distribution, we calculated the Gini-Simpson index of unevenness, which measures the inequality in the relative representation of species observed in a given sample, so that the higher the Gini Simpson index, the more unequal the distribution of individual clonotypes. An uneven distribution of both *IGH* and *TRB* clonotypes was observed in RAG-deficient patients vs. controls (Fig. 1, E and F). This difference was statistically significant for OS patients, but a clear trend was observed for the *TRB* repertoire of CID-G/AI and LS patients (fig. S1, C and D).

To analyze further the presence of clonotypic expansions, we estimated the diversity 50 (D50) index (14), which corresponds to the percentage of unique CDR3 sequences that account for 50% of the total number of sequences observed. Less than 10% of the unique clonotypes accounted for 50% of the total number of *IGH* sequences in patients with OS (Fig. 1G). Clonotypic expansions, resulting in markedly reduced D50, were observed in the *TRB* of all *RAG*-mutated patients, irrespective of their clinical phenotype (Fig. 1H). The top 100 most abundant IGH and TRB CDR3 (CDR-H3, CDR-B3) clonotypes accounted for less than 0.3% of all transcripts in healthy donors. With the exception of a single clonotype in patient CID3, all patients with CID-G/AI showed a similar representation of CDR-H3 clonality, whereas a significant expansion of CDR-H3 clonotypes was detected in patients with OS and in patient LS3 (Fig. 1I). A different pattern was observed for the top 100 CDR-B3 clonotypes. In particular, two CDR-B3 clonotypes accounted for more than 50% of all total sequences in patient OS4, and a significant expansion of CDR-B3 clonotypes was also observed in patients with CID and LS (Fig. 1J). Overall, these data demonstrate that restriction and clonotypic expansions characterize both the *TRB* and the *IGH* repertoires of patients with OS, whereas in patients with CID-G/AI and LS abnormalities of repertoire diversity are comparatively subtle and largely confined to T cells.

Finally, to assess whether analysis of the T and B cell repertoire may distinguish RAG-deficient patients from healthy controls, we used Principal Component Analysis (PCA) based on five variables: the number of total and unique sequences, Shannon's H index, Gini-Simpson index, and recombination activity of the mutant RAG protein. PCA successfully segregated healthy donors from the patients (Fig. 1, K and L), and permitted discrimination amongst different groups (CID-G/AI, LS and OS) of RAG-deficient patients, especially with respect to the *IGH* repertoire (Fig. 1K).

Non-stochastic restriction of *IGH* and *TRB* repertoires and skewed usage of *V*, *D*, and *J* genes in RAG-mutated patients

The analysis of repertoire diversity and composition among unique and total sequences permits distinguishing between constraints that occur during generation of the primary repertoire vs. secondary effects that occur in the periphery, such as clonotypic expansions in response to non-self or self-antigens. To determine whether *RAG* mutations alter targeting of

individual *V*, *D*, and *J* genes, we generated heat maps comparing usage of these genes in unique *IGH* (Fig. 2A) and *TRB* (Fig. 2B) sequences from healthy controls and patients. In these panels, coding genes are ordered according to their location along the chromosome, making it possible to ascertain whether skewed gene usage could reflect topological constraints. The summary for the Chi-squared test for goodness of fit shows that in most *RAG*-mutated patients the distribution of *V, D* and *J* gene usage among unique *IGH* and *TRB* sequences was distinct from that observed in healthy controls (Fig. 2, A and B). Results were similar when the same analysis was applied to total sequences (fig. S2, A and B).

The frequency of individual *V, D*, and *J* gene usage among unique sequences of the *IGH* repertoire was different in healthy controls and *RAG*-mutated patients (Fig. 2C, upper panel). For example, the *IGHJ3* gene was the second most frequently utilized *IGHJ* gene in patients with RAG deficiency, whereas it was only fourth in order in healthy controls. Conversely, *IGHJ6* was the second most commonly utilized gene in controls, but was only fourth in order in RAG-mutated patients. The *IGHD6* gene was the third in frequency among healthy controls, but it was the most commonly used in patients LS1 and OS5. Similar abnormalities were also observed in the frequency of usage of *TRBV*, *TRBD*, and *TRBJ* genes in patients vs. controls (Fig. 2D, upper panel). Moreover, for both *IGH* and *TRB* repertoires, the distribution of gene usage varied among *RAG*-mutated patients. The observation that such differences were present when analyzing unique sequences suggested that hypomorphic *RAG* mutations may alter selection of genes involved in V(D)J recombination during generation of the primary immune repertoire. To test this hypothesis, we analyzed the pattern of usage of individual *V*, *D*, and *J* genes at early stages of B and T lymphocyte development. In particular, we compared the frequency of *IGHV*, *IGHD*, and *IGHJ* gene usage in *IGH* transcripts from immortalized *Rag1*$^{-/-}$ pro-B cells engineered to express wild-type or mutant human RAG1 (12). As shown in Fig. 2C, lower panel, the frequency of usage of individual *V*, *D*, and *J* genes among rearranged *IGH* products was different when comparing cells reconstituted with wild-type or mutant RAG1. To assess whether RAG mutations affect composition of T cell repertoire at early stages of T cell development, we analyzed the frequency of usage of *TRBV*, *TRBD*, and *TRBJ* genes among productive *TRB* rearrangements during in vitro T cell differentiation of induced pluripotent stem cells (15). A different pattern of gene usage was observed in control vs. *RAG1* mutant cells (Fig. 2D, lower panel). Altogether, these data confirm what observed in vivo in the patients (Fig. 2C and 2D, upper panels) and indicate that hypomorphic *RAG* mutations affect not only the efficiency, but also the quality of the V(D)J recombination process.

To further illustrate this, PCA of individual *IGHV, D* and *J* gene usage clearly segregated patients from controls, and even distinguished among patients with different phenotypes (Fig. 2E, left panels). In the variable plot analysis (Fig. 2E, right panels), the distribution of the ten most abundantly used *IGHV*, *IGHD*, and *IGHJ* genes along PC1 and PC2 is shown for the entire population of subjects analyzed. By overlaying sample and variable plots, it was possible to define which genes are preferentially utilized in each subgroup of patients. Thus, the T and B cell repertoire of CID-G/AI patients included over-representation of *IGHV 3-9*; *IGHD 2-2, 3-9, 4-11, 4-17, 7-27*; and, *IGHJ3* genes. The *IGHJ 1*, *2* and *4* genes were more abundantly used in patients with OS and LS, and *IGHD 2-8* and *IGHD6-25* were

preferentially used in OS patients. Similarly, PCA segregated the patients from the controls based on usage of *TRBV*, *TRBD*, and *TRBJ* genes (Fig. 2F), with clear distinction among the various groups of *RAG*-mutated patients when the analysis was conducted on *TRBV* and *TRBJ* gene usage. Upon overlaying sample plots with variable plots, patients with CID-G/AI were found to have increased usage of *TRBJ 1-1, 1-3* and *1-5*, whereas overutilization of *TRBV5-8, 11-2, 4-1* and *18* was detected in patients with LS, and preferential usage of *TRBJ 2-7* was observed in patients with OS. Similar results were obtained when PCA analysis of *IGH* and *TRB* gene usage was conducted on total sequences (fig. S2, C, D, E and F). Altogether, these data demonstrate that usage of individual *V*, *D*, and *J* genes distinguishes *RAG*-mutated patients from controls, with a specific signature of gene expression among patients with distinct phenotypes.

Abnormalities of CDR3 length and amino acid composition

Abnormalities of the Complementarity Determining Region 3 (CDR3) of immunoglobulin and T cell receptor molecules CDR3 length and composition often have a significant impact on the ability to mount effective adaptive immune responses to a wide range of non-self antigens, and may also contribute to increased recognition of self antigens in patients with autoimmune diseases (16, 17). Analysis of CDR3 length distribution of both unique (Fig. 3, A and B) and total (fig. S3, A and B) *IGH* and *TRB* transcripts demonstrated progressive skewing of the CDR3 length profile from patients with less severe to patients with more severe clinical phenotype. In order to better define abnormalities of CDR3 length distribution, we calculated the CDR3 complexity score, and measured CDR3 skewness and kurtosis (18–20). In particular, the complexity score takes into account the number of major peaks of CDR3 length (defined as those with amplitudes of at least 10% of the sum of all peak heights), and their height contribution to the sum of all peak heights. Skewness measures the asymmetry of CDR3 length distribution above and below the mean. Finally, kurtosis measures the amount of events in the central part of the CDR3 distribution as opposed to the tails, and therefore defines the degree of peakedness. Significant differences in the complexity score, skewness and kurtosis of the CDR-H3 length profiles of *IGH* repertoire were observed for both unique and total sequences from OS patients, whereas the CDR-H3 profile of CID-G/AI and LS patients was similar to the profile observed in healthy controls (Fig. 3, C, D, and E; fig. S3, C, D, and E). The CDR-H3 length was not significantly different in patients vs. controls (Fig. 3F and fig. S3F). No significant differences were observed for the *TRB* repertoire (Fig. 3, G to J; fig. S3, G to J), although there was a trend towards reduced kurtosis in all patient groups. Overall, the observation that a similar pattern was detected in unique and total sequences from *RAG*-mutated patients further indicates that abnormalities of *IGH* and *TRB* repertoires in these patients can be predominantly attributed to the effect of *RAG* mutations in shaping the primary repertoire.

The CDR3 length is determined not only by the length of V, D and J gene sequences that are part of it, but also by the addition of palindromic (P) and ‘N’ nucleotides that contribute to its junctional diversity. The germline index (GI) can be used to estimate the abundance of P and N nucleotides, and is calculated by dividing the number of nucleotides in the CDR3 that are encoded by *V*, *D* and *J* genes by the total number of nucleotides contained in the CDR3, generating a value between 0 and 1 (21). A GI value of 1 indicates lack of P and N

nucleotide addition; thus, the higher the GI, the lower the junctional diversity. With the exception of a lower proportion of unique sequences containing P nucleotides in patients with OS, no significant differences were observed in P and N nucleotide addition (fig. S4) and in GI value (fig. S5, A and B) within *IGH* and *TRB* sequences from patients with *RAG* mutations and controls. Furthermore, CDR-B3 sequences from RAG-mutated patients had a GI value between 0.8 and 0.85, which is similar to the GI value observed in healthy controls in this study (fig. S5A) and in previous reports (21). Finally, when comparing the proportion of unique (fig. S5C) and total (fig. S5D) CDR-B3 sequences with a GI=1, there was a trend towards a reduced proportion of these among total sequences from patients with *RAG* mutations, irrespective of the disease phenotype. The average hydrophobicity of the amino acids within the CDR-H3 loop of circulating B cells forms a Gaussian distribution centering on neutrality to mild hydrophilicity (22, 23). In patients with RAG deficiency, skewed usage of *V*, *D*, and *J* genes may affect amino acid composition and the hydrophobicity profile of the CDR-H3 region, with potentially important consequences for antigen binding. We have observed increased usage of *IGHJ3* in patients with CID-G/AI (Fig. 2A) and decreased usage of *IGHJ6* in all of the patients (Fig. 4A). The *IGHJ6* gene encodes for five tyrosine residues (Y) in the CDR-H3, whereas *IGHJ3* and *IGHJ5* do not encode for any tyrosine residue (Fig. 4B). Indeed, a low content of tyrosine residues was detected in the CDRH3 region of immunoglobulin transcripts from patients with *RAG* mutations, reaching statistical significance in patients with CID-G/AI (Fig. 4C). This decreased presence of tyrosine residues was associated with abnormalities of the hydrophobicity profile of the CDR-H3 region in seven out of eight *RAG*-mutated patients, as measured by the normalized Kyte-Doolittle index of hydrophobicity of unique (Fig. 4D) and total (Fig. 4E) CDR-H3 sequences, suggesting that enrichment for immunoglobulin transcripts with an altered hydrophobicity profile in patients with *RAG* mutations occurs at the level of primary repertoire generation and is not simply due to expansion of selected clonotypes in the periphery.

OS is characterized by infiltration of peripheral tissues by activated, possibly self-reactive T cells. Recent data have shown that the hydrophobicity of amino acids (aa) at positions 6 and 7 of the 13 aa-long CDR-B3 promote the development of self-reactive T cells (24). Amino acid composition at positions 6 and 7 of the 13 aa-long CDR-B3 was very conserved in healthy controls, but not in *RAG*-mutated patients (Fig. 4F and 4G, upper panels). Furthermore, increased usage of hydrophobic amino acids was observed at both positions 6 and 7 in patient OS2, whereas the CDR-B3 of patient OS4 was enriched for hydrophobic aa at position 7 (Fig. 4F and G, lower panels).

Abnormalities of immunoglobulin class switching and somatic hypermutation in patients with OS

Abnormalities of T and B cell development in patients with *RAG* mutations compromise immune responses in peripheral lymphoid organs, including production of antibodies of various isotypes. In particular, low IgG and low IgA, but elevated IgE serum levels, are typically seen in patients with OS, and may be observed also in LS (25). However, virtual lack of B cells in patients with OS has so far precluded analysis of the distribution of B cells expressing various isotypes in the periphery of these patients. NGS analysis of the B cell

repertoire, with use of RNA as a template and reverse primers in the *IGHC* region, permitted analysis of the relative abundance of immunoglobulin transcripts containing various heavy chain isotypes in the peripheral blood of patients with *RAG* mutations and healthy controls. Consistent with the notion that unswitched cells comprise the majority of circulating B cells in normal individuals, switched transcripts represented less than 5% of all *IGH* productive transcripts detected in healthy controls both within both unique (Fig. 5, A and B) and total (fig. S6, A and B) sequences, as also previously reported (26). By contrast, an increased frequency of *IGHG* transcripts was observed within unique and total sequences of CID3, LS3, OS1, OS3 and OS5 patients, and increased frequency of *IGHE* transcripts was observed in OS1 and OS3 patients (Fig. 5A and fig. S6). Overall, patients with OS showed an increased frequency of *IGHG* and *IGHE* transcripts (Fig. 5B and fig. S6B). The most abundant CDR-H3 clonotypes accounted for a large proportion of total sequences in patients with OS (Fig. 5C), and were mainly represented by *IGHE* and/or *IGHG* transcripts (Fig. 5D). These data suggest that the few circulating B cells in patients with OS are represented by oligoclonal populations that have switched to IgE and IgG.

Somatic Hyper Mutation (SHM) introduces additional diversity in the *IGH* repertoire of mature B cells and allows selection of high-affinity antibodies. The SHM rate per 1000 nucleotides (nt) for all combined isotypes was higher in RAG-deficient patients than in controls, and this increase was more pronounced in patients with OS (Fig. 5E). When SHM was analyzed separately for the various isotypes, healthy controls showed a lower rate of SHM in *IGHM* transcripts than in switched transcripts, as expected. As compared to healthy controls, patients CID1, CID2, LS1 and OS5 had a higher SHM rate among *IGHM* transcripts (Fig. 5E). As for *IGHG* transcripts, a lower rate of SHM was observed in most of the patients compared to controls, with the exception of patients CID2 and OS5 (Fig. 5E). In the latter, a higher rate of SHM rate was observed also among *IGHE* transcripts. To determine whether SHM detected in Ig transcripts from patients with RAG deficiency reflects in vivo antigen-mediated selection, we have assessed the distribution of replacement and silent mutations based on the Lossos multinomial model (27) (fig. S7). A clear evidence for antigen-mediated selection was observed in switched transcripts from patients OS3 and OS5 (Fig. 5F). Furthermore, both in patients with CID-G/AI and in controls, the rate of SHM was slightly increased in the preferentially rearranged *IGHV3-9* gene as compared to the mean in all other genes, and this phenomenon was associated with clear evidence of antigen-mediated selection (Fig. 5G). Overall, these data indicate that despite restriction of primary repertoire generation, B cell function in patients with RAG deficiency remains intact with respect to CSR and SHM.

Mapping the disease-related mutations onto the synaptic-RAG complex models

Recently, crystallography and cryo-electron microscopy have allowed high resolution determination of the RAG1/RAG2 heterotetrameric complex (28, 29). We have mapped the twelve missense mutations from our patient cohort onto the cryo-electron microscopy structure in complex with synapsed RSS DNAs (28), and compared the predicted structural and functional effects (Fig. 6A) with analysis of recombination activity (12). The *RAG1* mutations at positions R396, R404, R737, R841 and R973 resulted in less than 10% recombination activity (12). While residues R396, R737 and R973 directly engage in RSS

binding using their side chains, residue R841 plays a role in stabilizing the closed conformation of the synaptic RAG complex by forming a salt bridge with the symmetric RAG1 (28) (Fig. 6B). The *RAG1* mutation at position R404 likely affects RSS binding and conformation of the nonamer binding domain dimer because it stacks with R443 of the other RAG1 monomer (28, 29) (Fig. 6B). In contrast, the *RAG1* mutation H612R did not compromise the recombination activity (12), likely due to the compensation of RSS binding by the mutated arginine (Fig. 6B).

Discussion

In the present study, we have demonstrated that the T and B cell repertoire of patients with OS is characterized by markedly reduced diversity, and a non-stochastic restriction of *V*, *D*, and *J* gene usage. Restriction of *TRB* repertoire diversity, with skewed V-J gene usage, in OS patients has also been reported by Yu et al. in four patients with OS (21). These data suggest that severe *RAG* mutations may impose constraints during generation of primary T and B cell repertoires, and that peripheral expansion in response to self or non-self antigens is not the only factor involved in the dominance of few T cell clonotypes.

In contrast to what is known about T and B cell abnormalities in patients with OS, limited information is available on the richness and complexity of T and B cell repertoire in patients with milder forms of RAG deficiency. Flow cytometric analysis of the expression of various TCR Vβ families and CDR3 spectratyping have revealed that patients with CID-G/AI and with CD4 lymphopenia often maintain a largely polyclonal T cell repertoire, although in some cases underand over-representation of individual TCR Vβ families have been documented (5, 7, 30, 31). However, these methods have limitations and do not permit analysis of CDR3 composition and use of individual V, D, and J genes. Here, we have shown that patients with CID-G/AI have reduced T cell repertoire diversity, with clonotypic expansions, and maintain a largely diversified B cell repertoire, with even distribution of individual clonotypes, but skewed usage of *IGHV*, *IGHD*, and *IGHJ* genes. We have also demonstrated that analysis of repertoire diversity and composition may distinguish not only *RAG*-mutated patients from controls, but correctly identifies patients with distinct clinical phenotypes. However, genotype-phenotype correlation in RAG deficiency is not absolute. Ijspeert et al. have demonstrated that patients carrying *RAG* mutations that affect the same region in the non-core domain of the RAG1 molecule, and allow similar levels of recombination activity, may present with distinct clinical and immunological features, thus emphasizing the role played by other genetic and epigenetic factors in determining the phenotype (32).

The catalytic core of RAG1 contains two coding flank-sensitive regions, at amino acid (aa) 609–614 (33) and aa 892–977 (34). Several *RAG*1 mutations associated with CID-G/AI fall within these coding flank-sensitive regions (11, 28), including the p.H612R mutation in patient CID1 and the p.F974L mutation in patient CID3. Studies in vitro had suggested that missense mutations in the coding flank-sensitive regions of RAG1 may perturb repertoire composition not just by affecting DNA cleavage, but also by preferentially targeting some coding elements (35). Indeed, patients CID1 and CID3 showed a skewed usage of individual *V*, *D*, and *J* genes. In particular, we have demonstrated increased usage of *IGHV3-9*,

IGHV4-31, and *IGHV3-23*, that are also expressed by autoantibody-secreting B cells in patients with tumors (36). These data suggest that perturbation of repertoire composition may contribute to the immune dysregulation observed in patients with CID-G/AI.

The length and amino acid composition of the immunoglobulin CDR3 region affect recognition of antigens. Progressive reduction of CDR3 length, and increase of highly hydrophobic and hydrophilic sequences during differentiation from immature to naïve and memory B cells are paralleled by a progressive decrease in the proportion of self-reactive B cell specificities during B cell ontogeny (37). We identified abnormalities of hydrophobicity profile of the CDR-H3 region in seven out of eight patients with hypomorphic *RAG* mutations, including all three patients with CID-G/AI tested. Close examination of amino acid composition revealed decreased frequency of tyrosine residues, which are abundant in CDR-H3 sequences of peripheral blood B cells from healthy controls (38). These abnormalities reflected markedly decreased usage of the *IGHJ6* gene, which encodes for five tyrosine residues. Furthermore, we have reported that the CDR-B3 of patients with OS is characterized by an increased frequency of hydrophobic amino acids at positions 6 and 7 of the CDR-B3, which has been previously associated with promotion of self-reactive T cells (24).

The analysis of the distribution of immunoglobulin heavy chain isotypes and somatic hypermutation has revealed unexpected features in patients with hypomorphic *RAG* mutations. In particular, we observed that the majority of immunoglobulin heavy chain transcripts in patients with OS were represented by switched transcripts, and IgE in particular. Direct μ to ε CSR has been previously reported in immature B cells from a mouse model with hypomorphic *Rag1* mutations (39). We have also demonstrated the presence of SHM in Ig transcripts from patients carrying hypomorphic *RAG* mutations, including patients with OS, who have very low to undetectable circulating B cells, disorganized secondary lymphoid organs, with lack of follicles and germinal centers, where SHM is actively induced (40). Our observation of CSR and SHM in peripheral blood B cells from patients with OS is consistent with previous evidence of Blimp1^{+} CD138^{+} plasma cells in lymph nodes from patients with OS, and with homeostatic expansion of immunoglobulin-secreting cells and increased expression of Activation-Induced Cytidine Deaminase (*AICDA*), in mouse models of this disease (41, 42).

Although this study offers insights into the mechanisms underlying the immunopathology and phenotypic heterogeneity of human RAG deficiency, it has important limitations, including the small sample size of each phenotypic subgroup of patients analyzed, and the inability to study both T and B cell repertoires in each patient. Despite these limitations we have provided a detailed analysis of T and B cell repertoire in patients with hypomorphic *RAG* mutations that are illustrative of the entire phenotypic spectrum of the disease. After introduction of newborn screening for SCID and related conditions, *RAG* mutations have emerged as the most common genetic defect associated with OS and atypical SCID (43). Detailed analysis of the immune repertoire in *RAG*-mutated patients may have important predictive implications, and may influence therapeutic interventions.

Material and Methods

Study Design

One to 5 ml of EDTA-blood samples were obtained upon written informed consent from patients diagnosed with RAG deficiency and a known clinical and immunological phenotype. The study was performed under the approval of the Institutional Review Board of Boston Children's Hospital, Harvard Medical School. For healthy controls, de-identified left-over blood samples were used, that had been obtained from children at the age of 9 months to 4 years at the time of regular well-child visits.

Generation and Analysis of T and B cell receptor repertoire by NGS

Equal amounts of total RNA extracted from peripheral blood of patients with RAG deficiency (n=5 for patients with OS; n=3 for patients with LS; and n=4 for patients with CID-G/AI) and from peripheral blood of healthy infants (n=3 for *TRB* and n=4 for *IGH*; age range: 9 months-4 years) were used as template to semi-quantitatively amplify the rearrangements at the endogenous T cell receptor beta (*TRB*) and immunoglobulin heavy chain (*IGH)* loci according to the manufacturer's protocol (iRepertoire, Inc) (44). PCR products were purified and sequenced using the GS Junior 454 platform (Roche, Inc).

Raw sequences were filtered for PCR errors, and resulting FASTA sequences were submitted to IMGT HighV-QUEST, and analyzed for V, D, and J gene usage, composition and length of the CDR-3, Kyte-Doolittle index of hydrophobicity, SHM and antigen-mediated selection using the IgAT software (45), as previously described (23). The sequences of *TRB* and *IGH* transcripts, upon processing through IMGT, are posted in the Supplementary Table 3. The diversity indices of Shannon's H entropy and Gini Simpson's D indexes were calculated using the VDJ statistics file from IgAT analysis and the PAST program, as described (23). D50 was calculated by determining the cumulative frequency of total sequences that constitute 50% of the cumulative unique sequences frequency (14). D50 is graphically represented by plotting the cumulative frequency of total sequences on the x-axis against cumulative frequency of unique sequences on the y-axis and then finding the intercept on the y axis of values that correspond to 50% on the×axis. Graphical representation of V-J pairing and the relative distribution of distinct rearrangements, hierarchical tree maps and isotype usages were generated using the iRepertoire software. All raw data used for the analyses represented in the various Figures are posted in Supplementary Table 3.

The complexity Score of the CDR3 length distribution was determined based on the followingcalculation (46):

$$\text{Complexity Score} = \frac{\sum_{1}^{\text{mp}} \text{nMP}}{N} (\text{NMPs})$$

where:

N = number of all the sequences studied (either unique or total

MP = major peak, defined by constituting at least 10% of N

mp = number of all the MP

nMP = number of sequences in each MP

NMPs = number of MP

RAG activity

The activity of various RAG mutant proteins was determined by a flow cytometric-based assay as previously described (12, 13), and expressed as percentage of the recombination activity of the wild type protein.

Statistical analysis

Unpaired *t*-test was used to compare the patient blood samples to infant controls for variables with normal distribution. For non-parametric variables, the Mann-Whitney test was used. The χ^2 test was used for categorical values. For all multiple t-tests, post hoc Bonferroni correction was applied. ANOVA with Dunnett's correction for multiple comparisons was used when comparing more than two groups. The analyses were performed using PRISM version 6 (Graph Pad). Non-hypothesis driven statistical analysis of Principal Component Analysis (PCA) was performed using the Excel add-in Multibase package (Numerical Dynamics, Japan).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We greatly thank the patients and patients' families for participating in this study.

Funding

This study was supported by grants from the National Institutes of Health (2R01AI100887, 2R01AI100887, and 2U54AI082973 to L.D.N.), and the March of Dimes (1-FY13-500 to L.D.N.).

References and notes

1. Villa A, Santagata S, Bozzi F, Giliani S, Frattini A, Imberti L, Gatta LB, Ochs HD, Schwarz K, Notarangelo LD, Vezzoni P, Spanopoulou E. Partial V(D)J recombination activity leads to Omenn syndrome. Cell. 1998; 93:885–896. [PubMed: 9630231]
2. Shearer WT, Dunn E, Notarangelo LD, Dvorak CC, Puck JM, Logan BR, Griffith LM, Kohn DB, O'Reilly RJ, Fleisher TA, Pai SY, Martinez CA, Buckley RH, Cowan MJ. Establishing diagnostic criteria for severe combined immunodeficiency disease (SCID), leaky SCID, and Omenn syndrome: the Primary Immune Deficiency Treatment Consortium experience. J Allergy Clin Immunol. 2014; 133:1092–1098. [PubMed: 24290292]
3. de Villartay JP, Lim A, Al-Mousa H, Dupont S, Dechanet-Merville J, Coumau-Gatbois E, Gougeon ML, Lemainque A, Eidenschenk C, Jouanguy E, Abel L, Casanova JL, Fischer A, Le Deist F. A novel immunodeficiency associated with hypomorphic RAG1 mutations and CMV infection. J Clin Invest. 2005; 115:3291–3299. [PubMed: 16276422]

4. Ehl S, Schwarz K, Enders A, Duffner U, Pannicke U, Kuhr J, Mascart F, Schmitt-Graeff A, Niemeyer C, Fisch P. A variant of SCID with specific immune responses and predominance of gamma delta T cells. J Clin Invest. 2005; 115:3140–3148. [PubMed: 16211094]
5. Schuetz C, Huck K, Gudowius S, Megahed M, Feyen O, Hubner B, Schneider DT, Manfras B, Pannicke U, Willemze R, Knuchel R, Gobel U, Schulz A, Borkhardt A, Friedrich W, Schwarz K, Niehues T. An immunodeficiency disease with RAG mutations and granulomas. N Engl J Med. 2008; 358:2030–2038. [PubMed: 18463379]
6. Henderson LA, Frugoni F, Hopkins G, de Boer H, Pai SY, Lee YN, Walter JE, Hazen MM, Notarangelo LD. Expanding the spectrum of recombination-activating gene 1 deficiency: a family with early-onset autoimmunity. J Allergy Clin Immunol. 2013; 132:969–971. e961–962. [PubMed: 23891352]
7. Kuijpers TW, Ijspeert H, van Leeuwen EM, Jansen MH, Hazenberg MD, Weijer KC, van Lier RA, van der Burg M. Idiopathic CD4+ T lymphopenia without autoimmunity or granulomatous disease in the slipstream of RAG mutations. Blood. 2011; 117:5892–5896. [PubMed: 21502542]
8. Abolhassani H, Wang N, Aghamohammadi A, Rezaei N, Lee YN, Frugoni F, Notarangelo LD, Pan-Hammarstrom Q, Hammarstrom L. A hypomorphic recombinationactivating gene 1 (RAG1) mutation resulting in a phenotype resembling common variable immunodeficiency. J Allergy Clin Immunol. 2014; 134:1375–1380. [PubMed: 24996264]
9. Geier CB, Piller A, Linder A, Sauerwein KM, Eibl MM, Wolf HM. Leaky RAG Deficiency in Adult Patients with Impaired Antibody Production against Bacterial Polysaccharide Antigens. PloS one. 2015; 10:e0133220. [PubMed: 26186701]
10. Patiroglu T, Akar HH, Gilmour K, Ozdemir MA, Bibi S, Henriquez F, Burns SO, Unal E. Atypical severe combined immunodeficiency caused by a novel homozygous mutation in Rag1 gene in a girl who presented with pyoderma gangrenosum: a case report and literature review. J Clin Immunol. 2014; 34:792–795. [PubMed: 25104208]
11. Notarangelo LD, Kim MS, Walter JE, Lee YN. Human RAG mutations: biochemistry and clinical implications. Nat Rev Immunol. 2016; 16:234–246. [PubMed: 26996199]
12. Lee YN, Frugoni F, Dobbs K, Walter JE, Giliani S, Gennery AR, Al-Herz W, Haddad E, LeDeist F, Bleesing JH, Henderson LA, Pai SY, Nelson RP, El-Ghoneimy DH, El-Feky RA, Reda SM, Hossny E, Soler-Palacin P, Fuleihan RL, Patel NC, Massaad MJ, Geha RS, Puck JM, Palma P, Cancrini C, Chen K, Vihinen M, Alt FW, Notarangelo LD. A systematic analysis of recombination activity and genotype-phenotype correlation in human recombination-activating gene 1 deficiency. J Allergy Clin Immunol. 2014; 133:1099–1108. [PubMed: 24290284]
13. Dutmer CM, Asturias EJ, Smith C, Dishop MK, Schmid DS, Bellini WJ, Tirosh I, Lee YN, Notarangelo LD, Gelfand EW. Late Onset Hypomorphic RAG2 Deficiency Presentation with Fatal Vaccine-Strain VZV Infection. J Clin Immunol. 2015; 35:754–760. [PubMed: 26515615]
14. Han, J. Google Patents. 2012.
15. Brauer PM, Pessach IM, Clarke E, Rowe JH, Ott de Bruin L, Lee YN, Dominguez-Brauer C, Comeau AM, Awong G, Felgentreff K, Zhang YH, Bredemeyer A, Al-Herz W, Du L, Ververs F, Kennedy M, Giliani S, Keller G, Sleckman BP, Schatz DG, Bushman FD, Notarangelo LD, Zuniga-Pflucker JC. Modeling altered T-cell development with human induced pluripotent stem cells from patients with RAG1 mutations and distinct immunological phenotypes. Blood. 2016
16. Schroeder HW Jr, Zhang L, Philips JB 3rd. Slow, programmed maturation of the immunoglobulin HCDR3 repertoire during the third trimester of fetal life. Blood. 2001; 98:2745–2751. [PubMed: 11675347]
17. Zemlin M, Schelonka RL, Bauer K, Schroeder HW Jr. Regulation and chance in the ontogeny of B and T cell antigen receptor repertoires. Immunol Res. 2002; 26:265–278. [PubMed: 12403364]
18. Miqueu P, Guillet M, Degauque N, Dore JC, Soulillou JP, Brouard S. Statistical analysis of CDR3 length distributions for the assessment of T and B cell repertoire biases. Mol Immunol. 2007; 44:1057–1064. [PubMed: 16930714]
19. Wu J, Liu D, Tu W, Song W, Zhao X. T-cell receptor diversity is selectively skewed in T-cell populations of patients with Wiskott-Aldrich syndrome. J Allergy Clin Immunol. 2015; 135:209–216. [PubMed: 25091438]

20. Bomberger C, Singh-Jairam M, Rodey G, Guerriero A, Yeager AM, Fleming WH, Holland HK, Waller EK. Lymphoid reconstitution after autologous PBSC transplantation with FACS-sorted CD34+ hematopoietic progenitors. Blood. 1998; 91:2588–2600. [PubMed: 9516161]
21. Yu X, Almeida JR, Darko S, van der Burg M, DeRavin SS, Malech H, Gennery A, Chinn I, Markert ML, Douek DC, Milner JD. Human syndromes of immunodeficiency and dysregulation are characterized by distinct defects in T-cell receptor repertoire development. J Allergy Clin Immunol. 2014; 133:1109–1115. [PubMed: 24406074]
22. Schroeder HW Jr, Ippolito GC, Shiokawa S. Regulation of the antibody repertoire through control of HCDR3 diversity. Vaccine. 1998; 16:1383–1390. [PubMed: 9711776]
23. Rechavi E, Lev A, Lee YN, Simon AJ, Yinon Y, Lipitz S, Amariglio N, Weisz B, Notarangelo LD, Somech R. Timely and spatially regulated maturation of B and T cell repertoire during human fetal development. Science translational medicine. 2015; 7 276ra225.
24. Stadinski BD, Shekhar K, Gomez-Tourino I, Jung J, Sasaki K, Sewell AK, Peakman M, Chakraborty AK, Huseby ES. Hydrophobic CDR3 residues promote the development of self-reactive T cells. Nat Immunol. 2016; 17:946–955. [PubMed: 27348411]
25. Villa A, Sobacchi C, Notarangelo LD, Bozzi F, Abinun M, Abrahamsen TG, Arkwright PD, Baniyash M, Brooks EG, Conley ME, Cortes P, Duse M, Fasth A, Filipovich AM, Infante AJ, Jones A, Mazzolari E, Muller SM, Pasic S, Rechavi G, Sacco MG, Santagata S, Schroeder ML, Seger R, Strina D, Ugazio A, Valiaho J, Vihinen M, Vogler LB, Ochs H, Vezzoni P, Friedrich W, Schwarz K. V(D)J recombination defects in lymphocytes due to RAG mutations: severe immunodeficiency with a spectrum of clinical presentations. Blood. 2001; 97:81–88. [PubMed: 11133745]
26. Rubelt F, Sievert V, Knaust F, Diener C, Lim TS, Skriner K, Klipp E, Reinhardt R, Lehrach H, Konthur Z. Onset of immune senescence defined by unbiased pyrosequencing of human immunoglobulin mRNA repertoires. PloS one. 2012; 7:e49774. [PubMed: 23226220]
27. Lossos IS, Tibshirani R, Narasimhan B, Levy R. The inference of antigen selection on Ig genes. J Immunol. 2000; 165:5122–5126. [PubMed: 11046043]
28. Ru H, Chambers MG, Fu TM, Tong AB, Liao M, Wu H. Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. Cell. 2015; 163:1138–1152. [PubMed: 26548953]
29. Kim MS, Lapkouski M, Yang W, Gellert M. Crystal structure of the V(D)J recombinase RAG1-RAG2. Nature. 2015; 518:507–511. [PubMed: 25707801]
30. De Ravin SS, Cowen EW, Zarember KA, Whiting-Theobald NL, Kuhns DB, Sandler NG, Douek DC, Pittaluga S, Poliani PL, Lee YN, Notarangelo LD, Wang L, Alt FW, Kang EM, Milner JD, Niemela JE, Fontana-Penn M, Sinal SH, Malech HL. Hypomorphic Rag mutations can cause destructive midline granulomatous disease. Blood. 2010; 116:1263–1271. [PubMed: 20489056]
31. Schuetz C, Pannicke U, Jacobsen EM, Burggraf S, Albert MH, Honig M, Niehues T, Feyen O, Ehl S, Debatin KM, Friedrich W, Schulz AS, Schwarz K. Lesson from hypomorphic recombination-activating gene (RAG) mutations: Why asymptomatic siblings should also be tested. J Allergy Clin Immunol. 2014; 133:1211–1215. [PubMed: 24331380]
32. H IJ, Driessen GJ, Moorhouse MJ, Hartwig NG, Wolska-Kusnierz B, Kalwak K, Pituch-Noworolska A, Kondratenko I, van Montfrans JM, Mejstrikova E, Lankester AC, Langerak AW, van Gent DC, Stubbs AP, van Dongen JJ, van der Burg M. Similar recombination-activating gene (RAG) mutations result in similar immunobiological effects but in different clinical phenotypes. J Allergy Clin Immunol. 2014; 133:1124–1133. [PubMed: 24418478]
33. Sadofsky MJ, Hesse JE, van Gent DC, Gellert M. RAG-1 mutations that affect the target specificity of V(D)j recombination: a possible direct role of RAG-1 in site recognition. Genes Dev. 1995; 9:2193–2199. [PubMed: 7657170]
34. Mo X, Bailin T, Sadofsky MJ. A C-terminal region of RAG1 contacts the coding DNA during V(D)J recombination. Molecular and cellular biology. 2001; 21:2038–2047. [PubMed: 11238939]
35. Wong SY, Lu CP, Roth DB. A RAG1 mutation found in Omenn syndrome causes coding flank hypersensitivity: a novel mechanism for antigen receptor repertoire restriction. J Immunol. 2008; 181:4124–4130. [PubMed: 18768869]

36. Wang J, Bu D, Zhu X. Immunoglobulin variable region gene analysis to the autoantibody-secreting B cells from tumors in association with paraneoplastic autoimmune multiorgan syndrome. International journal of dermatology. 2007; 46:1146–1154. [PubMed: 17988333]

37. Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, Nussenzweig MC. Predominant autoantibody production by early human B cell precursors. Science. 2003; 301:1374–1377. [PubMed: 12920303]

38. Mroczek ES, Ippolito GC, Rogosch T, Hoi KH, Hwangpo TA, Brand MG, Zhuang Y, Liu CR, Schneider DA, Zemlin M, Brown EE, Georgiou G, Schroeder HW Jr. Differences in the composition of the human antibody repertoire by B cell subsets in the blood. Frontiers in immunology. 2014; 5:96. [PubMed: 24678310]

39. Wesemann DR, Magee JM, Boboila C, Calado DP, Gallagher MP, Portuguese AJ, Manis JP, Zhou X, Recher M, Rajewsky K, Notarangelo LD, Alt FW. Immature B cells preferentially switch to IgE with increased direct Smu to Sepsilon recombination. J Exp Med. 2011; 208:2733–2746. [PubMed: 22143888]

40. Facchetti F, Blanzuoli L, Ungari M, Alebardi O, Vermi W. Lymph node pathology in primary combined immunodeficiency diseases. Springer seminars in immunopathology. 1998; 19:459–478. [PubMed: 9618768]

41. Cassani B, Poliani PL, Marrella V, Schena F, Sauer AV, Ravanini M, Strina D, Busse CE, Regenass S, Wardemann H, Martini A, Facchetti F, van der Burg M, Rolink AG, Vezzoni P, Grassi F, Traggiai E, Villa A. Homeostatic expansion of autoreactive immunoglobulin-secreting cells in the Rag2 mouse model of Omenn syndrome. J Exp Med. 2010; 207:1525–1540. [PubMed: 20547828]

42. Walter JE, Rucci F, Patrizi L, Recher M, Regenass S, Paganini T, Keszei M, Pessach I, Lang PA, Poliani PL, Giliani S, Al-Herz W, Cowan MJ, Puck JM, Bleesing J, Niehues T, Schuetz C, Malech H, DeRavin SS, Facchetti F, Gennery AR, Andersson E, Kamani NR, Sekiguchi J, Alenezi HM, Chinen J, Dbaibo G, ElGhazali G, Fontana A, Pasic S, Detre C, Terhorst C, Alt FW, Notarangelo LD. Expansion of immunoglobulin-secreting cells and defects in B cell tolerance in Ragdependent immunodeficiency. J Exp Med. 2010; 207:1541–1554. [PubMed: 20547827]

43. Dvorak CC, Cowan MJ, Logan BR, Notarangelo LD, Griffith LM, Puck JM, Kohn DB, Shearer WT, O'Reilly RJ, Fleisher TA, Pai SY, Hanson IC, Pulsipher MA, Fuleihan R, Filipovich A, Goldman F, Kapoor N, Small T, Smith A, Chan KW, Cuvelier G, Heimall J, Knutsen A, Loechelt B, Moore T, Buckley RH. The natural history of children with severe combined immunodeficiency: baseline features of the first fifty patients of the primary immune deficiency treatment consortium prospective study 6901. J Clin Immunol. 2013; 33:1156–1164. [PubMed: 23818196]

44. Wang C, Sanders CM, Yang Q, Schroeder HW Jr, Wang E, Babrzadeh F, Gharizadeh B, Myers RM, Hudson JR Jr, Davis RW, Han J. High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. Proc Natl Acad Sci U S A. 2010; 107:1518–1523. [PubMed: 20080641]

45. Rogosch T, Kerzel S, Hoi KH, Zhang Z, Maier RF, Ippolito GC, Zemlin M. Immunoglobulin analysis tool: a novel tool for the analysis of human and mouse heavy and light chain transcripts. Frontiers in immunology. 2012; 3:176. [PubMed: 22754554]

46. Krell PF, Reuther S, Fischer U, Keller T, Weber S, Gombert M, Schuster FR, Asang C, Stepensky P, Strahm B, Meisel R, Stoye J, Borkhardt A. Next-generationsequencing- spectratyping reveals public T-cell receptor repertoires in pediatric very severe aplastic anemia and identifies a beta chain CDR3 sequence associated with hepatitis-induced pathogenesis. Haematologica. 2013; 98:1388–1396. [PubMed: 23716544]

47. Delano, WL. The PyMol Molecular Graphics System, Version 1.8. Schrodinger, LLC; 2002.

48. Buchbinder D, Baker R, Lee YN, Ravell J, Zhang Y, McElwee J, Nugent D, Coonrod EM, Durtschi JD, Augustine NH, Voelkerding KV, Csomos K, Rosen L, Browne S, Walter JE, Notarangelo LD, Hill HR, Kumanovics A. Identification of Patients with RAG Mutations Previously Diagnosed with Common Variable Immunodeficiency Disorders. J Clin Immunol. 2015; 35:119–124. [PubMed: 25516070]

49. Walter JE, Rosen LB, Csomos K, Rosenberg JM, Mathew D, Keszei M, Ujhazi B, Chen K, Lee YN, Tirosh I, Dobbs K, Al-Herz W, Cowan MJ, Puck J, Bleesing JJ, Grimley MS, Malech H, De Ravin SS, Gennery AR, Abraham RS, Joshi AY, Boyce TG, Butte MJ, Nadeau KC, Balboni I,

Sullivan KE, Akhter J, Adeli M, El-Feky RA, El-Ghoneimy DH, Dbaibo G, Wakim R, Azzari C, Palma P, Cancrini C, Capuder K, Condino-Neto A, Costa-Carvalho BT, Oliveira JB, Roifman C, Buchbinder D, Kumanovics A, Franco JL, Niehues T, Schuetz C, Kuijpers T, Yee C, Chou J, Masaad MJ, Geha R, Uzel G, Gelman R, Holland SM, Recher M, Utz PJ, Browne SK, Notarangelo LD. Broad-spectrum antibodies against self-antigens and cytokines in RAG deficiency. J Clin Invest. 2015

50. Lev A, Simon AJ, Ben-Ari J, Takagi D, Stauber T, Trakhtenbrot L, Rosenthal E, Rechavi G, Amarigilio N, Somech R. Co-existence of clonal expanded autologous and transplacental-acquired maternal T cells in recombination activating gene-deficient severe combined immunodeficiency. Clin Exp Immunol. 2014; 3:380–386.

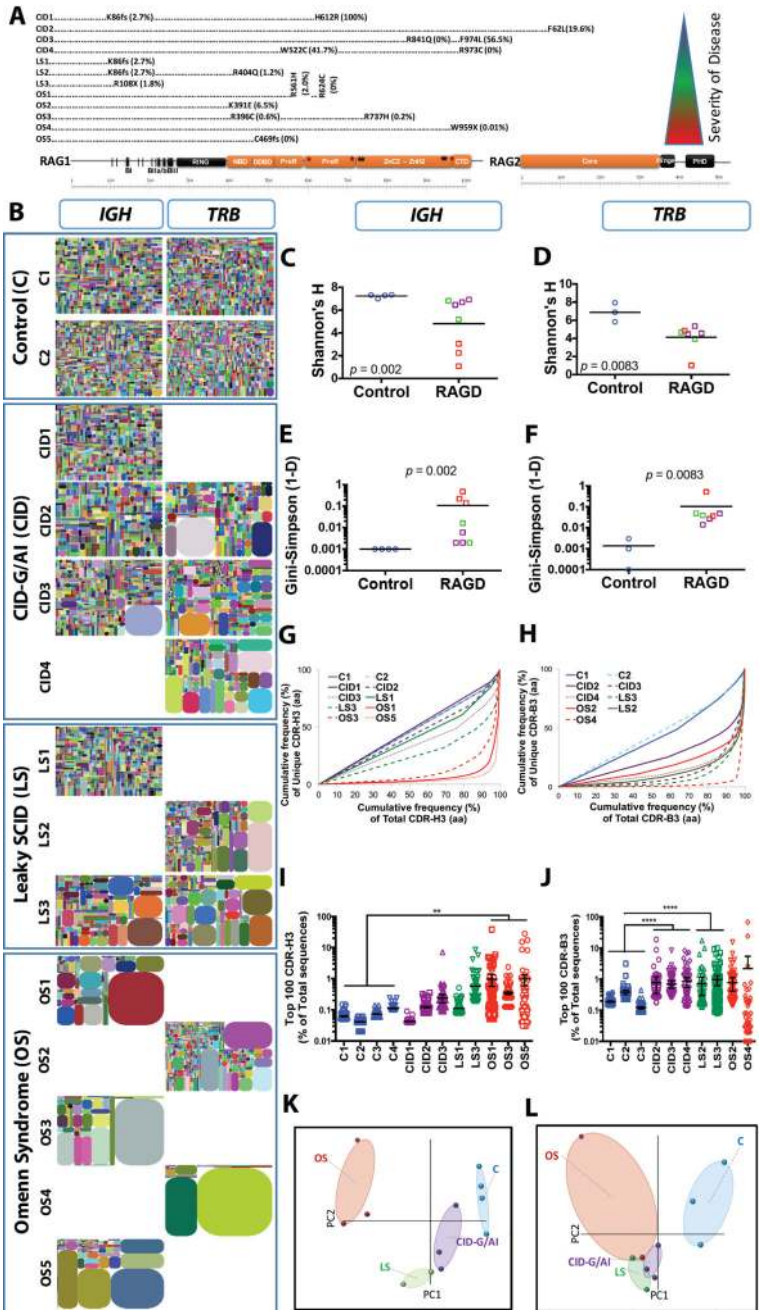

Fig. 1. Progressive *IGH* and *TRB* repertoire restriction with increased clonality in the patients with RAG deficiency

Schematic representation of RAG1 and RAG2 protein with the mutations of the 12 patients according to the severity of clinical presentation from top to bottom (**A**). Tree maps representing the diversity and clonality of *IGH* and *TRB* (**B**) repertoires from healthy donor controls (representative data from two subjects are presented) and patients with *RAG* mutations. Each dot represents a unique V to J joining, and the size of the dot represents the relative frequency of that rearrangement in the entire population. No amplification products were obtained for *IGH* repertoire from patients CID4, LS2, OS2, and OS4, and for *TRB* repertoire for patients CD1, LS1, OS1, OS3, and OS5. Quantification of the diversity (**C, D**)

and unevenness (**E, F**) of the *IGH* (**C, E**) and *TRB* (**D, F**) repertoires using Shannon's H index of diversity and Gini- Sipmson's index of unevenness in healthy controls (blue circles), and patients with CID-G/AI (purple boxes), LS (green boxes), and OS (red boxes). The cumulative frequencies of unique versus total CDR3 clonotypes are shown for *IGH* (**G**) and *TRB* (**H**) repertoires (CDR-H3 and CDR-B3, respectively). Mean values ± SE are shown; t-test was used for statistical analysis. Representation of the frequency of the top 100 most abundant clones for *IGH* (**I**) and *TRB* (**J**) sequences in *RAG*-mutated patients and healthy controls (mean ± SE; ANOVA with post hoc test of Dunnett's multiple comparisons with *** $0.001 < p < 0.01$ and * $p < 0.05$). Sample plots illustrating the segregation of the various patient groups from healthy controls based on primary component (PC) 1 and 2 determined by five variables (RAG recombination activity, Shannon's H, Gini-Simpson, number of total and unique sequences) for the *IGH* (**K**) and *TRB* (**L**) repertoires.

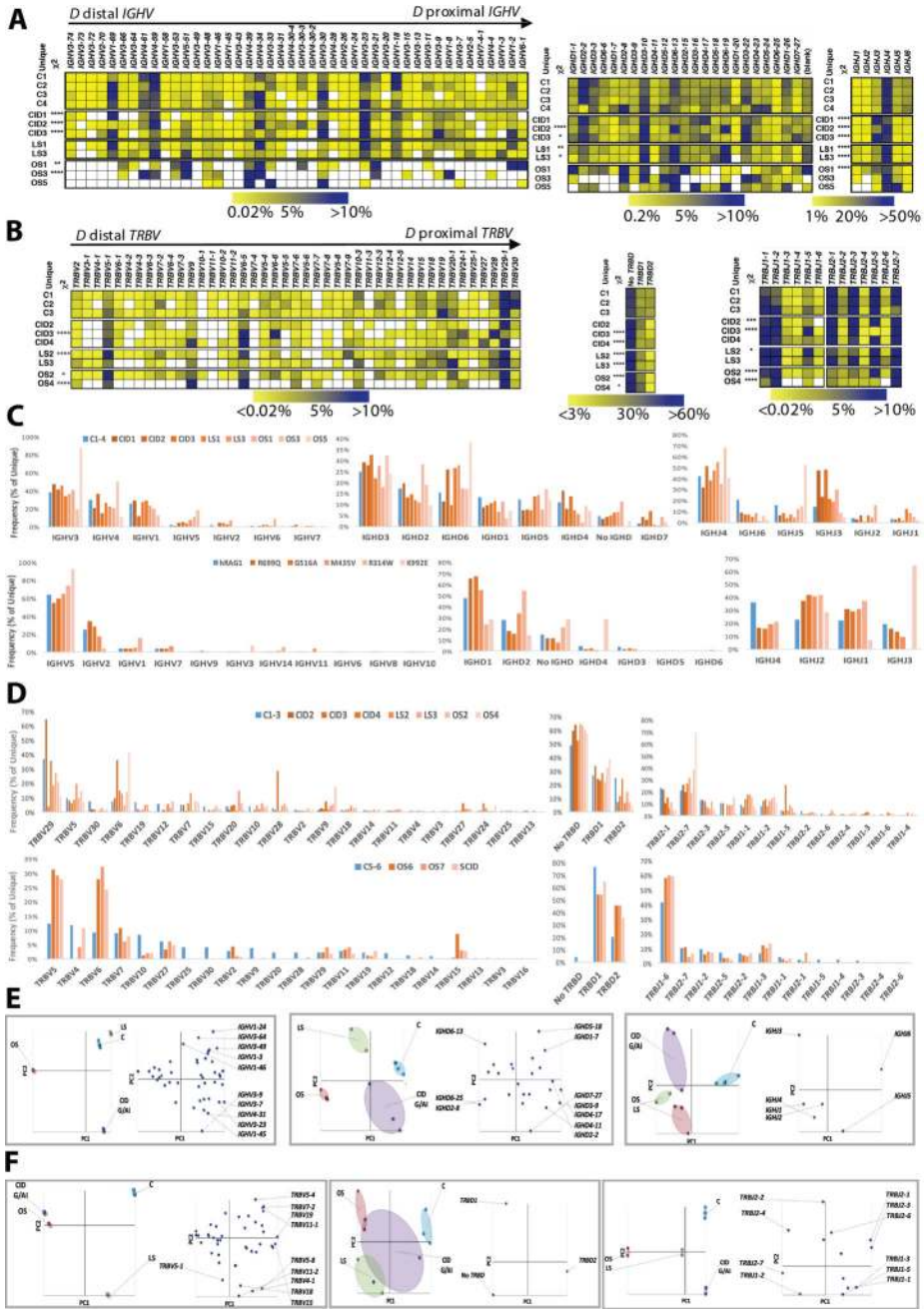

Fig. 2. Differential usage of V, D and J genes in the *IGH* and *TRB* repertoires of patients with RAG deficiency

Heat map representing the frequency of V, D and J gene usage among unique *IGH* (**A**) and *TRB* (**B**) sequences from healthy controls and *RAG*-mutated patients. Relative frequency of usage of *IGHV* and *IGHD* gene families, and of individual *IGHJ* genes, in healthy controls and in patients (**C**, upper panel) and in Abelson virus-transformed pro-B cell lines expressing various *RAG1* mutations (**C**, lower panel). Relative frequency of usage of *TRBV* gene family and of *TRBD* and *TRBJ* genes in healthy controls and in patients (**D**, upper panel) and in iPS-derived thymocytes (**D**, lower panel). In panel **E**, differential usage of

IGHV, *IGHD* and *IGHJ* genes, segregating control and patient samples and the various genes according to PC1 and PC2 is shown as sample plots (left panels) and variable plots (right panels). In panel **F**, differential usage of *TRBV*, *TRBD* and *TRB* genes, segregating control and patient samples and the various genes according to PC1 and PC2 is shown as sample plots (left panels) and variable plots (right panels).

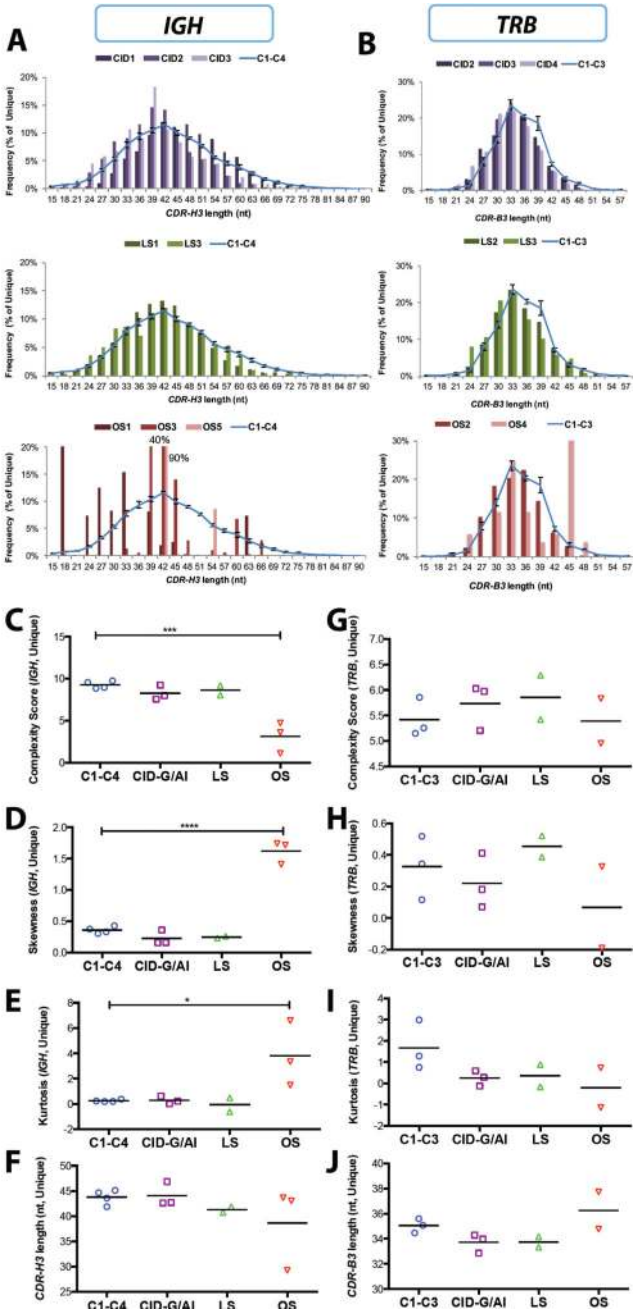

Fig. 3. Characteristics of the CDR3 region of *IGH* and *TRB* unique sequences in peripheral blood lymphocytes

Distribution of the length of the CDR3 region of *IGH* (CDR-H3) (**A**) and *TRB* (CDR-B3) (**B**) unique sequences from peripheral blood of patients with RAG deficiency and healthy controls (C1-C4; C6-C8). In panels **A** and **B**, the distribution of the CDR3 length in healthy controls is depicted as a blue line (representing mean values ± 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.

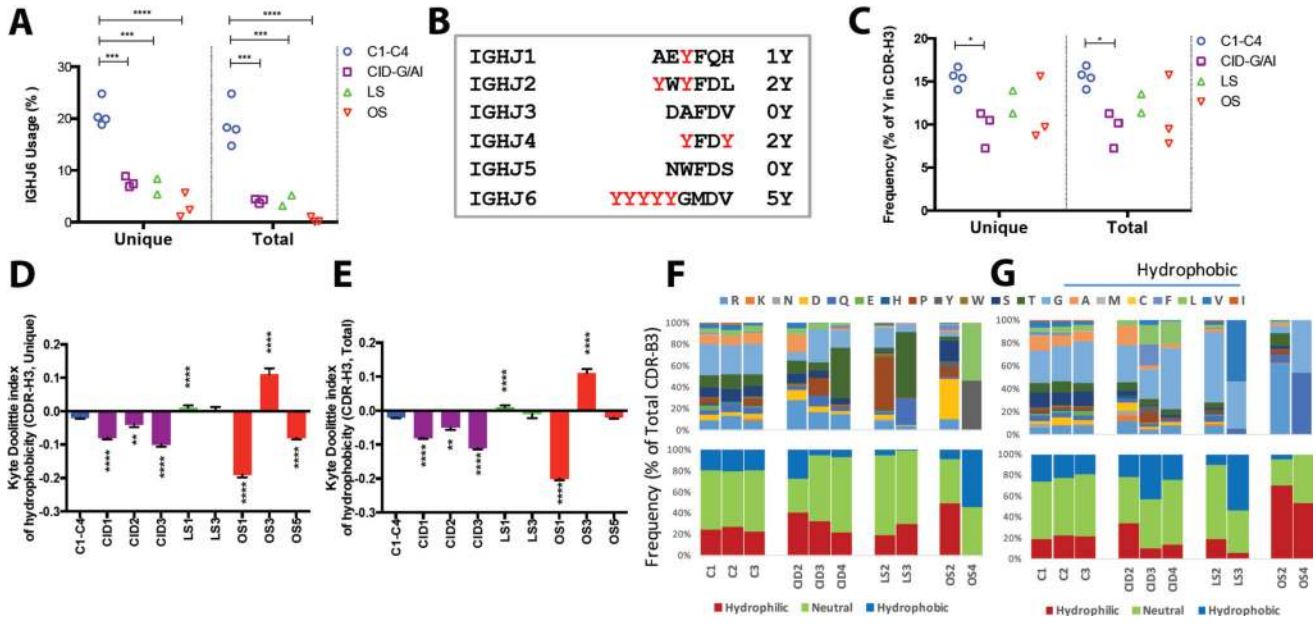

Fig. 4. Abnormal amino acid composition of CDR3 in the *IGH* and *TRB* sequences

Frequency of usage of the *IGHJ6* gene among unique CDR-H3 sequences (**A**). Summary of the Y content in the *IGHJ* genes (**B**). Percentage of tyrosine residues in the CDR-H3 of unique and total sequences (**C**). Summary of CDR-H3 hydrophobicity profile depicted as average Kyte-Doolittle index of hydrophobicity (mean ± SE) in patients and healthy control blood samples for unique (**D**) and total (**E**) sequences. (**, p<0.01; ***, p <0.001; ****, p < 0.0001; one-tail unpaired t-test). Amino acid composition of CDR-B3 in patients and healthy control for amino acid positions 6 (**F**) and 7 (**G**) of the 13 aa-long CDR-B3.

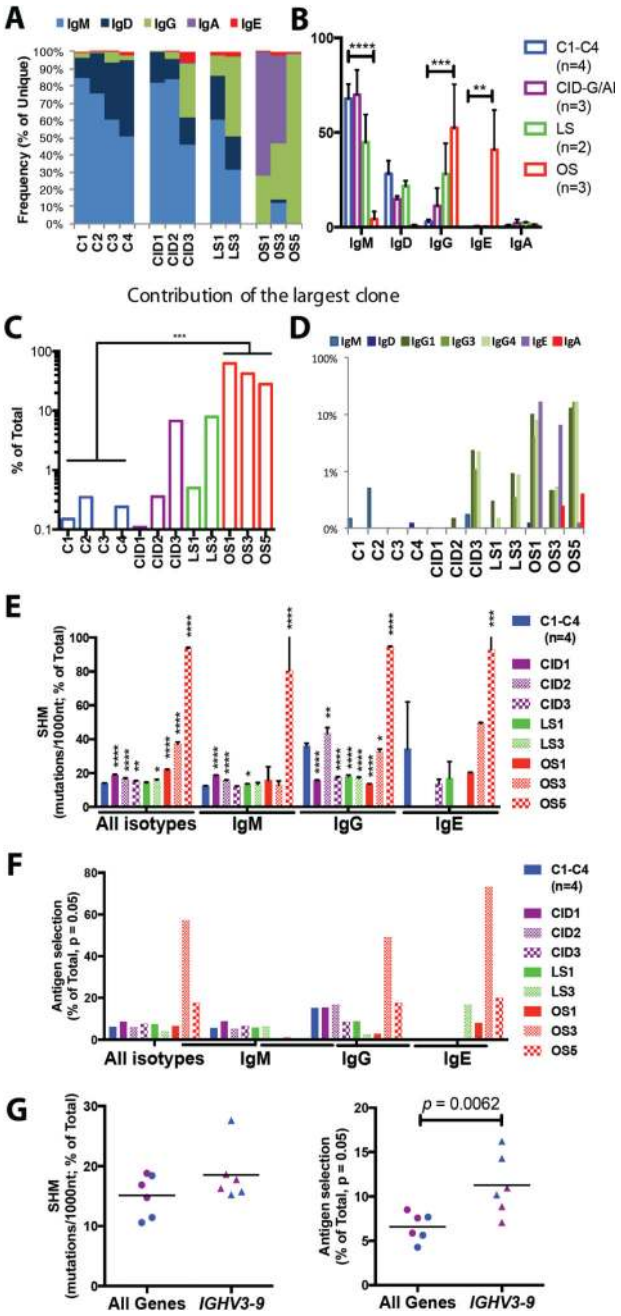

Fig. 5. Distribution of immunoglobulin heavy chain isotypes, somatic hypermutation and antigen-driven selection in peripheral blood B cells of patients with RAG deficiency

Frequency of immunoglobulin heavy chain constant gene usage among unique *IGH* sequences from peripheral blood lymphocytes of RAG-deficient patients and healthy controls (**A** and **B**). In panel **B**, mean values ± SE are shown (one-tail, unpaired t-test). Contribution of the most abundant clonotype to the total number of *IGH* sequences in patients and controls (**C**). Distribution of various isotypes among the most abundant *IGH* transcript (**D**). Rate of somatic hypermutation (SHM) in *IGH* transcripts (**E**, mean ± SE; unpaired t-test). Frequency of unique *IGH* transcripts displaying evidence of antigen-

mediated selection based on the distribution of replacement and silent mutations (**F**). Rate of SHM and antigen mediated selection in *IGHV3-9* for patients with CID-G/AI and healthy controls with line at the mean (**G**).

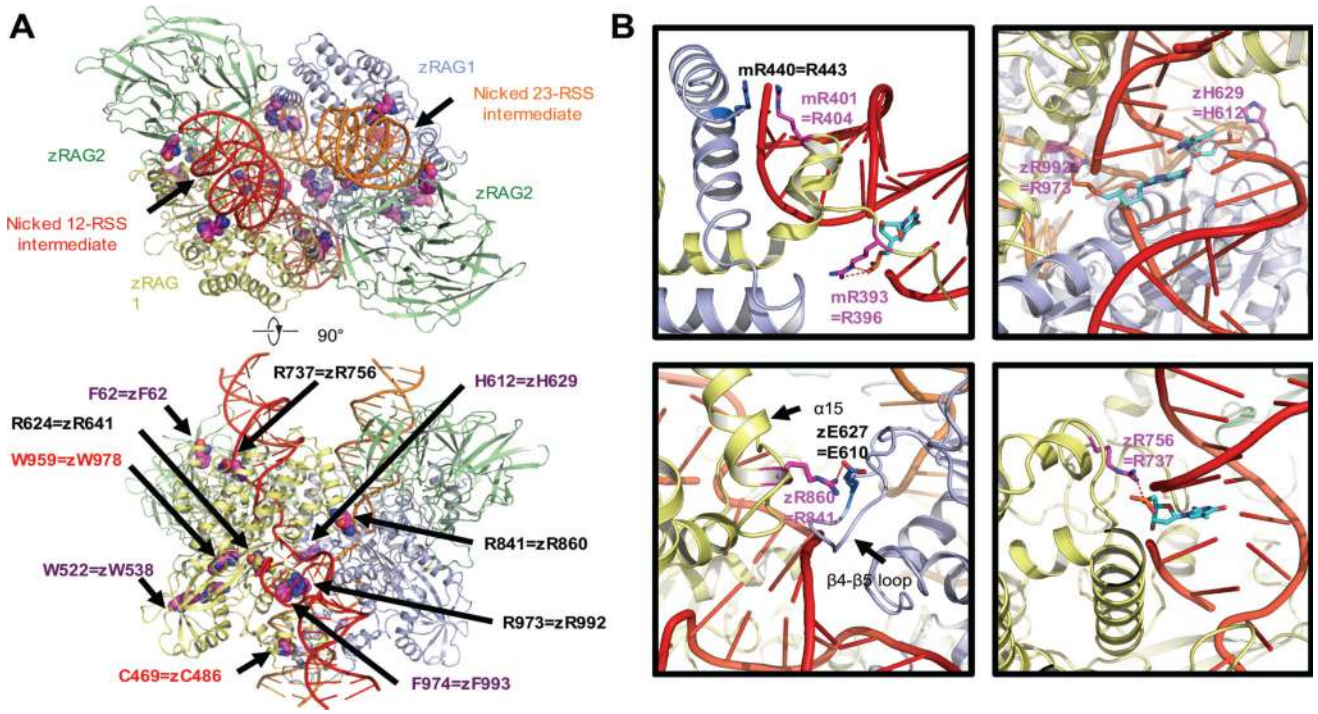

Fig. 6. Mapping the disease-related mutations onto the synaptic-RAG complex models

Overview of the disease-related mutations shown as space filling models mapped onto the ribbon diagram of the synaptic RAG complex structure (PDB ID 3JBY, top and bottom view) (**A**). Residues in zebrafish rag1 and rag2 and the equivalent residues that have been mutated in patients are labeled. Only one RAG1-RAG2 subunit is labeled for explicitness on the side view. Labeled in purple and red are residues that are mutated in patients with CID-G/AI and OS, respectively. Residues affected by mutations that correspond to the allele with lower recombination activity in compound heterozygous patients are labeled in black. Examples of the detailed interactions between the equivalent residues from patients and the RSS intermediates or partner residues (PBD ID 3GNA and 3JBY) (**B**). Equivalent residues that have interaction with RSS intermediates are shown as sticks and highlighted in magenta. The nucleotides in the RSS intermediates that have interaction with protein residues are shown as sticks and highlighted in cyan. The partner residues are shown as sticks and highlighted in marine. Potential interactions are displayed as red dashed lines. z is abbreviated for zebrafish and m is for mouse. All molecular representations were generated in PyMOL (http://www.pymol.org) (47).