Dominantly inherited combined immunodeficiency as a consequence of heterozygous germline mutations in V(D)J recombination enzymes

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie

vorgelegt der Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von
Annaïse Julie Jauch
Aus Silenen UR/Basel-Stadt BS
2021

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von
Prof. Dr. Christoph Hess
Referent und Fakultätsverantwortlicher

Prof. Dr. Annette Oxenius
Korreferentin

Prof. Dr. Mike Recher
Dissertationsleiter

Basel, den 2. März 2021

Prof. Dr. Marcel Mayor Dekan

Summary

Human immunodeficiencies associated with biallelic mutations in genes required for the generation of the B and T lymphocyte antigen receptors are prototypic inborn errors of immunity. We detected novel heterozygous missense mutations in critical genes implicated in the B and T cell receptor generation, termed *RAG1/RAG2* and *LIG4* and demonstrated that those can be associated with clinically severe immunodeficiency and immune-dysregulation. By analyzing murine models with titrable *in vivo* RAG activity we experimentally linked *Rag* gene dosage to lymphocyte development, anti-viral immunity and immune dysregulation.

Our study widens the spectrum of immunological and clinical phenotypes associated with RAG and LIG4 deficiency, which might have important implications for future molecular diagnosis and enabling effective targeted therapeutic interventions.

Acknowledgments

Table of Contents

General Introduction

The adaptive immune system can be regarded as the body's sixth sense[1] – responding to any conceivable antigenic structure, aiming for pathogen elimination while preserving the host tissue integrity. The main cell subsets of adaptive immunity – B and T lymphocytes – recognize foreign and self-antigens by specific surface receptors, B cell receptors (BCR) and T cell receptors (TCRs)[2]. Germline DNA cannot encode for the entire diversity of BCRs and TCRs, as it would exceed the genomic capacity of ~25'000 genes[3]. However, lymphocytes ensure the receptor diversity by somatic DNA recombination, which rearranges BCR and TCR gene segments in a process called V(D)J recombination[4].

Inborn errors of immunity in enzymes involved in V(D)J recombination

Inborn errors of immunity (IEI) or primary immune deficiencies (PIDs) are genetically determined diseases affecting the development, maintenance and/or function of the immune system[5]. Patients suffering from IEI were initially believed to present with infection susceptibility only, similar to Bruton's first description of a agammaglobulinemic patient suffering from nearly 20 episodes of *Streptococcus pneumoniae* infections[6]. It has later been demonstrated that IEI's might clinically present very variably comprising, besides susceptibility to infection[24], lymphoproliferation, autoimmunity, allergy, autoinflammation or malignancy[7,8]. Further, the field of IEI was widened after the recognition that non-immune cells are involved in host protection[9] and that metabolic alterations[10] might underly IEI.

 Thanks to the low costs and rapidity of next-generation DNA sequencing, the field of IEI is fast evolving and 431 different IEI entities have been described by the year 2020[11]. IEI can be subcategorized either by the affected cellular compartment (innate, adaptive, complement, non-immune cells), the infectious disease phenotype (PID, mendelian infections, monogenetic infections), the mode of inheritance (X-linked, autosomal recessive *vs.* dominant) or the genotype (mono-/biallelic, loss of function *vs.* gain-of-function, dominant-negative *vs.* haploinsufficiency)[11,12].

 IEI caused by mutated V(D)J recombination associated genes are typically accompanied by combined immunodeficiency (CID), characterized by compromised T cell numbers and/or functioning. B cells can be intrinsically affected or the absence of CD4+ T helper cells can prevent B cells from generating a normal immunoglobulin response, adding to the "combined immunodeficiency" phenotype[13]. The recombination-activating gene 1 (RAG1), RAG2 and DNA ligase 4 (LIG4) are three enzymes with central roles in orchestrating the V(D)J recombination[13]. Mutations in the same gene can underlie a vast range of clinical and immunological phenotypes, as exemplified by RAG[4] and LIG4[14] deficiencies. While complete loss of function mutations in *RAG* or *LIG4* lead to severe combined immunodeficiency associated with lethal infections early after birth or intrauterine death, hypomorphic mutations in *RAG* or *LIG4* are associated with residual enzyme function and late-onset combined immunodeficiency and autoimmunity. So far, mostly biallelic mutations have been linked to clinically relevant RAG or LIG4 deficiency.

Diversity originates from DNA damage

B and T cell lymphocytes originate both from a common lymphoid progenitor in the bone marrow (BM). Later, the development is compartmentalized for B cells in the BM and for T cells in the thymus[15–17].

 The generation of antigen receptor diversity (recognition of a wide variety of antigens) and the immune repertoire regulation (autoimmunity avoidance) are opposing demands – mirrored in the antigen receptor structural asymmetry. The antigen receptors are heteromers, one chain contributes mainly to diversity, while the other chain allows for the improvement of specificity[18]. The heteromers are composed of a constant domain paired with a variable antigen recognition domain[18]. The variable domain is encoded by dispersed gene segments (named variable *V*, diversity *D*, joining *J*), which rely on somatic V(D)J recombination for the gene

segment rearrangement[19]. The recombination *per se* is initiated in *cis* by two lymphocyte-specific endonucleases, namely recombination-activating gene 1 (RAG1) and RAG2. The human *RAG1/2* locus architecture is noteworthy, mirroring its transposable element origin. The two loci are in a tail towards tail configuration juxtaposed on chromosome 11p13, each gene harbors only one protein coding exon and both loci are separated by only ~8kb[20].

RAG recombination expression comes in two waves during lymphocyte development for the subsequently recombination of the seven mammalian antigen receptor-chain loci[21,22]. *RAG* expression is tightly regulated because aberrant RAG endonuclease activity represents a threat to the lymphocyte genome integrity[23]. There are multiple layers of RAG expression regulation, fine-tuned via post-transcriptional regulation, *cis*-regulatory elements within the *RAG* loci and the cell-cycle dependent degradation of the RAG2 protein[20]. The RAG proteins form a heterotetramer with two RAG1 subunits associated with two RAG2 proteins[24]. The RAG2 c-terminus will bind H3K4 trimethylated histones - an epigenetic marker of active transcription - regulating the RAG heterotetramer chromatin accessibility[25]. The RAG complex will further be guided by recombination signal sequences (RSS), adjacent to the antigen-receptor gene segments[4]. There, RAG1 - which harbors a catalytical domain - introduces a precise DNA strand break between the RSS and the gene segment. After cleavage, the DNA ends form a sealed hairpin, opened by the endonuclease Artemis[26], which is recruited with other non-homologous end joining (NHEJ) DNA repair proteins[13]. After different modifications of the broken DNA ends, the sequence in-between the two gene segments is either looped out or deleted. The final joining of DNA ends is mediated by the error-prone DNA ligase 4 (LIG4)[27]. The asymmetric hairpin opening allows for the palindromic sequence incorporation and/or the introduction of additional nucleotides and/or exonucleolytic cleavage[2].

The first *RAG* expression wave results in the recombination of the antigen receptor chains incorporating all three gene segments, thus *V*, *D* and *J* genes (TCRβ, TCRδ, immunoglobulin heavy chain (IGH)), whereas the second RAG activity wave recombines only *V* and *J* gene elements (TCRα, TCRγ, both Ig-light chains (IGL))[2].

Every antigen-receptor gene rearrangement is unique because the gene element recombination is "quasi- stochastic". The RSS variability, previous rearrangements and additional *cis*-acting elements (such as promoters and enhancers) orchestrate the process[28].

The somatic assembly for the *TRD, TRG* and *TRB* loci (loci for the TCRδ-, TCRγ-, TCRβ-chain) starts in CD4⁻CD8⁻ double-negative (DN2) thymocytes and is completed in the DN3 stage during the initial phase of *RAG* expression. The successful recombination of *TCRD* with *TCRG* supports the generation of γδ TCR T cells. The TCRβ allows for a first thymocyte selection, as thymocytes with a productive rearrangement will continue in their maturation, while thymocytes with non-productive rearrangements will undergo apoptosis[29]. The TCRβ forms together with a scaffold-chain (pre-Tα) and the CD3 molecule the pre-TCR, stimulating survival and substantial proliferation[30]. Both the expression of the pre-TCR or γδTCR lead to the downregulation of *RAG* expression. After several proliferation rounds, the thymocytes carrying a pre-TCR differentiate into intermediate immature CD8^lowCD4⁻CD3⁻ thymocytes, followed by a stage where thymocytes upregulate both CD4⁺ and CD8⁺ (double-positive DP). The second wave of *RAG* expression results in the DP stage in the *TRA* recombination (TCRα chain locus)[30]. The CD4/CD8 lineage decision is governed by the signaling strength of the TCR. Strong interaction with co-receptor-self-peptide:self-MHC will induce a transcription factor landscape driving the DP thymocytes towards the CD4⁺ single-positive (SP) lineage, while a less intense TCR signaling will commit the DP thymocytes to the SP CD8⁺ lineage[30].

As TCRs are randomly rearranged, each T cell clone needs validation for its usefulness and safety. The TCR on DP thymocytes is tested for its affinity to self-peptide:self-MHC complex (major histocompatibility)[31]. TCRs together with the co-receptors (CD4, CD8) which are unable to bind self-peptide:self-MHC die by apoptosis (~90% of the thymocytes), while those which bind too strongly undergo receptor editing/lineage

deviation/apoptosis (negative selection) and only low-affinity TCR will induce further maturation (positive selection)[31]. The success of positive selection relies only on downstream TCR signaling and the unresponsiveness of DP thymocytes to other survival factors[32].

BCR recombination is a similar two-step process. Progenitor B cells (pro-B cells) will join *D* to *J* genes within the *IGH* locus. Later in development, the *V* segment is joined to the *DJ*. Precursor B cells (pre-B cells), which express the variable domain with the constant µ domain have to associate with two surrogate light chains (λ5 and VpreB) and two signal-transduction molecules (Igα, Igβ) to form a pre-BCR[33]. Only successful assembly of the pre-BCR allows for the stimulation of proliferation, paralleling the DN thymocytes stages. After the proliferation phase, the surrogate light-chains are lost, and the RAG-dependent rearrangement within the Ig light chain locus is initiated. Successful rearrangement is marked by surface IgM expression[33]. If the BCR is tested to be self-reactive, RAG is reactivated for the replacement of the light-chain (receptor editing)[34]. Immature B cells then leave the BM for the periphery. After B cell activation through the BCR and CD40 co-stimulation, two RAG-independent maturation steps occur[13]. Class-switch recombination (CSR), mediates the change of the constant BCR region by genomic deletion from IgM to IgG/IgA/IgE, while keeping the same variable region[35] and occurs outside the germinal center and before the next maturation step[36]. The second maturation step introduces non-template mutations in the variable and flanking region (somatic hypermutation SHM). Both CSR and SHM rely on DNA breaks, the nicks are sealed by DNA Ligase 1 (LIG1) and LIG3, depending on the break length[37]. SHM is followed by positive selection, where clones with a higher BCR affinity are selected for survival[13].

BCR/TCR engagement with its cognate antigen is mediated via structural loops called complementarity-determining region (CDR). The V genes encode for the CDR1 and CDR2 loops, while the CDR3 is the most variable loop as it is encoded by all three gene segments and the two junctions[18].

In summary, antigen receptor diversity is achieved by permutations of V(D)J gene segments, the orientation of gene segments in germline configuration, junctional diversity (imprecise joining and addition of non-template N nucleotides or deletion of nucleotides), allelic exclusion (mostly heterozygous gene loci) and finally the pairing of the two chains *per se*. Keeping in mind that each recombination step involves DNA breakage and re-ligation, each round of recombination harbors the risk of genotoxicity.

The final DNA break sealing

Genotoxic assaults are a constant threat to cellular organisms. One of the most toxic DNA lesions are DNA double-strand breaks (DSBs)[27,38], which can be of endogenous origin (reactive oxygen species (ROS), V(D)J recombination) or of exogenous source (irradiation, exposure to ultra-violet light or alkylating agents). Several conserved repair mechanisms have evolved to safeguard genomic stability. As discussed above, the very same repair mechanisms are employed by higher organisms to create genetic diversity during lymphocyte receptor development[39]. Mammalian cells rely on two main DSB repair pathways, NHEJ and homologous recombination (HR). NHEJ functions throughout the cell cycle, is independent of a repair template[40] and re-joins DNA ends during V(D)J recombination[13]. The DNA damage response machinery first needs to sense the DNA damage, stabilize the broken DNA ends (synapsis formation), transduce the signal, and activate the appropriate repair pathway. RAG generates four broken DNA ends in the G1 phase of the cell cycle, two of which end with sealed hairpins and two blunt ends[41]. The DNA breaks will activate the DNA damage sensor Ataxia Telangiectasia Mutated (ATM). The phosphorylation of p53 by the ATM kinase will mediate cell cycle arrest. ATM will subsequently phosphorylates chromatin-associated proteins, such as histone family member H2Ax (forming γH2Ax)[42] and p53 binding protein 1 (53BP1) at the DNA breaks[43]. As the broken ends are now marked, proteins are recruited to protect the DNA ends (mediated by Ku70/Ku80 heterodimer, DNA dependent protein kinase catalytic subunit DNA-PKcs). Often DSBs have two incompatible DNA ends that would fail direct ligation[40].

Therefore, in NHEJ the ends are resected (via the nuclease-complex Artemis-DNA-PKcs), the process being rapid but error-prone. Gaps are filled by the DNA polymerases (Pol λ, Pol μ) and can result in a diverse DNA sequence at the repair site[27]. During RAG-induced DSB the DNA ends are stabilized/tethered by XRCC4, XLF and RAG2 (X-ray repair cross complementing protein 4, XRCC4-like factor)[44,45]. Ultimately the DNA breaks are sealed by DNA ligase 4 (LIG4) together with XRCC4 and their activity is enhanced by XLF and/or PAXX (paralogue of XRCC4 and XLF)[46,47]. Three mammalian DNA ligases are known (LIG1, LIG3, LIG4). LIG1 is described as ATP-dependent high fidelity ligase[48], whereas LIG4 is less stringent on correct DNA base pairing during nick sealing[49]. The NHEJ apparatus's importance is perceived in murine and human diseases caused by loss of function mutations[26,50–58]. Only biallelic mutations in DNA ligase 1 (LIG1) and ligase 4 (LIG4) have been associated with disease so far[14,59,60].

Severe loss of function mutations affecting genes involved in the V(D)J recombination drastically reduces the immune system`s functionality. Whether haploinsufficiency in these V(D)J related proteins may cause immune-dysregulation is ill-defined.

Aims of the conducted projects

We speculate that haploinsufficiency in V(D)J related genes may be associated with immunodeficiency and/or immune dysregulation. We aim at characterizing dominantly inherited heterozygous hypomorphic mutations in three pivotal enzymes associated with V(D)J recombination (RAG1/RAG2/LIG4). The aims of the thesis are threefold.

First, we aimed at titrating RAG activity *in vivo* and linking it with alternations in lymphocyte development and repertoire.

Second, we aimed at characterizing the *in vivo* anti-pathogen immune response and infection-associated immune dysregulation in murine models with titrated levels of RAG activity.

Third, we intended to characterize human patients with previously undescribed haploinsufficiency in genes affecting the V(D)J recombination machinery. We aimed at analyzing the impact on the lymphocyte repertoire as well as immune regulation, in patients with double heterozygous *RAG1/RAG2* hypomorphic mutations and patients with heterozygous *LIG4* mutation.

Manuscript 1: Heterozygous hypomorphic mutations in *Rag1/Rag2* alter the adaptive immune responses following infection and the susceptibility to immune dysregulation

Annaïse J. Jauch[1], Olivier Bignucolo[2,3], Marianne Dölz[4], Marie Ghraichy[5,6], Rebecca Higgins[7], Adhideb Gosh[7,8], Ingmar Heijnen[9], Valentin von Niederhäusern[5,6], Jolan E. Walter[10], Lukas Jörg[11,12], Alexander Navarini[5], Johannes Trück[5,6], Luigi D Notarangelo[12], Mike Recher[1*]

[1]Immunodeficiency Laboratory, Department of Biomedicine, University and University Hospital of Basel, Basel, Switzerland.
[2]Department of Pharmacology and Toxicology, University of Lausanne, Lausanne, Switzerland.
[3] SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland
[4]formerly Molecular Immune Regulation Laboratory, Department of Biomedicine, University and University Hospital of Basel, Basel, Switzerland.
[5]Division of Immunology, University Children's Hospital, University of Zurich, Zurich, Switzerland.
[6]Children's Research Center, University of Zurich, Zurich, Switzerland.
[7]Division of Dermatology and Dermatology Laboratory, Department of Biomedicine, University and University Hospital of Basel, Basel, Switzerland.
[8]Competence Center for Personalized Medicine University of Zürich/Eidgenössische Technische Hochschule, Zurich, Switzerland.
[9]Division of Medical Immunology, Laboratory Medicine, University Hospital Basel, Basel, Switzerland.
[10]University of South Florida and Johns Hopkins All Children's Hospital, Saint Petersburg, FL, USA
[11]Division of Allergology and Clinical Immunology, Department of Pneumology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
[12]Allergy Unit, Department of Dermatology, University Hospital Zurich, Zurich, Switzerland
[13]Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA.

*Correspondence Should be addressed to MR: e-mail: mike.recher@usb.ch
The authors declare no conflicts of interest

Abstract

Patients harboring homozygous or compound heterozygous hypomorphic mutations in the recombination-activating genes 1 and/or 2 (RAG1/2) suffer from delayed-onset combined immunodeficiency with granulomas and/or autoimmunity (CID-G/AI), mainly related to compromised but residual B and T lymphocyte development. The B and T cell receptor repertoire requires the RAG1/2 endonucleases, to initiate somatic *V, D* and *J* gene rearrangements – known as V(D)J recombination. Here we present three patients suffering from combined immunodeficiency (CID) with multiple infectious and non-infectious complications carrying double heterozygous missense mutations in *RAG1/RAG2*. So far it has not been studied in detail which impact heterozygous hypomorphic mutations in either *RAG* gene or double heterozygous *RAG1/RAG2* mutations have on the immune system. We therefore analyzed two hypomorphic RAG1^{mut} and RAG2^{mut} murine models in homozygous, heterozygous or double heterozygous state thereby creating an *in vivo* RAG activity gradient. RAG2$^{mut/mut}$, heterozygous RAG2$^{mut/+}$ and RAG1$^{mut/+}$ mice had only subtle changes in lymphocyte development and demonstrated a conserved innate and adaptive immune response to lymphocytic choriomeningitis virus (LCMV) infection. In contrast, similar to the index patient, double heterozygous *Rag* mutant mice (RAG1$^{mut/+}$RAG2$^{mut/+}$) displayed a significant block in lymphocyte development at RAG dependent steps, T cells being more profoundly affected than the B cells. High-throughput T cell receptor sequencing of double heterozygous *Rag* mutant thymocytes revealed a higher abundance of productive sequences with an augmented oligoclonality. Following LCMV infection, fewer LCMV-specific CD8$^+$ T cells of reduced affinity expanded in double heterozygous *Rag* mutant mice. Inversely, LCMV-neutralizing antibody titers, as well as IgG autoantibodies were augmented in double heterozygous *Rag* mutant mice. This study for the first time documents an additive-immunodeficiency and immune-dysregulation caused by double heterozygous *RAG1/RAG2* hypomorphic mutations.

Keywords
Inborn errors of immunity, haploinsufficiency, RAG, LCMV, immune dysregulation

Introduction

The adaptive immune system has evolved to clonally respond to pathogens *via* recognition by specific B and T cell receptors. Receptor diversity requires the recombination-activating gene 1 (RAG1) and RAG2 to initiate the stochastic assembly of variable (*V*), diversity (*D*) and joining (*J*) antigen receptor gene segments. These rearranged gene segments encode for the variable domain - determining antigen specificity - of the antigen receptors found on T and B cells[1,2]. The clinical spectrum and immunological phenotypes associated with human RAG deficiency are broad[3–5]. A complete absence of B and T lymphocytes is caused by completely abolished RAG function, leading to severe combined immunodeficiency (SCID) manifesting with life-threatening infections within the firsts months of life[6–8]. Hypomorphic *RAG* mutations causing reduced but not completely absent RAG activity are associated with various levels of residual B and T cells in the peripheral blood[9,10]. Clinically, such patients may present with delayed-onset combined immunodeficiency with granulomas and/or autoimmunity (CID-G/AI)[11,12]. Virtually all CID-G/AI patients described to date carried homozygous or compound heterozygous *RAG* mutations[3,13]. It is ill-defined whether heterozygous or double heterozygous *RAG* mutations have may result in clinically manifest RAG insufficiency. As heterozygous *RAG* missense mutations are found in as many as 1/425 individuals, this question is highly relevant[14,15]. We aimed at analyzing the impact of a single or double heterozygous hypomorphic *RAG* mutations on the lymphocyte development, the pathogen-specific *vs.* auto-reactive adaptive immune response.

Here we describe three individuals from the same Caucasian family carrying double heterozygosity for two missense mutations in the *RAG1* and *RAG2* genes. One of the three individuals suffered from a complicated form of CID ultimately requiring allogeneic bone marrow transplantation, while the other individuals were clinically more silent but shared some of the phenotypical immune dysregulations. To examine the *in vivo* impact of single or double heterozygous mutations in the *RAG1/2* genes, we characterized two mouse models carrying a hypomorphic missense mutation in *Rag1* (RAG1^mut^) or *Rag2* (RAG2^mut^) with regard to their impact on the lymphocyte development, the T cell repertoire composition, the capacity to respond to lymphocytic choriomeningitis virus (LCMV) infection and the infection-induced immune dysregulation.

Added value of this study to the field of inborn errors of immunity

We demonstrate for the first time that single heterozygous hypomorphic *Rag* mutations, even by decreasing RAG protein functionality by ~ 25%, can be compensated in the context of an acute viral infection. On the contrary decreasing RAG functionality by 40% seen in mice with double heterozygous *Rag1/2* mutations leads to alterations in lymphocyte development, the T cell receptor repertoire and the adaptive immune response towards virus infection. Several manifestations of immune-dysregulation observed in the double heterozygous *Rag* mutant mice were mirrored in the human patients carrying two heterozygous *RAG1/RAG2* mutations.

Results

Combined immunodeficiency associated with double heterozygous *RAG1/RAG2* missense mutations

A 51-year old Caucasian male (index patient) was referred with a seven-year history of gastrointestinal symptoms of cramping abdominal pain, large volume watery stools and progressing to protein-losing enteropathy. No causative pathogen could be identified. Histologic findings of the small intestine comprised increased intraepithelial lymphocytes, villous atrophy and cryptic hyperplasia. He was sequentially treated with systemic steroids, azathioprine, mycophenolate, infliximab (monoclonal anti-TNFα antibody) and vedolizumab (monoclonal anti-integrin $\alpha_4\beta_7$ antibody), all without clinical response. When the patient's weight reached the nadir of 54kg (loss of 20kg within two years), accompanied by malabsorption of vitamins and trace elements, as well as hyporegenerative microcytic anemia an intravenous Hickman line for total parental nutrition was installed. Due to profound antibody deficiency (see below), intravenous immunoglobulin (IVIG) substitution was initiated without amelioration of the gastrointestinal symptoms.

Past medical history was noteworthy for recurrent autoimmune hemolytic anemia, autoimmune neutropenia as well as recurrent pneumonia. Transbronchial biopsy showed dense polyclonal lymphocytic infiltration of the respiratory mucosa. Computer tomography (CT) of the lungs demonstrated bronchiectasis, nodules and ground glass phenomena (fig. 1a). Positron emission tomography (PET)-CT scans documented metabolic active bi-pulmonary consolidations as well as mediastinal and hilar lymphadenopathy (fig. 1b). The carbon monoxide diffusion capacity of the lung was decreased to 53% of normal. Bronchoalveolar lavage could not identify causative pathogens and the diagnosis of granulomatous-lymphocytic interstitial lung disease (GLILD) was made. Additionally, a pronounced splenomegaly was noted (fig. 1b). Cutaneous involvement included erythema nodosum on both lower legs and vitiligo. Besides recurrent pneumonia, the patient developed a distinct varicella-zoster virus reactivation with bilateral zona and suffered from a sepsis episode due to the opportunist bacteria *Morganella morganii* and bacteremia with *Citrobacter freundii*.

The patient's mother was affected during young adulthood from multiple cutaneous warts, suffered from one life-threatening pneumonia episode and had a history of poorly defined pneumopathy that however did not progress over time. The 16-year-old daughter of the index patient had a history of prolonged airway infections and recently onset diarrhea.

Immunologic investigations of the index patient revealed serum IgA and IgM below the detection limit (fig. 1c), whereas total serum IgG was 3.5g/L [normal range 7-16 g/L] despite 50g intravenous immunoglobulin supplementation twice per month. The index patient's mother displayed reduced serum levels for IgM and IgG1 (fig. 1c, fig. 1d). Also, the daughter of the index patient presented IgG subclass deficiency (fig. 1c, fig. 1d). More pronounced alterations within the peripheral blood B cell subsets were found in the index patient than in his mother and daughter (suppl. fig. S1a). Nevertheless, an increased frequency of transitional B cells ($IgD^+IgM^+CD38^+CD19^+CD20^+$) was found in all three diseased family members (fig. 1e). The frequency of $CD21^{low}$ B cells was increased in the index patient (suppl. table S1, suppl. fig. S1c), while the ratio of B cells expressing κ- *vs*. λ-light chains was normal and comparable in the three diseased family members (suppl. fig. S1d). Because of the observed alterations within the B cell subsets and immunoglobulins, we performed a B cell receptor high-throughput sequencing experiment focusing on the immunoglobulin heavy chain locus (*IGH*) of the index patient and his mother. We did not observe altered *V* gene family usage within naïve B cells (data not shown). However, both patients displayed an inversed utilization of the joining genes 4 and 6 (*IGHJ4*, *IGHJ6*) in naïve B cell transcripts, compared to healthy controls (fig. 1f).

Absolute numbers of peripheral blood T cells were reduced in the index patient and his daughter (suppl. table S1). We found that the index patient and his mother had a low frequency of T cells expressing the TCR Vα7.2 (fig. 1g), while the daughter's frequency was in the lower quartile of values found in healthy controls (fig. 1g). The mother and the index patient displayed an augmented $CD4^+/CD8^+$ ratio (suppl. fig. S1e). A virtually absence of naïve ($CD27^+CD45R0^-$) T cells (fig. 1h, 1i) and regulatory $CD4^+$ T cells (T_{regs}, $CD25^+CD127^{low}$) was noted in the index patient (suppl. fig. S1f). Further, he displayed an elevated fraction of $CD4^+$ T cells with a chemokine expression pattern corresponding to pro-inflammatory T_H1 and T_H1^* cells (suppl. fig. S1h). We documented in the index patient and his mother a decrease frequency of circulating mucosa-associated invariant T cells (MAIT, $Vα7.2^+CD161^+$) as well as a low percentage of $γδTCR^+$ T cells (suppl. fig. S1g). NK cell subsets were comparable in the index patient and his mother with healthy controls (suppl. fig. S1i).

In summary, in the index patient, his daughter and his mother the B and T cell compartments showed multiple anomalies, whereas the NK cell compartment was preserved.

The index patient was diagnosed with a combined immunodeficiency (CID) associated with recurrent infections, enteropathy and GLILD. Whole exome sequencing of the index patient's blood mononuclear cells (PBMC-) derived DNA revealed two heterozygous missense mutations in *RAG1* and *RAG2* (table 1, suppl. table S2). Both variants were confirmed by Sanger sequencing in the index patient and were present in the mother

and daughter (fig. 1j, 1k). In both RAG variant proteins, a methionine was replaced by a valine at positions p.1006 (RAG1) and p.502 (RAG2) correspondingly (fig. 1j, 1k).

Table 1 | Whole-exome sequencing information regarding the missense mutations in *RAG1* and *RAG2*.

Genetic location	Nucleotide change (c.DNA)	Zygosity	Gene symbol	AA change	rsID	ESP	Poly Phen2	SIFT	CADD-PHRED	gnomAD allele freq
11_36576320_A_G	c.A3019G	Het	*RAG1*	p.M1006V M[ATG] > V[GTG]	rs139113046	0.002	Possibly damaging 0.61	Deleter-ious, 0	24.4	1.472×10^{-3}
11_36592665_T_C	c.A1504G	Het	*RAG2*	p.M502V M[ATG] > V[GTG]	rs145614809	0.003	Benign 0.026	Deleter-ious, 0	19.3	1.479×10^{-3}

AA amino acid, Allele freq allele frequency, CADD-PHRED Combined Annotation Dependent Depletion, ESP Exome Sequencing Project, gnomAD Genome Aggregation Database v3.1, Het heterozygous, PolyPhen2 Polymorphism Phenotyping2, rsID single nucleotide polymorphism database identifier, SIFT Sorting for Intolerant from Tolerant.

Different algorithms estimating the impact of the genetic variant on the respective proteins, classified the missense mutations as benign (RAG2, PolyPhen2) or possibly damaging (RAG1, PolyPhen2) to deleterious (RAG1/RAG2, SIFT score). The whole sequence of AA1000 to AA1006 amino acid positions in RAG1 is strictly conserved across different species, while the RAG2 position Met502 is well-conserved (fig. 1l). The two variants have been previously tested separately in an *in vitro* RAG-recombination assay, demonstrating normal RAG activity[11,16,17]. Inspection of the available RAG2 deposited structures (PDB RCSB) reveals that the residues of the C-terminus (AA>350) seemed to escape crystallization (as PDB code 6OEO), suggesting that the RAG2 wild-type as well as mutated Met502 lies within an intrinsically disordered segment (fig. 1m). The missense mutation found in *RAG1* locates within the carboxy-terminal domain (CTD, fig. 1m), a domain that displays a mutation rate of approx. 20%[18]. Structural investigations, suggest that the long hydrophobic side-chain of the wild-type Met1006 interacts with the alkyl chain of the Lys486 side-chain. This interaction restricts the mobility of Lys486 in such a way that it allows interaction with phosphate groups of the DNA backbone (fig. 1n). With the residue change from Met1006 to Val1006, the Lys486 stabilization is not assured because of the smaller size of the valine (fig. 1n). So far, the *RAG1* and *RAG2* variants have not been functionally assessed in combination.

a

b

c

d

e

f

g

h

i

j

RAG1 c.A3016G

RAG1 c.A3016G

p.1006M

p.1006M

k

RAG2 c.A1504G

RAG2 c.A1504G

p.502V

p.502V

l

RAG1

H. sapiens	K Y L Q K F M N A H - N A L K T S G F	1017
M. musculus	K Y L Q K F M N A H - N A L K S S G F	1014
S. scrofa	K Y L Q K F M N A H - K A F K N S G F	1017
O. cuniculus	K Y L Q K F M N A H - N A V K N S G F	1016
X. laevis	K Y L Q K F M N A H - K T L R S Q G F	1020
G. gallus	K Y L Q K F M N A H - K T L R S Q G F	1015
D. rerio	K Y L Q K F M E A H K N S V K A M Q A	1037
O. mykiss	K Y L Q K F M E A H K D S A K A L Q A	1050

RAG2

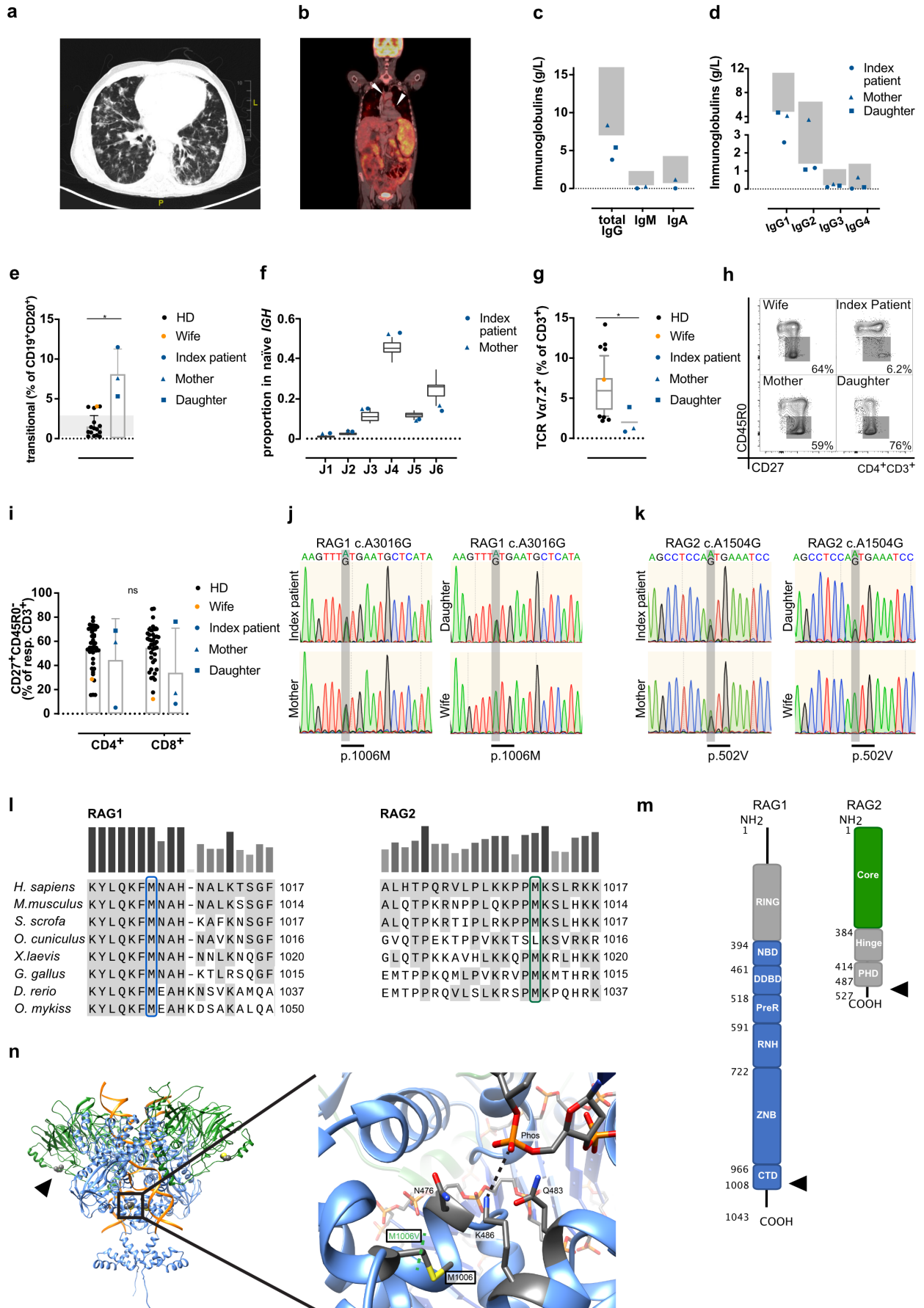|		
A L H T P Q R V L P L K K P P M K S L R K K	1017
A L Q T P K R N P P L Q K P P M K S L H K K	1014
A L Q T P K R T I P L R K P P M K S L H K K	1017
G V Q T P E K T P V K K T S L K S V R K R	1016
G L Q T P K K A V H L K K Q P M K R L H K K	1020
E M T P P K Q M L P V K R V P M K M T H R K	1015
E M T P P R Q V L S L K R S P M K P Q H R K	1037

m

RAG1

RAG2

n

Figure 1: Qualitative changes in the B cell and T cell compartment in patients carrying two heterozygous mutations in *RAG1* and *RAG2*. a) Computer tomography (CT) scan of the index patient, axial plane of the chest. The representative scan displays bronchiectasis, nodules and ground glass phenomena. **b)** Positron emission tomography (PET)-CT scan coronal plane of the index patient, documented bi-pulmonary consolidations with metabolic activity and mediastinal/hilar lymphadenopathy (white arrows) as well as splenomegaly. **c+d)** Immunoglobulin measurements in the three patients (gray box indicates the respective in-house reference range). **e)** Transitional B cells (CD38+IgM+IgD+) were enumerated within the B cell population by flow cytometry. Gray indicates the in-house reference range of healthy donors (HD). **f)** High-throughput sequencing of peripheral blood-derived B cell RNA. Naïve B cells were characterized regarding the *J* gene usage withing the immunoglobulin heavy chain (*IGH*) locus. **g)** Assessment of the proportion of peripheral blood T cells expressing the TCR Vα7.2, box plot with n = 47 controls indicates median, box 25-50% and whiskers 5-95%. **h)** Representative flow cytometric contour plots regarding naïve CD4+ T cells (CD27+CD45R0-, gray box). Number next to the box indicates percentage within the CD4+CD3+ cell population. **i)** Enumeration of naïve CD4+ and CD8+ T cells in the index patient, his mother, his daughter and the healthy wife. Sanger sequencing confirmed the heterozygous missense mutation in **j)** recombination-activating gene 1 (*RAG1*) and **k)** *RAG2* in the three patients. Heterozygosity is seen as superimposed peaks in the chromatogram. **l)** Multiple protein alignment for RAG1 resp. RAG2 from different vertebrates. **m)** Schematic illustration of the human RAG1 and RAG2 proteins. The two mutations are indicated with black arrows. The catalytic core domain is colored in blue for RAG1 and green for RAG2. Amino acid numbers refer to[19]. CTD (carboxy-terminal domain), DDBD (dimerization and DNA-binding domain), NBD (nonamer-binding domain), PHD (plant homeodomain domain), preR (pre-RNase H), RING (really interesting new gene, zinc-finger domain), RNH (catalytic RNase H). **n)** Molecular representation of the structure of the human RAG1-RAG2 heterotetramer in ribbons bound to DNA. Blue colored ribbons represent RAG1 (AA338-1010, the structure was constructed with SWISS-MODEL[20]), green represents RAG2 (AA1-527 , the structure was constructed using the ROBETTA algorithm[21]) and orange highlights the DNA helix. In RAG2 the M502 is indicated by a black arrow. The zoom-in view for RAG1 indicates the proximity of M1006 to Lys486 and the DNA (phosphate). The green dashed line represents the limit of a much smaller Val side-chain at position 1006. By the residual change to M1006V, the interaction between Lys486 and the DNA might be destabilized. Significance was computed with a non-parametric Mann-Whitney test in (f) and with multiple comparison correction for (g/h/i).

Analysis of lymphocyte development in monoallelic and bi-allelic murine *Rag* mutant mice

The crucial importance of the RAG proteins in the immune system development has been demonstrated in detail in homozygous RAG1[mut 22,23] and *Rag1* and *Rag2* knockout[24,25] mice. Our aim was to investigate in detail the in vivo impact of heterozygous *Rag1/Rag2* variants, alone or in combination, on the development and the function of the immune system. We studied two murine models, one bearing a missense mutation in the *Rag1* catalytic domain (resulting in p.S723C, referred to RAG1[mut])[22]. The second model carrying a missense mutation in the Rag2 core domain (leading to p.F62L, referred to as RAG2[mut], unpublished, fig. 2a, suppl. fig. S2a), which corresponds to the orthologous human mutation (RAG2 p.F62L) identified in a patient with combined immunodeficiency with granuloma and/or autoimmunity (CID-G/AI)[26]. By crossing homozygous *Rag1* and *Rag2* mutant mice with either wild-type mice or with each other, we aimed at investigating qualitative and quantitative changes in the lymphocyte development in the resulting Rag genotypes. Mice were viable and generated according to Mendelian inheritance for the RAG2[mut] strain, while a slightly less than expected proportion of homozygous RAG1[mut/mut] mice was noted amongst newborns (suppl. fig. S2b).

Lymphocytic developmental stages were analyzed in the thymus and the bone marrow (BM) of various *Rag* mutant or wild-type RAG[+/+] mice. Thymic lymphopenia was only observed in homozygous RAG1[mut/mut] mice (approx. 9-fold reduction compared to RAG[+/+] mice) and RAG[KO] mice (fig 2b, suppl. table S3). The degree of thymic lymphopenia in RAG1[mut/mut] mice was comparable with the initial description of the model on a SV129 background[22]. A trend for augmented thymocyte numbers was observed in the heterozygous RAG1[mut/+] mice (fig 2b, suppl. table S3). BM cellularity was reduced in RAG2[mut/+] mice compared to *Rag* wild-type mice (fig 2c, suppl. table S3).

We observed a decreased frequency and number of B cells at the pre-pro B cell stage in RAG2[mut/mut] and RAG2[mut/+] mutant mice (fig. 2d, suppl. fig. S2d). Only RAG1[mut/mut] and RAG[KO] mice demonstrated a significant developmental block at the transition of the progenitor B (pro B cell, *D* to *J* gene segment joining in the immunoglobulin heavy chain *Igh* locus) to the precursor (pre) B cell stage (fig. 2d). Double heterozygous RAG1[mut/+]RAG2[mut/+] mice, had proportionally more pro B cells than *Rag* wild-type mice (fig. 2d) and displayed an

absolute reduction of immature B cells (B220$^+$IgM$^+$IgD$^-$) – the B cell stage following successful *Igh* recombination (suppl. fig. S2d).

Two waves of RAG recombination occur during thymocyte development, namely at the double-negative (DN, CD4$^-$CD8$^-$ for the T cell receptor β chain TCRβ) and the double-positive (DP, CD4$^+$CD8$^+$ for the TCRα chain) stage. A relative accumulation of thymocytes at the DN stage was only observed for RAG1$^{mut/mut}$ and RAG^{KO} animals but not in the other *Rag* mutant models (fig. 2e, suppl. fig. S2e). In contrast, we observed a relative accumulation of thymocytes at the DP stage for single heterozygous RAG1$^{mut/+}$ and double heterozygous *Rag* mutant mice (fig. 2e, suppl. fig. S2e). DN thymocytes of RAG2$^{mut/+}$, RAG1$^{mut/+}$ and, double heterozygous *Rag* mutant mice accumulated at the DN3 stage, characterized by lower CD44 expression, while CD25 expression was retained[27] (fig. 2f, suppl. fig. S2f). Next, the T cell receptor β (TCRβ) chain rearrangement step occurring within the DN3 stage was investigated, which is associated with the upregulation of the co-stimulatory molecule CD28[28,29]. Thymocytes of the single heterozygous RAG1$^{mut/+}$ and double heterozygous RAG1$^{mut/+}$RAG2$^{mut/+}$ mice demonstrated a partial arrest at the DN3a (CD27^{high}) to DN3b (CD27^{high}CD28^{high}) transition compared to thymocytes of RAG$^{+/+}$ mice (fig. 2g, 2h, suppl. fig. S2g). Double heterozygous *Rag* mutant mice showed no detectable alteration in the positive selection process, seen at the transition of TCRβ^{int}CD69$^+$ to the TCRβ$^+$CD69$^+$ stage (fig. 2i). Analyzing more mature thymocytes at the single positive (SP) stages, a subtle increase was noted in the SP CD8$^+$ subset in RAG2$^{mut/+}$ mice, on the contrary we observed a significant reduction of the SP CD8$^+$ thymocytes in RAG1$^{mut/+}$ mice compared to *Rag* wild-type mice (fig. 2j, suppl. fig. S2e). The double heterozygous Rag mutant mice displayed a significantly decreased frequency and trend for a reduced number of SP CD4$^+$ thymocytes compared to RAG$^{+/+}$ mice (fig. 2j, suppl. fig. S2e).

To obtain an estimate for the impact of each *Rag* missense mutation - alone or in combination - on the thymocyte compartment composition, a ratio was calculated in which the fraction of thymocytes at RAG independent stages (both SP) were divided by the fraction of thymocytes at RAG dependent developmental stages (DN and DP) and then normalized to the average ratio observed in RAG$^{+/+}$ mice. This calculation illustrates the productivity of RAG-dependent rearrangement processes. Heterozygous and homozygous RAG2^{mut} mutant mice displayed a ratio close (99% and 92%, respectively) to RAG$^{+/+}$ mice. In contrast, the ratio calculated in RAG1$^{mut/+}$ and double heterozygous *Rag* mutant mice was 77% and 59%, respectively compared to wild-type RAG$^{+/+}$ mice. RAG1$^{mut/mut}$ mice displayed a ratio below 10% of RAG$^{+/+}$ mice (fig. 2k). Importantly, the ratio of the double heterozygous RAG1$^{mut/+}$RAG2$^{mut/+}$ mice was significantly reduced compared to both the single heterozygous RAG2$^{mut/+}$ and RAG1$^{mut/+}$ mice. Based on this mathematical estimate of *in vivo* RAG activity, a step-gradient could be generated identifying RAG2$^{mut/+}$ and RAG2$^{mut/mut}$ mice as very mildly RAG insufficient, while RAG1$^{mut/+}$ and even more the double heterozygous RAG1$^{mut/+}$RAG2$^{mut/+}$ mice were intermediately RAG insufficient. In keeping with published data, RAG1$^{mut/mut}$ were severely RAG insufficient (fig. 2k).
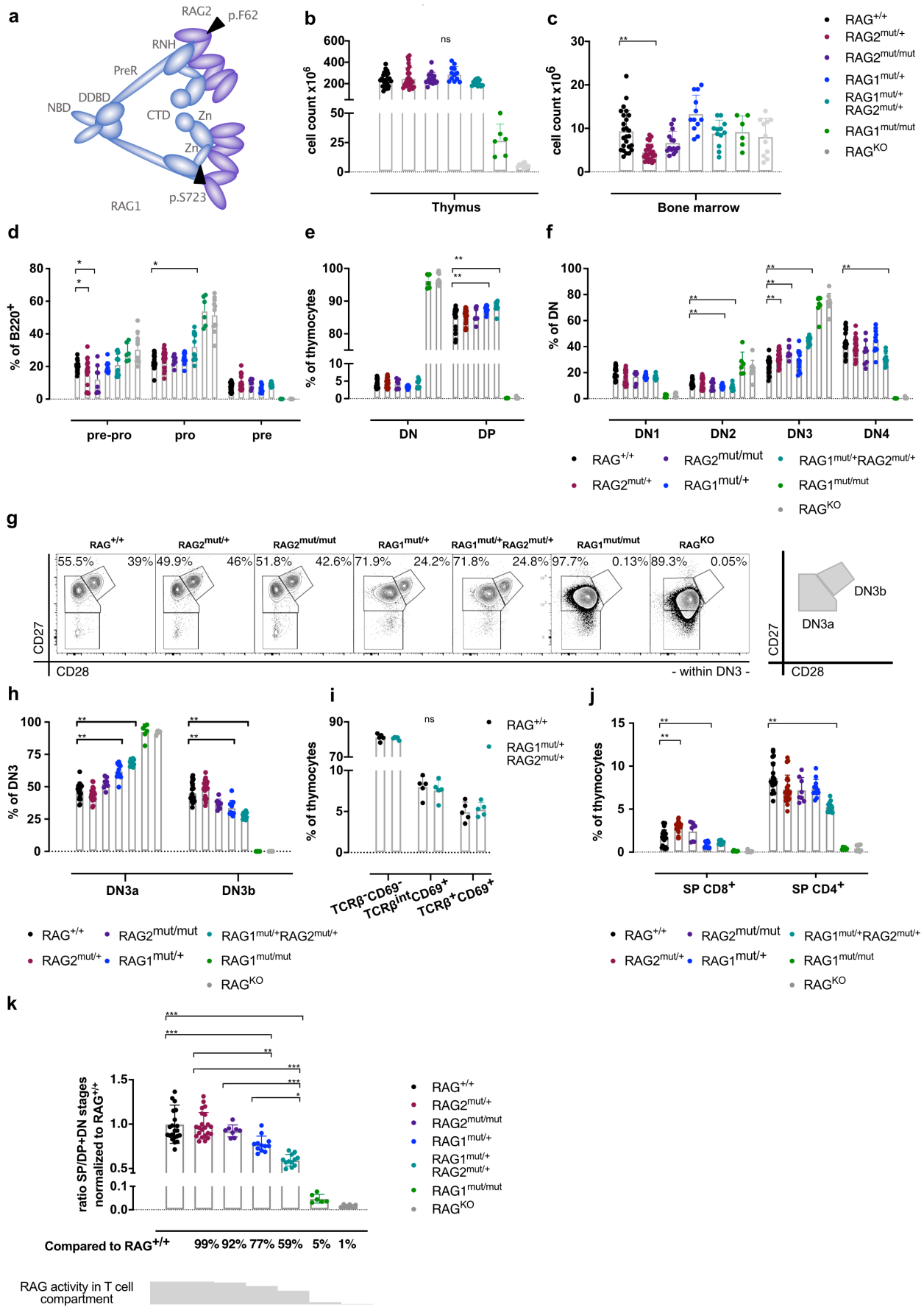
Figure 2: B cell and thymocyte development in *Rag* mutant mice. a) Illustration of the heterotetramer formed by the murine recombination-activating gene (RAG1) and RAG2 proteins bound to DNA. The RAG2 p.F62L mutation is found at the end of the first beta-propeller structure and one residue before the first interface with RAG1. The RAG1 p.S723 residue is found in the first Zinc-binding domain (black arrows indicate the mutation sites). Adapted from[3]. CTD (carboxy-terminal domain), DDBD (dimerization and DNA-binding domain), NBD (nonamer-binding domain), preR (pre-RNase H), RNH (catalytic RNase H), Zn (zinc-binding residue region). Quantitative and qualitative analyses of lymphocyte development were performed in the indicated *Rag* mutant mice at the age of eight to twelve days, using flow cytometry. **b)** Absolute bone marrow (BM) cellularity resp. **(c)** thymocyte count. RAG1$^{mut/mut}$ and RAG^{KO} thymocyte counts were significantly different from RAG$^{+/+}$ but statistics were omitted from the graphic for clarity reasons. **d)** Immunophenotyping of B cell developmental stages in the BM, including pre-pro B cells (B220$^+$CD19$^-$IgM$^-$IgD$^-$CD43$^+$CD117$^-$), pro B cells (B220$^+$CD19$^+$IgM$^-$IgD$^-$CD43$^+$CD117^{int}), pre B cells (B220$^+$CD19$^+$IgM$^-$IgD$^-$CD43$^-$CD25^{var} CD117$^-$). RAG1$^{mut/mut}$ and RAG^{KO} were significantly different from RAG$^{+/+}$ but statistics were omitted from the graphic for clarity reasons. **e)** Analysis of double-negative (DN, CD4$^-$CD8$^-$) and double-positive (DP, CD4$^+$CD8$^+$) thymocytes. RAG1$^{mut/mut}$ and RAG^{KO} thymocytes were significantly different from RAG$^{+/+}$ but statistics were omitted from the graphic for clarity reasons. **f)** Investigation of the double-negative 1 DN1 (CD44^{high}CD25^{low}), DN2 (CD44^{high}CD25^{high}), DN3 (CD44^{low}CD25^{high}) and DN4 (CD44^{low}CD25^{low}) thymocyte maturation stages. RAG1$^{mut/mut}$ and RAG^{KO} were significantly different from RAG$^{+/+}$ for the four DN stages analyzed but statistics were omitted from the graphic for clarity reasons. **g)** Representative flow cytometric contour plots showing the transition of DN3a (CD27^{high}) to DN3b (CD27^{high}CD28^{high}) stage, gated on DN3 cells. Frequencies above the gates indicate proportion within the DN3 thymocyte population. **h)** Relative distribution of thymocytes at the developmental stage DN3a and DN3b. RAG1$^{mut/mut}$ and RAG^{KO} were significantly different from RAG$^{+/+}$ for the subsets displayed. **i)** Thymocytes undergoing positive selection were analyzed. The transition is marked by the up-regulation of the T cell receptor (TCR)-β chain (prior selection CD69$^+$TCRβ^{int} to post selection CD69$^+$TCRβ$^+$). **j)** The proportional abundance of single-positive (SP) CD8$^+$ and CD4$^+$ thymocytes. RAG1$^{mut/mut}$ and RAG^{KO} were significantly different from RAG$^{+/+}$ but statistics were omitted from the graphic for clarity reasons. **k)** Mathematical illustration of the RAG-dependent alterations in the thymocyte compartment composition. The calculation of the is ratio based on the percentage of thymocytes at RAG-independent thymocyte stages (SP) divided by the percentage of thymocytes at RAG-dependent thymocyte stages (DN and DP) and normalized to the average ratio calculated in RAG$^{+/+}$ mice (set to 100%). Lower part: Percentage indicates perseverance of thymocyte development. A gradient of RAG activity in the various *Rag* genotypes based on the ratio calculation. Statistical comparisons used wild-type RAG$^{+/+}$ as a reference in (b-j). A Kruskal-Wallis test with multiple comparison correction was used for variance testing in (b-h/j), for (i) a Mann-Whitney test with post-hoc test was used. Merged data from six independent experiments are displayed except for i), where one experiment was performed with five mice per genotype. Otherwise three to eleven mice per experiment and genotype were used. Absolute numbers of BM cells are calculated based on pooled cells from two femurs and not normalized to weight.

Enhanced productive clonality in *Trb* sequences in SP RAG1$^{mut/+}$RAG2$^{mut/+}$ thymocytes

To address how the two heterozygous Rag missense mutations in the RAG1$^{mut/+}$RAG2$^{mut/+}$ mice would affect the primary T cell repertoire clonality, we performed next-generation sequencing of the T cell receptor β locus (*Trb*) in flow-sorted SP CD8$^+$ thymocytes (fig. 3a).

The *Trb* productive clonality was significantly higher in RAG1$^{mut/+}$RAG2$^{mut/+}$ mice compared to wild-type mice, reflecting augmented oligoclonality (fig. 3b). Computationally, we excluded the outgrow of one aberrant clone as an explanation for the oligoclonality (suppl. fig. S3a). Further, we found significantly less unique rearrangements in double heterozygous *Rag* mutant mice in total and productive sequences compared to RAG$^{+/+}$ mice (fig. 3c). RAG1$^{mut/+}$RAG2$^{mut/+}$ SP CD8$^+$ thymocytes favored productive over unproductive (out-of-frame or stop) rearrangements, while *Trb* sequences of *Rag* wild-type thymocytes harbored a greater proportion of unproductive rearrangements (fig. 3d). The augmented abundance of productive sequences has been described previously in lymphocyte repertoires of human patients and murine *Rag* mutated mice carrying homozygous *RAG* missense mutations[30–32].

The overall Trb diversity was comparable, in *Rag* wild-type *vs*. double heterozygous *Rag* mutant thymocytes, estimated using Shannon's Entropy index[33] (suppl. fig. S3b). In accordance, we observed a bell-shaped distribution of the complementary determining region 3 (CDR3) length in RAG$^{+/+}$ and RAG1$^{mut/+}$RAG2$^{mut/+}$ thymocytes (fig. 3e). During the T cell receptor rearrangement, the RAG proteins bind first to the *J* gene segments[34]. We did not observe a preferential utilization of either of the two J gene families in *Rag* wild-type *vs*. double heterozygous *Rag* mutant thymocytes tested (data not shown). Investigating the individual gene

usage in unique *Trb* sequences, we did not observe preferential usage of individual *J* and *V* gene segments in the two murine models (suppl. fig. S3c, S3d). However, we found a significantly more pronounced usage of the $D_{\beta 1}$ gene segment in unique productive *Trb* sequences of RAG1$^{mut/+}$RAG2$^{mut/+}$ thymocytes, whereas in RAG$^{+/+}$ thymocytes the utilization of $D_{\beta 1}$ and $D_{\beta 2}$ was more balanced (fig. 3f).

The gene segment composition in unique sequences revealed that double heterozygous *Rag* mutant thymocytes fostered not only more VDJ rearrangements *per se* but harbored also more aberrant VJ and VDDJ compositions in unique productive sequences, compared to RAG$^{+/+}$ thymocytes (fig. 3g). In line with the augmented aberrant compositions in RAG1$^{mut/+}$RAG2$^{mut/+}$ thymocytes, we found a higher sum of *V* pseudo-genes utilized in unique productive sequences compared to *Rag* wild-type (fig. 3h).
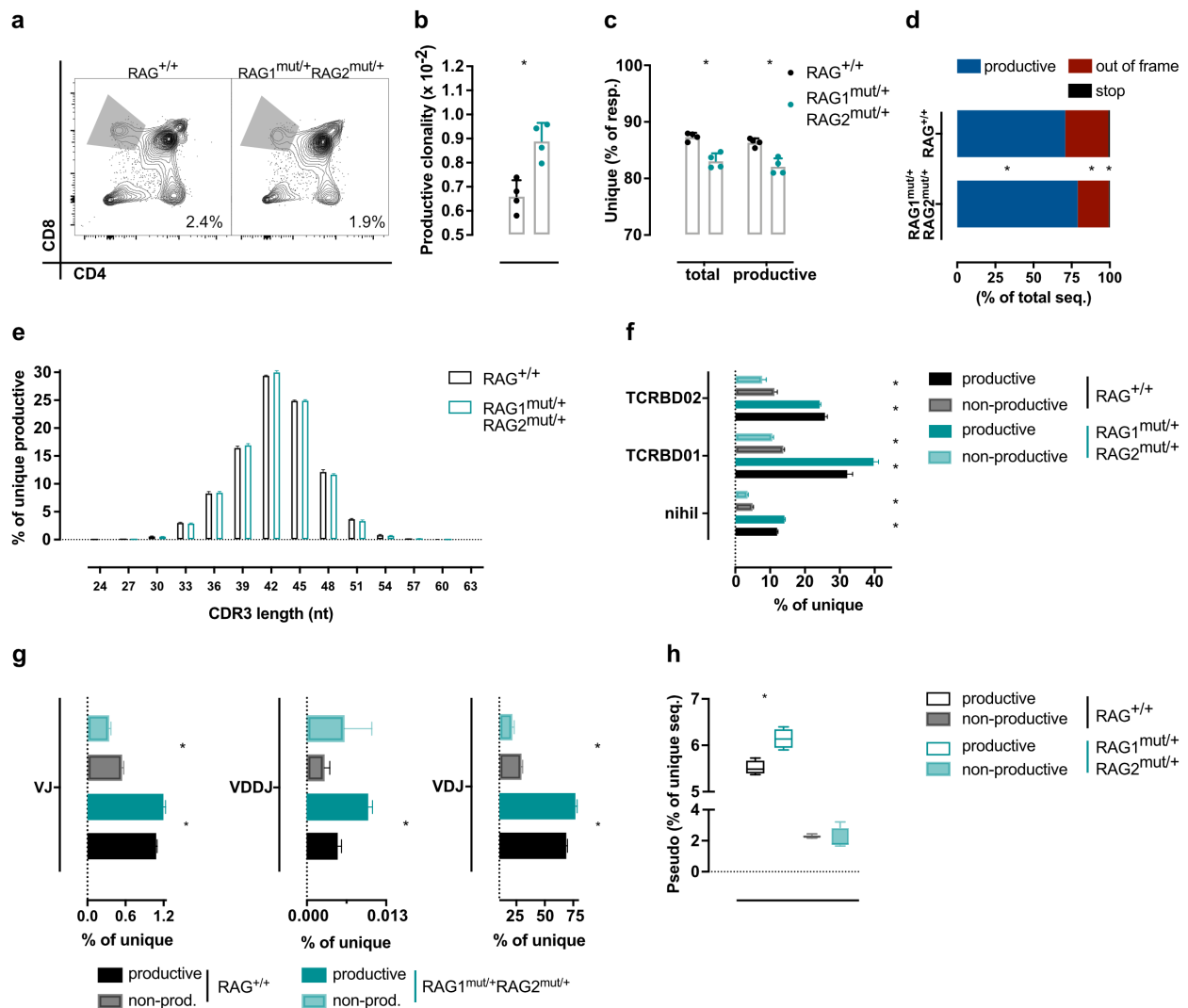
Figure 3: Enhanced productive clonality in *Trb* sequences in SP RAG1$^{mut/+}$RAG2$^{mut/+}$ thymocytes. **a)** Depiction of the sorting strategy for single-positive (SP) CD8$^+$ thymocytes, which were used for next-generation T cell receptor β locus (*Trb*) DNA sequencing. Frequency next to the gate refers to the bulk thymocyte population. **b)** The Trb productive clonality was quantified in SP CD8$^+$ thymocytes. **c)** Unique *Trb* sequences were computed within total vs. productive sequences. **d)** The proportion of productive, out-of-frame and stop sequences was assessed within unique SP CD8$^+$ thymocyte-derived *Trb* sequences. **e)** The relative distribution of the complementarity determining region 3 (CDR3) length in nucleotides (nt) was computed in unique productive *Trb* sequences. **f)** Unique *Trb* sequences were investigated for the relative usage of *D* genes within productive and non-productive sequences. **g)** In unique, productive and non-productive (out-of-frame and stop) *Trb* sequences, the proportion of *VJ, VDDJ* and, classical *VDJ* gene joining was analyzed. **h)** The sum of all *V* pseudogenes is displayed as frequency of total unique sequences. One experiment was performed with four mice per

genotype. Statistical significance was deduced using a non-parametric Mann-Whitney test in (b/h) and with additional post hoc correction in c/d/f/g.

Subtle accumulation of regulatory CD4$^+$ T cells and transitional B cells in RAG1$^{mut/+}$RAG2$^{mut/+}$ mice

To investigate if the changes observed during lymphocyte development would translate to altered lymphocytic subpopulations in adult mice, splenic T and B cell subsets were characterized, in mice with various *Rag1* and/or *Rag2* missense mutations.

RAG-independent NK cell numbers were preserved in all *Rag* mutant mice, whereas B and T cells numbers were reduced to absent in RAG1$^{mut/mut}$ and *Rag* knock-out mice[35] (fig. 4a, suppl. fig. S4a). A higher number of B220$^+$CD19$^+$ B cells and a trend for a higher number of CD3$^+$ T cells were observed in double heterozygous RAG1$^{mut/+}$RAG2$^{mut/+}$ animals compared to RAG$^{+/+}$ mice (fig. 4a, suppl. table S4).

To examine if a specific T cell subpopulation was responsible for the slightly increased numbers of bulk T cells in the RAG1$^{mut/+}$RAG2$^{mut/+}$ mice, we enumerated $\gamma\delta^+$, helper CD4$^+$ and cytotoxic CD8$^+$ T cells. Even though $\gamma\delta^+$ T cells have a reduced RAG-dependency[36,37], similar numbers were observed across mild and intermediate *Rag* insufficient mice. CD4$^+$ T cell numbers were increased in double heterozygous RAG1$^{mut/+}$RAG2$^{mut/+}$ mice compared to RAG$^{+/+}$ mice (fig. 4b).

As bulk T cell numbers were slightly reduced in RAG2$^{mut/mut}$ and RAG1$^{mut/+}$ mice but were marginally augmented in double heterozygous *Rag* mutated mice, we wondered whether this might be caused by altered homeostatic proliferation, which is associated with an increased frequency of T cells with an activated effector phenotype[22,38–40]. Probing for naïve, effector and central memory CD4$^+$ T cells was performed based on the expression of the activation molecule CD44 and the leucocyte adhesion molecule CD62L. No altered abundance of naïve, effector and central memory CD4$^+$ T cells was observed in mutant *Rag* mice except for RAG1$^{mut/mut}$ mice, which virtually completely lacked naïve CD4$^+$ T cells (fig. 4c, 4d). By comparing RAG1$^{mut/+}$RAG2$^{mut/+}$ with RAG$^{+/+}$ mice, no different proportion of CD4$^+$ follicular helper T cells (T$_{FH}$) was observed (fig. 4e). In contrast, a relative increase of CD4$^+$ regulatory T cells (T$_{reg}$) and a trend for higher absolute T$_{reg}$ numbers was enumerated in the double heterozygous RAG1$^{mut/+}$RAG2$^{mut/+}$ mice (fig. 4e, 4f, 4g). RAG$^{+/+}$ and double heterozygous mice had similar frequencies of thymic derived T$_{regs}$, which are marked by the expression of the transcription factor Helios[41] (suppl. fig. S4b). Further, no difference between RAG$^{+/+}$ and RAG1$^{mut/+}$RAG2$^{mut/+}$ T$_{regs}$ could be detected regarding their activation status (CD44 ICOS, PD1) and proliferation (ki67, suppl. fig. S4b).

Absolute B cell numbers were increased in the spleens of RAG1$^{mut/+}$RAG2$^{mut/+}$ mice compared to *Rag* wild-type mice, and this was paralleled by a trend for a higher frequency of B cells expressing the proliferation marker ki67 (suppl. fig. S4c). Analysis of transitional (transB CD24$^+$IgM$^+$IgD$^-$), follicular (follB CD21/CD35^{int} CD23^{high}) and marginal-zone (MZB CD21/CD35^{high}CD23$^{low/-}$) B cells showed, that the increased B cell numbers in the RAG1$^{mut/+}$RAG2$^{mut/+}$ animals were caused by increased numbers in the transitional B cell subset (fig. 4h).

Patients with RAG dependent leaky severe combined immunodeficiency (SCID) or RAG dependent Omenn syndrome suffer from profound hypogammaglobulinemia, while patients with clinically milder forms of RAG-associated combined immunodeficiency show variable levels of serum IgG and IgM[5]. Serum IgM levels were preserved in RAG1$^{mut/+}$ and RAG1$^{mut/+}$RAG2$^{mut/+}$ mice compared to wild-type RAG$^{+/+}$ animals. In naïve RAG1$^{mut/+}$ mice, total serum IgG levels were significantly elevated compared to RAG1$^{+/+}$ mice, mainly due to the increase of the IgG3 and IgG1 subclasses (fig. 4i). RAG1$^{mut/+}$RAG2$^{mut/+}$ mice had a 25% reduction of total serum IgG levels compared to RAG$^{+/+}$ mice (fig. 4i). Owing to the reduced total IgG levels, IgM was the predominant immunoglobulin isotype in the serum of RAG1$^{mut/+}$RAG2$^{mut/+}$ mice, while IgG was predominant in RAG1$^{mut/+}$ and RAG$^{+/+}$ mice (fig. 4j).
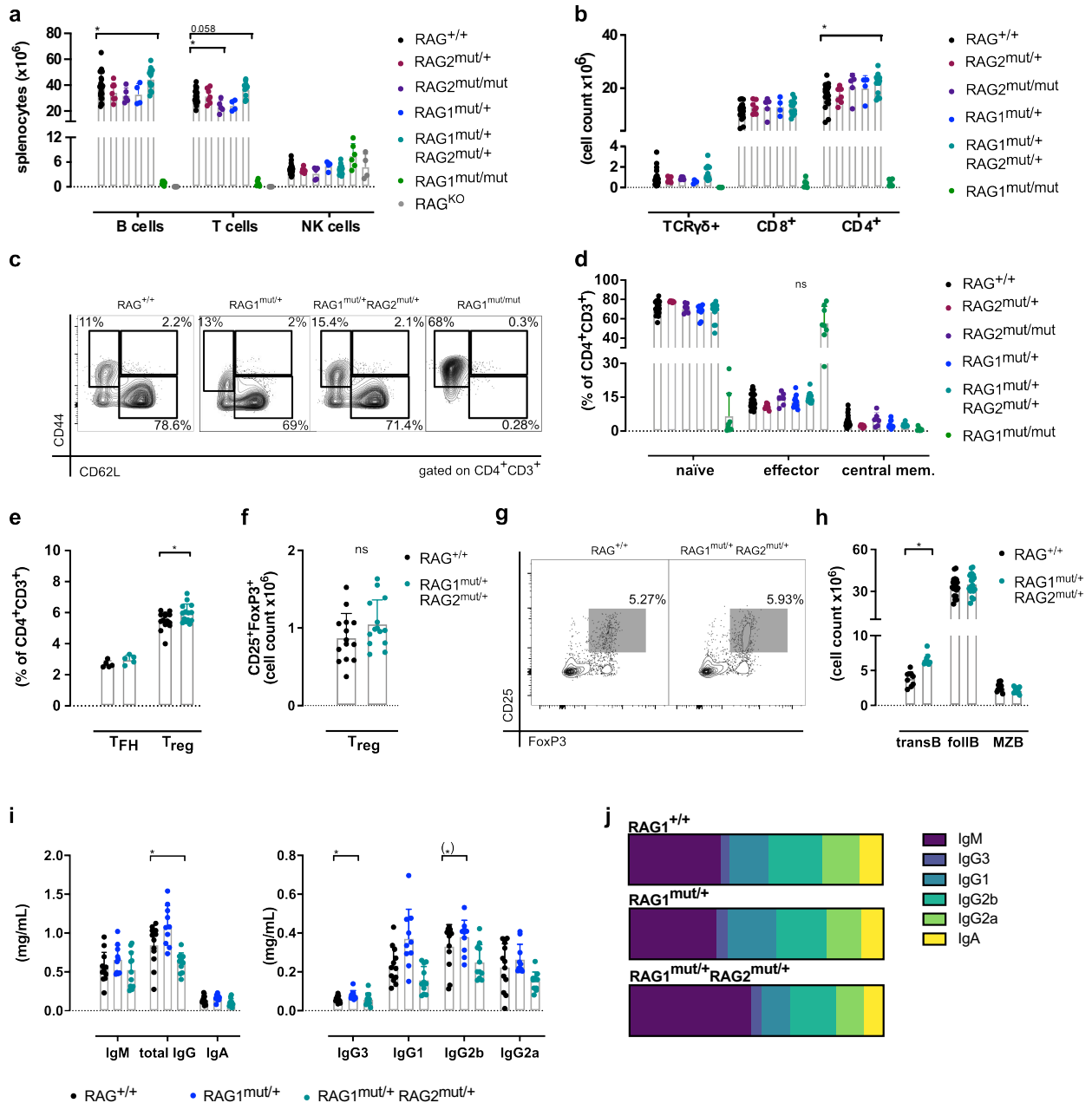
Figure 4: Subtle accumulation of T regulatory CD4$^+$ T cells and transitional B cells in RAG1$^{mut/+}$RAG2$^{mut/+}$ mice. The adult lymphocyte compartment was investigated focusing on the splenic lymphoid tissue. **a)** Quantitative analysis of the two RAG-dependent lymphocyte subpopulations - B cells (B220$^+$CD19$^+$) and T cells (CD3$^+$) as well as the RAG-independent NK cells (NK1.1$^+$). Numbers for B and T cells were significantly different in RAG1$^{mut/mut}$ and RAG^{KO} compared to RAG$^{+/+}$, but statistics were omitted from the plot for clarity motives. **b)** Enumeration of γδ$^+$ T cells (TCRγδ$^+$TCRβ$^-$CD19$^-$B220$^-$), CD4$^+$ helper T cells (CD4$^+$CD8$^-$CD3$^+$) and cytotoxic CD8$^+$ T cells (CD8$^+$ CD4$^-$CD3$^+$). Numbers for the T cell subsets were significantly different in RAG1$^{mut/mut}$ compared to RAG$^{+/+}$ but statistics were omitted from the graphic for clarity reasons. CD4$^+$ helper T cells were subdivided in naïve (CD62L$^+$CD44$^-$), effector (CD62L$^-$CD44$^+$) and central memory (central mem. CD62L$^+$ CD44$^+$) T lymphocytes. **c)** Representative flow cytometric contour plots and gating strategy and **d)** proportional distribution within CD4$^+$ T cells. Naïve and effector CD4$^+$ T cells were significantly different in RAG1$^{mut/mut}$ compared to RAG$^{+/+}$ but statistics were omitted from the graphic for clarity reasons. **e)** Relative enumeration of splenic follicular helper CD4$^+$ T cells (T$_{FH}$, bcl6$^+$ICOS$^+$) and regulatory CD4$^+$ T cells (T$_{reg}$, FoxP3$^+$CD25$^+$), assessed by flow cytometry. **f)** Absolute abundance of splenic T$_{reg}$ CD4$^+$ T cells (CD25$^+$FoxP3$^+$) in naïve mice. **g)** Representative flow cytometric plots for T$_{reg}$ CD4$^+$ T cells (CD25$^+$FoxP3$^+$). Frequency beside the quadrant indicates the number within the CD4$^+$ T cell population. **h)** Quantitative analysis of splenic transitional (transB CD24$^+$IgM$^+$IgD$^-$B220$^+$CD19$^+$), follicular (follB CD21/CD35^{int}CD23^{high}B220$^+$CD19$^+$), and innate-like marginal-zone (MZB CD21/CD35^{high}CD23$^{low/-}$B220$^+$CD19$^+$) B cell subsets in RAG$^{+/+}$ and RAG1$^{mut/+}$RAG2$^{mut/+}$ mice.

Investigation of unswitched and class-switched immunoglobulins (Ig) in the serum of naïve mice. Immunoglobulins are ordered according to their immunoglobulin constant region position on chromosome 12. **i)** The absolute level and **j)** relative distribution of the immunoglobulin subclasses are reported. Significance testing was performed using a Kruskal-Wallis test with multiple comparison correction in (a/b/d/i). For (f) the non-parametric Mann-Whitney test was performed and an additional post-hoc test was used for (e + h). (a-d) Merged data from four independent experiments are displayed, with two to nine mice per genotype and per experiment. (e-j) two independent experiments are merged with five to nine mice per genotype and experiment.

RAG1$^{mut/mut}$ mice have a profoundly impaired innate and adaptive immune response against lymphocytic choriomeningitis virus (LCMV) infection

It has been demonstrated previously that RAG1$^{mut/mut}$ mice do mount an adaptive immune response to vaccination with the T cell-independent trinitrophenyl (TNP)-Ficoll but only a minimal response to the T cell help dependent antigen TNP-keyhole limpet hemocyanin (TNP-KLH)[42]. Immune responses of RAG1$^{mut/mut}$ mice against a replicating pathogen have not yet been characterized. We aimed at assessing the functionality of the T and B compartments of RAG1$^{mut/mut}$ animals following infection of mice with the poorly cytopathic, persistence-prone lymphocytic choriomeningitis virus (LCMV), which has the mouse as natural host. The initial immune response against LCMV is dominated by a strong CD8$^+$ T cells expansion, comparable to hepatitis B and C viruses or human immunodeficiency virus (HIV)[43–47]. Virus-neutralizing antibodies are of pivotal importance at later time points of the infection[48].

The early innate antiviral response is decisive to contain viral-spread, while the adaptive immune system is being recruited. A crucial part of the innate immune response to an infection is the secretion of anti-viral type-I interferons (IFN)[49,50]. Previous characterization of the immune response in RAG^{KO} mice following LCMV infection revealed that those animals were unable to mount a normal type-I IFN repost[51], even though that type I interferon-producing cells develop independent from RAG expression[50]. It was unclear whether residual lymphocytes in RAG1$^{mut/mut}$ mice may sustain the cellular network needed to mount a type-I IFN response to LCMV. No crescendo of the proinflammatory INFα and IFNγ were measurable in the sera of RAG1$^{mut/mut}$ mice after LCMV low dose infection (LCMV$_{WE}$ strain, 200pfu), which is in sharp contrast to what is observed in *Rag* wild-type mice (fig. 5a, 5b). However, RAG1$^{mut/mut}$ mice were *per se* able of type-I IFN secretion, tested by challenging mice with a toll-like receptor 3 (TLR3) agonist (fig. 5c). We did not observe systemic immune reactions after LCMV infection in RAG1$^{mut/mut}$ animals, such as fluctuations of the surface body temperature (fig. 5d) or weight loss (data not shown).

We measured an increased proportion of effector CD8$^+$ T cells (CD44$^+$CD62L$^-$) nine days after LCMV infection in the spleens of RAG1$^{mut/mut}$ mice compared to pre-infection levels (fig. 5e), however no LCMV specific CD8$^+$ T cells recognizing the immunodominant glycoprotein gp33 (gp$_{33-41}$) were detected in the blood over time (fig. 5f). The lack of LCMV specific CD8$^+$ T cells was paralleled by high blood viral titers in RAG1$^{mut/mut}$ animals beyond four weeks after infection (fig. 5g). Despite elevated viral titers, no measurable LCMV-specific B cell response was noted in RAG1$^{mut/mut}$ mice, including the emergence of LCMV-glycoprotein (GP1) binding IgG (fig. 5h) and LCMV-neutralizing antibodies (fig. 5i).
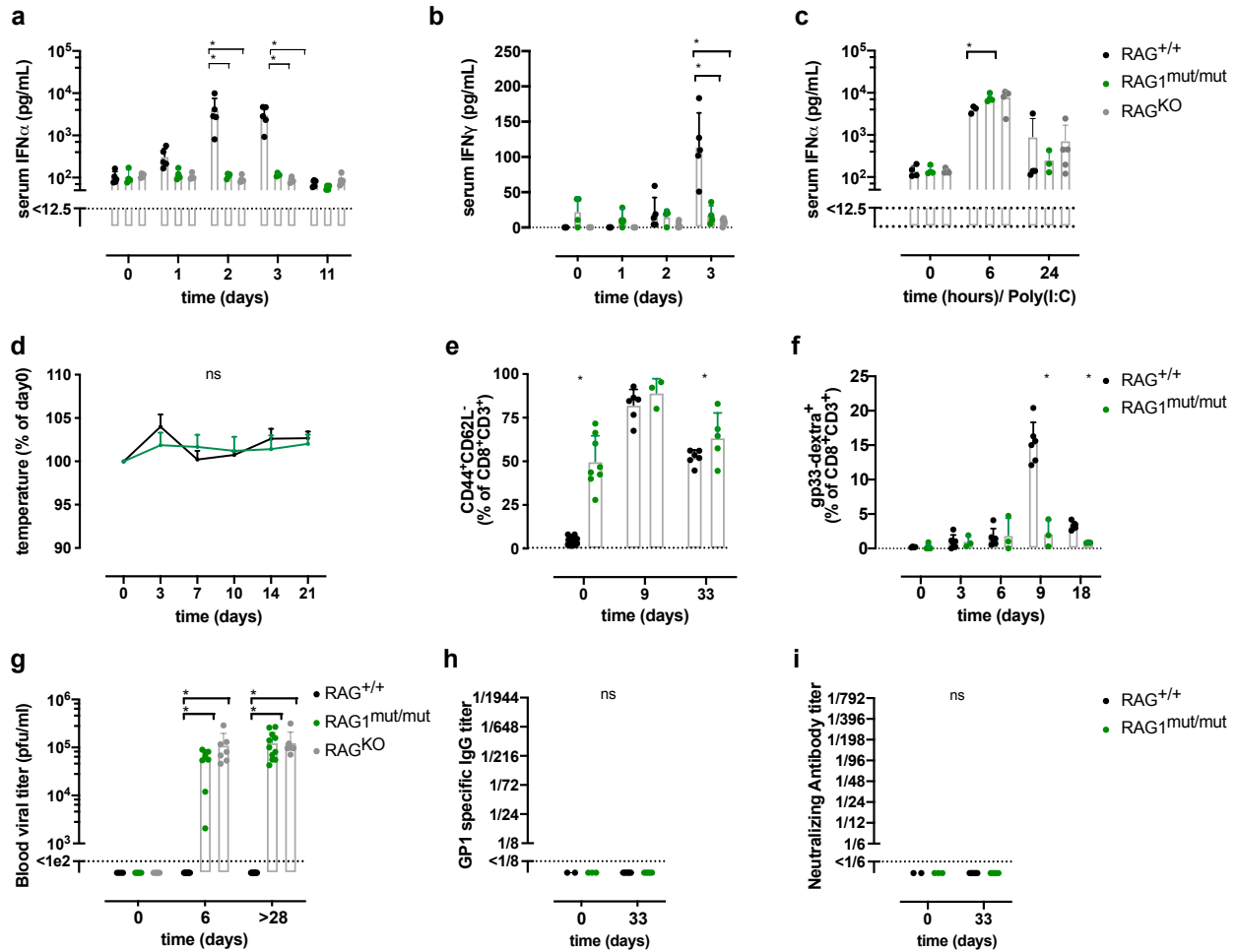
Figure 5: RAG1$^{mut/mut}$ have a profoundly impaired innate and adaptive immune response to LCMV infection. Mice were infected with LCMV$_{WE}$ low dose (200 pfu) in a,b and d-i. Innate cytokines were measured shortly prior to and after LCMV-infection in the sera of mice. **a)** Total serum IFNα levels and **b)** IFNγ levels were assessed at the indicated time points. **c)** The toll-like receptor agonist poly(I:C), a syntenic form of double-stranded RNA was injected once intraperitoneally and total serum IFNα levels were measured at indicated time points. **d)** Measurement of surface body temperature at indicated time points. **e)** Proportional distribution of effector CD8$^+$ T cells (CD44$^+$CD62L$^-$) was measured in the spleens, at three time points following LCMV$_{WE}$ infection with flow cytometry. **f)** CD8$^+$ T cells specific for the LCMV immunodominant epitope gp$_{33-41}$ were measured in the blood after LCMV infection by dextramer-staining and flow cytometry. **g)** Blood LCMV titers were quantified at indicated time points by functional plaque-forming assay after LCMV infection. LCMV specific B cell response was measured in the serum post-infection at the indicated time points. **h)** LCMV-glycoprotein (GP1)-specific IgG titers and **i)** LCMV-neutralizing antibody titers were assessed in a focus reducing assay. (a-i) Experiments were performed two to three times with three to six mice per genotype and per experiment, one experiment is displayed. Significance was computed with a 2wayANOVA and post hoc correction in (a-i) using the RAG$^{+/+}$ mice as a reference group.

Immunocompetence of RAG2$^{mut/mut}$ and RAG1$^{mut/+}$ mice following systemic LCMV infection

RAG2$^{mut/mut}$ and RAG1$^{mut/+}$ mice were infected with high dose LCMV (LCMV$_{WE}$ strain, 2 x10^6 pfu). Ten days post-infection, the amount of T and B cells in the spleen was comparable between the two *Rag* mutant mice and wild-type littermates (fig. 6a). Cytotoxic CD8$^+$ T cells recognizing the immunodominant virus epitopes (gp33 gp$_{33-41}$ and np396 np$_{396-404}$) were found in similar numbers and frequency in RAG1$^{mut/+}$ and RAG2$^{mut/mut}$ as in RAG$^{+/+}$ wild-type animals (fig. 6b, 6c).

As LCMV$_{WE}$ clearance mainly depends on CD8$^+$ T cells, RAG$^{+/+}$, RAG1$^{mut/+}$ and RAG2$^{mut/mut}$ were infected with the highly proliferative LCMV strain LCMV$_{clone-13}$[52], whose delayed clearance involves additional CD4$^+$ T

helper cell and B cell response[53–55]. The initial weight loss and recovery following infection was similar in all three genotypes (fig. 6d). Expansion of bulk CD8$^+$ T cells, LCMV-specific CD8$^+$ T cells (binding the gp33 or np396 LCMV derived peptides) and upregulation of the inhibitory surface molecule LAG3 were comparable in RAG$^{+/+}$ vs. the two *Rag* mutant mouse models (fig. 6e, 6f, 6g, 6h). Viral load was repetitively measured during the first month in the blood and was similar in mice of all three genotypes (fig. 6i). After ten weeks, the virus was cleared from all investigated organs in all three genotypes (fig. 6i). LCMV glycoprotein-specific (GP1) IgG appeared in the circulation with a comparable kinetic in RAG1$^{mut/+}$ and RAG2$^{mut/mut}$ mice compared to RAG$^{+/+}$ mice (fig. 6j).

In recapitulation, we did not observe a compromised adaptive immune response to LCMV infections in RAG2$^{mut/mut}$ and RAG1$^{mut/+}$ mice.

Figure 6: Immunocompetence of RAG2$^{mut/mut}$ and RAG1$^{mut/+}$ mice following systemic LCMV infection. Mice were challenged with intravenous infection of either LCMV$_{WE}$ (2 x10^{6}pfu, a-c) or the highly proliferative LCMV$_{clone13}$ (2 x10^{6}pfu, d -j). **a)** Quantitative enumeration of total T and B cells in the spleen ten days after LCMV infection, the measurement was
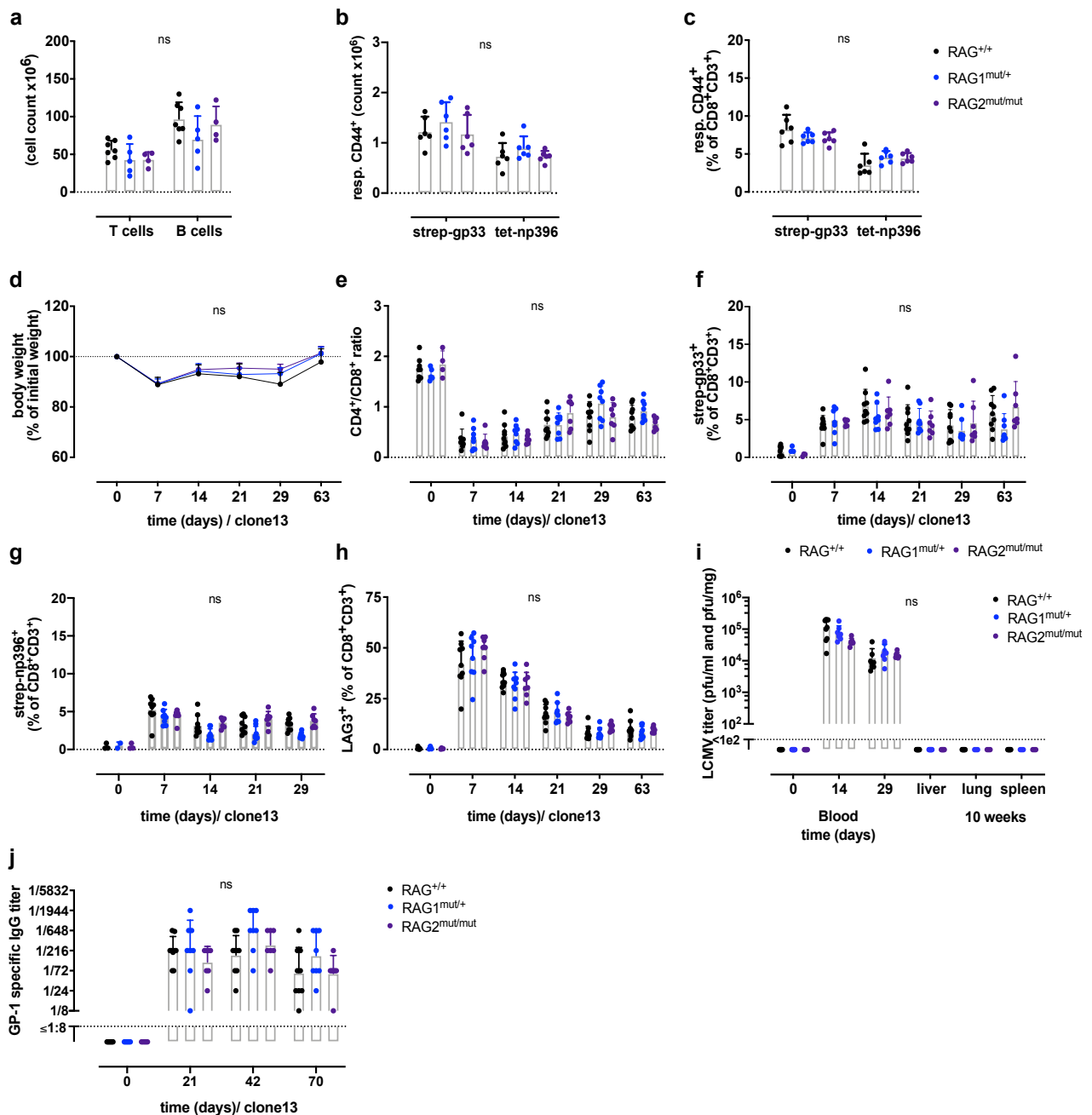
performed using flow cytometry. LCMV-specific CD8$^+$ T cells were assessed for their binding to the LCMV-derived gp$_{33-41}$ and np$_{396-404}$ peptides by multimer staining. Ten days after LCMV infection **b)** the quantitative and **c)** the relative abundance of splenic LCMV specific cytotoxic CD8$^+$ T cells was measured. **d)** Body weight changes (computed as a percentage from the initial weight) and **e)** expansion of CD8$^+$ T cells (expressed as CD4$^+$/CD8$^+$ T cell ratio) in the peripheral blood were monitored at indicated time points after LCMV$_{clone13}$ infection. The relative abundance of LCMV-specific cytotoxic CD8$^+$ T cells binding the **f)** gp33 or **g)** np396 peptides in the peripheral blood were measured by streptamer staining and flow cytometry at indicated time points following infection. **h)** Flow cytometric measurement of the surface expression of the inhibitory molecule LAG3 on CD8$^+$ T cells at indicated time points following post LCMV infection. **i)** LCMV titers were measured in the blood at indicated time points and in indicated organs by plaque-forming assay ten weeks after LCMV$_{clone13}$ infection. **j)** The LCMV specific B cell response was measured by monitoring LCMV-glycoprotein (GP1)-specific serum IgG titers at indicated time points, following LCMV infection. (a-j) Experiments were performed on two independent occasions with four to eight mice per genotype and experiment. A non-parametric Kruskal-Wallis test with post hoc correction was used for the significance testing in a-c. Significance was computed using a 2wayANOVA with multiple comparison correction in (d-j).

Robust T and B cell response to systemic LCMV infection in RAG2$^{mut/mut}$ und RAG1$^{mut/+}$ mice after transient CD8 T cell depletion

To unmask subtle differences in the immune response to LCMV infection, CD8$^+$ T cells were depleted prior to infection using monoclonal antibodies. This would facilitate viral spread during the transient absence of CD8$^+$ T cell response[56,57]. We sought to monitor the CD8$^+$ T cell reemergence as an approximation of the ability of cytotoxic T lymphocyte (CTL) generation and/or mobilization. The CD8$^+$ T cell reemergence kinetic was similar in RAG1$^{mut/+}$ and RAG2$^{mut/mut}$ compared with RAG$^{+/+}$ after high dose LCMV infection (LCMV$_{WE}$ 2x 10^{6}pfu) (fig. 7a). The kinetic for the down-regulation of the IL-7 receptor (CD127) - a marker previously demonstrated to negatively correlate with antigen load[58] - on CD8$^+$ T cells was comparable (fig. 7b). CD8$^+$ T cell recognizing the LCMV epitopes gp33 (gp$_{33-41}$) and gp276 (gp$_{276-284}$) tested by multimer staining, were found in similar abundance in the spleens of RAG$^{+/+}$, RAG1$^{mut/+}$ and RAG2$^{mut/mut}$ mice at day14/17 after LCMV infection (fig. 7c, suppl. fig. S5a). CD8$^+$ T cells nutrient uptake - as an indirect correlate for metabolic activity - was similar in mice of all three genotypes (fig. 7d). Restimulation of CD8$^+$ and CD4$^+$ T cells with LCMV derived peptides revealed a comparable capacity for IFNγ and TNFα production (suppl. fig. S5b). Immunopathology especially hepatitis is often mediated by cytotoxic CD8$^+$ T cells in non-cytopathic virus infections[59,60]. Levels of liver alanine transaminase (ALT) and aspartate transaminase (AST) were reduced in RAG2$^{mut/mut}$ mice 14 to 17 days after LCMV infection compared to *Rag* wild-type animals (fig. 7e, suppl. fig. S5c).

The relative abundance of CD4$^+$ follicular helper T cells (ICOS$^+$bcl6$^+$) was similar in RAG$^{+/+}$, RAG1$^{mut/+}$ and RAG2$^{mut/mut}$ mice (fig. 7f). The distribution of B cell subsets (follicular FollB CD21/CD35^{int}CD23^{high}B220$^+$CD19$^+$, marginal-zone MZB CD21/CD35^{high}CD23$^{low/-}$B220$^+$CD19$^+$ and germinal center B cells GC GL-7$^+$CD95$^+$IgM$^-$IgD$^-$B220$^+$CD19$^+$) was comparable in mice of all three genotypes following LCMV-infection (fig. 7g, suppl. fig. S5d). B cell functionality was preserved, as similar kinetics of LCMV-induced hypergammaglobulinemia were observed (fig. 7h). Besides, levels of LCMV specific GP1 IgG serum titers (fig. 7i) and neutralizing antibody titers were comparable in mice of all three genotypes (fig. 7j).

LCMV viral load 14 days after infection was comparable in all three genotypes, as tested in the spleen and different non-lymphoid organs (fig. 7k). LCMV titers in the blood decreased with similar kinetics in *Rag* wild-types, RAG1$^{mut/+}$ and RAG2$^{mut/mut}$ mice (suppl. fig. S5e). All animals had cleared the virus from their spleens by day 35 post-infection (fig. 7k).
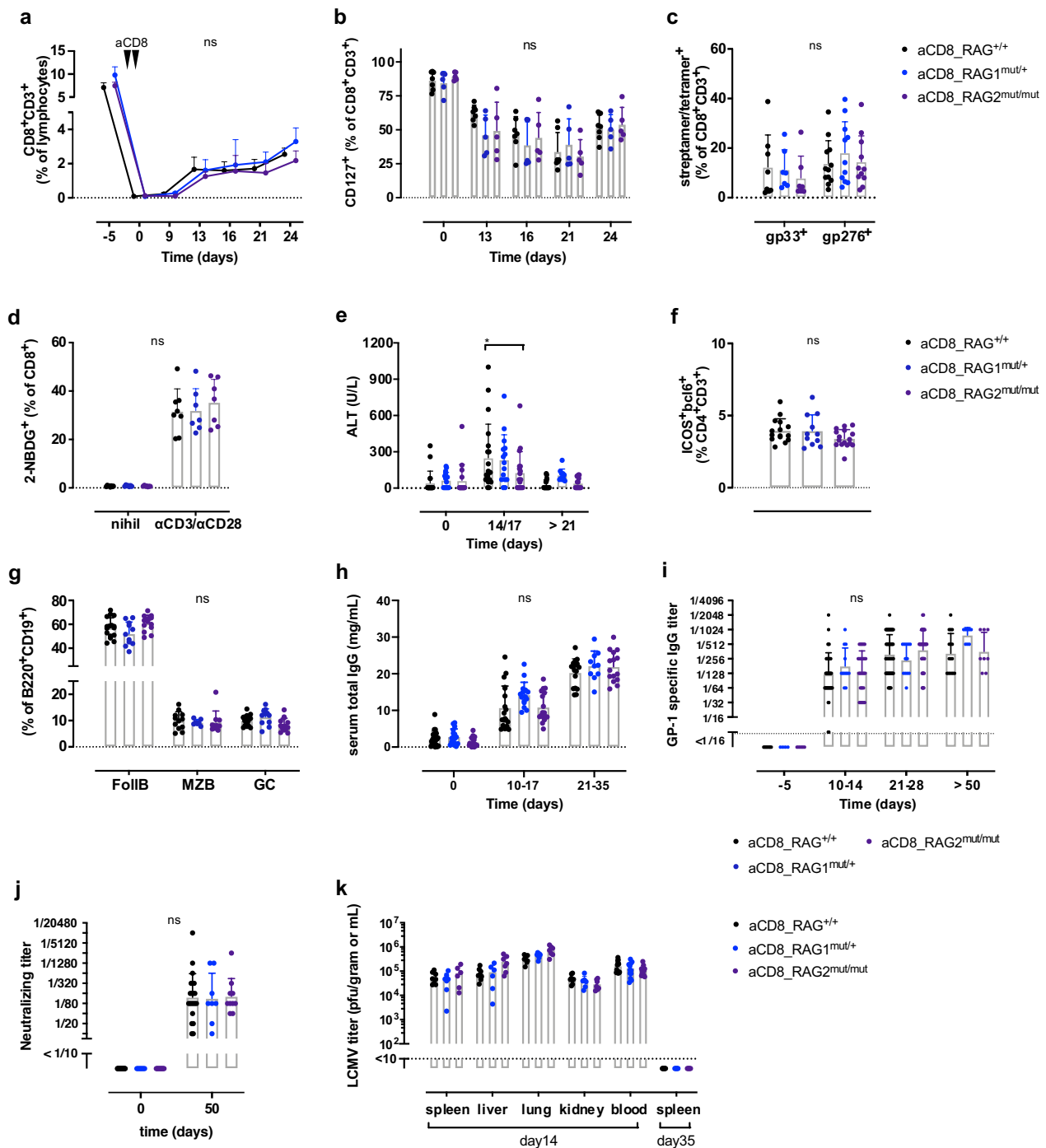
Figure 7: Robust T and B cell response to systemic LCMV infection in RAG2$^{mut/mut}$ und RAG1$^{mut/+}$ mice after transient CD8$^+$ T cell depletion challenge. RAG$^{+/+}$, RAG1$^{mut/+}$ and RAG2$^{mut/mut}$ mice were transiently depleted of cytotoxic CD8$^+$ T cells with monoclonal antibodies at day-4 and -1 (12.5 μg/g body weight, black triangles (a) before intravenous infection with a-h/k) LCMV$_{WE}$ (2x 10^{6}pfu) or i/j low dose LCMV$_{WE}$ (200pfu). **a)** Reemerging CD8$^+$ T cells were monitored in the peripheral blood at indicated time points and the frequency was determined using flow cytometry. **b)** Quantitative assessment of the IL-7 receptor (CD127) surface expression on CD8$^+$ T cells in the peripheral blood at indicated time points after infection. **c)** Reemerging CD8$^+$ T cells in the spleen were tested for their LCMV specificity via multimer staining using two of the immunodominant epitopes (gp$_{33-41}$, gp$_{276-284}$). The analysis was performed 14/17 days after infection. **d)** Two weeks after LCMV infection, reemerging CD8$^+$ T cells were tested for their ability for nutrient uptake measured after anti-CD3/anti-CD28 (αCD3/αCD28) stimulation. Measurement was performed using flow cytometry. **e)** Hepatic immunopathology was measured by assessing the serum liver alanine-aminotransaminase (ALT) activity in the serum after LCMV infection at indicated time points. **f)** The proportional abundance of splenic follicular helper CD4$^+$ T cells (ICOS$^+$bcl6$^+$) was evaluated 14/17 days post LCMV infection. **g)** B cell subsets were tested for the relative amount of follicular (FollB CD21/CD35^{int}

CD23[high]B220[+]CD19[+]), marginal-zone (MZB CD21/CD35[high]CD23[low/-]B220[+]CD19[+]) and germinal center B cells (GC GL-7[+]CD95[+]IgM[-]IgD[-]B220[+]CD19[+]), 14 to 17 days after LCMV infection. Serum levels of h) total IgG and i) GP1 binding IgG serum titers were evaluated by enzyme-linked immunosorbent assay (ELISA) prior to and after LCMV infection, at indicated time points. j) Neutralizing antibodies were measured with an immunofocus reducing assay, 50 days after LCMV infection. k) LCMV viral titers were measured in different organs and the blood, two weeks after LCMV infection and after 35 days in the spleen. Merged data from two to five independent experiments are displayed except for (d), each using three to eight mice per experiment and genotype. d) one experiment was performed with seven to eight mice per genotype. Statistical testing used a 2wayANOVA with multiple comparison correction for (a/b/d/e/h/i/j). For statistical testing of (f) a non-parametric Kruskal Wallis was used, and with an additional post-hoc correction for (c/g/k).

Reduced numbers and affinity-impaired LCMV specific CD8[+] T cells in double heterozygous *Rag* mutant mice

As we observed a heightened disturbance in thymic development in the double heterozygous *Rag* mutant *vs.* RAG[mut/+] mice, we next experimentally assessed the *in vivo* immune response against LCMV.

RAG1[mut/+]RAG2[mut/+] mice were infected with high dose LCMV$_{WE}$ (2x 10^6 pfu), which induced comparable levels and kinetics of systemic innate type I interferons release (suppl. fig. S6a). Kinetic analysis of the emergence of the immunodominant LCMV-glycoprotein (gp$_{33-41}$) specific CD8[+] T cells in the blood demonstrated a reduced frequency in double heterozygous *Rag* mutant mice compared to RAG[+/+] mice, persisting beyond day 20 post-infection (fig. 8a). Ten days after LCMV infection, at the peak of the CTL response, we enumerated significantly less splenic CD8[+] T cells in double heterozygous *Rag* mutant mice (fig. 8b). It was previously described that at the peak of the CD8[+] T cell immune response against LCMV virtually, all splenic CTLs are LCMV specific[61]. In keeping, we found a significantly reduced proportion of CD8[+] T cells recognizing the immunodominant peptides gp$_{33-41}$ and np$_{396-404}$, but not gp$_{276-284}$ in the spleens of double heterozygous *Rag* mutant mice (fig. 8c, 8d). The absolute number of all three LCMV-peptide-binding CD8[+] T cells were significantly reduced in double heterozygous *Rag* mutant mice (suppl. fig. S6d). This reduction of absolute numbers of LCMV-specific CD8[+] T cells in RAG1[mut/+]RAG2[mut/+] mice was independent of age and gender (suppl. fig. S6e, S6f, S6g). We observed no difference in the expression of various activation markers and inhibitory molecules on gp$_{33-41}$ binding CD8[+] T cells in RAG[+/+] compared to RAG1[mut/+]RAG2[mut/+] (suppl. fig. S6h).

A semi-quantitative estimation of TCR affinity can be achieved by tetramer titration and calculating the median fluorescent intensity (MFI) of the tetramer-binding T cell population[62,63]. A significantly lower MFI for the gp$_{33-41}$ tetramer binding (fig. 8e) as well as gp$_{276-284}$ tetramer binding (fig. 8f) CD8[+] T cells was documented in double heterozygous *Rag* mutant mice compared to RAG[+/+] mice for serval tetramer concentrations tested. To assess T cell functionality, splenic CD8[+] T cells isolated ten days after LCMV infection were restimulated *ex vivo* with titrated concentrations of LCMV-glycoprotein derived gp$_{33-41}$ peptide and intracellular cytokine production was measured. A significantly higher frequency of IFNγ and a trend for higher TNFα production was measured within the LCMV specific gp33-tetramer binding CD8[+] T cells of RAG1[mut/+]RAG2[mut/+] mice at the highest gp$_{33-41}$ peptide concentration tested (fig. 8g, suppl. fig. S6i).

To address viral-specific CD8[+] T cell cytotoxicity *in vivo*, we measured LCMV titers in different lymphoid and non-lymphoid organs. LCMV was still detectable in various organs 20 days after infection, with comparable viral titers in *Rag* wild-type vs. *Rag* double heterozygous mice (suppl. fig. S6j). Ten days after LCMV infection, we measured similar virus titers in different tissues tested in *Rag* wild-type vs. RAG1[mut/+]RAG2[mut/+] mice (suppl. fig. S6k), albeit a non-significant trend for higher splenic viral load in the latter was observed (fig. 8h).

As seen after infection with human hepatitis virus [64,65], delayed LCMV control in the liver is associated with increased liver damage and paralleled by augmented serum levels of liver enzymes, illustrating CD8[+] T cell-mediated immunopathology[66]. Hepatosplenomegaly was similar in Rag wild-type and double heterozygous *Rag* mutant mice ten days after infection (suppl. fig. S6l). Serum levels of aspartate-aminotransferase (AST) and alanine-aminotransferase (ALT) activity peaked at day ten after LCMV infection. Subgroup analysis demonstrated that three to four weeks old RAG1[mut/+]RAG2[mut/+] mice had a significantly higher peak of ALT and slightly higher AST levels compared to age-matched RAG[+/+] mice (fig. 8i). Higher serum liver enzyme levels in

double heterozygous *Rag* mutant mice were paralleled by a significantly higher frequency and a trend to higher absolute numbers of hepatic CD8$^+$ effector memory T cells (Tem CD69$^-$CD44$^+$CD62L$^-$TCRβiv^{neg}) compared to *Rag* wild-type mice (fig. 8j, 8k). While, the number of tissue resident CD8$^+$ T cells (Trm CD69$^+$CD44$^+$CD62L$^-$TCRβiv^{neg}) was comparable between the two genotypes tested.

To address whether the reduced numbers of LCMV specific CD8$^+$ T cells were a consequence of a suboptimal *in vivo* milieu for T cell proliferation in double heterozygous *Rag* mutant mice, we adoptively transferred naïve LCMV-specific CD8$^+$ T cells derived from 327^{tg} mice[67–69] (TCR-transgenic P14 CD8$^+$ T cells specific for LCMV-gp33/H-2D^b) into RAG$^{+/+}$ *vs*. RAG1$^{mut/+}$RAG2$^{mut/+}$ mice prior to LCMV infection. Following LCMV infection we measured a similar absolute and relative abundance of the LCMV-TCR transgenic CD8$^+$ T cells in *Rag* wild-type mice and RAG1$^{mut/+}$RAG2$^{mut/+}$ mice, (fig. 8l, suppl. fig. S6m). In addition, program cell death 1 (PD-1) surface expression on adoptively transferred LCMV-TCR transgenic CD8$^+$ T cells was similar in mice of both recipient genotypes (fig. 8m).

In summary, a reduction of the LCMV-specific CD8$^+$ T cell response was observed in double heterozygous *Rag* mutant mice, that was paralleled by lower CD8$^+$ T cell TCR affinity and higher immune–pathology at the peak of the CTL response.
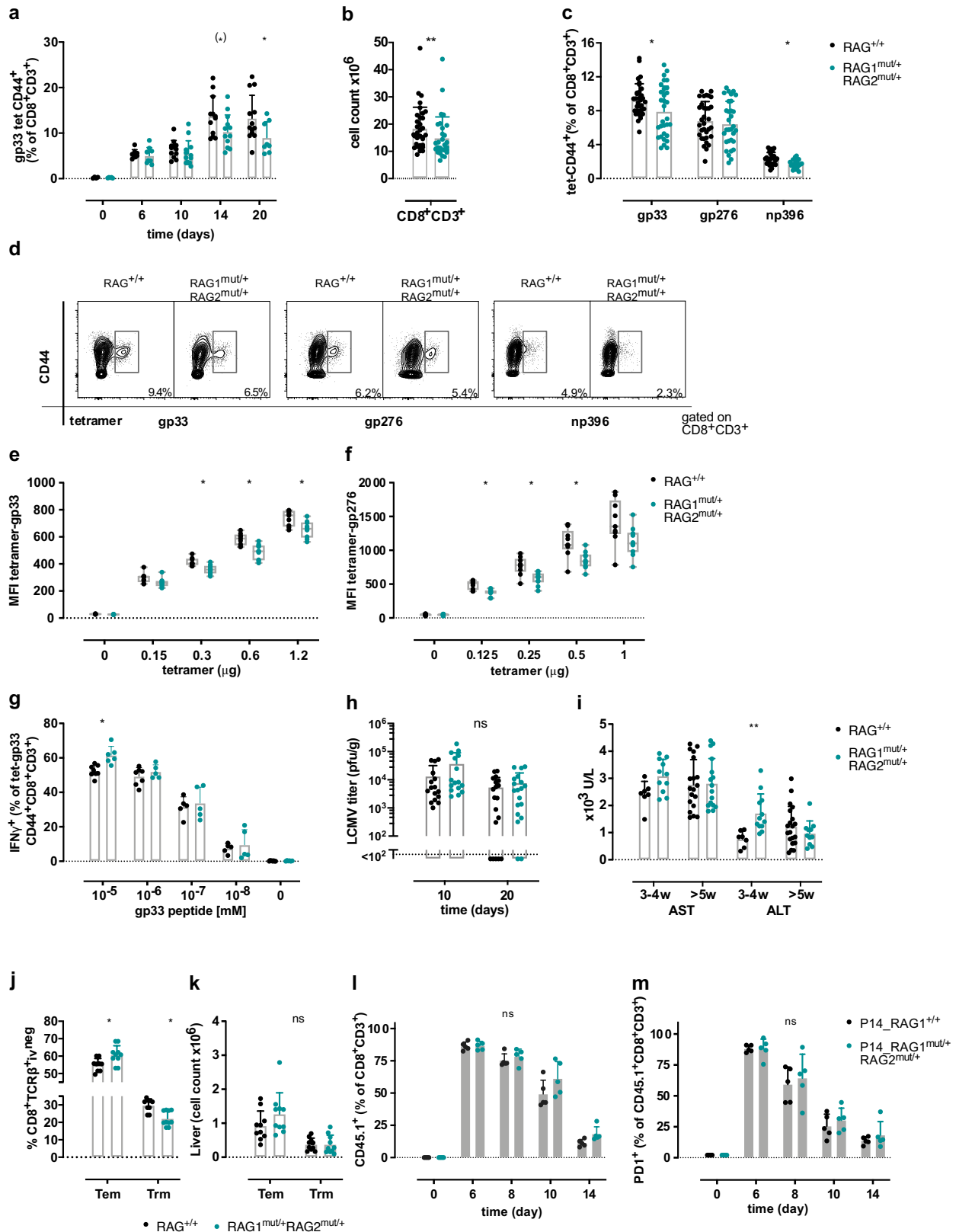
Figure 8: RAG1$^{mut/+}$RAG2$^{mut/+}$ mice have fewer virus-specific CD8$^+$ T cells with reduced TCR affinity after LCMV infection. (a-m) *Rag* wild-type or double heterozygous *Rag* mutant mice were infected intravenously with high dose LCMV$_{WE}$ (2×10^6 pfu). **a)** Time course analysis of LCMV-specific gp$_{33-41}$ binding CD8$^+$ T cell frequency found in the peripheral blood at indicated days after LCMV infection. **b)** Quantitative enumeration of splenic CD8$^+$ T lymphocyte ten days after LCMV infection. **c)** Splenic LCMV-specific CD8$^+$ T cells binding the gp$_{33-41}$ (gp33), gp$_{276-284}$ (gp276) and np$_{396-404}$ (np396) epitopes were identified

by tetramer staining and the frequency was enumerated at day ten post LCMV infection. **d)** Contour plots are displaying representative flow cytometric analysis of splenic gp33, gp276 and np396 tetramer-binding CD44$^+$ CD8$^+$ T cells, ten days after LCMV-infection. Numbers indicate frequency within the CD8$^+$CD3$^+$ T cell population. The TCR affinity for the **e)** tetramer-gp33 and **f)** tetramer-gp276 binding CD8$^+$ T cells was estimated ten days post LCMV-infection via tetramer-titration and calculation of the median fluorescence intensity (MFI). **g)** Splenic CD8$^+$ T cells were isolated and *ex vivo* re-stimulated with the immunodominant LCMV derived peptide gp$_{33-41}$ and intracellular IFNγ production was measured by flow cytometry ten days after LCMV-infection. **h)** LCMV titers were enumerated in the spleen ten *vs.* 20 days post LCMV infection by plaque-forming assay. **i)** Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) activity levels were measured in the serum ten days after LCMV-infection. The measurements were grouped regarding the age of mice at infection (3-4w 3-4 weeks old *vs.* >5w >5 weeks old). The hepatic lymphocyte infiltration was investigated ten days after LCMV infection for **j)** the frequency and k) the absolute cell count of effector memory (Tem CD69$^-$CD44$^+$CD62L$^-$TCRβ^{iv-neg}) and tissue-resident (Trm CD69$^+$CD44$^+$CD62L$^-$ TCRβ^{iv-neg}) CD8$^+$ T cells by flow cytometry. (l – m) Naïve LCMV-specific TCR-transgenic CD8$^+$ T cells recognizing the LCMV gp$_{33-41}$ epitope (P14 cells 1x10^6, CD45.1) were adoptively transferred into CD45.2 RAG$^{+/+}$ or RAG1$^{mut/+}$RAG2$^{mut/+}$ recipient mice one day prior to LCMV-infection. **l)** Frequency of LCMV-TCR transgenic CD8$^+$ T cell expansion in the peripheral blood was assessed at the indicated timepoints by gating on CD45.1$^+$ cells. **m)** Surface expression of PD-1 on peripheral blood-derived LCMV TCR-transgenic CD45.1$^+$ CD8$^+$ T cells at the indicated time points post LCMV infection. Merged data from two to six independent experiments are displayed except for (l-m) where one experiment was performed. Three to eight mice were used per experiment and genotype. Statistical testing employed a non-parametric Mann-Whitney in b, with additional post hoc correction for (c/h/j/k). A 2way ANOVA with Šídák multiple comparison correction was computed for (a/e/f/g/i/j/l/m), (*) adj.p 0.054.

Higher neutralizing antibody titers and higher IgG autoantibodies in double heterozygous *Rag* mutant mice following LCMV infection

Next, we analyzed LCMV-specific CD4$^+$ T cells in *Rag* wild-type vs. double heterozygous *Rag* mutant mice. This was assessed by isolating splenic CD4$^+$ T cells ten days after LCMV infection and *ex vivo* re-stimulation with the LCMV-glycoprotein derived immunodominant MHC class II-restricted gp$_{61-80}$ peptide. Intracellular production of IFNγ, TNFα and IL-2 was comparable in RAG$^{+/+}$ and double heterozygous *Rag* mutant mice (fig. 9a). Ten days following LCMV infection the proportion and numbers of CD4$^+$ helper T cells belonging to the T$_{H1}$ (SLAM$^+$CXCR5$^+$CD44$^+$) subset were comparable in RAG1$^{mut/+}$RAG2$^{mut/+}$ mice *vs.* wild-type RAG$^{+/+}$ animals (fig. 9b). Similarly, we observed that CD4$^+$ follicular T helper cells (T$_{FH}$ CXCR5$^+$PD1$^+$) were present in the spleen with a comparable frequency and absolute numbers in RAG$^{+/+}$ mice *vs.* RAG1$^{mut/+}$RAG2$^{mut/+}$ mice (fig. 9b).

Ten days after LCMV infection we observed similar numbers and frequency of follicular (follB CD21/CD35^{int}CD23^{high}B220$^+$CD19$^+$) and memory (CD38$^+$CD95^{var}IgM$^-$IgD$^-$B220$^+$CD19$^+$) B cells in *Rag* wild-type *vs.* double heterozygous RAG1$^{mut/+}$RAG2$^{mut/+}$ mutant mice (fig. 9c, 9f). We measured a non-significant trend for higher marginal zone B cell frequency and numbers (MZB CD21/CD35^{high}CD23$^{low/-}$B220$^+$CD19$^+$) in the RAG1$^{mut/+}$RAG2$^{mut/+}$ mice compared to RAG$^{+/+}$ animals (fig. 9c, 9d, 9f). Inversely, germinal center (GC CD95$^+$GL-7$^+$IgM$^-$IgD$^-$B220$^+$CD19$^+$) B cells in double heterozygous *Rag* mutated mice were found in slightly reduced absolute and relative numbers (fig. 9c, 9e, 9f).

LCMV is known to induce hypergammaglobulinemia[70]. Indeed, total serum IgG levels were more than 15-fold increased compared to the naïve state in both RAG$^{+/+}$ mice and RAG1$^{mut/+}$RAG2$^{mut/+}$ mice at day 14 post LCMV infection (fig. 9g). This increase of total IgG was mainly attributable to increased serum concentrations of IgG1 and IgG2a, which were measured in similar proportions in *Rag* wild-type and double heterozygous *Rag* mutant mice (fig. 9h).

IgG levels binding the LCMV glycoprotein GP1, tested by ELISA with recombinant GP1 protein, were significantly lower in the RAG1$^{mut/+}$RAG2$^{mut/+}$ mice compared to the wild-type RAG$^{+/+}$ mice (fig. 9i) from day 14 post-infection onwards. However, significantly higher titers of LCMV neutralizing antibodies were measured in RAG1$^{mut/+}$RAG2$^{mut/+}$ mice compared to RAG$^{+/+}$ mice, ten weeks after LCMV infection (fig. 9j, 9k).

It has been described previously that patients with homozygous or compound heterozygous hypo-morphic *RAG* mutations as well as RAG1$^{mut/mut}$ mice, displayed augmented levels of autoantibodies[23,26], implying reduced immune tolerance. Besides, there is growing evidence that virus infections are linked with the

emergence of autoimmunity[71–73]. We thus aimed to test whether LCMV infection might favor the production of autoantibodies in RAG[+/+] *vs*. double heterozygous *Rag* mutant mice. Sera were probed in an unbiased proteomic approach against 115 autoantigens coated on a microchip. Sera from naïve mice of both genotypes did not display a significant difference regarding IgG autoantibody abundance (suppl. fig. S7a). In contrast, two weeks following LCMV infection, sera of RAG1[mut/+]RAG2[mut/+] displayed a higher average change of the IgG autoantibody score compared to RAG[+/+] mice (fig. 9l, suppl. fig. S7b). Following LCMV infection, nineteen different anti-self IgGs specificities displayed a 1.5-fold higher difference (compared to pre-infection sera) in the RAG1[mut/+]RAG2[mut/+] mice compared to *Rag* wild-type mice (fig. 9m). Among those, we found IgGs directed against GP210 (glycoprotein of the nuclear envelope) and ssDNA (suppl. fig. S7c). In keeping, we calculated a higher number of autoantibodies with a median antibody score difference of ≥4 in double heterozygous *Rag* mice than in RAG[+/+] (fig. 9n). By analyzing the ten cellular categories which are represented on the microchip, while focusing on IgGs with an antibody score difference of ≥ 4, we found a higher proportion of IgGs directed against circulating proteins, cell and extracellular matrix proteins (ECM) in the double heterozygous *Rag* mutant mice compared to *Rag* wild-type mice (fig. 9o).
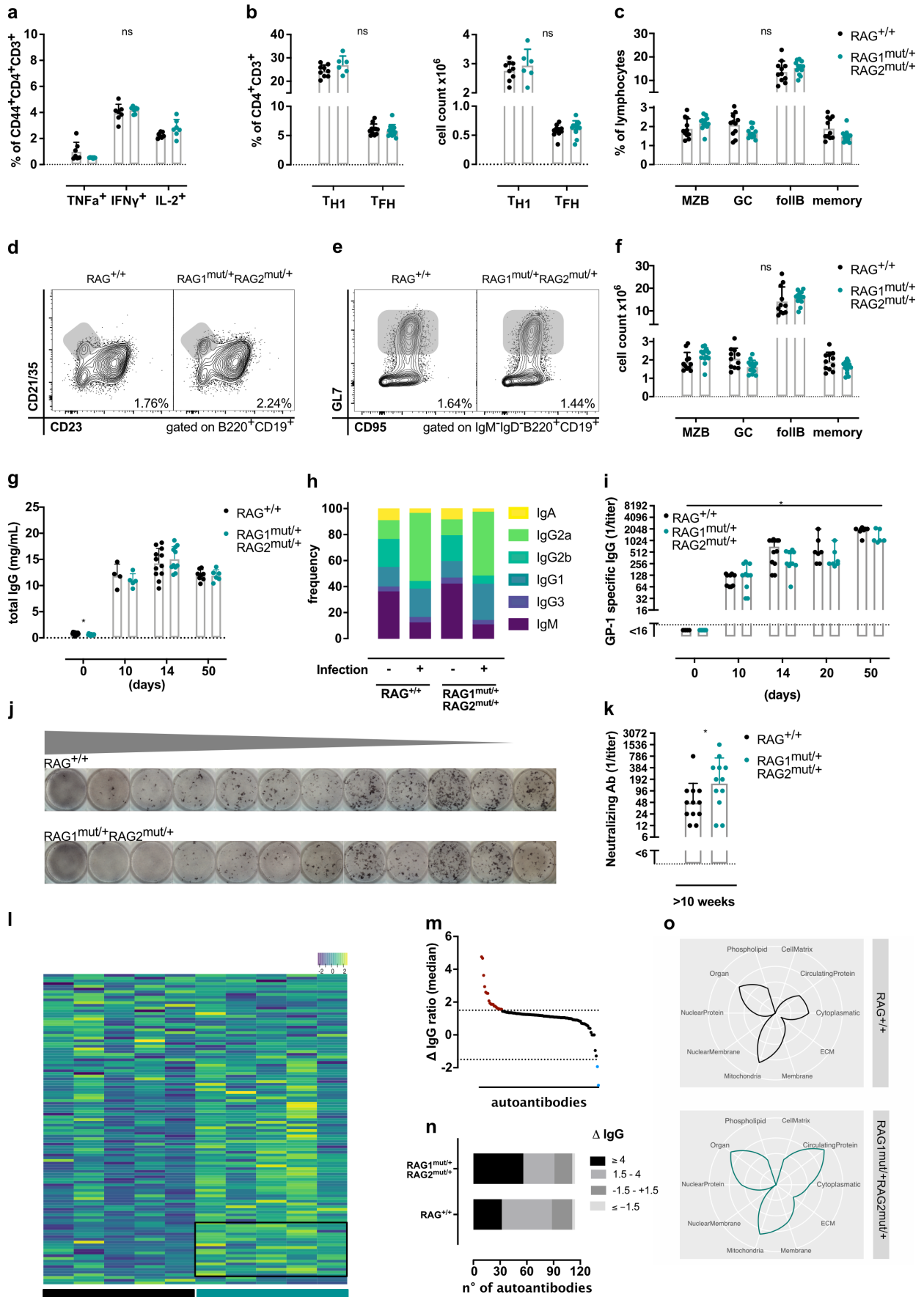
Figure 9: Dysregulated LCMV specific B cell response and a greater increase of autoantibodies in double heterozygous *Rag* mutant mice after infection. (a-j) *Rag* wild-type or double heterozygous *Rag* mutant mice were infected intravenously with high dose LCMV$_{WE}$ (2 x10^{6}pfu). **a)** Quantitative analysis of the LCMV-specific CD4$^+$ T cell response. Ten days after LCMV infection, we isolated CD4$^+$ helper T cells and re-stimulated those, *ex vivo* with the LCMV immunodominant MHC class II-restricted peptide gp$_{61-84}$ and consecutively the TNFα, IFNγ and IL-2 production was measured using flow cytometry. **b)** Qualitative and quantitative analysis of splenic CD4$^+$ T cells for the presence of T$_{H1}$ CD4$^+$ T cells (SLAM$^+$CXCR5$^+$CD44$^+$) and T follicular helper CD4$^+$ T cells (T$_{FH}$ CXCR5$^+$PD-1$^+$) ten days post LCMV infection. Quantification was performed with flow cytometry. **c)** Investigation of splenic B cells ten days post LCMV infection, for the following subsets: marginal-zone (MZB CD21/CD35^{high}CD23$^{low/-}$B220$^+$CD19$^+$), germinal center (GC CD95$^+$GL-7$^+$IgM$^-$IgD$^-$B220$^+$CD19$^+$), follicular (follB CD21/CD35^{int} CD23^{high}B220$^+$CD19$^+$) and memory (CD38$^+$CD95^{var}IgM$^-$IgD$^-$B220$^+$CD19$^+$) B cells. Flow cytometric analysis for **(d)** MZB cells and **(e)** GC B cells are depicted with representative contour plots, measured in the spleen ten days post LCMV infection. The percentage next to the grey gate indicates frequency within the lymphocyte population. **f)** Enumeration of MZB, GC, follB and memory B cells day ten after LCMV infection. **g)** Total serum IgG was measured at the specified time points after LCMV infection. **h)** Before and two weeks post LCMV-infection the immunoglobulin isotypes present in the serum were analyzed. Depicted is the proportional occurrence for each immunoglobulin isotype within the total measured immunoglobulin pool. **i)** LCMV specific IgG - binding the GP-1 protein - were measured at indicated time points in the sera of mice post LCMV infection. LCMV neutralizing antibody titers were measured by an immunofocus reduction assay using LCMV$_{WE}$ as a test substrate. **j)** Representative immunofocus reduction assay, with log$_2$ serum dilution from left to right is shown. **k)** Neutralizing antibody titers were measured ten weeks post LCMV infection in the indicated mice. **l)** Autoantigen-specific IgGs were assessed using an autoantibody microarray coated with 115 defined self-antigens and probing sera obtained from the indicated mice before and two weeks after LCMV infection. For five individual LCMV infected RAG$^{+/+}$ mice (black label at the bottom) and five LCMV infected RAG1$^{mut/+}$RAG2$^{mut/+}$ mice (turquoise label at the bottom) a heatmap was computed indicating IgG autoantibody binding difference between post compared to pre-infection and then normalized to the average of wild-type delta values. Black square indicates the IgG autoantibodies which were 1.5-fold higher in the sera of RAG1$^{mut/+}$RAG2$^{mut/+}$. Rows indicate self-antigens, columns indicate individual mice. **m)** For each auto-antigen IgG specificity, the ratio for RAG1$^{mut/+}$RAG2$^{mut/+}$ to RAG$^{+/+}$ were calculated (using the median difference between post to prior LCMV infection difference). Red points indicate autoantigens recognized 1.5-fold more often by RAG1$^{mut/+}$RAG2$^{mut/+}$ IgGs compared to RAG$^{+/+}$ IgGs, blue points indicate the opposite. **n)** IgG autoantibodies were classified according to the level of the median post to prior LCMV infection change (Δ IgG). **o)** Autoantigens were sub-grouped into ten categories, according to the cellular compartment to which those belonged. IgGs with the highest post to prior LCMV infection difference (alias group Δ IgG ≥ 4, in panel n) were grouped according to the cellular compartments those would recognize. Radar plot, circular lines indicate the intervals of 20%, 40%, 60%, and 80%. Merged data from two independent experiments are displayed, except for the T$_{H1}$ analysis and the autoantibody microarray (one experiment was performed), four to eight mice per experiment and genotype were used. Statistical significance was inferred using the non-parametric Mann-Whitney test for (j) and an additional post hoc correction was used for (a/b/c/f). A 2wayANOVA with a Šídák test was used for significance analysis in (g + i).

Discussion

Reduced but residual RAG activity may underly adult-onset combined immunodeficiency with a pronounced immune dysregulation phenotype (granulomas and/or autoimmunity, CID-G/AI)[11,26]. It is of pivotal importance to correctly diagnose CID-G/AI patients, as such patients have reduced overall survival[26]. Currently, only homozygous or compound heterozygous *RAG* mutations are generally considered to cause clinically relevant proteins loss of function. Previous studies have correlated the clinical phenotype with the genetically determined reduction of RAG enzymatic activity [11,16]. Heterozygous *RAG* missense mutations are estimated to be found in as many as 1/425 individuals[14,15], however their impact on the enzymatic *in vivo* RAG activity is unclear. We aimed at studying the effect of a single or double heterozygous hypomorphic *RAG* mutation(s) on the lymphocyte development, and the consequences on the pathogen-specific *vs.* auto-reactive adaptive immune response.

Three patients – the index patient, his mother and his daughter - with double heterozygous hypo-morphic *RAG1/2* mutations presented with different degrees of B and T cell anomalies but preserved NK cells – a typical immunological constellation for *RAG* deficiencies[19]. All three harbored specific immunoglobulin subclass deficiencies and an increased frequency of transitional B cells. Both features were observed in the

double heterozygous *Rag1/Rag2* mutant mice. Analyzing the naïve B cell receptor *IGH* repertoire for the *J* gene usage, we found a preference for *IGHJ4* over *IGHJ6* in the index patient and the mother. This is somewhat surprising, as the reduced *IGHJ6* fraction has been documented before in memory B cells[74]. The index patient and his mother displayed low Vα7.2$^+$ CD3$^+$ T cells, a diagnostic indication pointing at impaired V(D)J recombination[75]. While the index patient showed severe defects in both the B and T cell compartment, his mother demonstrated important alterations within the T cell subsets, while the B cell compartment was more preserved. The contrary was observed in his daughter which displayed subtle alterations in the B cell compartment, while the T cell subpopulations were well conserved. Studying murine models with hypomorphic *Rag* mutations we noted that the B and T lymphocyte development were not equally affected. Selective lineage deficiencies have been reported in *RAG* mutated patients, which displayed hyper-IgM syndrome[76], IgA deficiency[77], common variable immunodeficiency (CVID)[78,79] or idiopathic CD4$^+$ T cell lymphopenia[80,81]. Within the T cell compartment, we observed in the two older patients an increased CD4$^+$/CD8$^+$ T cell ratio in the peripheral blood, which was found to be in the normal range for the daughter. Similarly, our mouse model with double heterozygous *Rag* mutations showed a higher number of splenic helper CD4$^+$ T cells.

Both augmented numbers of peripheral CD4$^+$ T cells and B cells present in the spleen of naïve RAG1$^{mut/+}$RAG2$^{mut/+}$ mice could be attributed to enhanced homeostatic lymphoproliferation indicated by a trend for higher ki67 expression and in keeping the reduced SP CD4$^+$ thymocyte numbers. The index patient and the RAG1$^{mut/+}$RAG2$^{mut/+}$ mice shared the phenotype of an immature B cell compartment, accompanied by a reduction for class-switched immunoglobulins and expansion of the CD4$^+$ T cell subset. In summary, the RAG1$^{mut/+}$RAG2$^{mut/+}$ mouse model might constitute a suitable model to recapitulate the phenotype of the index patient.

Reducing RAG activity by 40% as seen in the double heterozygous *Rag* mutant mice was sufficient to observe significant, albeit a leaky blockage in thymocyte development, higher Trb repertoire clonality and a dysregulated adaptive immune response following acute LCMV infection. Previous work has shown that human patients suffering from a severe form of RAG deficiency (Omenn syndrome) with minimal residual RAG activity showed a very restricted *TRB* repertoire[82,83]. In contrast, patients presenting with CID-G/AI generally displayed a preserved TCR repertoire, with only mild changes in the gene segment usage[12,83,84]. Primary lymphoid tissue is rarely accessible in a human patient, making it difficult to dissect within the TCR repertoire analysis which alterations are due to a priori V(D)J rearrangement defects or the consequence of qualitative tolerance alterations. Thus, we analyzed the primary Trb repertoire in mouse thymocytes, to study the impact of the heterozygous Rag1/Rag2 mutations on the naïve repertoire. The Trb repertoire in RAG1$^{mut/+}$RAG2$^{mut/+}$ thymocytes displayed a significant preference for productive rearrangements. The observation of an augmented productive frequency in the context of *RAG* deficiency was found earlier in human patients and murine models with biallelic *RAG* mutations[30,32,85]. During wild-type RAG activity, V(D)J recombination is initiated on one allele, and in case of failure the second allele is rearranged[86]. Accordingly, the wild-type repertoire harbors a proportion of non-productive out-of-frame sequences together with productive in-frame sequences. We speculate that in thymocytes of double heterozygous *Rag* mutant mice only one allele was recombined due to RAG insufficiency while the second allele remained in germ-line configuration. Failure of a productive rearrangement on the first allele without the ability to rearrange the second allele will lead to apoptosis. This might explain the higher clonality and the augmented productive frequency in the *Trb* repertoire of double heterozygous *Rag* mutant mice.

We noted a reduced CD8$^+$ T cell expansion after LCMV infection in double heterozygous *Rag* mutant mice. In the spleen, CD8$^+$ T cell numbers had only increased 1.14 fold in RAG1$^{mut/+}$RAG2$^{mut/+}$ mice but 1.64 fold in RAG$^{+/+}$ mice ten days after infection. We also noted decreased numbers and frequency of CD8$^+$ T cells recognizing the LCMV-immunodominant epitopes, paralleled by a trend of higher splenic viral titers. A simple explanation of the observed reduced CD8$^+$ T cell expansion following LCMV infection is a lower precursor

frequency of LCMV-specific CD8[+] T cells in RAG1[mut/+]RAG2[mut/+] mice. In chronic polyomavirus and LCMV infection, the pool of CD8[+] T cells has to be replenished by newly emigrating T cells from the thymus[87,88]. The reduction of CD8[+] T cells in RAG1[mut/+]RAG2[mut/+] mice during LCMV infection might thus hint at an accentuated CD8[+] T cell production deficiency due to the destruction of the thymic architecture and/or suboptimal rearrangement[88]. The detected LCMV gp33- and gp276-specific CD8[+] T cells moreover demonstrated a reduced tetramer-binding affinity and the tetramer-gp33[+] CD8[+] T cells produced more IFNγ after *in vitro* restimulation with LCMV-derived peptide. IFNγ and TNFα are effector molecules secreted after TCR engagement[89]. Elevated cytokine production might be, as previously described, caused by an altered threshold of antigen-induced exhaustion of virus-specific T cells[90,91] or a reduced T cell suppressive environment in RAG1[mut/+]RAG2[mut/+] mice[92]. An additional putative explanation is that virus-specific CD8[+] T cells are provisioned by IL-21 derived from CD4[+] T cells, a T cell help that might be reduced in RAG1[mut/+]RAG2[mut/+] mice. This IL-21 dependent T cell help has been shown to be crucial in chronic LCMV infection and melanoma to sustain antigen-specific CD8[+] T cells[93]. The slight trend for reduced viral control in RAG1[mut/+]RAG2[mut/+] was paralleled by augmented hepatic immunopathology, indicated by a higher frequency of effector CD8[+] T cells infiltrating the liver tissue and higher ALT serum activity levels. Previous work by Lang et al. has established that reduced expansion kinetics of cytotoxic CD8[+] T cells was linked with an increased liver immunopathology[66]. By comparing the *V* gene utilization of the *Trb* repertoire found in the RAG1[mut/+]RAG2[mut/+] mice with elegantly performed recent work focusing on the *Trb* of memory gp33[+] specific CD8[+] T cells after acute LCMV infection[94], we found that the top three *Trb V* genes (TRBV29, TRBV14, TRBV1) used in gp33-specific CD8[+] T cells were found at similar frequency in the double heterozygous *Rag* mutated SP CD8[+] thymocytes as in RAG[+/+] thymocytes.

Several months after infection, a higher titer of neutralizing antibodies was measured in double heterozygous *Rag* mutant mice. This seems to be counter-intuitive, however higher neutralizing antibody titers could compensate for the reduced numbers of LCMV-specific CD8[+] T cells, which has been described by others before[56,95–97]. Higher neutralizing antibody titers have also been attributed to a reduction of CD8[+] T cell-mediated killing of infected neutralizing antigen-specific B cells[98,99]. Alternatively, the augmented neutralizing antibody titers measured in double heterozygous *Rag* mutant mice could hint at an altered CD4[+] T$_{FH}$ TCR repertoire. Tubo et al. and Keck et al. had elegantly demonstrated, that naïve CD4[+] T cells would not only differentiate into CD4[+] T$_{FH}$ cells based on the cytokine milieu[100] but are influenced by the affinity of their TCRs[101,102]. Future experiments might characterize the T$_{FH}$ repertoire and which immunoglobulin isotype dominates the higher neutralizing antibody titers found in RAG1[mut/+]RAG2[mut/+] mice. As the deletional recombination needed for isotype switch and the affinity maturation of the immunoglobulin variable domain via somatic hypermutation (SHM) are both RAG independent, the difference in the neutralizing titers might point at differences in the cytokine milieu[103] and increased *Igh* transcription[104], respectively.

Autoimmunity is a hallmark in patients with inborn errors of immunity (IEI). According to the French National Primary Immunodeficiency Registry (CEREDIH) and the UK National Institute for Health Research (NIHR) BioResource-Rare Diseases programme (NBR-RD) which together analyzed more than thousands IEI patients, every fourth IEI patient suffered from a least one autoimmune complication[105,106]. In keeping with the mathematical estimation that 1/425 individuals might carry at least one *RAG* missense mutation[14,15], we mined publicly available genome-wide association study (GWAS) metanalysis for the most common human autoimmune diseases[107–114]. We could not find a reproducible association of single or multi nucleotide polymorphism (SNPs or MNVs) in *RAG1/RAG2* and autoimmune diseases. The lacking association in multiple GWAS analyses between *RAG* variants and autoimmune diseases could indicate that only one mutated *RAG* allele is not sufficient to drive overt disease. This is in line with our experiments with RAG1[mut/+] mice which displayed no altered autoimmune response. RAG[+/+] and RAG1[mut/+]RAG2[mut/+] mice did not significantly differ in the amount of IgG autoantibodies in a naïve state. However, associations have been described in patients with hypomorphic *RAG* mutations between autoantibodies and a history of varicella-zoster virus (VZV), adenovirus

and cytomegaly virus (CMV) infection[115]. This was paralleled in our experiments, as we observe a significant post-infection augmentation in IgG autoantibodies in RAG1$^{mut/+}$RAG2$^{mut/+}$ mice. The break of tolerance during infection, could either be attributed to elevated IFNα levels[116,117], the pathogen *per se* acting as an adjuvant for the autoreactive immune response[118], molecular mimicry[119,120], bystander activation[121], epitope spreading[122] and/or a reduction of negative T cell selection as a consequence of thymic architecture destruction[88]. The measured IgG autoantibodies had a broad specificity in terms of recognition of cellular compartments. Interestingly, a broad spectrum of clinical autoimmune manifestation is observed in published patients with CID-G/AI (granulomatosis, myasthenia gravis, and different skin manifestation such as vitiligo, psoriasis, alopecia areata)[123], some of them were also present in the double heterozygous *RAG* mutated patients presented here.

Our study demonstrates for the first time that double heterozygous *RAG1/RAG2* mutations may cause altered lymphocyte development and T cell repertoire in mice and is replicated in human patients. In this mildly RAG insufficient state, the virus-specific CD8$^+$ T cell responses was reduced *in vivo* and infection-induced IgG auto-antibodies were augmented. It is the first report that suggests dominant inheritance of a *RAG*-dependent immunodeficiency in humans. Immune-dysregulations associated with heterozygous *RAG* mutations and subtle RAG insufficiency might be much more abundant as previously appreciated and may underlie severe infections, as seen nowadays during the COVID-19 pandemic.

Limitations

Only one family was studied, which does not allow to fully exclude other genetic or environmental factors as disease modifiers. Since the *RAG1* and *RAG2* of *H.sapiens* and *M.musculus*, show a > 80% sequence similarity, murine models represent a valuable tool to study *RAG* deficiencies. The here studied murine CID-G/AI RAG1$^{mut/+}$RAG2$^{mut/+}$ model did not harbor the exact orthologue *Rag* missense mutations identified in the three human patients. We used two existing mouse models with hypomorphic missense mutations in *Rag1* and *Rag2*, respectively.

It is delicate to translate findings regarding the different lymphocyte subset composition from the murine *Rag* mutated models to the human patients, the analysis focused on different immune compartments (primary/secondary lymphoid organs in the mice *vs.* peripheral blood in the human patients).

An important caveat is that the intrinsic affinity of the TCR to the peptide-MHC ligand cannot be directly measured through tetramers (which would require monovalent interaction studies on a single cell level with surface plasmon resonance). The tetramer-binding was not normalized to surface TCR expression. Further, the TCR signaling might be better indicated by the measurement of the t_a dwell time of the TCR-pMHC interaction.

We cannot definitely attribute the observed increase in the post-infection IgG autoantibody production to a break in central *vs*. peripheral tolerance. We did not clinically observe overt autoimmune disease in any of the murine *Rag* mutated models with age or following LCMV infection. Additional genetic and/or environmental factors might be required to induce overt autoimmune disease *in vivo*.

Author contributions

AJ and MR conceived the project. AJ planned and performed all the murine *in vivo* and *in vitro* experiments including their analysis, AJ was in charge of the mouse license administration and the genotyping, AJ performed the immunophenotyping/Sanger sequencing of the human patients, AJ wrote the manuscript and conceived all figures. OB constructed the molecular RAG1/RAG2 model and gave scientific input. MD contributed to data analysis and interpretation. RH/AG/AN performed the WES experiment and its analysis. IH performed the human Ig measurements. JEW provided the RAG2$^{mut/mut}$ mice. JL helped with patient care. LDN gave scientific input and provided mice. MR supervised the project, corrected the manuscript, organized and supervised the patient cohort as well as financial support.

Acknowledgment

Experimental Procedures

Ethics approval and human subjects

Following informed consent, the patients were enrolled in a prospective cohort of patients with primary inborn errors of immunity, the study was ethically approved (EKNZ 2015-187) according to Swiss law.

Blood samples and single-cell suspension from murine organs

Healthy blood donors as well as patient-derived peripheral blood mononuclear cells (PBMCs) were obtained after signed informed consent. Ficoll density gradient separation used Lymphoprep™ (density of 1.077g/mL, Axonlab) for PBMCs isolation from whole blood.

For murine single-cell suspension, we gently mashed spleens, lymph nodes, thymus, liver through 70µm cell strainers (Corning). For bone marrow cells, two femurs were flushed with staining media, erythrocytes contained in the bone marrow were lysed subsequently. For the isolation of peripheral blood lymphocytes, the containing erythrocytes were lysed with the ACK lysis buffer for 3-10 min at room temperature (RT) and then washed with staining media.

Liver single cell suspension were resuspended in 40% isotonic Percoll (Sigma), which was overlaid with 70% isotonic Percoll, centrifuged for 30min at 1800 rpm without acceleration/break.

Genetic investigation, variant calling, annotation, and filtering

The genetic sequencing was carried out following written consent. Genomic DNA was extracted from cultured T cell blasts or PBMCs using the QIAamp DNA Blood Mini Kit (Qiagen). Whole exome sequencing was carried out as described earlier[124]. In short, DNA was fractionated, followed by a pull-down of the coding sequences, adapter ligation, and massively parallel sequencing on an Illumina HiSeq 2000 at the Functional Genomics Center (Zurich, Switzerland). Read lengths of 2 x100 base pairs (bp) were produced, aiming for average target sequence coverage of more than 60x and generating more than 20 reads for 90% of the GENCODE exome[125]. The raw reads were quality controlled and aligned to the reference sequence (GRCh38), genotypes were called with Genome Analysis Toolkit, and variants were annotated with the position of nucleotide change with respect to coding genes. The results were filtered according to an in-house list of known primary immunodeficiency genes, which is updated annually. Variants giving rise to non-/synonymous amino acid substitutions, aberrant splicing, or protein truncation events were filtered for functional impact based on PolyPhen2[126] and (CADD)-PHRED[127] scores, and on a minor allele frequency of less than 0.001 in public databases (1000 Genomes, NHLBI GO Exome Sequencing Project (Exome Variant Server) and Exome Aggregation Consortium (Exome Aggregation Consortium ExAC) and our in-house database of more than 2700 exomes. The *RAG1* and *RAG2* variants were confirmed by Sanger sequencing of PCR amplicons of PBMCs derived genomic DNA using GoTaq® G2 Master mix (Promega). After running the amplicon on a 1.5% agarose gel, the amplified DNA was extracted with QIAquick Gel Extraction Kit (Qiagen). The purified PCR products were then bidirectionally sequenced by Microsynth (Switzerland).

Next-generation Trb sequencing

The DNA high throughput sequencing for the T cell receptor β (*Trb*) was performed by Adaptive Biotechnologies (Seattle WA, USA). The DNA was extracted from FACS-sorted SP CD8⁺ thymocytes of eight to ten days old pups. A bias-controlled multiplex polymerase chain reaction (PCR) was used with a blend of all *V* and *J* gene primers for the amplification of the rearranged CDR3β. The amplicon sequencing was performed on the Illumina HiSeq platform (Illumina, San Diego, USA). After read alignments to a reference genome, the gene annotation was based on the ImMunoGeneTics (IMGT) reference system[128].

Next-generation IGH-sequencing

The naïve and antigen-experienced B cell IgH repertoires were determined in PBMCs derived RNA and has been described previously[129]. In summary, cells were lysed in RLT buffer (Qiagen, #79216). Reverse transcription relayed on SuperScript™ III/IV (Invitrogen, #18090010) and a 14nt unique molecular identifier (UMI) was integrated the IgH constant region primers with partial p7 adaptors. Two rounds of reverse transcription were performed – once with primers with IgM and IgD specificity and once with IgA-/IgG-/IgE-specific reverse primers. The 1st step of the multiplex PCR amplified FR1 *V* gene families with specific primers with partial p5 adaptors. The following amplification completed the p7 and p5 adaptor sequences. Amplicons separation was performed by gel-based electrophoresis, and cDNA products were extracted. Each library was concentration-normalized and sequenced on the Illumina MiSeq platform. The Immcantation platform[130,131] was used for sample processing after being de-multiplexing based on the Illumina indices. In summary, the raw FASTQ files were filtered (quality score > 20, paired reads joining if they had a least 10nt length, max. error rate of 0.3, α threshold of 0.001) and reads with the same UMI were collapsed. A data frame containing unique sequences (per sample and isotype) was created by further collapsing reads with identical full-length and equal constant primer sequences. The read assignment to VDJ used the IgBlast[132] and IMGT[128] reference system. The reads were mapped to the IgH constant region using the Stampy[133] algorithm (which allowed to keep only reads with well mapped constant region sequence). B cell subpopulations were discriminated based on the constant region sequence and the mutational load. A *V* gene was classified as naïve/unmutated if <2nt mutations were found across the entire IgD and IgM sequence. The R package SHazaM[131] was used for the determination of *V* gene mutation amount and type.

Experimental model

The following mouse strains were rederived (Center for Transgenic Models Basel) and housed in individually ventilated cages in the specific pathogen-free Animal Facility University Hospital Basel.

Model	Full Name	CD45	Background	Reference
C57BL/6J		CD45.2		Jackson Laboratory
RAG1^{mut}		CD45.2	C57BL/6J	22
RAG2^{mut}		CD45.2	C57BL/6J	Jolan Walter, unpublished
RAG1^{mut} RAG2^{mut}		CD45.2	C57BL/6J	this work
P14	B6;D2-Tg (TcrLCMV)327$^{Sdz/Jdvs}$	Ly5.1	C57BL/6J	134
RAG1^{KO}	B6.129S7-*Rag1^{tm1Mom}*	CD45.2	C57BL/6J	8
RAG2^{KO}	B6(Cg)-Rag2$^{tm.1Cgn}$	CD45.2	C57BL/6J	135

Statistical modeling was used to predetermine sample size (G*Power v3.1.9.6). The experimental mice were randomized and littermates when possible. All mouse experiments have been performed in the accordance with the veterinary office Basel-Stadt (licenses #2772 and 2773).

Virus infection and toll-like receptor challenge

The LCMV isolates WE[136] and clone-13[52] were a gift by Daniel Pinschewer (University of Basel). The viruses were propagated by infection of BHK-21 cells with a low multiplicity of infection (moi) of 0.01, and supernatants were collected after two days. The viral stocks were titrated by plaque-forming assay. Mice were infected via intravenous (i.v.) injection (LCMV strain and infectious dose can be found in the corresponding figure legends.

For toll-like receptor 3 stimulation, the mice receive a single i.p. injection (10µg/g body weight, poly(I:C) HMW VacciGrade™ InvivoGen).

Adoptive transfer, cell depletion, and in vivo treatments

Naïve CD8$^+$ T cells were isolated from Ly5.1$^+$ P14 mice spleen and lymph nodes (cervical, brachial, axillary, inguinal) and negatively enriched (Stemcell #19858,), 1 x10^6 cells were adoptively transferred into recipient mice. Prior to the transfer, the purity of the isolated cells was tested by flow cytometry.

Depletion of CD8$^+$ T cells was achieved with a rat monoclonal IgG2b κ against CD8α (BioXcell, clone YTS 169.4, #BE0117 -12.5ug/g body weight), two injections were given - four and one day prior LCMV infection. The transient depletion was confirmed by flow cytometric assessment of peripheral CD8 T cells.

For the labeling of circulating cells in contrast to tissue-resident cells, we injected mice i.v. with the fluorochrome-conjugated anti-CD45 (3μg, Alexa Fluor700, clone 30-F11, BioLegend #103127), 5 minutes before sacrifice. The procedure has been described before[137].

Experimental cell lines

	Usage	Organism, Tissue	Source
3T3	Plaque-assay, immunofocus reduction assay	*Mus musculus* Embryo	ATCC CRL-1658™
BHK-21	LCMV propagation	*Mesocricetus auratus,* Kidney	ATCC CCL-10™
HEK-GP1 T 239	Production GP1 fusion protein	*Homo sapiens,* Kidney	A gift from Daniel Pinschewer

Peptides, recombinant proteins, and chemicals for the in vitro stimulation

	Sequence/clone	Working concentration	Source
LCMV-derived gp_{33-41}	KAVYNFATM, H-2D^b	10^{-6}M or indicated in the figure legend	AnaSpec #AS-61296
LCMV-derived $gp_{276-284}$	SGVENPGGYCL, H-2D^b	10^{-6}M	AnaSpec #AS-62539
LCMV-derived gp_{61-80}	GLKGPDIYKGVYQFKSVEFD, I-A^b	10^{-6}M	AnaSpec #AS-62851
anti-CD3 ULTRA-LEAF	17A2	5μg/ml	BioLegend #100239
anti-CD28 ULTRA-LEAF	37.51	10μg/ml	BioLegend #102115
Brefeldin A		5ug/ml	BioLegend # 420601
2-NBDG	2-(N-(7-Nitrobenz-2-oxa-1,3-diazol-4-y) Amino) 2-deoxyglucose	20μM	ThermoFisher Scientific #N13195
PMA	Phorbol-12-myristate-13-acetate	50ng/ml	Sigma # P1585
Iono	Ionomycin	500ng/ml	Sigma # I9657

To assess nutrient uptake, 0.4×10^6 splenocytes were isolated and T cells were consequently activated with plate-bound anti-CD3 and anti-CD28 for 18h. For the last 45min of incubation, 2-NBDG was added to the cell media, the analysis was performed using flow cytometry.

For the quantification of intracellular cytokine production, isolated T cells were stimulated *in vitro* for four hours with LCMV derived peptides or PMA/Iono with the addition of BrefeldinA.

Serum analytes

IFNα levels were measured with an enzyme-linked immunosorbent assay (ELISA, PBL Assay Science, # 42120) and serum IFNγ levels were measured with a bead-based assay (BioLegend #740150), both were performed according to the protocol of the manufacturers.

Total serum IgG was enumerated by ELISA according to the manufacturer protocol (ZeptoMetrix, #0801180), IgG subclasses and IgM/IgA were measured by multiplex bead-based assay (BioLegend, #740493). A standard curve was integrated with each assay and serum concentration could be determined. For the GP1-binding IgG ELISA, the GP1-fusion protein was produced as described previously[138]. In short, HEK-GP1 T293 cells, which express the LCMV glycoprotein (GP), were cultured in media containing Hygromycin B (Thermo Fisher Scientific, #10687010), and the supernatant was harvested. For the ELISA, 96-well plates were coated with goat anti-human IgG Fcγ (Jackson ImmunoResearch #109-005-098), after the blocking the plates were incubated with the GP1 T293 cell-supernatant. Further, the GP-1 IgG saturated plates were incubated with the serial-dilutions of mouse serum, next the plates were incubated with HRP-coupled rabbit anti-human IgG Fcγ (Jackson

ImmunoResearch # 309-035-082). HRP was detected by an ABTS™ (2,2'azinobis (3-ethyl-) benzthia-zolinesulfonic acid) color reaction (Roche #10102946001). In-between the different incubations, the plates were washed.

Biochemical analysis for alanine-transaminase and aspartate-transaminase were conducted by the Labormedizin, University Hospital Basel, Switzerland.

Autoantibody microarray
Screening for IgM and IgG autoantibodies was performed using the previously described autoantibody microarray (University of Texas Southwestern Medical Center, Microarray Core Facility)[139]. In brief, recombinant or purified proteins were printed on a microchip, which was then incubated with the sera of interest. Bound antibodies were detected with secondary anti-IgG-Cy5 or anti-IgM-Cy3 and read with the Genepix scanner. The PBS negative control was deducted from the net fluorescence intensity (NFI) for each antigen. To obtain the true signal, the signal-to-noise ratio (SNR) was calculated, which was further used for antibody-score estimation (ln(NFI*SNR+1)).

Plaque forming assay and immunofocus reduction assay
Viral titers were determined by a plaque-forming assay as previously described[140]. As LCMV does not lyse its host cells (aka non-cytolytic virus), plaques are not created *per se.* However, units in this manuscript will be expressed with the frequently utilized term "plaque-forming units (pfu)", instead of the accurate term focus forming units (ffu). 3T3 mouse fibroblasts were seeded at 0.1×10^6 cells/ml per well (96-well plate) and infected with 8-12 dilutions of tissue or blood containing LCMV. Blood samples were diluted 1 in 40 with BSS plus heparin, organs were homogenized with microbeads (Qiagen # 69989). After, two to four hours of incubation at 37°C, the wells were recovered with methylcellulose in 2x supplemented DMEM to restrain viral spreading. Two days later, the supernatant was removed and cells were fixed with 4.2% formaldehyde for 30min at RT, followed by permeabilization with BSS/0.1% TritonX-100 for 20min. After blocking, the intranuclear virus was detected with anti-LCMV nucleoprotein (rat IgG2a, clone VL-4, BioXcell #BE0106), bound antibodies were detected with the secondary antibody HRP goat anti-rat IgG (JacksonImmuno Research #112-035-003). The color reaction mixture comprised DAB (3,3'-Diaminobenzdine tetrahydrochloride hydrate 0.5g/l, Sigma #D5637), ammonium nickel(II) sulfate (0.5g/l, Sigma #A1827) in PBS supplemented with 0.015% H_2O_2. Stained foci were counted manually and the final titer was corrected for the dilution and organ weight.

Neutralizing antibodies against $LCMV_{WE}$ or $LCMV_{clone-13}$ were measured by immunofocus reduction assay. In short, serial dilutions of sera were incubated for 2 hours together with the originally used viral strain as a test substrate. The 3T3 fibroblasts were seeded as above and incubated with the virus+sera mixture for two days. The staining procedure was the same as stated for the plaque-forming assay. The neutralizing titer was defined as the $\log_2$ dilution resulting in a half-maximal reduction of plaques compared with the same amount of virus incubated with naïve mouse sera or medium alone.

Multimers, antibody staining and, flow cytometry
The staining was performed on isolated human PBMCs, on murine peripheral blood lymphocytes, or single-cell suspension from the indicated organs. Human cells were stained in PBS containing 2.5% human AB serum (Blood Donation Center University Hospital Basel, Switzerland), NaH3 0.01%, HEPES 25mM, and Fc block (BioLegend #422301) for 30min at 4°C. Chemokine receptor staining was performed at 37°C for 20 min. For murine cells, the staining was performed in PBS with 2.5% FCS, HEPES 25mM, EDTA 1mM, and Fc block (BioLegend, #101319).

Cell viability was assessed using Live/Dead Fixable NIR dye (# L34975, Invitrogen™, ThermoFisher Scientific) or Zombie-UV™ dye (BioLegend #423107).

Antigens	Reactivity	Clone	Fluorochrome	Source, Identifier
CD3	Human	OKT3	BV711™, PE	BioLegend # 317327, 317307
		UCHT1	PE-Cy7, PB,	BioLegend # 300419, 300418
			BUV395	BD # 563548
CD3	Mouse	17A2	PE/Dazzle™ 594, BV711™	BioLegend #100245, 100241
CD3ε	Mouse	145-2C11	PE/Cy5	BioLegend # 100309
CD4	Human	OKT4, SK3	Ax700, BV510™	BioLegend # 317425, 344634

CD4	Mouse	RM4-5	FITC, Ax700, BV510™	BioLegend # 100509, 100536, 100553
CD8	Human	SK1 HIT8a	PE/Dazzle™ 594, APC BUV395	BioLegend # 344743, 344721 BD # 740303
CD8	Mouse	53-6.7	FITC, PE, BV785™	BioLegend #100705, 100707, 100749
CD11b	Mouse	M1/70	PerCP/Cy5.5	BioLegend #101227
CD11c	Mouse	N418	PE/Cy7	BioLegend #117317
CD16/32 FcγR	Mouse	S17011E	PE	BioLegend #156605
CD19	Human	HIB19	Ax700	BioLegend #302225
CD19	Mouse	6D5	Ax700, PE/Cy5	BioLegend # 115527, 115509
CD20	Human	2H7	BV510™, PE/Cy5	BioLegend # 302339, 302307
CD21/CD35 Complement rec2/ complement rec 1	Mouse	Ebio4E3	PE	eBioscience™ #12-0212-82
CD24	Human	ML5	PE	BioLegend # 311105
CD24/ heat-stable antigen	Mouse	M1/69	Ax488	BioLegend #101815
CD23	Mouse	B3B4	BV510™	BioLegend # 101623
CD25/IL-2Rα	Human	BC69	Ax488	BioLegend # 302615
CD25/IL-2Rα	Mouse	PC61	BV510™, BV605™, APC	BioLegend # 102041, 102035, 102012
CD27	Human	O323	BV650™	BioLegend # 302827
CD27	Mouse	LG.3A10	PE	BioLegend # 124209
CD28	Human	28.2	BV785™	BioLegend # 302949
CD28	Mouse	37.51	BV421™	BioLegend # 102127
CD38	Human	HIT-2	APC	BioLegend # 303509
CD43	Mouse	S11	PE-Cy7	BioLegend # 143209
CD44	Mouse	IM7	AxF488, PerCpCy5.5 PE/Cy5, APC BUV496	BioLegend # 103015, 103031, 103009, 103011 BD #741057
CD45.1	Mouse	A20	APC, Pacific Blue, BV510™	BioLegend # 110713, 110721, 110741
CD45.2	Mouse	104	BV605™ BUV395	BioLegend # 109841 BD # 564616
CD45R/B220	Mouse	RA3-6B2	FITC, Ax488, Ax647, Ax700 BUV395	BioLegend # 103205, 103228, 103229, 103231 BD #56793
CD45RA	Mouse	HI100	FITC	BioLegend # 983002
CD45R0	Mouse	UCHL1	BV510™	BioLegend #
CD62L	Mouse	MEL-14	PE, PE/Dazzle™ 594 BUV395	BioLegend # 104407, 104447 BD # 740218
CD69	Human	FN50	BV421™, BV605™	BioLegend # 310929, 310937
CD69	Mouse	H1.2F3	AxF488, PE/Cy5, BV605™	BioLegend # 10516, 104509, 104529
CD73	Mouse	TY/11.8	PE/Cy7	BioLegend # 127223
CD95/ Fas	Mouse	Jo2	FITC	BD # 554257
CD117/ c-Kit	Mouse	2B8	APC	BioLegend # 105811
CD127/ IL-7Rα	Human	A019D5	BV650™, BV711™, PE/Cy7	BioLegend # 351325, 351327, 351320
CD127/ IL-7Rα	Mouse	A7R34	BV510™, BV650™	BioLegend # 135033, 135043
CD138/ Syndecan-1	Mouse	281-2	BV711™	BioLegend # 142519
CD150/ SLAM	Mouse	TC15-12F12.2	PE	BioLegend # 115903
CD161	Human	HP-3G10	BV605™	BioLegend # 339915
CD183/ CXCR3	Human	G025H7	Alexa Fluor 647	BioLegend # 353711
CD185/ CXCR5	Human	RF8B2	BUV395	BD # 740266
CD185/ CXCR5	Mouse	L138D7	BV421™	BioLegend # 145511
CD194/ CCR4	Human	1G1	PE-Cy7	BD # 557864
CD196/ CCR6	Human	11A9	PE	BD # 551773
CD199/CCR9	Human	L053E8	BV421™	BioLegend # 358913
CD223/ LAG-3	Human	7H2C65	APC	BioLegend # 369211
CD223/ LAG-3	Mouse	C9B7W	BV650™	BioLegend # 125227
CD278/ ICOS	Mouse	7E.17G9	BV510™	BioLegend # not available
CD279/ PD-1	Human	EH12.2H7	BV785™	BioLegend # 329929
CD279/ PD-1	Mouse	29F.1A12	PE-Cy7, BV785™	BioLegend # 135216, 135225
CD366/ Tim-3	Human	F38-2E2	BV421™	BioLegend # 345018

CD366/ Tim-3	Mouse	RMT3-23	BV421™	BioLegend # 119723
Dextramer®-gp$_{33-41}$ H-2D^b	Mouse		APC	Immunodex # JA2160
CXCR5	Human	RF8B2	BUV395	BD # 740266
CXCR5	Mouse	L138D7	BV421™	BioLegend # 145511
F4/80	Mouse	BM8	FITC	BioLegend # 123107
FR4/ folate receptor 4	Mouse	12A5	FITC	BioLegend # 125005
GL-7	Mouse	GL-7	Pacific Blue™	BioLegend #
IgD	Human	IA6-2	PE-Cy7	BioLegend # 348209
IgD	Mouse	11-26c.2a	Ax700, BV650™	BioLegend # 405729, 405721
IgM	Human	MHM-88	BV605™	BioLegend # 314523
IgM	Mouse	RMM-1	BV605™	BioLegend # 406523
Light-chain κ	Human	MHK-49	FITC	BioLegend # 316507
Light-chain λ	Human	MHL-38	PE	BioLegend # 316607
Ly-6C	Mouse	HK1.4	BV605™	BioLegend # 128035
Ly-6G/ Ly-6C Gr1	Mouse	RB6-8C5	APC	BioLegend # 108411
MHC-I Streptamer® – gp33 H2-D^b	Mouse	KAVYNFATM	PE, APC	iba # 6-7016-001
MHC-I Streptamer® – np396 H2-D^b	Mouse	FQPQNGQFI	PE, APC	Iba # 6-7098-001
MHC II/ I-A^b	Mouse	KH74	FITC	BioLegend # 115305
NK1.1	Mouse	PK136	APC	BioLegend # 108709
Tetramer – gp$_{33-41}$ H2-D^b	Mouse	KAVYNFATM	APC, PE	TCMetrix, UNIL
Tetramer – gp$_{276-284}$ H2-D^b	Mouse	SGVENPGGYCL	PE	TCMetrix, UNIL
Tetramer – np$_{396-404}$ H2-D^b	Mouse	FQPQNGQFI	PE	TCMetrix, UNIL
TCRα/β	Human	IP26	Ax488	BioLegend # 306711
TCRβ	Mouse	H57-597	FITC, Ax700	BioLegend # 109205, 109223
TCRγδ	Human	B1	PE	BioLegend # 331209
TCRγδ	Mouse	GL3	PE	BioLegend # 118107
TCR Vα7.2	Human	3C10	PE	BioLegend # 351705
TCR Vα2	Mouse	B20.1	PE	BioLegend # 127807
TCR Vβ5.1, 5.2	Mouse	MR9-4	APC	BioLegend # 139505
TCR Vβ8.1, 8.2	Mouse	KJ16-133.18	FITC	BioLegend # 118405
Intracellular/Intranuclear				
Bcl-6	Mouse	K112-91	PE	BD # 561522
FoxP3	Mouse	clone FJK-16s	FITC, PE	eBioscience # 25-5773-82
HELIOS	Mouse	22F6	eFluor450	eBioscience # 48-9883-41
ki67	Mouse	11F6	AxF647	BioLegend # 151206
IFNγ	Mouse	XMG1.2	PE, BV510™	BioLegend # 505807, 505842
TNFα	Mouse	MP6-XT22	FITC	BioLegend # 506303
IL-2	Mouse	JES6-5H4	APC, BV605™	BioLegend # 503809, 503829
IL-4	Mouse	11B11	BV605™	BioLegend # 504126
IL-10	Mouse	JES5-16E3	PE/Cy7	BioLegend # 505025

The measurement for intracellular cytokine production was performed with isolated murine T cells after stimulation with either LCMV-derived peptides or PMA/Iono for four hours in the presence of BrefeldinA. First, the surface, and viability staining were performed (see above), this was followed by fixation with permeabilization (BD Cytofix/Cytoperm™ (# 554722), third intracellular antibodies were diluted in BD Perm/Wash Buffer (# 554723) following the manufacturer's protocol.

For the determination of transcription factors via intranuclear staining, we relied on the FoxP3/ Transcription Factor Staining Buffer Set (eBioscience™ # 00-5523-00).

Samples were acquired on a LSRII-Fortessa™ (BD Bioscience – equipped with 355nm, 405nm, 488nm, 561nm, and 640nm laser lines). The PMT voltages were manually calibrated prior to sample acquisition, with the help of single stain controls, this allowed to minimize fluorescence spillover and the computation of a compensation matrix. Data were analyzed using the FlowJo software (Version 10.5.2, TreeStar, USA).

Buffers and media

	Components	working concentration	Identifier
ACK (ery lysis buffer) pH adjusted to 7.2-7.4	ddH20 – 1L		
	NH$_4$Cl – 8.3g	1.5M	Sigma # A9434
	KHCO$_3$ – 1g	100mM	Sigma # 237205

	EDTA 200uL of 0.5M	0.1mM	Gibco # AM9260G
BHK-21 media	DMEM with 4.5g/L Glucose		Gibco # 11965084
	FCS HI	10%	Gibco # 16000044
	HEPES	1mM	Gibco # 15630080
	Sodium Pyruvate	0.2 mM	Gibco # 11360070
	Tryptose Phosphate Broth	0.02%	Gibco # 18050039
BSS pH adjusted to 7.2	ddH2O	5L	
	Phenol Red	28.2 mM (0.05g)	Sigma # P3532
	$CaCl_2 \cdot 2H_2O$	952.38 uM (0.7g)	Sigma # 1086436
	NaCl	136.89 mM (40g)	Sigma # S7653
	KCl	5.36 mM (2g)	Sigma # P3911
	$MgSO_4 \cdot 7H_2O$	811.45 uM (1g)	Sigma # 63138
	$MgCl_2\ 6H_2O$	985.2 uM (1g)	Sigma # M2670
	KH_2PO_4 (0.3g)	440.88 uM (0.3g)	Sigma # P5379
	$Na_2HPO_4\ 2H_2O$	1.34 mM (1.2g)	Sigma # 71643
	D-Glucose	M (5g)	Sigma # G8270
BSS + Heparin	BSS – see above		
	HEPARIN Bichsel 50 IE/ml	0.01 IE	Pharmacode # 2764204
Cell dissociation	TrypLE™ Express Enzyme		Gibco # 12605010
Mouse media	RPMI-1640 with HEPES		Sigma # R5886
	FCS HI	10%	Gibco # 16000044
	2-Mercaptoethanol	50 uM	Gibco # 21985023
	Penicillin / Streptomycin	100U/mL	Gibco # 10378016
	GlutaMAX™	2mM	Gibco # 35050061
Plaque-assay media	MEM		Gibco # 31095029
	FCS HI	2.5%	Gibco # 16000044
	HEPES	1 mM	Gibco # 15630080
	GlutaMAX™	2mM	Gibco # 35050061
	Penicillin / Streptomycin	100U/mL	Gibco # 10378016
Overlay media for plaque assay	2x DMEM		Sigma # D5030
	FCS HI	10%	Gibco # 16000044
	Penicillin / Streptomycin	200 U/ml	Gibco # 10378016
	GlutaMAX™	4mM	Gibco # 35050061
	$NaHCO_3$		Sigma # S5761
Methylcellulose - overlay media for plaque assay	ddH20 viscosity 400 cP	2%	Sigma # M0262
3T3 media	DMEM with 4.5g/L Glucose and L-glutamine		Gibco # 11965084
	FCS HI	10%	Gibco # 16000044
Infection media	Opti-MEM™		Gibco # 31985062
	FCS HI	2%	Gibco # 16000044

Oligonucleotides

The oligonucleotides utilized for Sanger sequencing or genotyping are listed below, all were ordered with Microsynth (Switzerland).

Gene	Forward	Reverse	Source
Human *RAG1*	TTC ACA AAA CCC TGG CCC AT	GCA ACC CAG AGG GAA ACT CA	This manuscript
Human *RAG2*	AAC CCG CCA TGA TCT ACT GC	TTT CTG GCC CTT AAT TCA TGT AAC	This manuscript
Mouse *Rag1* p.S723	TGG GTT GCA GTG AGG TCT TCT CCT AGC ACC	CCT GTT GCT CTT CTG GCC TCA CTA AAC GGC	[141]
Mouse *Rag2* p.F62L	AGG CTC TAC TGC TAA GTC CTG AAC TG	TTG AAC AAC GAT GGG GAC TCA GGT AC	Jolan Walter
Mouse 327[tg]	CAT GGA GGC TGC AGT CAC CC	GTT TGT TTG CGA GCT CTG TTT TGA TGG CTC	[68]
Mouse 327[tg] internal control	CAA ATG TTG CTT GTC TGG TG	GTC AGT CGA GTG CAC AGT TT	[68]

Multiple protein alignment

The UniProt database was used, results were filtered for reviewed proteins (Swiss-Prot) or predicted proteins with an annotation score of at least 3, and length RAG1 > 800 or RAG2 >450. The alignment was performed using the MUSCLE algorithm[142].

Statistical analysis

The statistical analysis relied on GraphPad Prism (v.9) for variance testing and data visualization. Individual tests are noted in the corresponding figure legends. Further, R studio was used for statistical testing and computation of differential expression with empirical Bayes statistics (**Limma package**[143]), graphics were generated with the R package ggplot2[144], RColorBrewer[145], viridis and heatmap.2.

Supplementary material
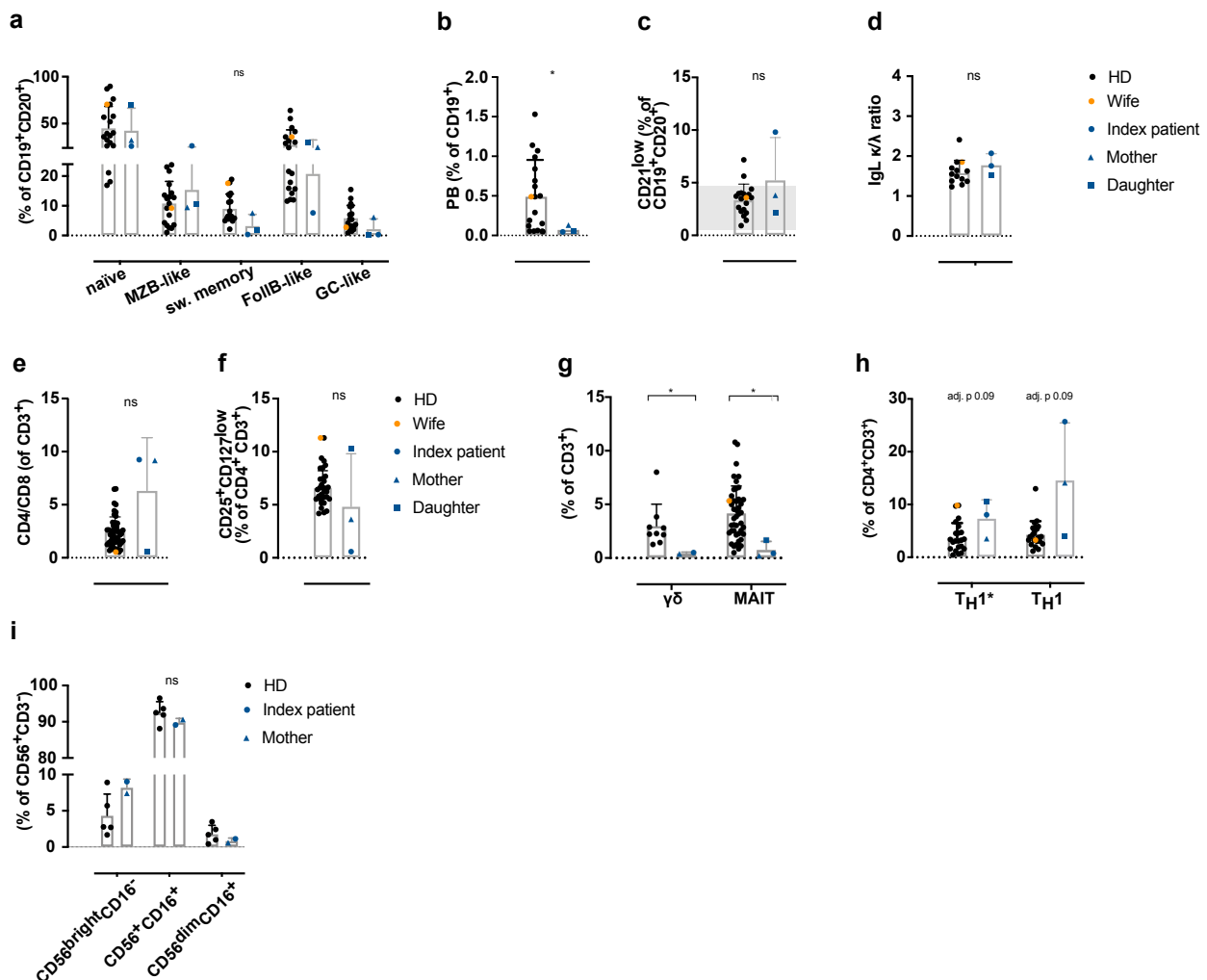
Supplementary figure S1: Qualitative changes in the B cell and T cell compartment in patients carrying two heterozygous mutations in *RAG1* and *RAG2*. a) Proportional distribution of peripheral blood B cell subsets, investigated in the index patient, his mother, daughter and healthy controls (HD) including the wife by flow cytometry. B cell subsets comprised naïve (CD27-IgD+), switched memory (sw. memory CD27+IgD-IgM-), marginal zone-like (MZB-like, CD21+CD27varIgM+IgDlow), follicular-like (FollB-like, CD21+CD27-IgM-IgD+) and germinal center like (GC-like, CD38+CD27+ IgM-IgD-). **b)** Enumeration of circulating plasma blasts (PB CD27+CD38+CD20lowCD19+). **c)** Comparison of CD21low B cells in the index patient, his mother and daughter and healthy subjects – measured by flow cytometry. Gray shading indicates the in-house reference range of healthy adults. **d)** Ratio for the surface expression of the B cell receptor light chains kappa and lambda (IgL κ/λ) on bulk CD19+CD20+ B cells, evaluated by flow cytometry. **e)** Analysis of the CD4+/CD8+ T cell ratio in the peripheral blood of the index patient, his mother, daughter and wife including healthy blood donors. **f)** Regulatory CD4+ T cells (Tregs, CD25+CD127low) were enumerated in the blood by flow cytometry. **g)** Percentage of γδ+ T cells (TCRγδ+TCRαβ-) and MAIT

cells (mucosa-associated invariant T cells, Vα7.2+CD161+) in the patients were compared to control blood donors. **h)** Investigation of CD4+ T_H1 subsets in the peripheral blood (T_H1CCR6-CD45RA-CCR4-CXCR3+, T_H1*CCR6+CD45RA-CXCR3+CCR4-). **i)** NK cell subsets of the peripheral blood were investigated, in the two patients and control subjects. For significance testing a non-parametric Mann-Whitney test was used in (b/c/d/e/f) and with multiple comparison correction in (a/g/h/i).
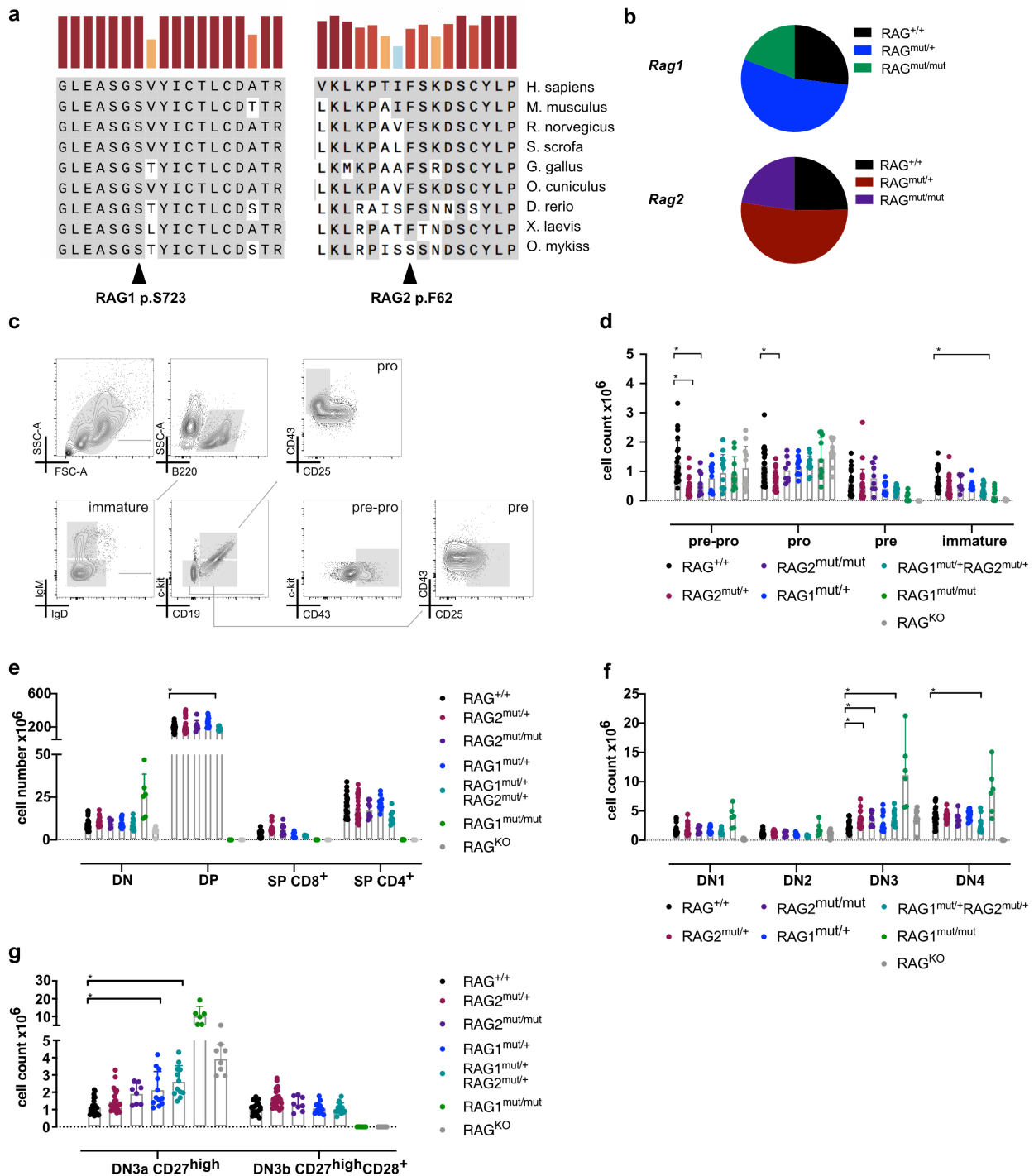
Supplementary figure S2: B cell and thymocyte development in Rag mutant mice. a) Protein alignment for RAG1 and RAG2 proteins in different species. The residues position RAG1 p.S723 and RAG2 p.F62 are indicated with triangles at the bottom. The height of bars on top of the alignment indicates the level of residue conservation across the species tested. **b)** Distribution of either RAG1mut or RAG2mut genotypes shortly after birth. *Rag1* mutant mice n = 483, *Rag2* mutant mice n = 455 mice. Enumerative and qualitative analyses of lymphocyte development were performed at the age of eight to twelve days, using flow cytometry. Bone marrow (BM) was analyzed for B cell development and T cell development was analyzed in the thymus. **c)** Representative flow cytometric plots of the B cell developmental stages in the bone marrow (BM) with

gating strategy. **d)** Quantitative analysis of B cell maturation stages in the BM, including pre-pro B cells (B220$^+$CD19$^-$IgM$^-$IgD$^-$CD43$^+$CD117$^-$), pro B cells (B220$^+$CD19$^+$IgM$^-$IgD$^-$CD43$^+$CD117^{int}), pre B cells (B220$^+$CD19$^-$IgM$^-$IgD$^-$CD43$^-$CD25^{var} CD117$^-$), immature B cells (IgM$^+$IgD$^-$). Significant changes were seen in RAG1$^{mut/mut}$ and RAG^{KO} regarding the pre-B cell and immature B cell stages, but statistics were omitted from the graphic for clarity reasons. **e)** Absolute thymocyte numbers at the double-negative (DN, CD4$^-$CD8$^-$), double-positive (DP, CD4$^+$CD8$^+$), and both single-positive (SP CD4$^+$, SP CD8$^+$) stages. Significant differences were seen in RAG1$^{mut/mut}$ and RAG^{KO} regarding all the stages except for DN, but statistics were omitted from the graphic for clarity reasons. **f)** Enumeration of thymocyte numbers at the four DN stages, including DN1 (CD44^{high}CD25^{low}), DN2 (CD44^{high} CD25^{high}), DN3 (CD44^{low}CD25^{high}) and DN4 (CD44^{low}CD25^{low}) thymocytes. Significant differences were seen in RAG1$^{mut/mut}$ and RAG^{KO} regarding all the stages except for DN2, but the depiction of significance was omitted in the graphic for clarity reasons. **g)** Absolute abundance of thymocytes at the DN3a (CD27^{high}) and DN3b stage (CD27^{high}CD28^{high}). Significant differences were seen in RAG1$^{mut/mut}$ and RAG^{KO} regarding the two stages, but the illustration of significance was omitted in the graphic for clarity reasons. Significance was calculated by using the RAG$^{+/+}$ mice as a reference group and using a Kruskal-Wallis test with posthoc correction in (d-g). Merged data from six independent experiments are displayed with three to eleven mice per experiment and genotype. Absolute numbers of BM cells are calculated based on pooled cells from two femurs.
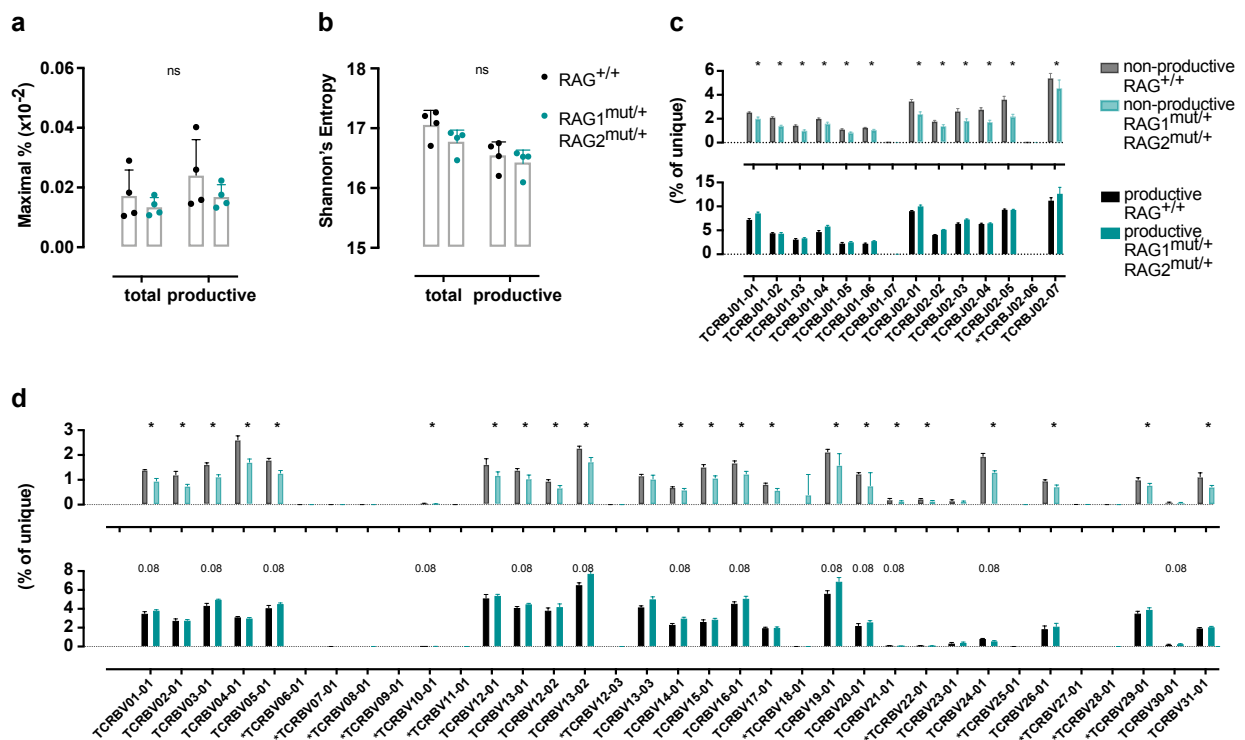
Supplementary figure S3: RAG1$^{mut/+}$RAG2$^{mut/+}$ thymocytes display an enhanced productive clonality. DNA derived from sorted single positive CD8$^+$ thymocytes was used in a T cell receptor β-chain locus (*Trb*) sequencing experiment. **a)** The maximal frequency of a single clone was assessed in total and productive *Trb* sequences. **b)** The Shannon's entropy index was calculated as an estimation for the *Trb* diversity. Unique *Trb* sequences were investigated regarding the relative utilization of **c)** *J* and **d)** *V* genes. *V* genes are ordered as on the chromosome, left displays 5' end, right side displays 3' end. * Asterisk demarks pseudogenes. One experiment was performed with four mice per genotype. The non-parametric Mann-Whitney test with multiple comparison correction was used in a-d.

a

b

c
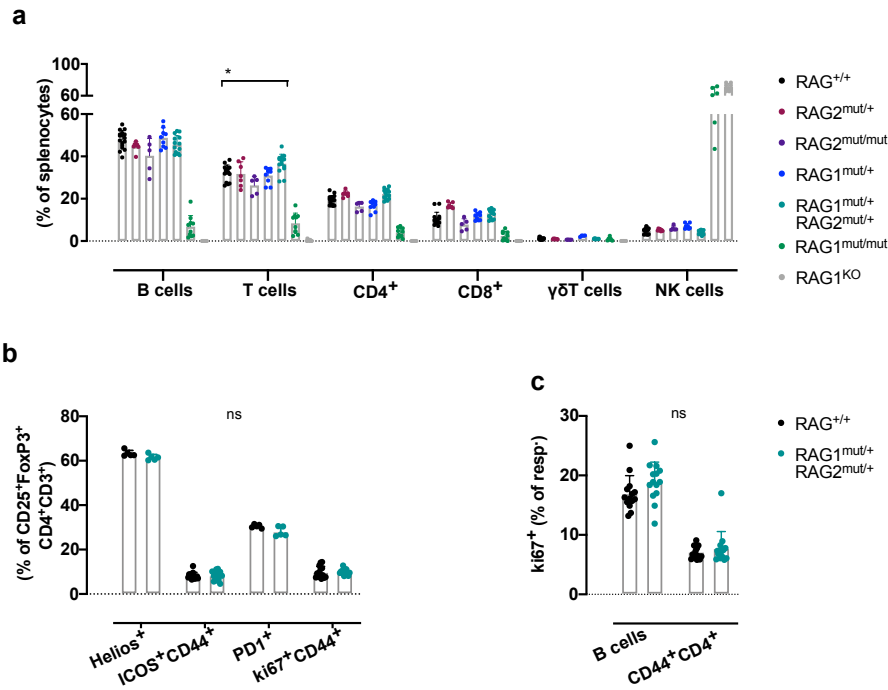
Supplementary figure S4: Subtle accumulation of T regulatory CD4$^+$ T cells and transitional B cells in RAG1$^{mut/+}$RAG2$^{mut/+}$ mice. a) Proportional distribution of splenic lymphocyte subpopulations was assessed by flow cytometry, including B cells (B220$^+$CD19$^+$), T cells (CD3$^+$), CD4$^+$ T cells (CD4$^+$CD8$^-$CD3$^+$), CD8$^+$ T cells (CD8$^+$CD4$^-$CD3$^+$), γδ$^+$ T cells (TCRγδ$^+$TCRβ$^-$CD19$^-$B220$^-$) and NK cells (NK1.1$^+$B220$^-$CD3$^-$). b) The origin and activation status of CD4$^+$ regulatory T cells T$_{reg}$ (FoxP3$^+$CD25$^+$) was assessed by probing for Helios, ICOS, PD1 and ki67. c) The proliferation marker ki67 was measured in activated CD44$^+$CD4$^+$ T cells and bulk B cells (CD19$^+$B220$^+$) by flow cytometry. Significance testing was performed using a Kruskal-Wallis test with correction in (a), for (b-c) we used the non-parametric Mann-Whitney test with multiple comparison correction. (a-c) Merged data from two to six independent experiments are displayed, with two to nine mice per genotype and per experiment.

a

b

c

d

e
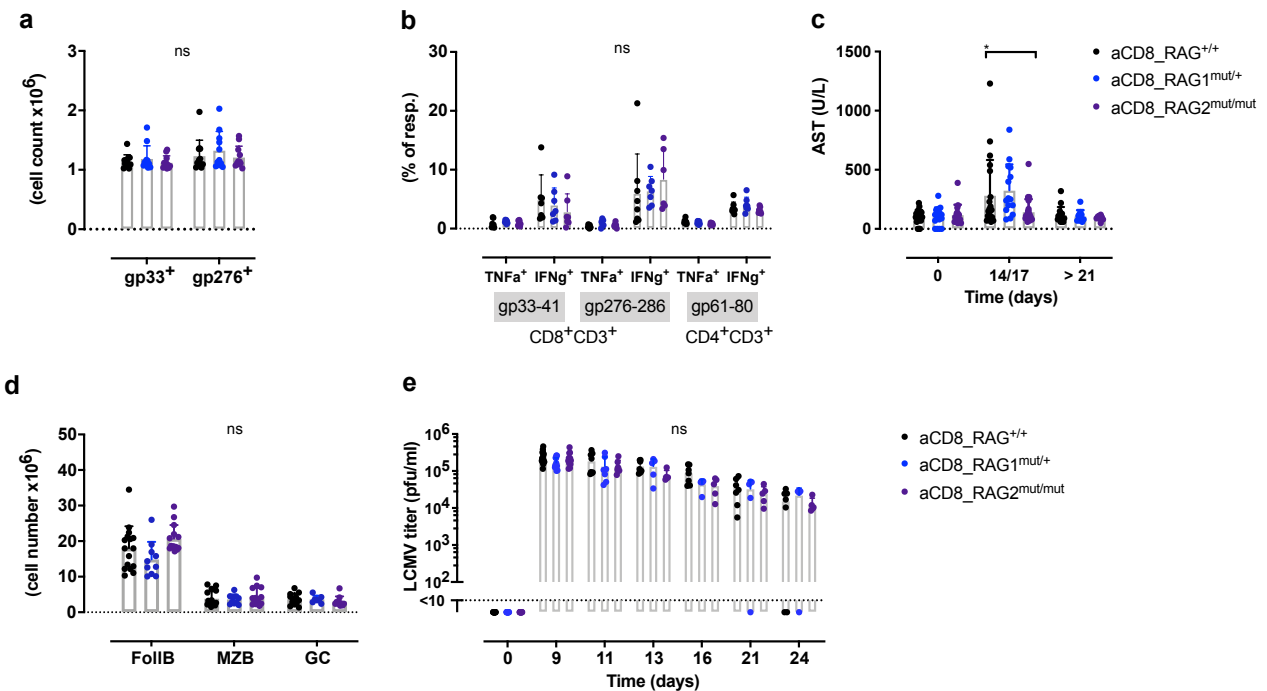
Supplementary figure S5: Robust T and B cell response to systemic LCMV infection in RAG2$^{mut/mut}$ und RAG1$^{mut/+}$ mice after transient CD8$^+$ T cell depletion challenge. RAG$^{+/+}$, RAG1$^{mut/+}$ and RAG2$^{mut/mut}$ mice were transiently depleted of cytotoxic CD8$^+$ T cells with monoclonal antibodies at day-4 and -1 (12.5 μg/g body weight) before intravenous infection with LCMV$_{WE}$ (2x 10^{6}pfu). **a)** Quantification of splenic LCMV-specific CD8$^+$ T cell binding the peptides gp33 and gp276 analyzed with streptamer resp. tetramer staining, 14 to 17 days after infection. **b)** Virus-specific effector function of CD8$^+$ and CD4$^+$ T cells was determined after T cell isolation 14 and 17 days after infection and *ex vivo* restimulation with the LCMV peptides (gp$_{33-41}$, gp$_{276-284}$, gp$_{61-80}$). Subsequently, the intracellular TNFα and IFNγ production was measured with flow cytometry. **c)** Hepatic immunopathology was quantified by serum aspartate-aminotransferase (AST) activity measurement at the indicated time points after infection. **d)** B cell subsets were characterized and quantified for follicular (FollB CD21/CD35^{int}CD23^{high}B220$^+$CD19$^+$), marginal-zone (MZB CD21/CD35^{high}CD23$^{low/-}$B220$^+$CD19$^+$) and germinal center B cells (GC GL-7$^+$CD95$^+$IgM$^-$IgD$^-$B220$^+$CD19$^+$), 14 to 17 days after infection. **e)** Peripheral blood LCMV titers were measured at indicated time points after infection. In a/b/d significance was assessed using a Kruskal-Wallis test with post hoc correction. Further, the statistical testing used a 2wayANOVA with multiple comparison correction for c/e. Merged data from two to three independent experiments are depicted with three to eight mice per genotype.
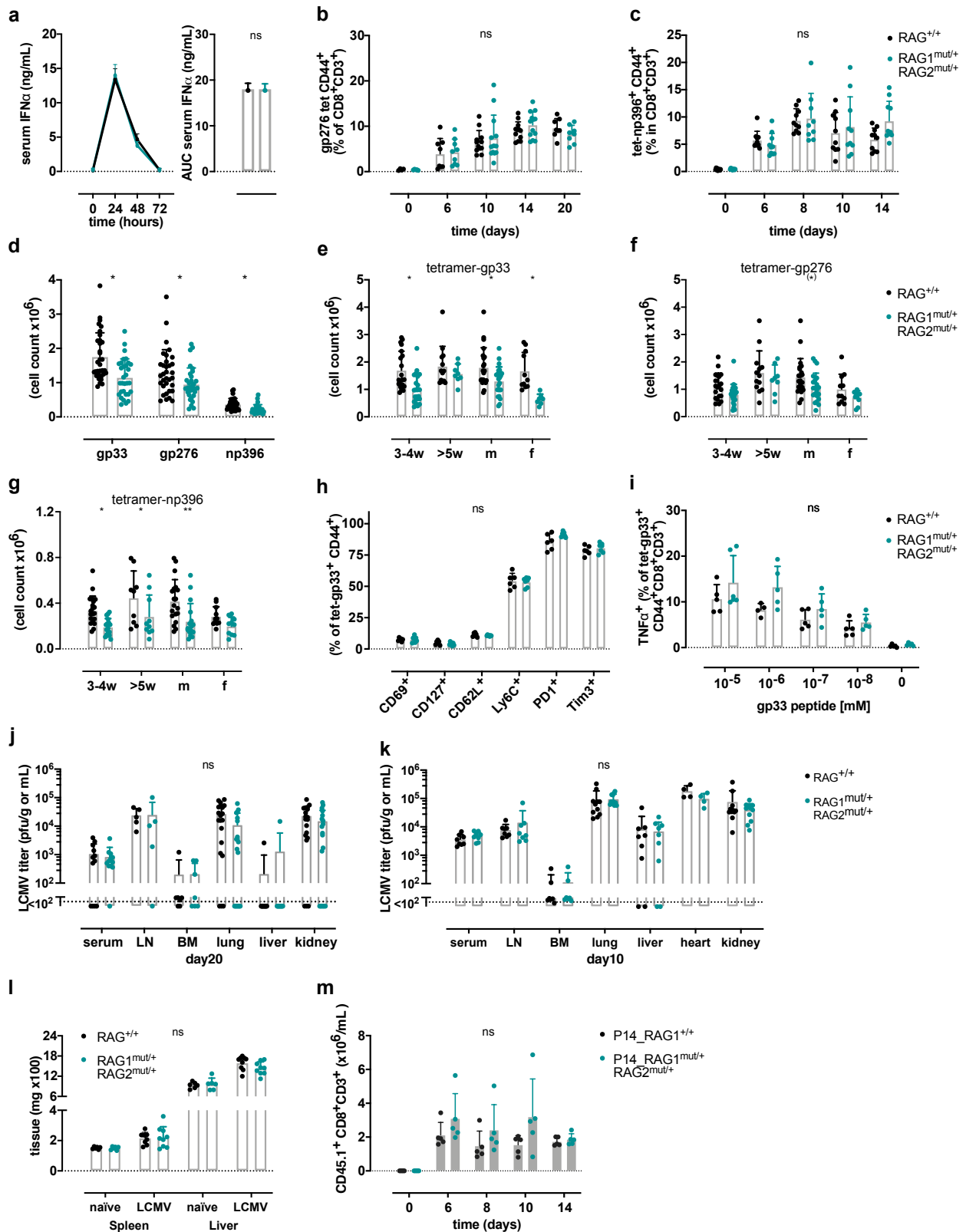
Supplementary figure S6: RAG1$^{mut/+}$RAG2$^{mut/+}$ mice have fewer virus-specific CD8$^+$ T cells with reduced TCR affinity after LCMV infection. a) Interferon-α (IFNα) was measured in the serum at the indicated time points after LCMV-infection. Left side: time course over three days, right side enumeration of the area under the curve (AUC). Temporal kinetic analysis for the frequency of **b)** gp276-tetramer and **c)** np396-tetramer binding CD8$^+$ T cells measured in the peripheral blood at indicated time points after LCMV-infection. **d)** Absolute count of splenic LCMV epitope-specific CD8$^+$ T cells was assessed by tetramer-staining with gp$_{33-41}$ (gp33), gp$_{276-284}$ (gp276) and np$_{396-404}$ (np396) peptides day ten after LCMV-infection. CD8$^+$ T cells recognizing the LCMV epitopes **(e)** gp33, **(f)** gp276, **(g)** np396 were measured in two age groups (3-4 weeks old (3-

- 50 -

4w) *vs.* > 5 weeks old (>5w)). LCMV-specific CD8$^+$ T cells were additionally grouped according to their sex (m male, f female). **h)** Characterization of the activation phenotype of splenic tetramer-gp33 binding CD8$^+$ T cells ten days after infection. The surface expression of the activation markers and inhibitory molecules was evaluated by flow cytometry. **i)** Ten days post-infection splenic CD8$^+$ T cells were isolated and re-stimulated *ex vivo* with different concentrations of the LCMV derived peptide gp$_{33-41}$ and the intracellular production of TNFα was measured *via* flow cytometry. LCMV titers were quantified **(j)** twenty and **(k)** ten days after LCMV infection by plaque-forming assay in different lymphoid and non-lymphoid organs and in the serum. **l)** Splenomegaly was assessed by weighing the spleen and liver of naïve mice *vs.* day ten post LCMV infection. **m)** Naïve LCMV-specific TCR-transgenic CD8$^+$ T cells recognizing the LCMV gp$_{33-41}$ epitope (P14 cells 1x10^6, CD45.1) were adoptively transferred into CD45.2 RAG$^{+/+}$ or RAG1$^{mut/+}$RAG2$^{mut/+}$ recipient mice one day prior to LCMV-infection. Post LCMV-infection, the quantity of transferred P14 cells was assessed in the peripheral blood at indicated time points. Statistical significance was assessed using a Mann-Whitney test in (a) and with posthoc test in d/h/j/k. A 2wayANOVA with Šídák multiple comparison correction was utilized for b/c/e/f/g/i/l/m. Merged data from two to six independent experiments are depicted except for (n) which depicts data from one experiment. All experiments were performed with three to eight mice per genotype.
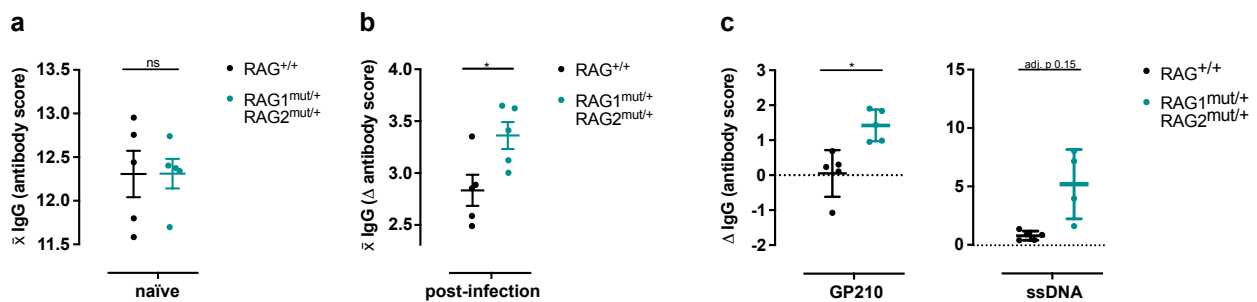
Supplementary figure S7: Dysregulated LCMV specific B cell response and a greater increase of autoantibodies in double heterozygous Rag mutant mice after infection. (b-c) *Rag* wild-type or double heterozygous *Rag* mutant mice were infected intravenously with high dose LCMV$_{WE}$ (2 x10^{6}pfu). **a)** The average IgG antibody score was computed across 115 autoantibodies in the sera of naïve mice. **b)** IgGs directed against 115 autoantigens were tested 14 days after LCMV infection and compared to before infection. The average was computed across all antigens regarding the post-pre infection antibody score differences. **c)** Representative post-pre infection antibody score changes are depicted for the nuclear membrane antigen GP210 and single-stranded DNA (ssDNA). In one experiment with five mice per genotype, statistical significance testing was performed using in (a+b) the Mann-Whitney test and in (c) with multiple comparison correction. Graphs in (a+b) depict the mean with the standard error of the mean.

Supplementary table S1 |Leukocyte subsets and surface markers used for the characterization of the patients with *RAG1/RAG2* haploinsufficiency. Blue refers to values below the in-house reference range, red refers to values above.

			Daughter	Index patient	Mother
	Gender		F	M	F
	Age		16	51	79
Hemoglobin	g/L	(males 140-180, females 120-160)	153	120	
Thrombocytes	X10^9/L	150-450	215	111	
WBC	Leukocytes	(3.5-10 x10^9/L)	8.26	2.3	
	Neutrophils	40-74% (1.3-6.7 x10^9/L)	6.85	59.3% (1.3 x10^9/L)	
	Monocytes	3.4-9% (0.12- 0.62 x10^9/L)	0.4	8.3% (0.19 x10^9/L)	
	Eosinophils	0-7% (0-0.3 x10^9/L))	0.07	3.8% (0.09 x10^9/L)	
	Basophils	0-1.5% (0-0.09 x10^9/L)	0.02	1% (0.02 x10^9/L)	
	Total lymphocytes	20-45% (0.9-3.3 x10^9/L)	0.9	25.9% (0.62 x10^9/L)	
	Total T cells	55-86% [742-2750/µL]	67% (613/µL)	77% (509/µL)	40.1%
	Total B cells	5-22% [80-616/µL]	11.8% (105/µL)	4% (28/µL)	3.4%
	Total NK cells	5-26%	21%	18%	35.5%

	Surface markers	% range [cells/uL]	84-724/µL	(184/µL)	(118/µL)
Helper/cytotoxic ratio	CD4$^+$/CD8$^+$	1.5-3.5	0.58	4.4	5.1
Total helper T	CD3$^+$ CD4$^+$	33-58% [404-1612/uL]	33.8%	86.3% [404/uL]	86.9%
Helper T, naïve	CD3$^+$ CD4$^+$CD27$^+$ CD45RO$^-$	15.7-54.7%	68.8%	5.66%	41.7%
Helper T, central memory	CD3$^+$ CD4$^+$CD27$^+$ CD45RO$^+$	8-28.8%	27.1%	67.5%	45.2%
Helper T, effector memory	CD3$^+$ CD4$^+$CD27$^-$ CD45RO$^+$	16.8-57.4%	3.52%	24.9%	11.2%
Helper T, follicular	CXCR5$^+$PD1$^+$	6.9-19.1%		27.7%	3.3%
Helper T, regulatory	CD25^{hi}CD127^{low}	6.1-11%	10.3%	0.6%	3.63%
Helper, recent thymic emigrants	CD31$^+$CD27$^+$CD45R0$^-$	14.1-37.2% (% of naïve)		1.6%	
Total cytotoxic T	CD3$^+$ CD8$^+$	13-39% [220-1129/uL]	57.9%	9.33% [93 /uL]	9.48%
Cytotoxic T, naïve	CD3$^+$ CD8$^+$CD27$^+$ CD45RO$^-$	7-62.5%	76.3%	10.37%	14.8%
Cytotoxic T, central memory	CD3$^+$ CD8$^+$CD27$^+$ CD45RO$^+$	0.6-4.4%	15.5%	51.4%	27%
Cytotoxic T, effector memory	CD3$^+$ CD8$^+$CD27$^-$ CD45RO$^+$	4.3-64.5%	2.6%	18.1%	19%
Cytotoxic, TMRA	CD45RA$^+$CD62L$^-$	8.1-60.5%	3.5%	28.7%	
Total double-negative T	CD3$^+$ CD8$^-$CD4$^-$	< 1.5%	1.23%	1.29%	0.72%
Total, γδ T	CD3$^+$ TCRβ$^-$TCRγδ$^+$	1.27-2.8%		0.51%	0.39%
Total, Vα7.2$^+$	Vα7.2$^+$ CD3$^+$	3.26-8.85% (of CD3$^+$)	3.9%	0.8%	1.27%
Total, MAIT	Vα7.2$^+$CD161$^+$CD3$^+$	1.55-6.77% (of CD3$^+$)	1.66%	0.35%	0.2%
Plasma blasts	CD27$^+$CD38$^+$CD19$^+$CD20$^-$	0.1-3% (of CD19$^+$ cells) [1-5/uL]	0.13%	<0.1% [<1/uL]	<0.1%
Naïve B	IgD$^+$CD27$^-$	25.1-92.4% [66-228/uL]	70%	60% [17/uL]	62%
Switched memory B	IgD$^-$CD27$^+$	2.4-32.6% [8-102/uL]	1.89%	2.28%	21.9%
Transitional B	IgD$^+$IgM$^+$CD38$^+$	0.3-2.9% [1-5uL]	5.32%	11.5%	7.59%
Marginal zone like B	IgD$^+$IgM$^+$CD27$^+$	3.1-59.7% [8-172/uL]	10.6%	26.3%	9.5%
CD21^{low} B	CD21^{low}	0.5-4.7%	2.1%	9.8%	3.8%
κ/λ ratio		1.28-2.4	1.75	2.05	1.51
Immunoglobulins IgG total		7.0 - 16.0 g/L	5.85	3.8 (with IVIG)	8.35
IgG1		4.9-11.4 g/L	4.68	2.59	4.14
IgG2		1.5-6.4 g/L	1.08	1.17	3.49
IgG3		0.2-1.1 g/L	0.184	0.11	0.275
IgG4		0.08-1.4 g/L	0.096	0.02	0.649
IgM		0.4-2.3 g/L		<0.05	0.26
IgA		0.7-4 g/L		<0.05	1.18
IgE		0-100 IU/mL	31	<5	317

WBC white blood count, Ig immunoglobulin, MAIT mucosa-associated invariant T cell, NK Natural killer.

Supplementary table S2 | Additional *in silico* predictor parameters regarding the missense mutations in *RAG1* and *RAG2*.

Genetic location	Nucleotide change (c.DNA)	Zygosity	Gene symbol	Exon	AA change	rsID	SpliceAI	PrimateAI	REVEL
11_36576320_A_G	c.A3019G	Het	*RAG1*	2	p.M1006V M[ATG] > V[GTG]	rs139113046	0, no consequence	0.645	0.396
11_36592665_T_C	c.A1504G	Het	*RAG2*	3	p.M502V M[ATG] > V[GTG]	rs145614809	0, no consequence	0.229	0.614

AA amino acid, Het heterozygous, PrimateAI Primate artificial intelligence[146], REVEL rare exome variant ensemble learner[147], rsID single nucleotide polymorphism database identifier, SpliceAI Splice artificial intelligence[148].

Supplementary table S3 | Thymic and bone marrow (BM) cellularity of various Rag mutant mice at the age of eight to twelve days. Numbers indicate the mean, SD x10^6 (n).

	RAG$^{+/+}$	RAG2$^{mut/+}$	RAG2$^{mut/mut}$	RAG1$^{mut/+}$	RAG1$^{mut/+}$ RAG2$^{mut/+}$	RAG1$^{mut/mut}$	RAG^{KO}
Thymus	237.7, 63.9 (25)	253.4, 97 (25)	248.5, 58.1 (14)	292.9, 67 (12)	207, 27.9 (12)	27.2, 13.7 (6)	5.6, 1.67 (11)
BM	9.4, 4.6 (25)	4.6, 1.82 (25)	6.7, 2.6 (14)	13.4, 4.2 (12)	8.89, 3 (12)	9.3, 3.6 (6)	8.1, 4.24 (11)

Supplementary table S4 | Splenic lymphocyte counts in various adult Rag mutants. Below detection limit (bdl). Numbers indicate mean, SD x10^6.

	RAG$^{+/+}$	RAG2$^{mut/+}$	RAG2$^{mut/mut}$	RAG1$^{mut/+}$	RAG1$^{mut/+}$ RAG2$^{mut/+}$	RAG1$^{mut/mut}$	RAG^{KO}
T cells	31.14, 5.3	31.6, 5.8	23.38, 4.8	24.24, 4.07	36.6, 5.52	0.844, 0.45	bdl
γδ T cells	0.57, 0.27	0.46, 0.1	0.9, 0.13	0.72, 0.24 *	0.906, 0.13	0.007, 0.007	bdl
B cells	39.14, 10.3	34.53, 7.1	32.4, 5.7	33, 8.2	44.63, 8.1	0.89, 0.45	bdl
NK cells	4.3, 1.18	3.97, 0.61	3.9, 1.42	5.07, 1.09	4.13, 1.15	7.54, 2.95	4.91, 3.06
(n)	(22)	(7)	(7)	(4), *(7)	(14)	(6)	(4)

References

1. Lieber, M. R. Transposons to V(D)J Recombination: Evolution of the RAG Reaction. *Trends Immunol.* **40**, 668–670 (2019).
2. Schatz, D. & Swanson, P. V(D)J recombination: mechanisms of initiation. *Annu. Rev. Genet.* **45**, 167–202 (2011).
3. Notarangelo, L. D., Kim, M.-S., Walter, J. E. & Lee, Y. N. Human RAG mutations: biochemistry and clinical implications. *Nat. Rev. Immunol.* **16**, 234–46 (2016).
4. Delmonte, O. M., Villa, A. & Notarangelo, L. D. Immune dysregulation in patients with RAG deficiency and other forms of combined immune deficiency. *Blood* **135**, 610–619 (2020).
5. Delmonte, O. M., Schuetz, C. & Notarangelo, L. D. RAG Deficiency: Two Genes, Many Diseases. *J. Clin. Immunol.* **38**, 646–655 (2018).
6. Schwarz, K. *et al.* RAG mutations in human B cell-negative SCID. *Science (80-.).* **274**, 97–99 (1996).
7. Shinkai, Y. *et al.* RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68**, 855–867 (1992).
8. Mombaerts, P. *et al.* RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* **68**, 869–877 (1992).
9. Villa, A. *et al.* Partial V(D)J recombination activity leads to omenn syndrome. *Cell* **93**, 885–896 (1998).
10. Schuetz, C. *et al.* An Immunodeficiency Disease with RAG Mutations and Granulomas. *NEJM* **358**, 2030–8 (2008).
11. Lee, Y. N. *et al.* A systematic analysis of recombination activity and genotype-phenotype correlation in human recombination-activating gene 1 deficiency. *J. Allergy Clin. Immunol.* **133**, 1099-1108.e12 (2014).
12. Lee, Y. N. *et al.* Characterization of T and B cell repertoire diversity in patients with RAG deficiency. *Sci. Immunol.* **1**, 1–13 (2016).
13. Abraham, R. S. *et al.* Adult-onset manifestation of idiopathic T-cell lymphopenia due to a heterozygous RAG1 mutation. *J. Allergy Clin. Immunol.* **131**, 1421–1423 (2013).
14. Chen, K. *et al.* Autoimmunity due to RAG deficiency and estimated disease incidence in RAG1/2 mutations. *J. Allergy Clin. Immunol.* **133**, 880–2.e10 (2014).
15. Kumánovics, A. *et al.* Estimated disease incidence of RAG1/2 mutations: A case report and querying the Exome Aggregation Consortium. *J. Allergy Clin. Immunol.* **139**, 690-692.e3 (2017).
16. Tirosh, I. *et al.* Recombination activity of human recombination-activating gene 2 (RAG2) mutations and correlation with clinical phenotype. *J. Allergy Clin. Immunol.* **2**, 1–10 (2018).
17. Lawless, D. *et al.* Predicting the Occurrence of Variants in RAG1 and RAG2. *J. Clin. Immunol.* **39**, 688–701 (2019).
18. Notarangelo, L. D., Kim, M. S., Walter, J. E. & Lee, Y. N. Human RAG mutations: Biochemistry and clinical implications. *Nat. Rev. Immunol.* **16**, 234–246 (2016).
19. Notarangelo, L. D., Kim, M.-S., Walter, J. E. & Lee, Y. N. Human RAG mutations: biochemistry and clinical implications. *Nat. Rev. Immunol.* **16**, 234–46 (2016).
20. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).
21. Yang, J. *et al.* Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci.* **117**, 1496–1503 (2020).
22. Giblin, W. *et al.* Leaky severe combined immunodeficiency and aberrant DNA rearrangements due to a hypomorphic RAG1 mutation. *Blood* **113**, 2965–2975 (2009).
23. Ott de Bruin, L. M. *et al.* Hypomorphic Rag1 mutations alter the preimmune repertoire at early stages of lymphoid development. *Blood* **132**, 281–292 (2018).
24. Schatz, D. G. & Oettinger, M. A. The V (D) J Recombination Activating Gene , RAG-1. *Cell* **59**, 1035–1048 (1987).
25. Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. RAG-1 and RAG-2, Adjacent Genes That Synergistically Activate V(D)J Recombination. *Science (80-.).* **248**, 1517–1523 (1990).
26. Walter, J. E. *et al.* Broad-spectrum antibodies against self-antigens and cytokines in RAG deficiency. *J. Clin. Invest.* **125**, 4135–4148 (2015).
27. Yui, M. A. & Rothenberg, E. V. Developmental gene networks: A triathlon on the course to T cell identity. *Nature Reviews Immunology* **14**, 529–545 (2014).
28. Abramowski, V. *et al.* PAXX and Xlf interplay revealed by impaired CNS development and immunodeficiency of double KO mice. *Cell Death Differ.* **25**, 444–452 (2018).
29. Klein, F. *et al.* The transcription factor Duxbl mediates elimination of pre-T cells that fail β-selection. *J. Exp. Med.* **216**, 638–655 (2019).
30. Ohm-Laursen, L., Nielsen, C., Fisker, N., Lillevang, S. T. & Barington, T. Lack of nonfunctional B-cell receptor rearrangements in a patient with normal B cell numbers despite partial RAG1 deficiency and atypical SCID/Omenn syndrome. *J. Clin. Immunol.* **28**, 588–592 (2008).
31. Ijspeert, H. *et al.* Similar recombination-activating gene (RAG) mutations result in similar immunobiological effects but in different clinical phenotypes. *J. Allergy Clin. Immunol.* **133**, 1124-1133.e1 (2014).
32. Ott de Bruin, L. M. *et al.* Hypomorphic Rag1 mutations alter the preimmune repertoire at early stages of lymphoid development. *Blood* **132**, 281–292 (2018).
33. Shannon, C. E. The mathematical theory of communication. *MD. Comput.* **14**, 306–17 (1963).

34. Ji, Y. *et al.* The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* **141**, 419–431 (2010).
35. Giblin, W. *et al.* Leaky severe combined immunodeficiency and aberrant DNA rearrangements due to a hypomorphic RAG1 mutation Leaky severe combined immunodeficiency and aberrant DNA rearrangements due to a hypomorphic RAG1 mutation. 2965–2975 (2011). doi:10.1182/blood-2008-07-165167
36. Krangel, M. S. Mechanics of T cell receptor gene rearrangement. *Curr. Opin. Immunol.* **21**, 133–139 (2009).
37. Vantourout, P. & Hayday, A. Six-of-the-best: unique contributions of γδ T cells to immunology. *Nat. Rev. Immunol.* **13**, 88–100 (2013).
38. Datta, S. & Sarvetnick, N. Lymphocyte proliferation in immune-mediated diseases. *Trends Immunol.* **30**, 430–438 (2009).
39. Marrella, V. *et al.* A hypomorphic R229Q Rag2 mouse mutant recapitulates human Omenn syndrome. *J. Clin. Invest.* **117**, 1260–1269 (2007).
40. Khiong, K. *et al.* Homeostatically proliferating CD4+ T cells are involved in the pathogenesis of an Omenn syndrome murine model. *J. Clin. Invest.* **117**, 1270–1281 (2007).
41. Thornton, A. M. *et al.* Expression of Helios, an Ikaros Transcription Factor Family Member, Differentiates Thymic-Derived from Peripherally Induced Foxp3 + T Regulatory Cells. *J. Immunol.* **184**, 3433–3441 (2010).
42. Walter, J. E. *et al.* Expansion of immunoglobulin-secreting cells and defects in B cell tolerance in *Rag* -dependent immunodeficiency. *J. Exp. Med.* **207**, 1541–1554 (2010).
43. Zinkernagel, R. M., Leist, T., Hengartner, H. & Althage, A. Susceptibility to lymphocytic choriomeningitis virus isolates correlates directly with early and high cytotoxic T cell activity, as well as with footpad swelling reaction, and all three are regulated by H-2D. *J. Exp. Med.* **162**, 2125–2141 (1985).
44. LEHMANN-GRUBE, F., MOSKOPHIDIS, D. & LÖHLER, J. Recovery from Acute Virus Infection. *Ann. N. Y. Acad. Sci.* **532**, 238–256 (1988).
45. Schmitz, J. E. Control of Viremia in Simian Immunodeficiency Virus Infection by CD8+ Lymphocytes. *Science (80-.).* **283**, 857–860 (1999).
46. Fukazawa, Y. *et al.* B cell follicle sanctuary permits persistent productive simian immunodeficiency virus infection in elite controllers. *Nat. Med.* **21**, 132–139 (2015).
47. Rehermann, B. & Nascimbeni, M. Immunology of hepatitis B virus and hepatitis C virus infection. *Nat. Rev. Immunol.* **5**, 215–229 (2005).
48. Hangartner, L., Zinkernagel, R. M. & Hengartner, H. Antiviral antibody responses: The two extremes of a wide spectrum. *Nat. Rev. Immunol.* **6**, 231–243 (2006).
49. Janeway, C. A. & Medzhitov, R. Innate Immune recognition. *Annu. Rev. Immunol.* **20**, 197–216 (2002).
50. Crouse, J., Kalinke, U. & Oxenius, A. Regulation of antiviral T cell responses by type i interferons. *Nat. Rev. Immunol.* **15**, 231–242 (2015).
51. Louten, J., van Rooijen, N. & Biron, C. A. Type 1 IFN Deficiency in the Absence of Normal Splenic Architecture during Lymphocytic Choriomeningitis Virus Infection. *J. Immunol.* **177**, 3266–3272 (2006).
52. Ahmed, R., Salmi, A., Butler, L. D., Chiller, J. M. & Oldstone, M. B. Selection of genetic variants of lymphocytic choriomeningitis virus in spleens of persistently infected mice. Role in suppression of cytotoxic T lymphocyte response and viral persistence. *J. Exp. Med.* **160**, 521–40 (1984).
53. Bergthaler, A. *et al.* Impaired Antibody Response Causes Persistence of Prototypic T Cell–Contained Virus. *PLoS Biol.* **7**, e1000080 (2009).
54. Matloubian, M., Concepcion, R. J. & Ahmed, R. CD4+ T cells are required to sustain CD8+ cytotoxic T-cell responses during chronic viral infection. *J. Virol.* **68**, 8056–8063 (1994).
55. Battegay, M. *et al.* Enhanced establishment of a virus carrier state in adult CD4+ T-cell-deficient mice. *J. Virol.* **68**, 4700–4 (1994).
56. Ciurea, A. *et al.* Viral persistence in vivo through selection of neutralizing antibody-escape variants. *Proc. Natl. Acad. Sci.* **97**, 2749–2754 (2000).
57. Recher, M. *et al.* Extralymphatic virus sanctuaries as a consequence of potent T-cell activation. *Nat. Med.* **13**, 1316–1323 (2007).
58. Lang, K. S. *et al.* Inverse correlation between IL-7 receptor expression and CD8 T cell exhaustion during persistent antigen stimulation. *Eur. J. Immunol.* **35**, 738–745 (2005).
59. Zinkernagel, B. Y. R. M. *et al.* T CELL-MEDIATED HEPATITIS IN MICE INFECTED WITH LYMPHOCYTIC CHORIOMENINGITIS VIRUS Liver Cell Destruction by H-2 Class I-restricted Virus-specific Cytotoxic T Cells as a Physiological Correlate of the " Cr-release assay ? (LCMV) usually causes a lethal . **164**, (1986).
60. Shin, E.-C., Sung, P. S. & Park, S.-H. Immune responses and immunopathology in acute and chronic viral hepatitis. *Nat. Rev. Immunol.* **16**, 509–523 (2016).
61. Gallimore, A. *et al.* Induction and Exhaustion of Lymphocytic Choriomeningitis Virus–specific Cytotoxic T Lymphocytes Visualized Using Soluble Tetrameric Major Histocompatibility Complex Class I–Peptide Complexes. *J. Exp. Med.* **187**, 1383–1393 (1998).
62. Slifka, M. K. & Whitton, J. L. Functional avidity maturation of CD8+ T cells without selection of higher affinity TCR. *Nat. Immunol.* **2**, 711–717 (2001).
63. Daniels, M. A. & Jameson, S. C. Critical Role for Cd8 in T Cell Receptor Binding and Activation by Peptide/Major Histocompatibility Complex Multimers. *J. Exp. Med.* **191**, 335–346 (2000).
64. Thimme, R. *et al.* CD8+ T Cells Mediate Viral Clearance and Disease Pathogenesis during Acute Hepatitis B Virus Infection. *J. Virol.* **77**, 68–76 (2003).
65. Bertoletti, A. The immune response during hepatitis B virus infection. *J. Gen. Virol.* **87**, 1439–1449 (2006).
66. Lang, P. A. *et al.* Aggravation of viral hepatitis by platelet-derived serotonin. *Nat. Med.* **14**, 756–761 (2008).
67. Pircher, H. *et al.* Characterization of virus-specific cytotoxic T cell clones from allogeneic bone marrow chimeras. *Eur. J. Immunol.* **17**, 159–166 (1987).
68. Pircher HP, Moskophidis D, Rohrer U, Bürki, K, Hengartner, H. & Zinkernage, R. M. Viral escape by selection of cytotoxic T cell-resistant virus variants in vivo. *Nature* **346**, 629–633 (1990).
69. Pircher, H. *et al.* Molecular analysis of the antigen receptor of virus-specific cytotoxic T cells and identification of a new Vα family. *Eur. J. Immunol.* **17**, 1843–1846 (1987).
70. Hunziker, L. *et al.* Hypergammaglobulinemia and autoantibody induction mechanisms in viral infections. *Nat. Immunol.* **4**, 343–349 (2003).
71. Vehik, K. *et al.* Prospective virome analyses in young children at increased genetic risk for type 1 diabetes. *Nat. Med.* **25**, 1865–1872 (2019).
72. Getts, D. R., Chastain, E. M. L., Terry, R. L. & Miller, S. D. Virus infection, antiviral immunity, and autoimmunity. *Immunol. Rev.* **255**, 197–209 (2013).
73. Münz, C., Lünemann, J. D., Getts, M. T. & Miller, S. D. Antiviral immune responses: triggers of or triggered by autoimmunity? *Nat. Rev. Immunol.* **9**, 246–258 (2009).
74. Wu, Y.-C. *et al.* High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood* **116**, 1070–1078 (2010).
75. Berland, A. *et al.* PROMIDISα: A T-cell receptor α signature associated with immunodeficiencies caused by V(D)J recombination defects. *J. Allergy Clin. Immunol.* **143**, 325-334.e2 (2019).
76. Riccetto, A. G. L. *et al.* Compound heterozygous RAG2 mutations mimicking hyper IgM syndrome. *J. Clin. Immunol.* **34**, 7–9 (2014).
77. Kato, T. *et al.* RAG1 Deficiency May Present Clinically as Selective IgA Deficiency. *J. Clin. Immunol.* **35**, 280–288 (2015).
78. Buchbinder, D. *et al.* Identification of Patients with RAG Mutations Previously Diagnosed with Common Variable Immunodeficiency Disorders.

J. Clin. Immunol. **35**, 119–124 (2015).

79. Abolhassani, H. *et al.* A hypomorphic recombination-activating gene 1 (RAG1) mutation resulting in a phenotype resembling common variable immunodeficiency. *J. Allergy Clin. Immunol.* **134**, 1375–1380 (2014).

80. Kuijpers, T. W. *et al.* Idiopathic CD4+ T lymphopenia without autoimmunity or granulomatous disease in the slipstream of RAG mutations. *Blood* **117**, 5892–5896 (2011).

81. Abraham, R. S. *et al.* Adult-onset manifestation of idiopathic T-cell lymphopenia due to a heterozygous RAG1 mutation. *J. Allergy Clin. Immunol.* **131**, 1421–1423 (2013).

82. Lee, Y. N. *et al.* Characterization of T and B cell repertoire diversity in patients with RAG deficiency. *Sci. Immunol.* **1**, 1–13 (2016).

83. Rowe, J. H. *et al.* Abnormalities of T-cell receptor repertoire in CD4+ regulatory and conventional T cells in patients with RAG mutations: Implications for autoimmunity. *J. Allergy Clin. Immunol.* **140**, 1739-1743.e7 (2017).

84. Yu, X. *et al.* Human syndromes of immunodeficiency and dysregulation are characterized by distinct defects in T-cell receptor repertoire development. *J. Allergy Clin. Immunol.* **133**, 1109-1115.e14 (2014).

85. IJspeert, H. *et al.* Similar recombination-activating gene (RAG) mutations result in similar immunobiological effects but in different clinical phenotypes. *J. Allergy Clin. Immunol.* **133**, 1124-1133.e1 (2014).

86. Bednarski, J. J. & Sleckman, B. P. At the intersection of DNA damage and immune responses. *Nat. Rev. Immunol.* **19**, 231–242 (2019).

87. Vezys, V. *et al.* Continuous recruitment of naive T cells contributes to heterogeneity of antiviral CD8 T cells during persistent infection. *J. Exp. Med.* **203**, 2263–2269 (2006).

88. Elsaesser, H. J. *et al.* Chronic virus infection drives CD8 T cell-mediated thymic destruction and impaired negative selection. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 5420–5429 (2020).

89. Sandu, I., Cerletti, D., Claassen, M. & Oxenius, A. Exhausted CD8+ T cells exhibit low and strongly inhibited TCR signaling during chronic LCMV infection. *Nat. Commun.* **11**, 1–11 (2020).

90. Moskophidis, D., Lechner, F., Pircher, H. & Zinkernagel, R. M. Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells. *Nature* **362**, 758–761 (1993).

91. Tay, S. S. *et al.* Antigen expression level threshold tunes the fate of CD8 T cells during primary hepatic immune responses. *Proc. Natl. Acad. Sci.* **111**, E2540–E2549 (2014).

92. McLane, L. M., Abdel-Hakeem, M. S. & Wherry, E. J. CD8 T Cell Exhaustion During Chronic Viral Infection and Cancer. *Annu. Rev. Immunol.* **37**, 457–495 (2019).

93. Zander, R. *et al.* CD4+ T Cell Help Is Required for the Formation of a Cytolytic CD8+ T Cell Subset that Protects against Chronic Infection and Cancer. *Immunity* **51**, 1028-1042.e4 (2019).

94. Yermanos, A. *et al.* Profiling Virus-Specific Tcf1+ T Cell Repertoires During Acute and Chronic Viral Infection. *Front. Immunol.* **11**, 1–15 (2020).

95. Ciurea, A., Hunziker, L., Zinkernagel, R. M. & Hengartner, H. Viral escape from the neutralizing antibody response: the lymphocytic choriomeningitis virus model. *Immunogenetics* **53**, 185–189 (2001).

96. Ciurea, A., Hunziker, L., Klenerman, P., Hengartner, H. & Zinkernagel, R. M. Impairment of Cd4+ T Cell Responses during Chronic Virus Infection Prevents Neutralizing Antibody Responses against Virus Escape Mutants. *J. Exp. Med.* **193**, 297–306 (2001).

97. Battegay, M. *et al.* Impairment and delay of neutralizing antiviral antibody responses by virus-specific cytotoxic T cells. *J. Immunol.* **151**, 5408–5415 (1993).

98. Barnaba, V., Franco, A., Alberti, A., Benvenuto, R. & Balsano, F. Selective killing of hepatitis B envelope antigen-specific B cells by class I-restricted, exogenous antigen-specific T lymphocytes. *Nature* **345**, 258–260 (1990).

99. Planz, O., Seiler, P., Hengartner, H. & Zinkernagel, R. M. Specific cytotoxic T cells eliminate cells producing neutralizing antibodies. *Nature* **382**, 726–729 (1996).

100. Crotty, S. T Follicular Helper Cell Differentiation, Function, and Roles in Disease. *Immunity* **41**, 529–542 (2014).

101. Tubo, N. J. *et al.* Single Naive CD4+ T Cells from a Diverse Repertoire Produce Different Effector Cell Types during Infection. *Cell* **153**, 785–796 (2013).

102. Keck, S. *et al.* Antigen affinity and antigen dose exert distinct influences on CD4 T-cell differentiation. *Proc. Natl. Acad. Sci.* **111**, 14852–14857 (2014).

103. Kinoshita, K. & Honjo, T. Unique and unprecedented recombination mechanisms in class switching. *Curr. Opin. Immunol.* **12**, 195–198 (2000).

104. Chaudhuri, J. *et al.* Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **422**, 726–730 (2003).

105. Fischer, A., Provot, J., Jais, J. P., Alcais, A. & Mahlaoui, N. Autoimmune and inflammatory manifestations occur frequently in patients with primary immunodeficiencies. *J. Allergy Clin. Immunol.* (2016). doi:10.1016/j.jaci.2016.12.978

106. Thaventhiran, J. E. D. *et al.* Whole-genome sequencing of a sporadic primary immunodeficiency cohort. *Nature* **583**, 90–95 (2020).

107. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).

108. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).

109. Ramos, P. S. *et al.* A Comprehensive Analysis of Shared Loci between Systemic Lupus Erythematosus (SLE) and Sixteen Autoimmune Diseases Reveals Limited Genetic Overlap. *PLoS Genet.* **7**, e1002406 (2011).

110. González-Serna, D. *et al.* A cross-disease meta-GWAS identifies four new susceptibility loci shared between systemic sclerosis and Crohn's disease. *Sci. Rep.* **10**, 1862 (2020).

111. Messemaker, T. C., Huizinga, T. W. & Kurreeman, F. Immunogenetics of rheumatoid arthritis: Understanding functional implications. *J. Autoimmun.* **64**, 74–81 (2015).

112. The Australo-Anglo-American Spondyloarthritis Consortium, (TASC) Reveille, J., Sims, A. & et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat. Genet.* **42**, 123–127 (2010).

113. Deng, Y. & Tsao, B. P. Advances in lupus genetics and epigenetics. *Curr. Opin. Rheumatol.* **26**, 482–492 (2014).

114. Grammatikos, A. P. & Tsokos, G. C. Immunodeficiency and autoimmunity: Lessons from systemic lupus erythematosus. *Trends Mol. Med.* **18**, 101–108 (2012).

115. Walter, J. E. *et al.* Broad-spectrum antibodies against self-antigens and cytokines in RAG deficiency. *J. Clin. Invest.* **125**, 4135–4148 (2015).

116. Baechler, E. C. *et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc. Natl. Acad. Sci.* **100**, 2610–2615 (2003).

117. Lombardi, A., Tsomos, E., Hammerstad, S. S. & Tomer, Y. Interferon alpha: The key trigger of type 1 diabetes. *J. Autoimmun.* **94**, 7–15 (2018).

118. Ishii, K. J., Koyama, S., Nakagawa, A., Coban, C. & Akira, S. Host Innate Immune Receptors and Beyond: Making Sense of Microbial Infections. *Cell Host Microbe* **3**, 352–363 (2008).

119. Oldstone, M. B. A. Molecular mimicry and immune-mediated diseases. *FASEB J.* **12**, 1255–1265 (1998).

120. Fujinami, R. & Oldstone, M. Amino acid homology between the encephalitogenic site of myelin basic protein and virus: mechanism for autoimmunity. *Science (80-.).* **230**, 1043–1045 (1985).

121.   Fujinami, R. S., von Herrath, M. G., Christen, U. & Whitton, J. L. Molecular Mimicry, Bystander Activation, or Viral Persistence: Infections and Autoimmune Disease. *Clin. Microbiol. Rev.* **19**, 80–94 (2006).
122.   Lehmann, P. V., Forsthuber, T., Miller, A. & Sercarz, E. E. Spreading of T-cell autoimmunity to cryptic determinants of an autoantigen. *Nature* **358**, 155–157 (1992).
123.   Walter, J. E. *et al.* Broad-spectrum antibodies against self- antigens and cytokines in RAG deficiency Find the latest version : Broad-spectrum antibodies against self-antigens and cytokines in RAG deficiency. **125**, 4135–4148 (2016).
124.   Navarini, A. A. *et al.* Vedolizumab as a successful treatment of CTLA-4–associated autoimmune enterocolitis. *J. Allergy Clin. Immunol.* **139**, 1043-1046.e5 (2017).
125.   Coffey, A. J. *et al.* The GENCODE exome: sequencing the complete human exome. *Eur. J. Hum. Genet.* **19**, 827–831 (2011).
126.   Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
127.   Kircher, M., Witten, D., Jain, P., O'Roak, B. & Cooper, G. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310–5 (2014).
128.   Alamyar, E., Duroux, P., Lefranc, M.-P. & Giudicelli, V. IMGT® Tools for the Nucleotide Analysis of Immunoglobulin (IG) and T Cell Receptor (TR) V-(D)-J Repertoires, Polymorphisms, and IG Mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. in 569–604 (2012). doi:10.1007/978-1-61779-842-9_32
129.   Ghraichy, M. *et al.* Maturation of the Human Immunoglobulin Heavy Chain Repertoire With Age. *Front. Immunol.* **11**, 1734 (2020).
130.   Vander Heiden, J. A. *et al.* pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–1932 (2014).
131.   Gupta, N. T. *et al.* Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data: Table 1. *Bioinformatics* **31**, 3356–3358 (2015).
132.   Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–W40 (2013).
133.   Lunter, G. & Goodson, M. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
134.   Pircher, H. *et al.* Viral escape by selection of cytotoxic T cell-resistant virus variants in vivo. *Nature* **346**, 629–633 (1990).
135.   Shinkai, Y. *et al.* RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68**, 855–867 (1992).
136.   Rivers, T. M. & Scott, T. F. M. MENINGITIS IN MAN CAUSED BY A FILTERABLE VIRUS. *J. Exp. Med.* **63**, 415–432 (1936).
137.   Anderson, K. G. *et al.* Intravascular staining for discrimination of vascular and tissue leukocytes. *Nat. Protoc.* **9**, 209–222 (2014).
138.   Eschli, B. *et al.* Early Antibodies Specific for the Neutralizing Epitope on the Receptor Binding Subunit of the Lymphocytic Choriomeningitis Virus Glycoprotein Fail To Neutralize the Virus. *J. Virol.* **81**, 11650–11657 (2007).
139.   Quan, L. Z. *et al.* Identification of autoantibody clusters that best predict lupus disease activity using glomerular proteome arrays. *J. Clin. Invest.* **115**, 3428–3439 (2005).
140.   Battegay, M. *et al.* Quantification of lymphocytic choriomeningitis virus with an immunological focus assay in 24- or 96-well plates. *J. Virol. Methods* **33**, 191–198 (1991).
141.   Giblin, W. *et al.* Leaky severe combined immunodeficiency and aberrant DNA rearrangements due to a hypomorphic RAG1 mutation. *Blood* **113**, 2965–2975 (2009).
142.   Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
143.   Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
144.   Wickham, H. ggplot2: Elegant Graphics for Data Analysis. in 1–7 (Springer New York, 2016). doi:10.1007/978-0-387-98141-3_1
145.   Spinner, M. A. *et al.* GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood* **123**, 809–821 (2014).
146.   Sundaram, L. *et al.* Predicting the clinical impact of human mutation with deep neural networks. *Nat. Genet.* **50**, 1161–1170 (2018).
147.   Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
148.   Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548.e24 (2019).

Manuscript 2: Dominant inheritance of life-threatening multi-organ autoimmunity associated with a novel monoallelic missense mutation in DNA ligase 4

Annaïse J. Jauch[1], Olivier Bignucolo[2,11], Marie Ghraichy[3,4], Ottavia Delmonte[5], Benedikt J. Meyer[1], Florian Marquardsen[1], Valentin von Niederhäusern[3,4], Rebecca Higgins[6], Adhideb Gosh[6,7], Stephen R. Daley[8], Stephan Ehl[9], Anne Rensing-Ehl[9], Helmut Hopfer[10], Thomas Daikeler[12], Tamara Diesch[13], Thomas Kühne[13], Alexander Navarini[6], Johannes Trück[3,4], Jean-Pierre deVillartay[14], Luigi D Notarangelo[5], Mike Recher[1*]

[1]Immunodeficiency Laboratory, Department of Biomedicine, University and University Hospital of Basel, Basel, Switzerland.
[2]Department of Biomedical Sciences, University of Lausanne, Lausanne, Switzerland. SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3]Division of Immunology, University Children's Hospital, University of Zurich, Zurich, Switzerland.
[4]Children's Research Center, University of Zurich, Zurich, Switzerland.
[5]Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA.
[6]Division of Dermatology and Dermatology Laboratory, Department of Biomedicine, University and University Hospital of Basel, Basel, Switzerland.
[7]Competence Center for Personalized Medicine University of Zürich/Eidgenössische Technische Hochschule, Zurich, Switzerland.
[8]Infection and Immunity Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia.
[9]Institute for Immunodeficiency, Center for Chronic Immunodeficiency, Medical Center, Faculty for Medicine, University of Freiburg, Germany.
[10]Institute for Pathology, University Hospital Basel, Basel, Switzerland.
[11]SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland
[12]Department of Rheumatology, University Hospital Basel, Switzerland
[13]Division of Pediatric Oncology/Hematology, University Children's Hospital Basel, Switzerland
[14]Laboratory of Genome Dynamics in the Immune System, INSERM UMR1163, Université Paris Descartes Sorbonne Paris Cité, Institut Imagine, Paris, France.

*Correspondence - Should be addressed to MR: e-mail: mike.recher@usb.ch
The authors declare no conflicts of interest

Abstract

Homozygous and compound heterozygous mutations in *LIG4* encoding DNA-ligase 4 cause an immuno-deficiency syndrome presenting typically early in childhood with life-threatening and/or opportunistic infections, skeletal malformation, radiosensitivity and neoplasia. LIG4 is pivotal during DNA repair *via* nonhomologous end-joining (NHEJ) as it performs the final DNA-break sealing. DNA repair is not only required to assure genome integrity after exogenous or endogenous genotoxic assaults but also physiologically during the generation of the B and T cell receptors after V(D)J recombination. Here, we describe a novel heterozygous *LIG4* mutation, that was associated with dominantly inherited occurrence of autoimmune cytopenia, lymphoproliferation and life-threatening infiltration of nonlymphoid organs. Peripheral blood immunophenotyping, revealed reduced naïve T cells in both patients, low TCR-Vα7.2$^+$ T cells and antibody deficiency. *LIG4* mutated T cells displayed reduced DNA repair capacity and increased susceptibility to genotoxic stress. In keeping with the clinically dominant autoimmune cytopenias, we demonstrate alterations in T$_{reg}$ phenotype and function, altered autoantibody profiles, while changes in the T and B cell receptor repertoire were minimal. The identified *LIG4* missense variant affected the *in vitro* DNA ligation capacity. Using molecular dynamics simulation, we demonstrated that the mutated LIG4 exhibited reduced DNA binding strength.

These results suggest LIG4-haploinsufficiency as a cause of life-threatening immune dysregulation and widen the phenotypic spectrum of the disease manifestation.

Introduction

The three mammalian DNA ligases (LIG1, LIG3, LIG4), are pivotal components for genomic recombination, replication and repair[1]. Especially LIG4 is involved in resolving DNA double-strand breaks (DSB) - the most noxious DNA lesions given their oncogenic potential[1]. DNA DSB can be resolved by the ubiquitous non-homologous end-joining (NHEJ) repair pathway, which utilizes LIG4 for the last step of DNA re-ligation[2]. During repair, the two broken DNA backbone termini encompassing a 3'-hydroxyl (OH) and a 5'-phosphate are fused to form a new phosphodiester bond[3]. LIG4 relies on a three-step Mg^{2+} dependent catalytic cycle, where an

active site lysine (p.K273) engages with the α-phosphate moiety of ATP to form a covalent intermediate. Consequently, the pyrophosphate is released, allowing the AMP moiety to be relocated from the LIG4 active site lysine to the 5'-phosphate of the DNA terminus. Lastly, the 3'-OH of the second DNA terminus attacks the phosphate atom at the 5'-end, releasing the AMP moiety and a new phosphodiester bond is formed[3].

NHEJ is preferentially used after genotoxic assaults like ionizing radiation and reactive oxygen species as well as physiologically during V(D)J recombination, a crucial step in the T and B cell receptor generation[2]. The non-redundancy of different NHEJ proteins in the development and/or functioning of the immune system is well evidenced by immunodeficient patients harboring loss of function mutations[4]. LIG4-dependent V(D)J recombination is mandatory for the adaptive immunity development, as the variability and consequently the recognition of a plethora of antigens is ensured by the semi-stochastic recombination of the variable (*V*), diversity (*D*) and joining (*J*) gene segments encoding the variable domains of both B cell resp. T cell receptors[5]. A well-regulated DNA-damage response is therefore imperative for immune homeostasis, to reduce the severe conditions of immunodeficiency and autoimmunity/autoinflammation.

Major defects in proteins involved in NHEJ result in severe combined immunodeficiency (SCID) associated with lethality in the first month of life. In contrast, hypomorphic mutations with residual protein functionality may manifest as "leaky SCID" and even combined immunodeficiency with adult-onset[6]. Although the first patient with LIG4 deficiency syndrome was characterized > 30 years ago, only 65 patients with either homozygous or compound heterozygous mutated *LIG4* have been published to date (suppl. table S3). The identification of additional patients is central for ameliorating our understanding of the natural history of this condition. *LIG4* haploinsufficiency caused by monoallelic *LIG4* mutations has not been reported in human patients, whereas murine data suggests that a single functional *LIG4* allele is not sufficient to protect from malignancy and may be associated with decreased survival[7–9]. Here we identified a novel mono-allelic missense mutation in *LIG4* associated with impaired DNA binding capacity in two patients (father and daughter). Although the B and T cell receptor repertoire showed only subtle changes in their diversity, both patients developed life-threatening autoimmunity associated with phenotypic and functional dysregulations in T and B cells. We utilized a novel ensemble of molecular dynamic simulation tools to characterize the dysfunctional interaction of the mutated LIG4 with the DNA on an atomic level.

Identification of dominantly inherited *LIG4* haploinsufficiency has important diagnostic, therapeutic and prognostic implications.

Added value of this study to the field of inborn errors of immunity

For the first time, we report a monoallelic *LIG4* haploinsufficiency in two patients with life-threatening autoimmunity associated with reduced DNA binding capacity. This widens the phenotypic spectrum of LIG4-dependent immunodeficiency and prompts to consider *LIG4* haploinsufficiency in patients with immune-dysregulation of unknown cause.

Results

Dominantly inherited immune-dysregulation

The index patient (P1), first child of non-consanguineous parents, presented at the age of three years with sequential development of autoimmune hemolytic anemia (AIHA) and immune thrombocytopenia (ITP) (fig. 1a), which was responsive to cycles of systemic steroids. P1 experienced lymphoproliferation (splenomegaly and lymphadenopathy) and multiple infections including opportunistic pathogens (fig. 1a). The patient developed biopsy-proven interstitial nephritis with T and B cell infiltration at the age of eleven years (fig. 1b). Basic immune-suppression with mycophenolate was installed at this time. At the transition into the adult immunology service, agammaglobulinemia was noted and immunoglobulin replacement therapy was started. Despite normalized serum IgG levels, the patient developed life-threatening autoimmune pneumonitis with

polyclonal B and T cell infiltration (fig. 1a, 1c, 1e), resolving under high dose systemic steroids, which could be fully tapered (fig. 1d). At 24 years, P1 developed sterile granulomatous parotitis and inflammation of both lacrimal glands (fig. 1a, 1f). The family history was noteworthy for the circumstance that her father and two paternal uncles experienced several steroid-sensitive adult-onset ITP episodes.

Given the history of lymphoproliferation, lymphocytic organ infiltration and familial occurrence of autoimmune cytopenia, a detailed immunological evaluation was performed in P1 and her father. In the absence of immune-modulating treatment, the father had mildly reduced lymphocytes (1.02 x10^9/L) (fig. 1g) and reduced thrombocyte counts (114 x10^9/L) (suppl. table S1). Analysis of peripheral blood-derived lymphocytes revealed a reduced frequency of naïve CD27$^+$CD45RO$^-$ T cells in both patients (fig. 1h, 1i), affecting both helper CD4$^+$ T cells (fig. 1h, 1i) and cytotoxic CD8$^+$ T cells (suppl. fig. S1a). CD8$^+$ T cells from both patients displayed an augmented *in vitro* proliferation capacity after anti-CD3/anti-CD28 stimulation, seen by cell trace violet (CTV^{TM}) dilution (fig. 1j). A reduction of T cells bearing the TCR Vα7.2$^+$ was noted in the daughter and the father (fig. 1k). Previous reports had described the diagnostic impact of low TCR Vα7.2$^+$ T cells specifically in patients with combined-immunodeficiency (CID) related to defects in the V(D)J recombination[10,11]. Concordantly, we observed a similar reduction of TCR Vα7.2$^+$ T cells in some of the CID patients in our inborn errors of immunity (IEI) cohort, while patients with primary antibody deficiency (PAD) and autoinflammation (Autoinflamm) showed median frequencies closer to the one observed in healthy controls (fig. 1l). In our patients, the TCR Vα7.2$^+$ T cell diminution affected both mucosal-associated invariant T cells (MAIT, CD161$^+$Vα7.2$^+$) and non-MAIT cells (CD161$^-$Vα7.2$^+$) (suppl. fig. S1b). Despite the reduced MAIT cell frequency in P1, MAIT cells showed a comparable activation phenotype after bacterial extract stimulation (suppl. fig. S1cs). The mother of P1 did not demonstrate any of the immunological abnormalities observed in P1 and her father.

In summary, a dominantly inherited familial immune dysregulation, clinically dominated by autoimmune cytopenias was linked with a shared immunological phenotype of low naïve T cells and low TCR Vα7.2$^+$ T cells. Since the latter are hallmarks observed in patients with V(D)J recombination defects[10,11], we performed high throughput sequencing experiments for the T cell receptor (TCR) resp. B cell receptor (BCR).
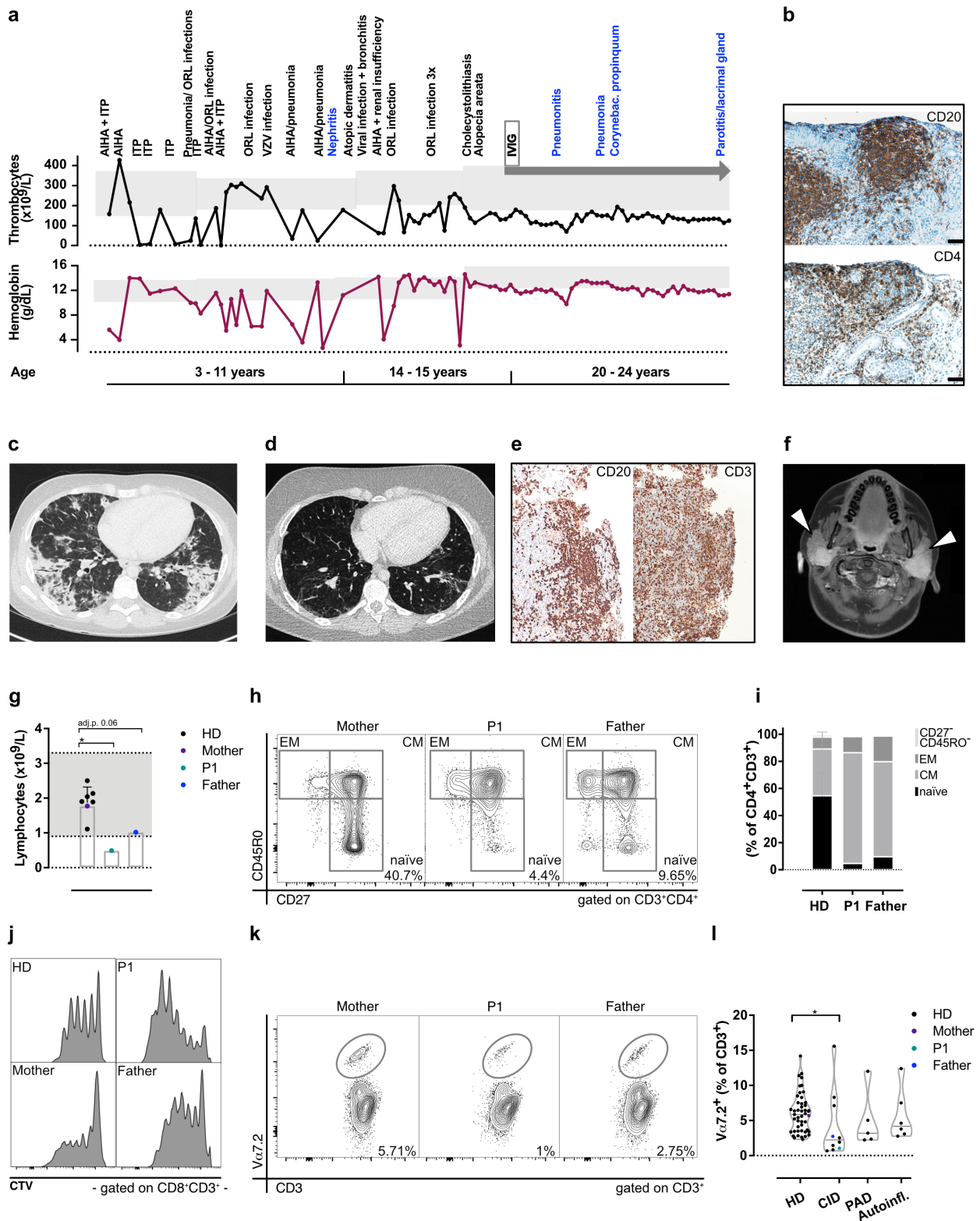
Figure 1| Multiple autoimmune manifestations and reduction of naïve T cells in the peripheral blood. a) The time course of the clinical manifestations in the index patient P1. Her thrombocyte counts (black line) and hemoglobin levels (red line) values are depicted over time. On top: Blue highlights life-threatening clinical episodes, grey background depicts age-matched reference range for thrombocyte counts and hemoglobin respectively, start of the intravenous immunoglobulin (IVIG) treatment is depicted. Otorhino-laryngo ORL, varicella-zoster virus VZV. b) Kidney biopsy from patient P1 during the interstitial nephritis episode at the age of eleven years, with decreased renal function. Immunohistochemistry staining with anti-CD20 and anti-CD4. Displaying tubular and interstitial lymphocytic infiltration (scale bars, 50μm). c) High-resolution chest computer tomography (CT) scan of the lungs of P1 during the pneumonitis episode. Bipulmonal peribronchovascular

and interstitial infiltration, with peripheral consolidations were seen. **d)** Transversal pulmonary CT scan after steroid treatment. **e)** Lung biopsy specimens were obtained during the pneumonitis episode and stained with anti-CD20 and anti-CD3. **f)** Cranial magnetic resonance imaging (MRI), showing bilateral swelling of the parotid gland (white arrowhead, image T1). **g)** Lymphocyte numbers were enumerated in whole blood obtained from P1 and her parents. Grey shading indicates the in-house reference range of healthy individuals. **h)** Representative flow cytometry plots for T cells from the peripheral blood of the mother, the father and P1 (gated on CD4$^+$CD3$^+$ T cells). Boxes indicate naïve (CD27$^+$CD45RO$^-$), effector memory (EM, CD27$^-$CD45RO$^+$) and central memory (CM, CD27$^+$CD45RO$^+$) CD4$^+$ T cells, numbers adjacent to the boxes indicate the frequency within the CD4$^+$CD3$^+$ T cell population. **i)** Enumeration of naïve, EM, CM and CD27$^-$CD45R0$^-$ T cell frequency within the CD4$^+$CD3$^+$ T cell population in P1 and her father compared to healthy controls (n=44). **j)** Representative T cell proliferation flow cytometric plots. T cells from peripheral blood were labeled with CellTrace^{TM} violet (CTV) and stimulated *in vitro* with anti-CD3/anti-CD28. Cell proliferation was determined on day 5 *via* the quantification of cell dye dilution. **k)** Enumeration of T cells bearing the TCR Vα7.2 segment were analyzed in P1 and her parents by flow-cytometry. Representative flow cytometric plots are depicted. The number indicates the frequency within the CD3$^+$ T cell population. **l)** Comparison of the TCR Vα7.2$^+$ T cell frequency in P1 and her father with patients affected by different entities of inborn errors of immunity (combined immunodeficiency CID, primary antibody deficiency PAD, autoinflammation Autoinfl. or to healthy donors (HD, n= 44). P1 and her father were categorized as CID patients. In (g+l) significance was computed using the non-parametric Kruskal-Wallis test with Dunn's correction. In (g) the SD of the HD group was added manually to the mean of P1 respectively her father to allow for variance testing.

TCR high-throughput sequencing reveals a preserved repertoire

The most common TCR loci (*TRA*, *TRB*) were sequenced by high-throughput sequencing of DNA derived from peripheral blood T cells from P1 and her parents.

Abnormalities in the TCR most variable region, the complementarity-determining region 3 (CDR3), may impact the recognition of foreign- and self-antigens[12] and irregularities can be assessed by analyzing its length[13]. The distribution of the CDR3 lengths in the T cell receptor α-chain (*TCRA*, fig. 2a) and β-chain (*TCRB*) sequences (suppl. fig. S2a) were comparable between P1, her father and mother. Additional metrics offer the opportunity to study repertoire skewing occurring during thymocyte development or in the periphery[14]. To account for repertoire diversity and clonality, the Shannon's (H) entropy[15] respectively Simpson's clonality[16] indices were computed, and found to be unchanged in P1's and her father's *TCRA* and *TCRB* sequences (fig. 2b, 2c).

We focused on the individual *V* gene segment usage within unique *TCRA* sequences, as the TCR-α-locus can adopt a directional multistage recombination, which is halted only upon positive thymocyte selection[17]. We found the V-gene segment *27-01-03* to be significantly overrepresented in the two patients compared to healthy controls (fig. 2d, suppl. fig. S2c). Additionally, trends were observed for a preferential usage for the rather distal *TRAV04-01-01* and a trend for lower abundance for *TRAV27-01* was noted (fig. 2d, suppl. fig. S2c).

To investigate the pairing of individual *V* with *J* gene segments, heatmaps were computed. The pairing of individual V with J gene segments was overall maintained and no significant clonal expansions were observed, neither in total (suppl. fig S2b) nor unique *TCRA* sequences (fig. 2e) of P1 and her father compared with the healthy mother. We assumed that especially distal *V* and *J* segments might be underrepresented in light of suboptimal distal VJ pairing capacity, as observed in conditions with V(D)J recombination defects[10]. An under-representation of distal gene segment usage was however not observed in total or unique *TCRA* transcripts of P1 and her father (fig. 2e, suppl. fig S2b, square).

Altogether, the *TRA* and *TRB* sequencing data underpins a diverse repertoire with no specific skewing of the CDR3 length and only subtle changes regarding the utilization of individual *TCRAV* genes.
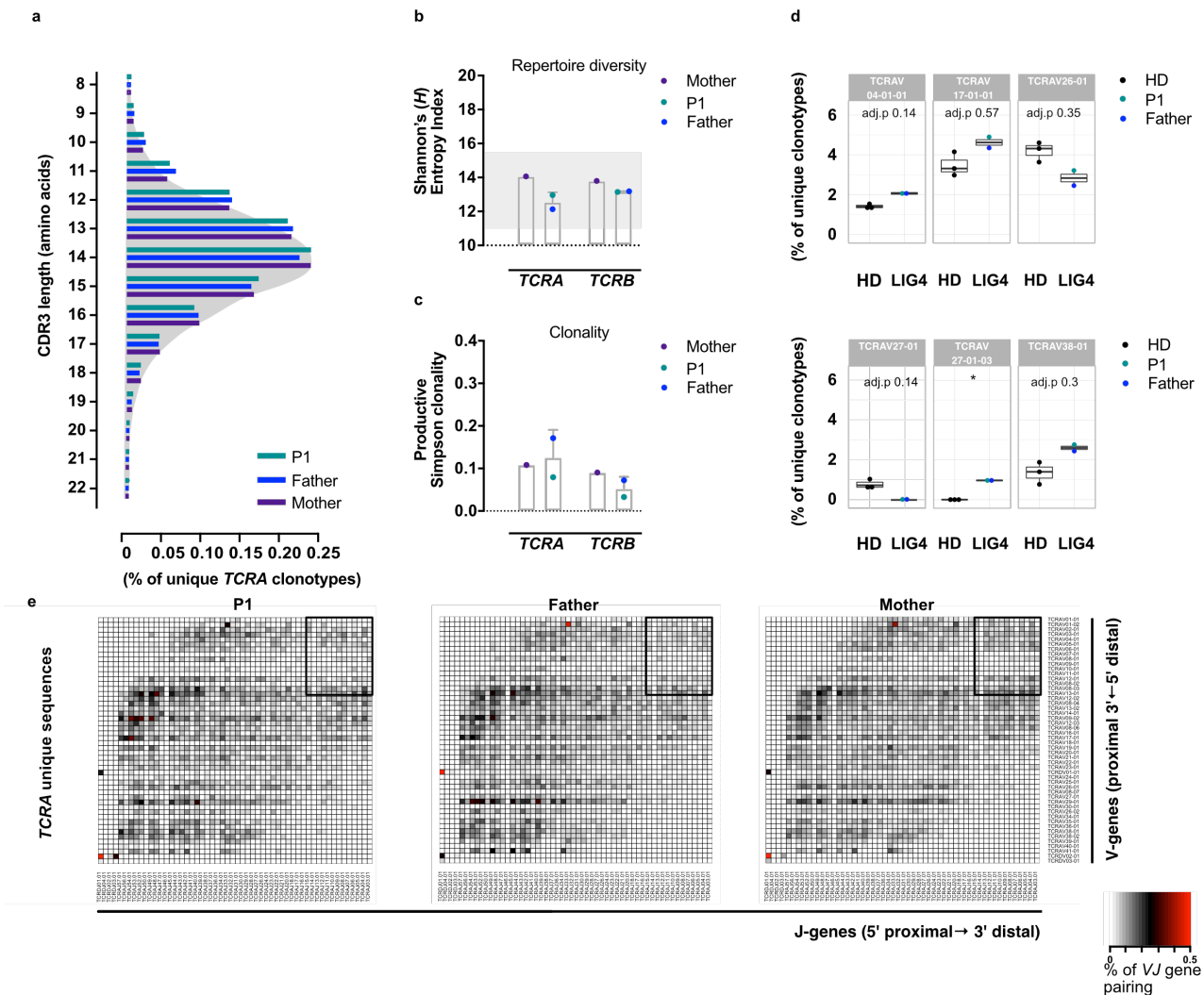
Figure 2| P1's and her father's peripheral T cells harbor preserved *TRA* and *TRB* repertoires. High throughput sequencing experiment of the T cell receptor α-chain (*TRA*) and β-chain (*TRB*) loci of peripheral blood-derived bulk T cells. **a)** The most variable non-germline region of the TCR, namely the (complementary determining region 3) CDR3 was investigated regarding the relative distribution of its length within unique *TCRA* sequences. The CDR3 length distribution of the healthy mother is highlighted as a grey shadow. **b)** The *TCRA* and *TCRB* repertoire diversity was calculated using Shannon's (H) entropy index, high values indicate a greater diversity (dashed lines represent values from healthy controls in[18]). **c)** The clonal expansion of clonotypes was computed with the productive Simpson clonality index. Values range from 0 to 1, with 0 being each clone appears only one time and 1 reflecting the presence of one unique clone. **d)** Within the unique *TCRA* clonotypes the differential expression was computed regarding the individual *V* gene segment usage in P1 and her father compared to three healthy controls (HD). **e)** For the analysis of *VJ* gene pairing within unique *TCRA,* heatmaps were computed displaying the *V*-gene (rows) and *J*-gene (columns) pairing frequency. Most distal *V* with distal *J* gene pairing is indicated by a box in the top right of individual heatmaps. Differential expression analysis in (d) was assessed with the linear modeling approach and an empirical Bayes method with a false discovery rate of 5%.

The B cell receptor repertoire harbors subtle quantitative and qualitative impairments

To investigate if the autoimmune disposition in P1 and her father could reflect differences in B cell subsets and/or B cell receptor (BCR) repertoire, B cells were immunophenotyped by flow cytometry and RNA-derived from peripheral blood B cells was sequenced using isotype-resolved barcode based adaptive immune receptor repertoire-sequencing (AIRR-seq) technology.

P1 displayed a reduction of absolute B cell numbers in the peripheral blood. Proportionally mainly memory B cells and plasma blasts were reduced (suppl. table S1). The utilization of the two BCR light chains (kappa κ *vs*. lambda λ) was inverted in P1's B cells compared to healthy controls (fig. 3a). Both patients had an

increased percentage of CD21^{low} B cells, known to be associated with autoimmunity[19] (suppl. table S1). Before immunoglobulin replacement therapy was started in P1, her serum IgM, IgA and IgG levels were all below the detection limit. Her father had normal serum IgM levels, but a slight reduction of the class-switched isotypes IgG and IgA, whereas the mother's immunoglobulins were within the normal range (fig. 3b, suppl. table S1).

The ¡ locus was investigated by high throuput sequencing, to study the maturation status of peripheral blood B cells on transcript level. The vast majority of P1's B cells included unmutated naïve and mutated IgM/IgD (*MD*) memory transcripts (fig. 3c). IgG (*IGHG*) and IgA (*IGHA*) transcripts were barely detectable (fig. 3c), paralleling P1's serum immunoglobulins. The father's B cells contained mostly *MD* and *IGHG* transcripts (fig. 3c). Analyzing more closely the *IGHA* and *IGHG* subclasses in the two patients, a normal distribution of *IGHA1* and *IGHA2* was seen in both patients (fig. 3d). Both patients displayed a tendency for a reduced frequency for the *IGHG2* subclass (fig. 3d). In addition, P1's B cells transcripts showed a skewing towards the utilization of the *IGHG3* subclass. The IgG3 constant region is the first class-switched constant region found on chromosome 14 (fig. 3e).

We observed that P1's MD memory B cells had an inverted usage of the V_H3 and V_H4 gene families (fig. 3f). In the father's transcripts an augmented utilization of the J_H6 gene segment in naïve, MD memory and IGHG memory B cells was detected (fig. 3g). In both patients the *MD* and *IGHG* memory B cell transcripts harbored less abundantly the J_H4 gene segment (fig. 3g).

After a B cell has encountered its antigen, three maturation steps are initiated, involving somatic hypermutation (SHM, affinity maturation of variable regions), class switch recombination (CSR, deletional change of the constant region) and antigenic selection. The first two steps involve DNA damage responses[20]. Affinity maturation was analyzed via the count of SHM enumerated in *V*-genes. The level of SHM detected in the *IGHG* and *MD* memory B cell transcripts were below the normal range for P1 and marginally low in the MD memory compartment for the father (fig. 3h). A process that points at antigen selection is an increased ratio of replacement mutations (R, change of amino acid) compared to silent mutations (S, no alteration of amino acid) (R/S ratio) in the complementary determining regions (CDR). If positive selection is undertaken, the R/S ratio is found to be higher as by chance [21,22]. P1's *IGHG* and *MD* memory B cell transcript showed a decreased R/S ratio compared to healthy age-matched controls (fig. 3i), while in the father's B cell transcripts the R/S ratio was only marginally low in the MD memory B cells (fig. 3i).

Altogether, in P1 we observed a reduction of memory B cells and plasma blasts, a lack of class-switched serum immunoglobulins, an inverted κ/λ ratio, a reduced number of SHM and antigen selection - pointing at B cell immaturity. Both patients shared a higher proportion of CD21^{low} B cells and a well-diversified IGH repertoire with subtle differences in the usage of the V_H3, V_H4 and J_H4 regions as well the J_H6 gene segment in the father.
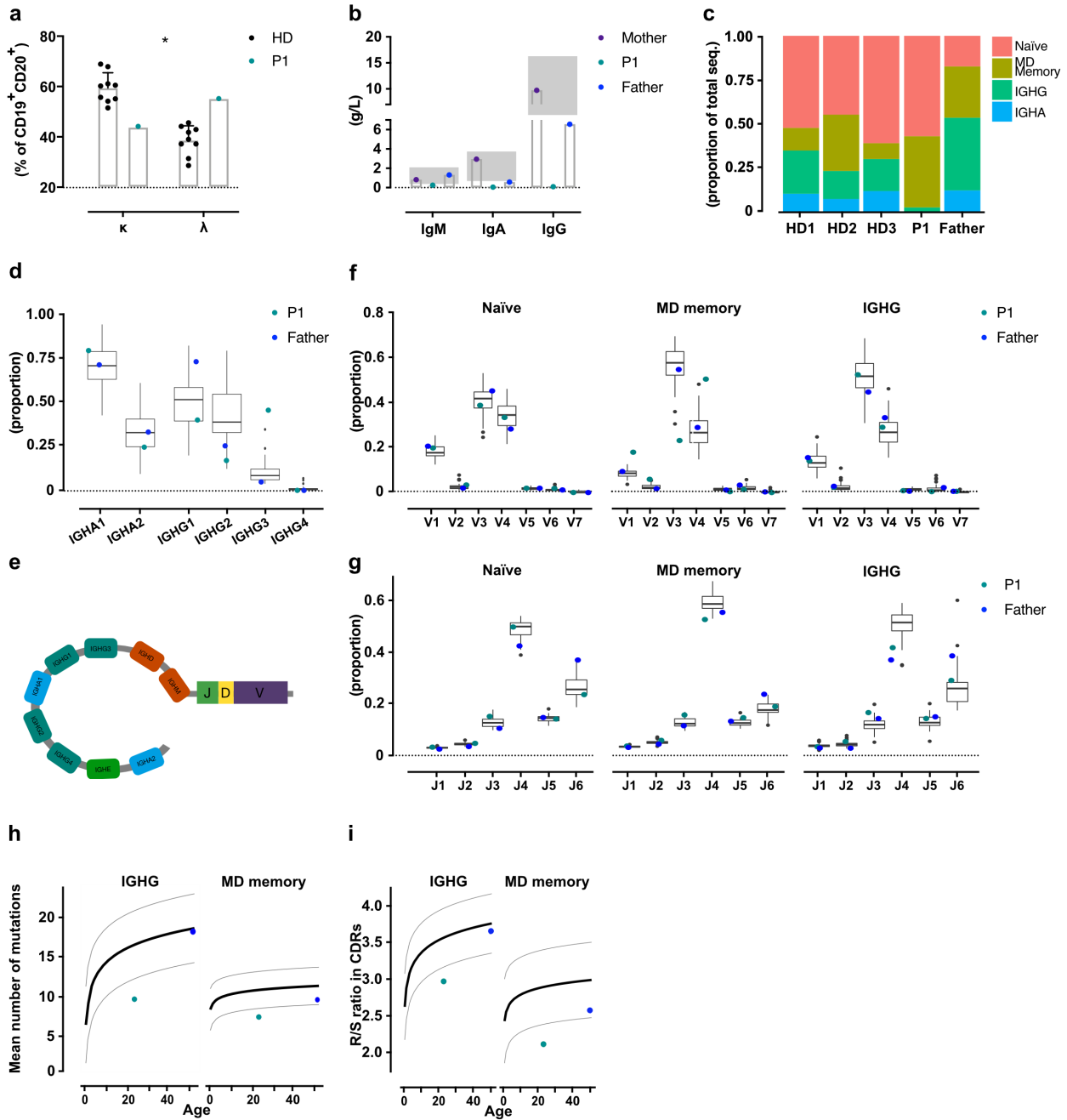
Figure 3 | Inversed light chain surface expression on P1's B cells and discrete changes in IGH *V* and *J* gene usage in both patients. a) Flow cytometric analysis of peripheral blood B cells, regarding the surface expression of the two BCR light chains κ and λ in P1 compared to control subjects. **b)** Serum IgM, IgA and total IgG immunoglobulin levels were measured in P1 and her parents. Grey shading indicates the in-house reference range. **c)** High-throughput RNA sequencing of the *IGH* locus derived from peripheral blood B cells. Transcripts of total sequences (seq.) were used to determine the maturation status and constant region gene usage within B cells in the two patients and healthy donors. **d)** Relative quantification of the IgA and IgG subclass utilization in P1 and her father, which was inferred from *IGH* sequences. Box-plot indicates values found in healthy donors, with median the box displaying 25-75% and error bars 5-95%. **e)** Cartoon for the constant region organization of the *IGH* locus on chromosome 14. Freely adapted from[23]. **f)** *V* family and **(g)** *J* gene segment usage were analyzed in *IGH* transcripts derived from naïve, IgM/IgD memory (*MD*) and IgG memory (*IGHG*) B cells of P1 and her father. Box-plot indicates values found in healthy donors. **h)** Average of somatic hypermutations (SHM) was determined by calculating the *V* gene mutations in unique IgM/IgD (*MD*) and IgG (*IGHG*) B cell memory transcripts of P1 and her father. The black line indicates the median of the respective age group, gray lines indicate the 95% confidence interval. **i)** Antigen selection was quantified by the computation of the mean replacement/silent (R/S) ratio in the *IGH* CDRs of IgG (IGHG) and IgM/IgD (MD) memory B cells of the two patients. The black line indicates the median of the respective age group, gray

lines indicate the 95% confidence interval. Significance was computed in (a) using a Mann-Whitney test with post-hoc correction, the SD of the controls was manually added to the value of P1 to allow for variance testing.

Identification of a novel heterozygous *LIG4* missense mutation

We next investigated PBMC-derived DNA of P1, her parents and the clinically healthy younger brother using whole-exome sequencing (WES), followed by filtering with a custom-designed IEI disease gene panel. We did not observe any shared homozygous variants in P1 and her father. However, in both diseased individuals we detected a novel heterozygous missense variant in DNA-ligase 4 (*LIG4*, c.G1739A) (table 1, suppl. table S2). Heterozygosity was confirmed by Sanger sequencing of PBMCs-derived cDNA (fig. 4a). The mother and brother of P1, who both did not display any sign of immune dysregulation, did not carry the *LIG4* variant (fig. 4a). Mammalian cells rely mainly on the DNA double-strand break (DSB) repair pathway called non-homologous end-joining (NHEJ), which requires the LIG4 enzyme for the final ligation of broken DNA ends[24,25]. The base triplet affected by the missense mutation at position c.G1739A encoded for the replacement of an arginine at position 580 by a glutamine (p.R580Q). The arginine at position 580 is highly conserved across various vertebrates (fig. 4b) and locates within the oligonucleotide/oligosaccharide-binding domain (OBD), which is crucial for complete LIG4 encirclement of the DNA during NHEJ[26] (fig. 4c). The missense mutation was predicted to have a severe impact on the LIG4 protein (CADD-PHRED score 33[27], PolyPhen2[28] score 1 and SIFT[29] score 0. According to the genome browser of the Exome Aggregation Consortium (ExAC), [01/2021]), the missense variant is a single nucleotide variant (SNV) with an allele frequency of 3.6^{-4} (table 1, suppl. table S2). This *LIG4* variant had not been described in the literature previously (suppl. table S3). In P1 and her father, we excluded the presence of the known N-terminal p.A3V/T9I *LIG4* polymorphisms, which have been described to reduce LIG4 function[30] (data not shown).

Table 1| Whole-exome sequencing information regarding the missense mutation in *LIG4* and *FAS*.

Genetic location	Nucleotide change (c.DNA)	Zygosity	Gene symbol	AA change	rsID	ESP	Poly Phen2	SIFT	CADD-PHRED	gnomAD allele freq
13_108209530	c.A1739G	Het	*LIG4*	p.R580Q R[CGA] > Q[CAA]	rs146616552	0.0	Deleter-ious, 1	Deleter-ious, 0	33	3.6^{-4}
10_90768694	c.G383A	Homo (Father) Het (P1, Brother)	*FAS*	p.R128K R[AGA] > K[AAA]	na	0.0	0.541	0.241	23.6	na

AA amino acid, Allele freq allele frequency, CADD-PHRED Combined Annotation Dependent Depletion, Chr. chromosome, ESP Exome Sequencing Project, gnomAD Genome Aggregation Database, Het heterozygous, Homo homozygous, PolyPhen2 Polymorphism Phenotyping, rsID single nucleotide polymorphism database identifier, SIFT Sorting for Intolerant from Tolerant.

In addition, a novel homozygous missense mutation in *FAS* (c.G383A, p.R128K) was detected in the father (table 1, suppl. table S2) and confirmed by Sanger sequencing (data not shown). Both children, P1 and her healthy brother, were heterozygous carriers for this *FAS* variant. Mutations in *FAS* are the most common genetic cause for autoimmune lymphoproliferative syndrome (ALPS), impairing FAS related apoptosis, resulting in chronic polyclonal lymphoproliferation and autoimmunity[31]. While P1 and her father showed a decreased proportion of naïve CD4[+] T cells, the brother who carried the same heterozygous *FAS* variant as P1 displayed a normal percentage of naïve CD4[+] T cells (suppl. fig. S4a). Diagnostic criteria for ALPS include persistent lympho-proliferation (after exclusion of secondary causes) and elevated frequency of double negative CD4[-]CD8[-] TCRαβ[+] T cells (DNT ≥ 1.5% of lymphocytes, ≥2.5% of CD3[+] T cells). Neither the father nor P1 displayed elevated absolute lymphocyte counts (fig. 1g), but P1 had suffered from lymphoproliferative episodes in the past (fig. 1a). DNTs were marginally elevated in the father (suppl. fig. S3b, S3c) but were within the normal range in P1. In keeping, soluble serum FAS ligand (FASL) and serum vitamin B12 levels were repetitively normal in P1, while those are typically elevated in ALPS[31] (suppl. fig. S3d). Nevertheless, T cell blasts-derived from the two patients *vs.* healthy

controls were tested for their *in vitro* resistance to FAS-induced apoptosis. By incubating the cells with either agonistic anti-FAS antibody or with recombinant FAS-ligand (FASL), we did not overserve a reduced FAS-related apoptosis-sensitivity in T cells of the father or P1 (suppl. fig. S3e). Based on these experimental observations and the fact that the healthy brother carried the same heterozygous *FAS* variant, we excluded a functionally relevant impact of the rare *FAS* variant and focused on the characterization of the *LIG4* variant.

Next, mRNA levels of *LIG4* and the additional mammalian DNA-ligases *LIG1* and *LIG3* were determined by reverse transcription quantitative PCR (RT-qPCR) in PBMCs of the two patients, the mother and healthy controls. Levels of *LIG4* mRNA showed a trend to be decreased in the father when compared to healthy controls including the mother (fig. 4d). In both patients we found preserved expression levels of *LIG1* and *LIG3* mRNA compared to control PBMCs (suppl. fig. S3f). Immunoblots derived from PHA T cell blasts of P1, her mother and brother revealed conserved LIG4 protein levels in P1's T cells (fig. 4e).
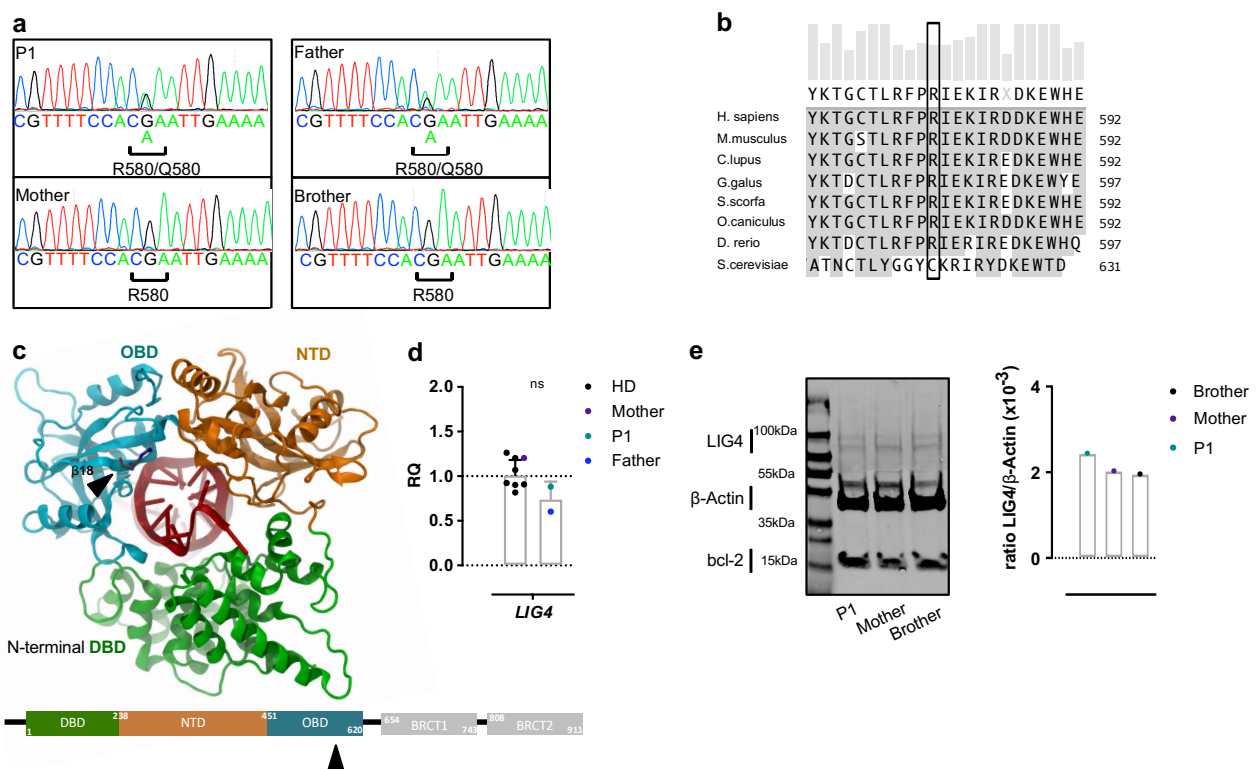
Figure 4 | Novel missense mutation affecting the catalytic core of DNA ligase 4. a) Confirmation of the missense mutation at c.A1739G by Sanger sequencing of bulk T cell-derived DNA, the resulting amino acid change at p.580 (R arginine to Q glutamine) is indicated. Heterozygous mutations are seen as two superimposed peaks in the chromatogram. **b)** Multiple LIG4 protein sequence alignment. The position of the arginine 580 is highlighted with a box. The height of bars on top of the alignment indicates the degree of conservation across the species. **c)** Molecular representation of the human LIG4 catalytic core bound to a DNA duplex is shown in ribbons. The wild-type arginine at position 580 is shown as stick and is demarked with an arrow. The corresponding β sheet 18 is indicated. At the bottom: Schematic of the entire LIG4 protein, the mutated residue resides in the catalytic oligonucleotide/oligosaccharide-fold domain (OBD, blue). Numbers indicate the amino acid position in NP_001091738. BRCT1 BRCA1 C terminus; BRCT2 BRCA2 C terminus; DBD DNA binding domain in green; NTD nucleotidyltransferase in orange. **d)** Reversed transcription RT-qPCR was used to measure *LIG4* mRNA levels in PBMCs of the two patients and healthy controls including the mother. The relative quantity (RQ) was normalized to different housekeeping genes and to the mean of the HDs. **e)** The LIG4 protein levels were quantified by separating PHA T cell blast cell lysates by SDS-PAGE electrophoresis and probed with rabbit-anti LIG4. Right side normalization of LIG4 protein levels to β-actin levels. Variance testing in (d) was performed using the non-parametric Mann-Whitney rank test.

The clinical phenotype of the two *LIG4* mutation carriers pointed to a protein loss of function associated with the R580Q mutation. To experimentally addressed this, we performed substrate ligation assays comparing the enzymatic activity of the mutant (R580Q) *vs*. the wild-type (WT) LIG4 protein. Recombinant WT and mutant LIG4 proteins (catalytic domain only) were produced and concentrations were normalized (fig. 5a). As ligation substrate, a 42bp nicked oligonucleotide duplex was used with an attached fluorescent dye (fig. 5b). Using increasing concentration of DNA substrate (fig. 5c) and reaction duration (fig. 5d), we observed reduced amounts of ligated products in presence of R580Q LIG4 as compared to WT LIG4.

The reduced biochemical ligation activity of the mutant R580Q LIG4 prompted us to study the LIG4-DNA interaction at the structural level. We performed molecular dynamics simulations, an ensemble of tools that proved efficient to help interpreting the effect of mutations on protein function and structural propensities[32–34]. The simulations were designed to investigate the catalytic domain of LIG4 in closed conformation with a nicked adenylated-DNA substrate (PDB 6BKG). Six independent unbiased trajectories of > 500ns each, for the WT *vs*. the R580Q mutant protein, were performed, totalizing in ~6μs of trajectory. The residue 580 is located in the β18 of the oligonucleotide/oligosaccharide-fold (OBD) subdomain[35]. The crystal structures suggest that the OBD domain (blue, fig. 5e) undergoes a complex conformational change to encircle the DNA during the open-to-close transition, whereas the neighboring nucleotidyltransferase domain (NTD) remains bound to the DNA (orange, fig. 5e). In the closed conformation, the arginine 580 interacts with the broken 5' AMP-carrying DNA strand, with its guanidium moiety within reach of a salt bridge formation (~ 5 Å) with two phosphate groups (fig. 5e). We reasoned that an arginine residue at position 580 would stabilize the protein-DNA complex and its replacement by a glutamine residue would weaken this interaction. Since DNA is mostly negatively charged, positively charged residues like arginine and lysine are abundant in DNA binding proteins. The molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA)[36–38] method was used to calculate the free binding energy between the WT *vs*. R580Q LIG4 to the DNA duplex. Lower binding energy values (BE) indicate a stronger association between the protein and the DNA. The WT LIG4 had significantly lower BE than the mutant enzyme (fig. 5f, suppl. fig S4a), confirming that the amino acid change from an arginine to a glutamine residue at position 580 would weaken the complex.

Intrigued by the BE weakening of the mutant LIG4, we hypothesized that the mutant glutamine would dislocate neighboring residues and their interaction with the DNA. We addressed the question, by analyzing each of the 632 residues of the catalytic domain for their binding energy with the DNA. In most cases, the BE difference of residues in the WT and the R580Q enzyme was negligible (suppl. fig S4b). Only six residues showed an absolute difference of >20 kJ/mol between the residue in the WT LIG4 and the corresponding residue in the mutant R580Q LIG4 (fig. 5g), of which only the BE difference for residue 580 reached statistical significance (fig. 5g). Thus, surrounding residues were not able to compensate for the lower binding at position 580.

We focused the conformational analysis on the interactions of the residue with the DNA backbone and on their torsion angles. The dihedral χ1 angle indicates the orientation of the side-chain with respect to the protein main chain. Minimal oscillations during the whole set of time series for the WT were seen, whereas we observed greater dihedral χ1 angle fluctuation for the mutant R580Q (fig. 5h), hinting at a lower stability of the glutamine 580 as compared to arginine 580. The χ1 angles of the mutated R580Q residue showed a bimodal distribution, with orientations either at +60 or -60°, while the WT residue had a unimodal distribution at +60° only (fig. 5i). The fluctuations of the mutant glutamine 580 affected the secondary structure, since the values of the torsion angles ϕ and ψ, which describe the secondary structure propensities, fluctuated more for the mutated glutamine 580 residue compared to the WT residue (suppl. fig. S4c – S4f). A χ1 angle of ~60° at position 580 corresponds to a side-chain oriented towards the 5' end of the nicked adenylated DNA strand in the WT situation (fig. 5j). Consequently, the WT residue engaged in strong interactions with oxygen atoms of the 3rd and 4th DNA backbone phosphate group of this strand (fig. 5j). The mutant residue R580Q was placed with the

same orientation as the WT residue at the beginning of the simulations. However, it adopted in most trajectories a $\chi 1$ angle of $\sim$ -60° degrees (fig. 5h, 5i), impairing most interaction with the 3rd DNA backbone group. The mutant glutamine 580 interacted mainly with the 4th DNA phosphate group but for only short durations (fig. 5k, 5l, suppl. video V1). Most importantly in the context of the BE, we observed that, despite its reorientation, the mutant R580Q could not engage in a substantial number of hydrogen bonds with any phosphate groups of the DNA backbone (fig. 5k, 5l, suppl. video V1). Quantification of either the salt bridges formed between WT arginine at position 580 and the DNA *vs*. the hydrogen bonds formed between mutant R580Q and the DNA, disclosed a higher abundance of salt-bridges being formed for the WT (fig. 5m, suppl. fig. S4g), which outnumbered significantly the weaker hydrogen bonds for the mutant R580Q with the DNA (fig. 5m, suppl. fig. S4h).

Since several mutations affecting the LIG4 catalytic domain have been reported, we wondered whether any of them would be related to DNA-binding, similarly to the R580Q mutation characterized in this study. The location of the reported human missense mutations affecting the LIG4 catalytic domain (suppl. table S3) was compared to those of the trajectories in which the enzyme and the DNA were ≤ 3Å (putatively enabling interaction). Using this filter criteria, we found three residues other than the arginine 580 that were close enough to allow an interaction with the DNA: p.278, p.447 and p.449 (suppl. fig. S4i). The positions p.278 and p.449 are very well described ATP-binding residues (not the DNA), the p.278 residue mutation being identified in the first described LIG4 deficient patient[39]. We did not find a biochemical characterization for the LIG4 mutation at position p.447 in the literature. Thus, the here described mutation at p.580 might be the first with experimental evidence for reduced LIG4-DNA binding.

In summary, a reduced enzymatic ligation activity was observed for the R580Q LIG4, paralleled *in silico* by a reduced DNA binding energy, most likely due to the replacement of the positively charged arginine, which is suitable for the interaction with the DNA backbone, by a neutral glutamine. In addition, the side-chain reorientation of the mutant glutamine led to lower number of electrostatic interactions with the DNA.
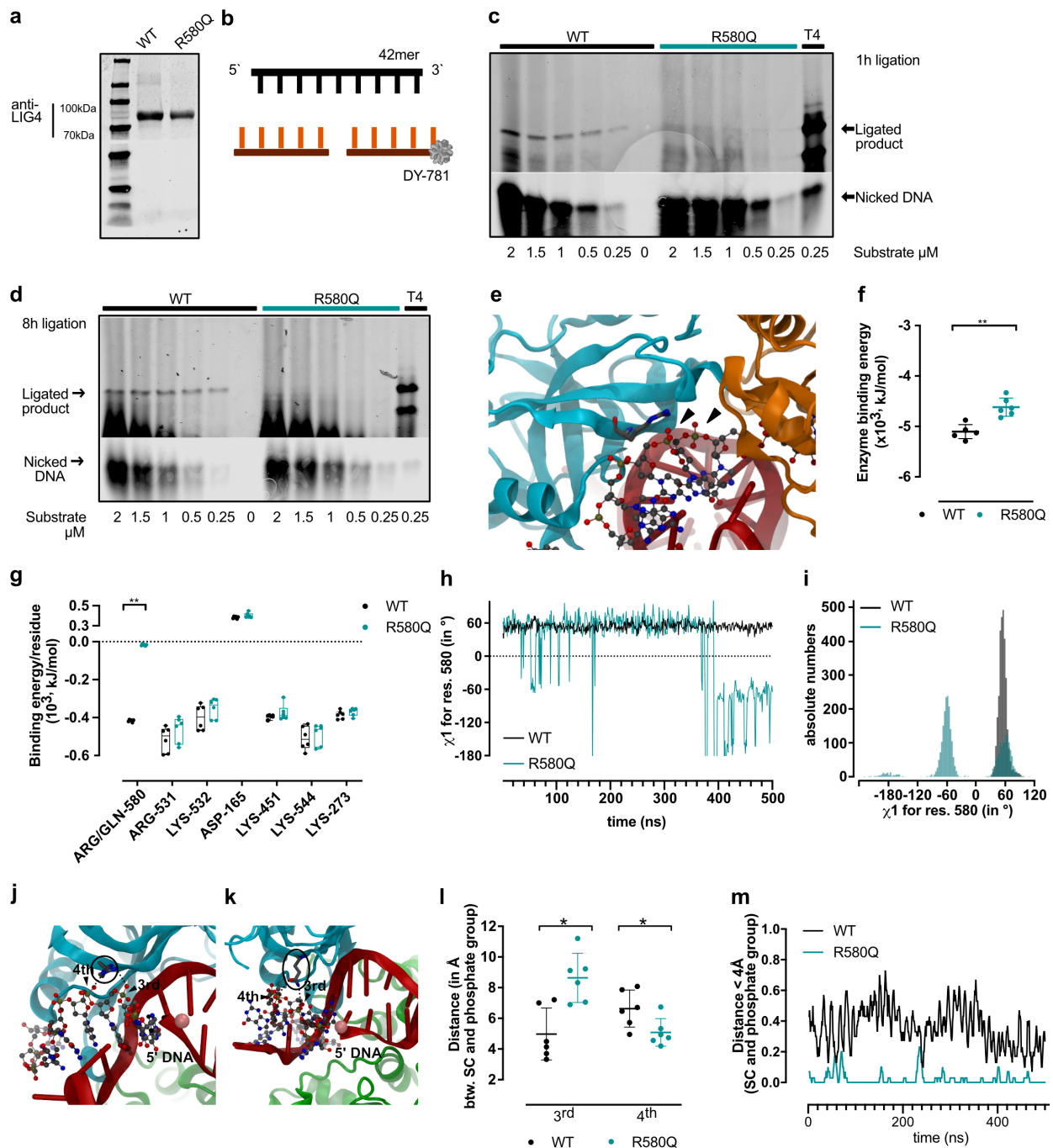
Figure 5| Decreased enzymatic LIG4 activity and DNA binding capacity of the mutant LIG4 R580Q. a) Recombinant wild-type (WT) or mutant (R580Q) LIG4 protein (amino-acids 1-620, representing the LIG4 catalytic core) were generated and analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and probed with rabbit anti-LIG4 antibody. **b)** Schematic representation of the 42mer nicked DNA duplex used for the enzymatic assay. Created partially with BioRender. The LIG4 WT *vs*. R580Q catalytic domain (2mM) was tested for its ligation activity on increasing concentration of unadenylated nicked DNA duplex. Multiple turnover ligation products were separated on a TBE-Urea polyacrylamide gel. Reactions were carried out at 37°C for **(c)** 1 hour *vs*. **(d)** 8 hours in the presence of Tris 40mM, 10mM $MgCl^{2+}$, 0.5mM ATP and 10mM DTT. T4 ligase (0.25U) served as a positive control. **e)** Molecular representation of the catalytic core of LIG4 shown in ribbons, bound to a DNA duplex. The wild-type arginine at position 580 within the OBD is shown as stick and the two nearby phosphorous atoms of the DNA backbone are highlighted with arrows. Oligonucleotide/ oligosaccharide-fold domain OBD in blue, NTD nucleotidyltransferase in orange. **f)** Enumeration of the LIG4 binding energy (kJ/mol) between either the WT or the R580Q mutant LIG4 and the adenylated-DNA complex. The binding of the enzyme to the DNA was quantified in six independent time series, each of >500ns. Lower values indicated stronger protein-DNA

interaction. Each point represents the average of one independent trajectory over the last 100ns of the trajectory. **g)** Binding energy of each residue was enumerated, next only the LIG4 residues displaying a binding energy difference between WT and R580Q LIG4 greater than 20 kJ/mol are plotted. **h)** One representative time series for the WT residue (arginine) *vs*. mutant R580Q (glutamine) regarding the dihedral χ1 angles (in °). **i)** Distribution of dihedral χ1 angle adopted by the WT *vs*. mutant residue R580Q. The dihedrals were extracted at intervals of 1ns over the last 100ns of six independent > 500ns long trajectories per enzyme. Molecular representation of the **(j)** WT LIG4 and **(k)** R580Q LIG4: DBD (green), OBD (blue) domains and DNA (red) shown in ribbons. The WT arginine 580 and mutated glutamine 580 are shown as stick (black circles) and the adenylated nicked-DNA strand is shown as ball and stick. 3rd and 4th phosphate group of the DNA backbone are indicated by black arrows, dashed lines indicate interaction of residue 580 and the oxygen atom of the phosphate group. The NTD domain has been omitted for clarity. **l)** The distance between the WT *vs*. R580Q residue side-chain and the oxygen atoms of the DNA backbone phosphate groups (3rd and 4th phosphate group according to the DNA 5' end). **m)** Reported is the fraction of time, when the distance between the respective residue 580 side-chain and the oxygen phosphate groups of the DNA strand were < 4 Å (moving average is displayed). A Mann-Whitney testing was performed to test for significance in (f) with multiple comparison correction in (l), in (g) we used a 2wayANOVA with Šídàk correction. In the structural representations, carbon atoms are in black, oxygen in red, nitrogen in blue and phosphorous in tan. Hydrogens are omitted for clarity.

Prolonged repair response to experimental DNA-damage in heterozygous LIG4 mutated primary cells

To experimentally test LIG4 functionality in the cellular context of a heterozygous missense mutation, we characterized the DSB response in T cells of P1 and her father at steady-state and after genotoxic stress.

After two days of i*n vitro* culture without additional manipulation, we observed in increased phosphorylation of two important DNA damage associated proteins H2Ax (γH2Ax) and 53BP1 (p53BP1)[40,41] in T cells of both *LIG4* mutation carriers (fig. 6a, 6b), compared to T cells from the mother and healthy controls. To functionally examine how the patient's cells would respond to DNA damaging agents, we measured the kinetics of nuclear γH2Ax levels after ionizing radiation (IR). P1's memory CD45R0[+]CD4[+] T cells had higher γH2Ax[+] levels beyond 24 hours after IR compared to cells from the mother and healthy donors (fig. 6c). The father's memory CD4[+] T cells had elevated γH2Ax levels at 48 hours post-IR compared to the healthy controls (fig. 6c). The father's memory CD45R0[+]CD4[+] T cells showed a trend and P1's memory CD4[+] T cells a distinctly augmented proportion of histone H2Ax phosphorylation after *in vitro* treatment of PBMCs with the chemotherapeutic drug Bleomycin sulfate compared to the mother and unrelated healthy controls (fig. 6d). In line with the augmented γH2Ax levels, we measured reduced cell viability after *in vitro* Bleomycin sulfate exposure in both patients naïve (CD45R0[-]) and memory (CD45R0[+]) CD4[+] T cells compared to the mother and healthy controls (fig. 6e).

Next, we investigated how IR would affect the T cell proliferation capacity. Therefore, peripheral blood-derived T cells were labeled with CellTrace[TM] violet (CTV), then exposed to different IR doses, and subsequently stimulated with anti-CD3/anti-CD28. Proliferation was quantified by assessing the CTV dye dilution. With rising IR-doses, we observed a trend for a decrease relative proliferation index in both CD4[+] and CD8[+] T cells of the two *LIG4* mutation carriers as compared to healthy controls including the mother (fig. 6f, 6g).
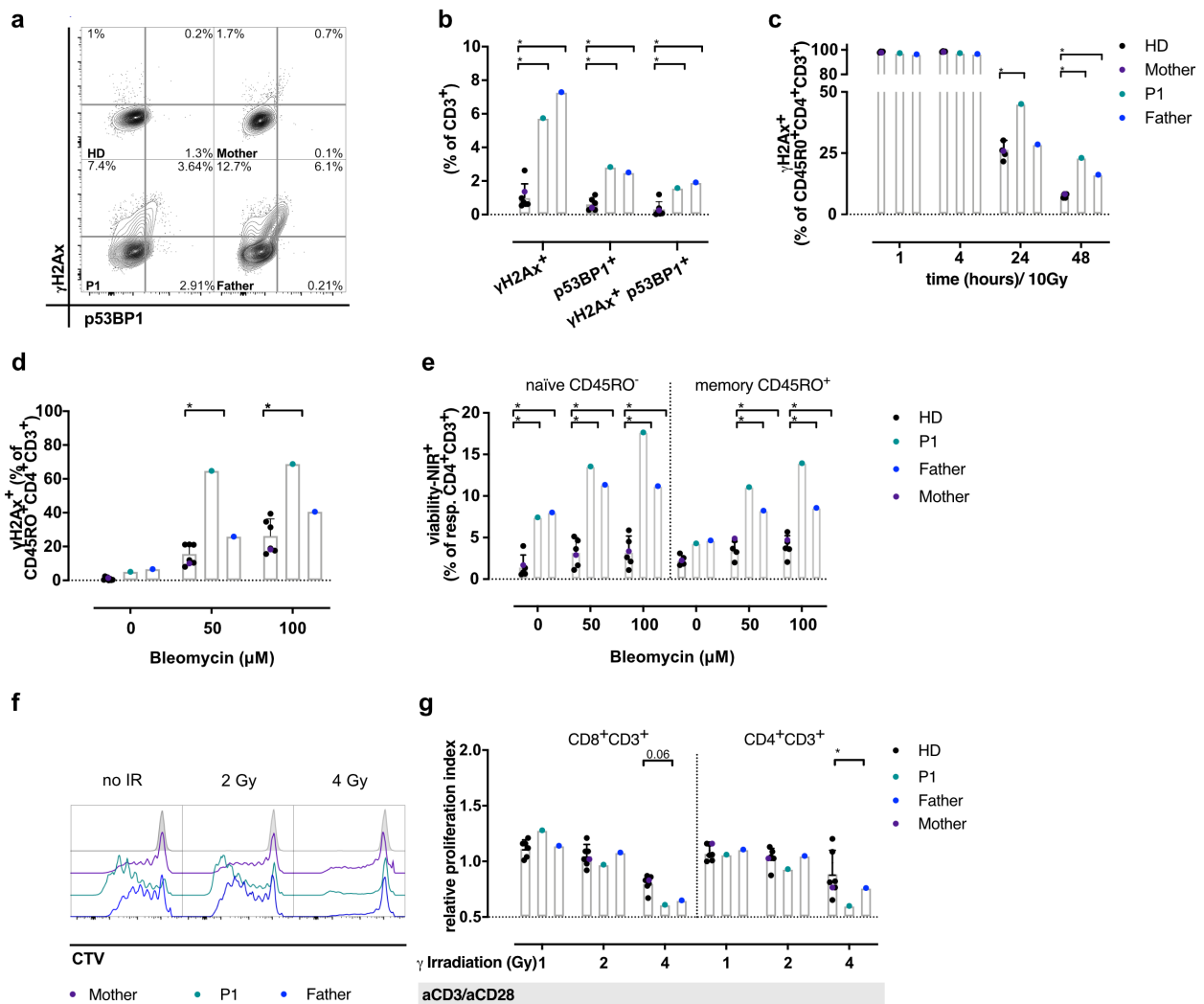
Figure 6| Augmented susceptibility to *in vitro* DNA-damage in the two *LIG4* mutated patients T cells. a) T cells derived from the peripheral blood of healthy controls, the two patients, as well as the mother were cultured for two days without stimulation. The phosphorylation of H2Ax (γH2Ax) and 53BP1 were assessed by flow cytometry, representative plots are depicted, numbers indicate the frequency within the CD3+ T cells population. **b)** Quantification of the H2Ax and 53BP1 phosphorylation in bulk CD3+ T cells, after two days of *in vitro* culture. **c)** Quantitative enumeration of the γH2Ax kinetic in CD45RO+ CD4+ helper T cells derived from PBMCs after IR with 10Gy in P1, her parents and healthy controls using flow cytometry. **d)** Analysis regarding the nuclear γH2Ax+ fraction in memory CD45RO+ CD4+ T cells after *in vitro* treatment of PBMCs with Bleomycin sulfate for 24 hours. **e)** The viability of naïve (CD45RO-) and memory (CD45RO+) CD4+ T cells was assessed 24 hours after *in vitro* Bleomycin sulfate exposure, measurement was performed using flow cytometry. **f)** Quantification of T cell proliferation after IR. T cells were labeled with CellTrace™ violet (CTV), followed by IR and subsequently stimulated for five days *in vitro* with anti-CD3/anti-CD28 (aCD3/aCD28). Representative flow-cytometric plots for the CTV dilution are depicted. Gray shaded population indicates the non-stimulated condition for the mother's T cells. **g)** The relative proliferation index was computed for CD4+ helper T cells and CD8+ T cells after different intensities of IR for P1, the father and the mother, as well as healthy controls. Significance was determined in (b) using a Kruskal-Wallis test (c/d/e/g) using 2way ANOVA with Šídàk correction.

Qualitative changes in T~regs~ and B cell tolerance in the two *LIG4* mutation carriers

Immune dysregulation is observed in approximately one third of all reported patients with LIG4 deficiency (suppl. table S3). Therefore, we investigated peripheral tolerance in the heterozygous *LIG4* mutated patients.

Peripheral blood-derived CD4+ T regulatory cells (T~reg~, CD25^hi^CD127^low^) were reduced in frequency in both *LIG4* mutation carriers compared to healthy blood donors (fig. 7a). A similar reduction was found in other

patients with the diagnosis of CID evaluated at our center, while the T_{reg} proportion was within the normal range in patients with a diagnosis of primary antibody deficiency (PAD) or autoinflammation (Autoinfl.) (fig. 7a). In line with the aforementioned observation, we found that the frequency of $CD4^+$ helper T cells expressing the transcription factor forkhead box protein 3 (FoxP3) and simultaneously displaying high levels of the IL-2 receptor α-chain (CD25) were reduced in both *LIG4* mutation carriers compared with the mother and in several healthy controls (fig. 7b). The inverse expression pattern for CD45R0 and CD27 is associated with recent T_{reg} activation[42,43], thus we wondered if the reduced T_{reg} cells in the two patients would show signs of augmented activation. T_{regs} in both patients displayed a reduced percentage expressing the co-stimulatory molecule CD27, while virtually all cells expressed CD45R0 (fig. 7c). Two inhibitory molecules, lymphocyte-activation gene-3 (LAG3) and programmed cell death-1 (PD-1) correlate with the suppressive ability of T_{regs}[44,45] – both were expressed at elevated levels on T_{regs} of P1 and her father (fig. 7c). In addition, a higher abundance of $CD25^{hi}CD127^{low}CD4^+$ T cells were found with a $CD161^+$ pro-inflammatory phenotype[46,47] (fig. 7c). *In vitro* activation of T cells with anti-CD3/anti-CD28 (αCD3/αCD28), induced higher expression of the co-stimulatory molecule 4-1BB (CD137) and inhibitory molecule PD-1 on $FoxP3^+CD25^{hi}CD4^+$ T cells, whereas expression levels of LAG3 were only elevated in P1 (suppl. fig S5a). Human $CD4^+$ T helper subsets can be characterized by a distinctive chemokine receptor expression pattern[48] (suppl. fig S5b). Investigating peripheral blood $CD4^+$ T cells, we measured an increase in T_{H1} ($CCR6^-CD45RA^-CCR4^-CXCR3^+$) proportion in P1 and her father, while T_{H1}*, T_{H2} and T_{H17} proportions were within the frequencies observed in healthy controls (fig. 7d). The higher T_{H1} $CD4^+$ T cell frequency was paralleled by a higher proportion of $TNFα^+$ producing $CD4^+$ and $CD8^+$ T cells in P1 and her father following *in vitro* stimulating (suppl. fig. S5c). Elevated levels of T_{H1} associated TNFα were detected in both the naïve $CD45R0^-$ and memory $CD45R0^+$ T cell subsets (suppl. fig. S5d, S5e). T follicular helper (T_{FH})-like $CD4^+$ T cells were relatively less abundant in P1 compared to the normal range of controls (suppl. fig. S5f).

The semi-stochastic process of V(D)J gene recombination generates TCRs with self-reactive potential, which escaped negative selection. Since we suspected a reduced DNA-repair capacity, potentially more self-antigens could be released by dying cells, stimulating self-reactive T cells. Thus, we investigated two self-reactive biomarkers within the TCR's CDR3 transcripts in bulk T cells[49,50]. Interfacial hydrophobic amino acids in the CDR3 of the respective TCR-chains at positions 6 and 7 promote the development of a self-reactive TCR[49]. The hydrophobic index was not elevated in P1 or her father (fig. 7e). The second biomarker reflects the frequency of TCR chain sequences encoding for cysteine residues within two positions of the CDR3 apex (cysteine index), this index was found unchanged in P1 and her father (fig. 7f). To elucidate subtle differences regarding the expansion of self–reactive T cell clones, both self-reactive indices were plotted as a function of total *vs*. unique *TCRA* respectively *TCRB* sequences. A deviation from the diagonal would indicate T cell clone expansion with autoreactive potential, which was however not observed in the two *LIG4* mutation carriers (suppl. fig. S5g, S5h).

B cells are strongly selected against self-reactivity features in the bone marrow and the germinal center (GC)[51,52], in a myriad of autoimmune diseases these immune tolerance checkpoints are defective[53]. Self-reactivity can so far not be extrapolated with certitude from AIRR-seq data yet, but specific sequence characteristics correlate with self-reactivity – such as lengthy and positively charged CDR3 and the utilization of the V gene segment V_H4-34[54–56]. Analysis of bulk V_H4-34 (mutated and germline) transcripts indicated similar frequency within naïve, IgM/IgD (*MD*) memory, *IGHG* and *IGHA* B cell transcripts of P1 and her father with healthy controls (suppl. fig. S5i). Interestingly, the IgM/IgD (MD) memory B cells of P1 had a trend for relatively more V_H4-34 sequences in germline configuration (fig. 7g). A lengthy CDR3 was computed in P1's MD memory B cells (fig. 7h). Autoreactivity of B cells was further investigated by probing the father's serum immunoglobulins against different self-antigens on a protein microarray and compared with gender-matched controls. There were no significant changes within the father's IgM immunoglobulins (suppl. fig. S5j). We found a reduction of several IgGs directed against self-proteins in the father's serum compared to controls (fig. 7i), probably

reflecting the father's slight reduced total serum IgG levels (fig. 3b). Notably, four of the tested IgG autoantibodies, were found to be significantly elevated in the serum of the father (fig. 7i). Those were directed against the extracellular matrix component chondroitin sulfate C, the cytoskeleton-associated protein α-actinin, the membrane-associated β2 microglobulin and against genomic DNA (fig. 7j, 7l). P1's autoantibodies were not investigated because of the intrinsic agammaglobulinemia and the continuous intravenous-immunoglobulin substitution (IVIG) therapy.

In summary, evidence was found for a compromised immune-tolerance in both *LIG4* mutation carriers – mostly strikingly indicated by a reduction of peripheral T_{regs}, evidence of recent activation as well as a skewing towards a pro-inflammatory phenotype of peripheral T_{regs} and a trend to proportionally more T_{H1} $CD4^+$ cells. P1's IgM/IgD memory B cells had a lengthy CDR3 and harbored an elevated frequency of the unmutated autoreactive V_H4-34 gene segment. The father displayed a restricted pattern of elevated serum IgG autoantibodies.
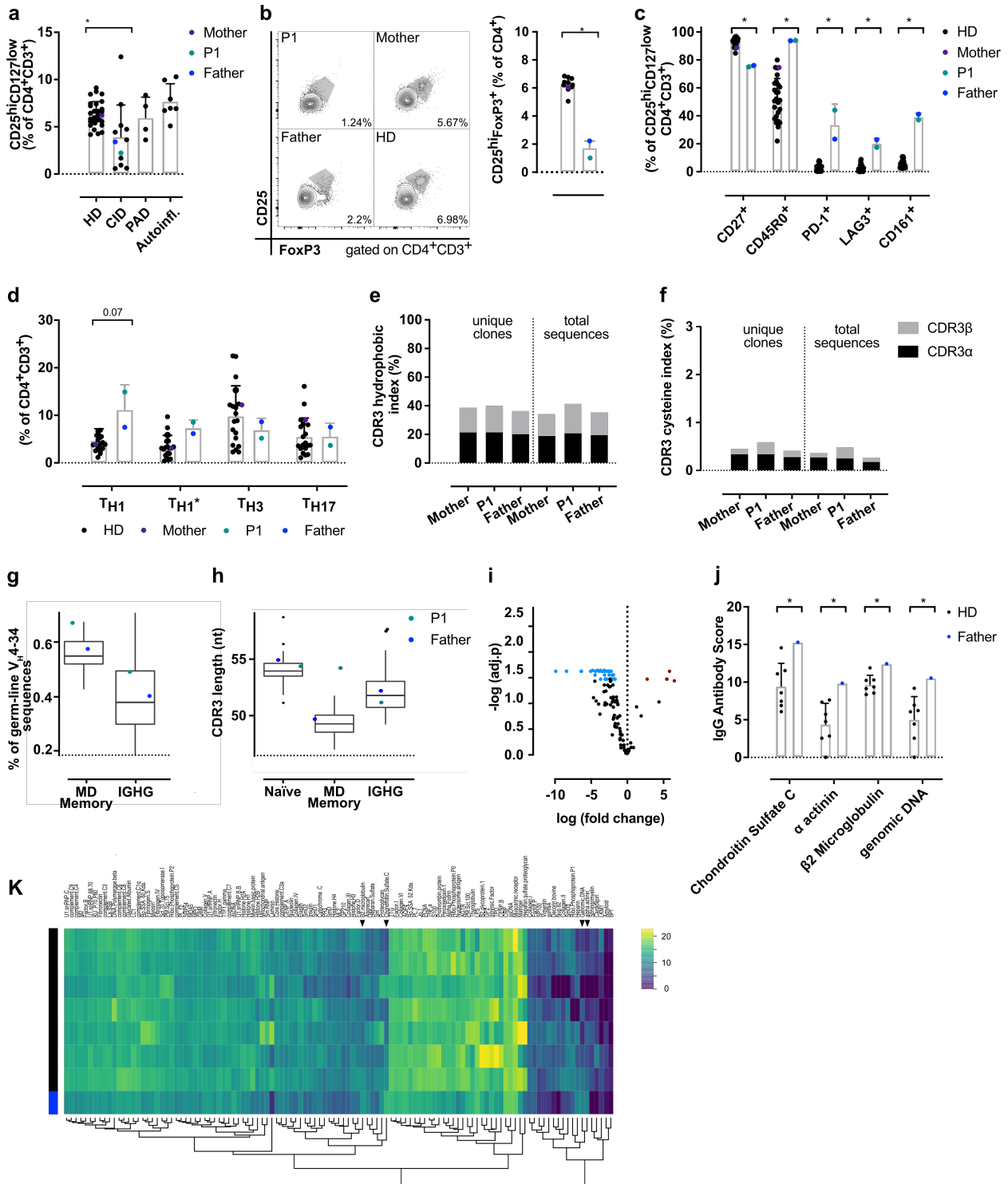
Figure 7| Qualitative changes in T_regs and B cell tolerance in the two *LIG4* mutation carriers. a) Evaluation of circulating CD4⁺CD25^hi CD127^low T regulatory cells (T_reg) in two patients carrying a heterozygous LIG4 mutation, either compared to different entities of inborn errors of immunity (combined immune deficiency CID, primary antibody deficiency PAD, autoinflammation Autoinfl.) or to healthy blood donors (HD, n= 29). **b)** Expression of the transcription factor Forkhead box protein 3 (FoxP3) and the IL-2 receptor α-chain (CD25) within CD4⁺ helper T cells were measured by flow cytometry in healthy donors or P1, her father and mother. Proportion of CD25^hiFoxP3⁺ CD4⁺ T cells among CD3⁺CD4⁺ T helper cells was quantified. Representative flow-plots are shown on the left side with the respective proportion of CD25^hiFoxP3⁺ cells within the CD4⁺ T cell population. **c)** Measurement of the surface molecule expression of CD27, CD45R0, LAG-3, PD-1 and CD161 within CD4⁺CD25^hiCD127^low T regulatory cells using flow cytometry. **d)** Characterization of different CD4⁺ T cell subsets based on the chemokine surface receptor expression pattern. T_H1 (CCR6⁻CD45RA⁻CXCR3⁺CCR4⁻), T_H1* (CCR6⁺CD45RA⁻

CXCR3$^+$CCR4$^-$), T$_{H2}$ (CCR6$^-$CD45RA$^-$CXCR3$^-$CCR4$^+$) and T$_{H17}$ (CCR6$^+$CD45RA$^-$CXCR3$^-$CCR4$^+$). CDR3 indices regarding the TCR self-reactivity were enumerated. **e)** Subgroup analysis for unique and total sequences of the TCR α- and β-chain hydrophobic index, carrying hydrophobic amino acids at positions 6 and 7 of the CDR3 apex and **f)** cysteine index - sequences with cysteines within two positions of the CDR3 apex). **g)** The usage of the intrinsically self-reactive V gene segment V$_H$4-34 in IgM/IgD (MD) and IgG (IGHG) memory B cell transcripts. The boxplots indicate the median, 25-75% (box) an 5-95% (bars) of the HD values. **h)** The CDR3 nucleotide length of naïve, IgM/IgD (MD) and IgG (IGHG) memory B cell transcripts were investigated in P1 and her father. Boxplot indicates the variance of HD. **i)** Volcano plot displaying the father's serum IgG fold change of autoantibody abundance detected by protein microarray, and compared to serum IgG of healthy gender-matched controls. Blue color indicates auto-antigens, which were significantly less abundantly recognized by the father's IgG, whereas red indicates auto-antigens which were significantly recognized at a higher level by the father's IgGs. **j)** The autoantibody score for IgGs, which were significantly higher in the serum of the father compared to gender-matched healthy controls. **k)** Heatmap with hierarchical clustering of all 119 tested IgG autoantibodies. The serum of the father (left column with a blue line) is compared to the serum of healthy gender-matched donors (n=7, columns indicated with a black line on the left). Antigens detected with a significant positive fold change by the father's serum IgGs are marked with arrows on the top. In (a) significance was computed using a non-parametric Kruskal-Wallis test with Dunn's correction. Statistical testing used the non-parametric Mann-Whitney test in (b) and with additional multiple comparison correction in (c/d). Estimation of fold change in (i/j/k) was achieved by fitting a linear model and empirical Bayes statistics.

Discussion

The clinical phenotype of human LIG4 deficiency is broad, ranging from an asymptomatic state to death at a young age (suppl. table S3). A complete LIG4 loss of function has not yet been described in humans and is therefore probably not compatible with life, similar to observations made in murine models[57]. To our knowledge, all LIG4 deficient patients described so far carried homozygous or compound heterozygous missense, premature stop-codons or frameshift mutations. However, Rucci et al. described reduced survival in mice carrying a heterozygous *Lig4* mutation (the orthologue missense mutation found in the first LIG4 deficient patient)[59]. To our knowledge, the immune-phenotype and clinical status of parents or siblings of published LIG4 deficient patients has not been studied systematically yet.

We present solid evidence that a heterozygous novel *LIG4* mutation segregated with combined immunodeficiency in an autosomal dominant manner. Variants in other known immunodeficiency associated genes (except for a rare *FAS* variant, discussed below) did not segregate with the familial immune dysregulation. P1 demonstrated preserved LIG4 protein levels, but the mutated R580Q LIG4 showed reduced *in vitro* ligation activity and has not been described in the literature (suppl. table S3). The immunologic phenotype observed in P1 with reduced naïve T cells and agammaglobulinemia has been described repetitively in patients with LIG4 deficiency (suppl. table S3).

The T cell intrinsic DNA repair response in the two heterozygous *LIG4* carriers points to an increased sensitivity and a delayed repair kinetics, key characteristics in LIG4 deficiency syndrome[60]. While both LIG4 mutation carriers suffered from autoimmune cytopenias, only the index patient P1 demonstrated infection susceptibility, lymphoproliferation and non-lymphoid tissue infiltration. Autoimmune pneumonitis, interstitial nephritis and granulomatous parotitis, all responsive to augmented immune-suppressive therapy, are rarely diagnosed in patients with inborn errors of immunity (IEI). We propose that in patients presenting with these autoimmune features, IEI in general and *LIG4* insufficiency in particular, should be actively searched for.

The clinical disease severity correlated with the observed immunological anomalies, which were more pronounced in P1. It is essential to state that only P1 was treated with immune-suppressive medication over several years which may have aggravated the immunodeficiency/immune-dysregulation. P1 was treated with azathioprine at pediatric age, which *per se* increases DNA DSB propability[61,62] and alters the DNA methylation state (via decreased levels of DNA cytosine-5 methyltransferase 1 DNMT1)[63,64]. Further, her current immune-suppressive treatment includes mycophenolate, an inhibitor of the *de novo* purine synthesis pathway[65].

It is plausible that the clinical phenotype is aggravated by her female gender, as the gender bias is overserved in many autoimmune diseases[66]. However, our literature search did not demonstrate a higher autoimmune/-inflammation incidence in female LIG4 deficient patients (suppl. table S3).

Mutations affecting the NHEJ pathway may perturb the rearrangement of V(D)J gene segments, which might alter the composition and diversity of both B and T cell repertoires[67]. However, the IgH B cell receptor (BCR) repertoire was overall well preserved. Differences were found in the utilization of individual V_H gene families and J_H genes. Both patients had reduced usage of the J_H4 gene segment in IgG B cells. Usually an increased J_H4 segment utilization is a hallmark of antigen-experienced B cells[68], which could point at a lower precursor frequency. Diversification of the BCR is achieved by pairing of the IgH with a light-chain. Interestingly, we observed in P1 an inverted Ig light-chain κ/λ usage. This could point to multiple rounds of receptor editing during B cell development[69] as the IgLλ would only be expressed after serial rounds of unsuccessful IgLκ locus recombination[70] or to resolve auto-reactivity[54,71,72].

We observed in both patients a drastic reduction of naïve T cells and of T cells bearing a TCR with the most distal V gene segment encoding for Vα7.2, both features observed previously in biallelic mutated LIG4 patients[10,73]. The decrease of TCR Vα7.2$^+$ T cells is suggestive of a thymocyte maturation problem and/or impaired survival, as distal *TCRA V* segments (such as *TRAV1* encoding for TCR Vα7.2) are only available after several rounds of unsuccessful recombination[74].

The TCR repertoire demonstrated to be diversified in both *LIG4* mutation carriers in keeping with TCR repertoire analysis in three siblings carrying compound heterozygous *LIG4* mutations[73]. Interestingly, both LIG4 mutation carriers showed similar preferential usage of TRAV segments. Especially an augmented usage of TRAV27-01-03 was observed in both patients. TRAV27-01 had been linked to staphylococcal enterotoxin exposure[75] and is predicted to bind Influenza peptide M[76]. Both are improbable explanations in our two patients as the augmentation was found in unique sequences (thus not representing clonal T cell expansion) and sequencing was performed on DNA, not on RNA, which does not represent TCR expression levels.

Besides NHEJ, DNA DSB breaks can also be resolved by alternative repair pathways[77]. An elegant study investigating the TCR of LIG4 deficient individuals, had demonstrated an increased usage of the alternative end-joining (a-EJ) directed repair pathway[73]. Incompatible DNA ends are found typically after DSB which precludes the direct repair. The a-EJ needs very short microhomology domains to initiate repair and relies on LIG1/LIG3[77]. Upcoming investigations might want to focus on which alternative repair pathways are preferentially used in patients harboring different *LIG4* mutations. The evaluation of the a-EJ pathway usage is important because it harbors a greater risk for chromosomal translocations[78].

We performed molecular dynamics simulations to gain mechanistic insight on the atomic level regarding the residue change at position 580 from arginine to glutamine. The binding energy between the mutated residue and the DNA was reduced, as shown in the lower abundance of electrostatic interactions formed with the DNA backbone. Potentially because the mutated glutamine was often reoriented in a position that weakened its interactions with the DNA. We investigated all published *LIG4* missense mutations for their putative DNA binding capacity. To our knowledge, it is for the first time that a biochemically verified LIG4 amino acid change is described to affect its DNA binding capacity.

Autoimmunity was a major clinical issue in both patients. Normally, immune tolerance is guarded by multiple mechanisms[79,80], including regulatory (T_{reg}) CD4$^+$ T cells[81] and inhibitory receptors[82]. We observed several immune cellular characteristics of immune dysregulation in the two *LIG4* mutated patients: T cell hyperproliferation, lower T_{reg} frequency, T_{H1} CD4$^+$ effector skewing and CD21^{low} B cell accumulation (suppl. fig. 6). T cell hyperproliferation was probably a consequence of the increased effector subset within the T cells. Autoimmunity can be the consequence of a reduced apoptosis capacity of auto-reactive lymphocytes as seen in autoimmune lymphoproliferative syndrome (ALPS)[83]. The novel *FAS* variant identified here locates to exon 4, an exon known to harbor less penetrant mutations[84]. The variant identified here has not been documented in

a well-established ALPS cohort consisting of more than 120 ALPS patients[84] nor in any publicly available databases. ALPS related serum biomarkers and low immunoglobulins did argue against ALPS as the underlying disease. Most importantly, FAS related apoptosis was not impaired in the three *FAS* variant carriers. In addition, immunophenotyping was normal in the healthy brother of P1, which carries the same *FAS* variant but lacks the *LIG4* variant. In summary, the data argues for the *LIG4* variant as the main driver of this familial disease.

Dysregulation within the helper T cell compartment may subsequently affect the B lymphocytes[85]. B cells encoding the autoreactive V-gene segment V_H4-34 were proportionally elevated in P1, those B cells are intrinsically self-reactive being directed against I/i carbohydrates expressed on erythrocytes[86,87]. V_H4-34 antibodies are a double-edged sword, as additionally, they cross-react with antigens found on commensal bacterial[88], thus helping to contain bacteremia due to a leaky gastrointestinal tract. The augmented proportion of V_H4-34 B cells might indicate a systemic immune response against commensal bacteria as described before in RAG deficiency[89], MYD88/IRAK4 deficient patients[24] and systemic lupus erythematosus[45]. In the father, we measured increased IgG autoantibodies directed against genomic DNA, a feature linked to enhanced cell death[91].

What are the consequences of the putative LIG4 dependent immunodeficiency for personalized treatment of the patients studied here? Genotoxic assaults must be minimized, including x-ray based imaging which is often repetitively performed in patients with immunodeficiency. Also, UV-light exposure of the skin needs to be reduced using wide application of UV-light active sun-blockers. There is currently no known compound specifically elevating/modulating LIG4 function. Nevertheless, the immune-dysregulation linked to a dysfunctional T_{reg} compartment could be restored using treatment with low dose interleukin-2 (IL-2)[92] or sirolimus[92] or abatacept[93] (CTLA-4 fusion protein). Vitamin D, likely to be low in a patient with extended sun-blocker use, has been demonstrated to support the immune-regulatory action of abatacept[94] and is currently supplemented in P1. Successful allogeneic bone marrow transplantation has been described in patients with LIG4 deficiency[95]. However, conditioning that includes DNA toxic medication is an issue due to the impaired DNA repair machinery and may impact on the outcome[96,97].

To our knowledge, this is the first report of a dominant inheritance of LIG4 dependent immune-dysregulatory human disease. In patients with immune-dysregulation of unknown cause, it is essential to consider *LIG4* haploinsufficiency as it may have specific prognostic and therapeutic consequences.

Limitations

There is no *in vivo* model available so far, carrying the same orthologous *LIG4* mutation as the patients described here. Thus, we are unable to fully exclude additional genetic and environmental factors on the clinical *in vivo* and *in vitro* phenotype.

The involved mechanisms underlying autoimmunity in the patients described here are speculative. We cannot distinguish whether the demonstrated T_{reg} numbers with an activation/proinflammatory phenotype and T_{H1} CD4+ T cell subset skewing are rather consequence than the cause of reduced central and/or peripheral tolerance. *TRA* and *TRB* sequencing was performed on bulk T cells. It is conceivable that focusing on the TCR repertoire of T_{regs} cells would have revealed more striking differences. Additionally, single cell-based approaches would provide the chance of paired *TCRA* with *TCRB* analysis.

Acknowledgments

Author contribution

AJ and MR conceived the project. AJ wrote the manuscript. AJ designed, performed, analyzed most experiments and conceptualized the figures and tables. OB carried out the computational simulations, their analysis, the generation of the video, and critically reviewed the manuscript. ODM/LDN performed and analyzed the *TRA/TRB* sequencing experiment, MG/VvN/JT performed and analyzed the AIRR-sequencing experiment. BM gave scientific input. FM performed the apoptosis assay. SE performed the sFASL and VitB12 measurement, and gave scientific input. ARE gave scientific input. IH performed the immunoglobulin measurements, SH analyzed the histology samples. TD/TK/TD were involved in patient care. MR supervised the study, corrected the manuscript and coordinated the patients.

Experimental Procedures

Ethics approval and human subjects

Following informed consent, the patients were included into a prospective cohort of subjects with primary immunodeficiency/immune dysregulation that was approved by the Ethikkommission Nordwest und Zentral-schweiz (EKNZ 2015-187), it complied with all important ethical regulations. Several patients and healthy controls were prospectively recruited between November 2018 and February 2021. Blood samples from healthy donors were obtained after informed consent from the Blood Donor Center, University Hospital Basel.

Genetic analysis, variant calling, annotation and filtering

Genetic sequencing was performed following informed consent. Genomic DNA was isolated from cultured T-cell blasts or PBMCs using the QIAamp DNA Blood Mini Kit (Qiagen). Whole exome sequencing was performed as described earlier[98,99]. In short, the DNA was fragmented, exon-specific biotinylated primers (SureSelect Human All Exon V6, Agilent Technologies) were hybridized to the DNA, then exons were pull-down using Dynabeads (MyOne Streptavidin T1 magnetic beads, Thermo Fisher Scientific). Paired-end 125bp sequencing was completed on Illumina HiSeq 4000 platform (Functional Genomics Center - Zurich, Switzerland). For each sample the paired-end sequencing resulted in two FASTQ files (fwd and rev). Those were subjected to quality control and aligned to the reference genome (GCRCh37 - hg19) using Burrows-Wheeler aligner, which generated sequence alignment map files. Picard tools (v.2.7.1) was used to make a binary version of the sequence alignment map files. Quality control and genotype calling was performed using the Genome Analysis Toolkit (gatk, Broad Institute, USA)[100] and variants were annotated with the position of nucleotide change with respect to coding genes. Variants were called using HaplotypeCaller from the gatk, resulting in a variant called file. CADD (v.1.3) and vVariant effect predictor programs were used for mutation annotation. To elucidate a common disease-driving variant in P1 and her father, variants with a coverage of < 10 were excluded, variants encoding synonymous or inframe- mutations were excluded, further shared variants with the mother and brother were excluded, the only mutation which was common in the father and P1 was LIG4.

The LIG4 variant was confirmed by Sanger sequencing of PCR amplification products of cDNA derived from PBMCs. After running the amplicon on an 1.5% agarose gel, DNA was extracted with QIAquick Gel Extraction Kit (Qiagen). The purified PCR products were then bidirectionally sequenced by Microsynth (Switzerland).

Histology and radiology

For P1 all the available histological specimens, immunohistochemical stains and radiological investigations were reviewed and performed as part of the clinical routine

Blood samples routine phenotyping and immunoglobulin quantification.

From all patients a differential white blood cell count was obtained. For patient P1, some of the T and B cell subsets were immunophenotyped during the clinical routine (surface markers for B and T cell subsets pheno-typing are found in suppl. table S1). Serum levels of total IgG, IgM and IgA were measured by the ISO 17025 accredited Medical Immunological Laboratory of the University Hospital Basel.

Cell isolation and in vitro T cell activation

Patient- and healthy control-derived peripheral blood mononuclear cells (PBMCs) were isolated from whole blood, via Ficoll density gradient separation using Lymphoprep™ (density of 1.077g/mL, Axonlab).

PBMCs derived T cell blasts were expanded with initial phytohaemaglutinin stimulation (5 mg/ml, Sigma Aldrich) and human IL-2 (300U/ml, Proleukin®, Novartis), cultured in media RPMI (Sigma-Aldrich) supplemented with 10% heat-inactivated FCS (v/v, Gibco), 1% penicillin-streptomycin (v/v, Gibco), 1% none-essential amino acids (v/v, Gibco) and 2mM Glutamax (v/v, Gibco) at 37°C in a humified 5% CO_2 incubator. IL-2 was renewed every 5-7 days.

Cell activation: PBMCS, either freshly isolated or thawed from liquid nitrogen-stored samples, were cultured in RPMI with 5% human AB-serum (blood donation center Basel, Switzerland), 1mM Sodium Pyruvate (Gibco), 1% penicillin-streptomycin (Gibco), non-essential amino acids (Gibco) and 1% Glutamax (Gibco). T cells were stimulated with anti-CD3 ([1mg/mL], OKT3, Ultra-LEAF™, BioLegend # 317325) and anti-CD28 ([2mg/mL], 28.2, Ultra-LEAF™, BioLegend # 302933), at a cell concentration of 1^6/mL.

For evaluation of T cell proliferation, CellTrace™ Violet Cell Proliferation kit (0.5μM, ThermoFisher, # C34557) was used for PBMCs labelling. Cell concentration was adjusted to 10^7/mL, then incubated at 37°C in the dark for 20min, followed by 10min incubation on ice. 10 volumes of media were added, followed by 2 washing steps. Cells were stimulated as described above for 5 days.

For the in vitro apoptosis assay PHA T cells blasts were incubated four hours with either anti-CD3 (BioLegend, Ultra-LEAF™ OKT3 #317325, 0.5 - 5μg/mL), anti-FAS (BioLegend, Ultra-LEAF™ EOS9.1 #305705, 0.5 - 5μg/mL), recombinant FASL (eBioscience, NOK-1 #16-9919-81, 0.5 - 5μg/mL) or staurosporine (Sigma #S5921, 0.1 − 0.5μM). The surface staining and the AnnexinV-binding coupled to fluorochrome (BioLegend, #640905) were performed in AnnexinV binding buffer (BioLegend, #422201) for 30min at RT.

Immunophenotyping
PBMCs were stained for cell surface markers with fluorochrome-conjugated antibodies to the following proteins (dilutions indicated)

1° antibodies				
Antigen	fluorochrome	clone	Dil.	Reference
CCR6	BV421™	G034E3	50	BioLegend # 353407
CD3	BV711™, PE PE-Cy7, PB, BUV395	OKT3 UCHT1	20	BioLegend # 317327, 317307 BioLegend # 300419, 300418 BD # 563548
CD4	Ax700, BV510™	OKT4 SK3	40	BioLegend # 31742, 344634
CD8	PE/Dazzle™, APC BUV395	SK1 HIT8a	40	BioLegend # 344743, 344721 BD # 740303
CD19	Ax700	HIB19	40	BioLegend # 302225
CD20	BV510™, PE/Cy5	2H7	40	BioLegend # 302339, 302307
CD21	BV711™	1048	50	BD Bioscience #742763
CD25	Ax488	BC69	50	BioLegend # 302615
CD27	BV650™	O323	50	BioLegend # 302827
CD38	APC	HIT-2	50	BioLegend # 303509
CD45RA	FITC	HI100	50	BioLegend # 983002
CD45R0	BV510™	UCHL1	50	BioLegend # 304245
CD69	BV421™, BV605™	FN50	50	BioLegend # 310929, 310937
CD127	BV650™, BV711™, PE/Cy7	A019D5	50	BioLegend # 351325, 351327, 351320
CD152/CTLA-4	PE, BV421™	BNI3	30	BD # 557301, 562743
CD137/4-1BB	PE/Cy7	4B4-1	50	BioLegend # 309818
CD161	BV605™	HP-3G10	20	BioLegend # 339915
CD183/CXCR3	Alexa Fluor 647	G025H7	50	BioLegend # 353711
CD194/CCR4	PE-Cy7	1G1	50	BD # 557864
CD196/CCR6	PE	11A9	50	BD # 551773

CD223/ LAG-3	APC	7H2C65	100	BioLegend # 369211
CD279/ PD-1	BV785 ™	EH12.2H7	100	BioLegend # 39929
CD366/Tim-3	BV605 ™	F38-2E2	50	BioLegend # 345018
IgD	PE-Cy7	IA6-2	50	BioLegend # 348209
IgM	BV605 ™	MHM-88	100	BioLegend # 314523
Ig light chain k	FITC	MHK-49	50	BioLegend # 316507
Ig light chain l	PE	MHL-38	50	BioLegend # 316607
TCRαβ	Alexa Flour 488	IP26	40	BioLegend # 306711
TCRγδ	PE	B1	20	BioLegend # 331209
Vα7.2	PE	3C10	20	BioLegend # 351705
2°Antibodies				
Goat anti-Rabbit IgG	Alexa Flour 488	Polyclonal (H+L), min X Hu Sr Prot	500	Jackson ImmunoResearch # 111-545-045
Goat anti-Rabbit IgG	Alexa Fluor 647	Polyclonal (H+L), min X Hu Sr Prot	500	Jackson ImmunoResearch # 111-605-045

Staining media: Cells were stained in PBS containing 2.5% human AB serum, NaH_3 0.01%, Hepes 25mM and Fc block (BioLegend # 426101) for 30min at 4°C. Chemokine receptor staining was performed at 37°C for 20 min.

Cell viability was assessed using Live/Dead Fixable NIR (# L34975, Invitrogen™, ThermoFisher Scientific). Data were acquired on an LSRII-Fortessa™ (BD Bioscience – equipped with 355nm, 405nm, 488nm, 561nm and 640nm laser lines). Before acquisition, PMT voltages were manually adjusted with single stain controls to minimize fluorescence spillover and calculation of compensation matrixes. Data analysis was performed using FlowJo software (Version 10.5.2, TreeStar, USA).

Intracellular cytokine production by T cells

For intracellular cytokine production, we incubated PBMCs for 4-5 hours at 37°C in media with PMA (50ng/mL, Sigma), ionomycin (500ng/mL Sigma) and BrefeldinA (BioLegend, 5ug/mL). Cells were first stained for surface marker and viability, then fixed and permeabilized using the BD Cytofix/Cytoperm™ (BD Bioscience), next intracellular staining was performed using BD Perm/Wash Buffer (BD Bioscience).

Antigen	Fluorochrome	Clone	Dil.	Reference
IL-2	APC	MQ1-17H12	50	BioLegend # 500311
IL-17A	PE	BL168	50	BioLegend # 512305
INF-γ	BV421™	4S.B3	50	BioLegend # 502531
TNF-α	BV421™, PE-Cy7	Mab11	50	BioLegend # 502931, 502929

Intranuclear transcription factor and phospho-protein staining

The expression of transcription factors was determined by intranuclear staining performed according to the manufacturers' Transcription Factor staining buffer protocols (eBioscience™).

For the detection of phosphor-proteins, PBMCs were fix and permeabilized in BD Cytofix/Cytoperm™ (BD Bioscience) for 20min at room temperature (RT), next the cells were permeabilized with BD Perm III for 30min at 4°C. Primary anti-phospho antibodies were added in staining media for 1h at RT, after washing secondary antibodies were added subsequently for 30min at RT.

Antigen	Fluorochrome	Clone	Dil.	Reference
FoxP3	PE	150D	20	BioLegend # 320007
γH2Ax (Ser139)	Alexa Fluor488	2F3	20	BioLegend # 613405
phospho-53BP1 (Ser1778)	unconjugated	Polyclonal	100	Cell Signaling # 2675

Radiosensitivity Testing and DNA-damaging compounds

PBMCs from patients and healthy controls were γ-irradiated with indicated Gy dose (^{137}Cs Gamma counter). Cells were harvested at indicated time-points, next cells were stained for phospho-proteins (see above).

Bleomycin sulfate (European Pharmacopoeia EP Reference Standard, cat# B1141000) was diluted in human cell culture media and added to the PBMCs for the duration of 24 hours.

Immunoblotting

Protein extracts were prepared from immune cells through lysis with RIPA buffer (Thermo Fisher Scientific) containing protease/phosphatase inhibitors (Roche). Protein denaturation was achieved with 1x Lämmli sample buffer (Bio-Rad) at 90°C for 10min. Proteins were separated by electrophoresis using 4-15% Mini Protean TGX Gel (Bio-Rad) and transferred to nitrocellulose membranes (semi-dry Trans-Blot Turbo Transfer, Bio-Rad). The membranes were incubated in blocking solution (TBST with 5% non-fat milk powder) for several hours at room temperature, followed by a 1° antibody directed against the antigen of interest in blocking solution. After washing, membranes were incubated with appropriate 2° antibodies and images were captured using LICOR Odyssey imaging system.

Antigen	Host, isotype	clone	Dil.	Reference
β-Actin	Rabbit, IgG	D6A8	1000	Cell Signaling Technology # 8457
Bcl-2	Rabbit, IgG	D55G8	1000	Cell Signaling Technology # 4223
LIG4	rabbit IgG	D5N5N	1000	Cell Signaling Technology #14649
2° antibodies Host	Reactivity	Fluorochrome	clone	
Goat	rabbit IgG (H+L)	IRDye-800CW	polyclonal	LICOR: 926-32211
Goat	mouse IgG (H+L)	IRDye-800CW	polyclonal	LICOR: 926-32219

Generation and analysis of TCR and IGH repertoire by NGS

Bulk T and B cells were analyzed by deep sequencing. In brief, for TCR repertoire analysis genomic DNA was extracted from PBMCs using QIAamp DNA Blood Mini Kit (Qiagen) the quantity and purity were assessed with spectrophotometric analysis. For the IgH repertoire analysis PBMCs were lysed in RLT buffer (Qiagen).

SuperScript III/IV (Invitrogen) was used for reverse transcription and the primers for the IgH constant region included 14nt unique molecular identifiers (UMI) and partial p7 adaptors. Per sample two reverse transcription were run, one with IgM and IgD specific reverse primers and another with IgA, IgG and IgE-specific reverse primers. Primer sequence and PCR conditions can be found in Ghraichy M., 2020[101]. Two-step multiplex PCR amplifications were used for IgH gene rearrangements, 1st step used a mix of FR1 V family specific primers with partial p5 adaptors, the 2nd amplification step completed the adaptor sequences. PCR amplicons were separated on a gel, products were gel-extracted – purified and quantified (following Illumina qPCR library quantification protocol). Each library was normalized for its concentration, followed by multiplexing in batches of 24 for sequencing on the Illumina MiSeq platform (2x 300bp paired-end chemistry). IgH samples were de-multiplex with help of their Illumina indices and processed using the Immcantation platform[102,103]. Briefly, raw FASTQ files were filtered if the quality score was >20, paired reads were joined if they had a at least a length of 10nt, a maximum error rate of 0.3 and the α threshold of 1^{-4}. Reads with identical UMI were collapsed. Reads with equal full-length sequence and equal constant primer but different UMI were further collapsed – resulting in a data frame containing unique sequences per sample and isotype. VDJ assignment was performed with IgBlast[104], IMGT germline database was taken as reference for sequence annotation. The Stampy[105] algorithm was used for mapping the constant region sequences to germline, only sequences with well-defined constant region were further analyzed. B cell populations were defined according to the constant region annotation and mutation number. V gene were classified as "unmutated" (naïve) if ≤ 2nt mutation were found across IgD and IgM sequences (correcting for sequencing bias and allelic variance). The amount and type of V gene mutations was calculated with the R package SHazaM[103]. SHM levels were computed by calculating V gene mutations in single sequences and average values were computed across samples and cell subsets. An effective representative sequence of each clonal group was determined with SHazaM R package, this allowed the computation of the selection pressure using BASELINe[106]. Selection was tested by calculating CDR_R/(CDR_R + CDR_S), which corrects age-dependency.

For the *TRA* and *TRB* loci analysis, rearranged products were amplified via multiplex PCR (Adaptive Biotechnologies, USA). In short, 52 forward primers were used for *TRBV* gene segments and 13 reverse primers to cover the *TRBJ* gene segments. The amplicons were sequenced using the Illumina HiSeq platform, the assay was designed at a survey level (detection limit 1 cell in 0.04e6). Amplification bias due to multiplex PCR were reduced by Adaptive Biotechnologies with assay-based and computational approaches. Custom algorithms were used for the alignment of the reads to reference genome sequences. ImmunoSEQ Analyzer (V.3.0), was used for data processing. One clonotype was defined as the unique arrangement of CDR3 amino acid sequence and its related *V* gene. Heatmaps for the visualization of *TRAV* to *TRAJ* gene pairing in total and unique sequences were produced with the Morpheus platform (Broad Institute, USA). The CDR3 self-reactivity indices were investigated, as described previously[50]. The *TRAV* gene differential expression was computed with the R platform, using the Limma package[107] (v3.36.5). Data was visualized with the ggplot2 package (version 3.3.2).

Molecular dynamic simulations

The atomic model of the human LIG4 catalytic domain was modelled using the SWISS-MODEL server[108] and was based on the crystal structure PDB code 6BKG, assumed to correspond to the closed state of the enzyme encircling an adenylated DNA strand, resolved at 2.4 Å. Whereas the deposited structure already corresponds to the protein of interest (no homology), this strategy allowed to automatically build the missing loops. The longest missing loop was three residues long. The R580Q mutation was introduced with VMD[109]. The adenylated DNA strand was built with Chimera[110] and the force field parameters were generated with the CGenFF web-service as previously described[111–113]. The systems containing the modelled protein, either the WT or the R580Q, the DNA and the adenylated DNA strands, were solvated with ~ 43,000 water molecules, represented by the TIP3 model[114] and further neutralized at a salt concentration of 150 mM with K^+ and Cl^- counterions. While experiments show that LIG4 requires magnesium for nick sealing, no divalent ions were observed in the crystal structure. However, metal binding residue candidates were proposed[35]. We added manually magnesium ions within coordination distance (4Å) of Asp_{275}, Glu_{331} and Glu_{427}. Since the distance between the Cβ atom of residue 580 is located at ~18 Å of the magnesium coordinating candidates, the dynamics of the divalent ions was not further studied. We noted however that they remained coordinated by these acidic residues during the whole length of the simulations. A typical system contained ~ 140,000 atoms. Molecular dynamics simulations were conducted with the GROMACS package (v.2018)[115] and the CHARMM force-field (v27)[116]. Bond lengths and angles involving hydrogen atoms were constrained using the LINCS algorithm[117]. Short-range electrostatics were cut off at 1.2 nm, and the particle mesh Ewald method was used for long-range electrostatic[118]. Van der Waals interactions were described with Lennard-Jones potential up to a distance of 1.2 nm. A constant pressure of 1 bar was maintained using the Parrinello-Rahman algorithm (time constant 5ps). The temperature was kept at 310 K using the Nose-Hoover algorithm (time constant 1ps). The systems were equilibrated following the CHARMM-GUI protocol[119].

For each investigated sequence (WT or R580Q), six independent individual trajectories of 500 ns length were generated, amounting for 6 μs of simulation. Structural biology analyses were performed using an ensemble of R, python, and tcl in-house scripts. Molecular representations were generated with VMD. The free energy of interactions was calculated by solving the adaptive Poisson-Boltzmann equation[38], using the g_mmpbsa module implemented in GROMACS[37].

Recombinant LIG4 and nick sealing assay

WT and mutant LIG4 proteins (aa 1-620) were produced by GeneScript.

To generate a nicked DNA duplex, three oligonucleotides (see below) were annealed and phosphorylated by mixing equimolar amounts in a buffer comprising 40mM Tris, 10mM $MgCl_2$, 10mM dithiothreitol (DTT), 0.5mM ATP and PNK kinase (New England Biolabs). Oligonucleotides were phosphorylated at 37°C for 30min, then heated to 95°C for 5min with sequential cooling steps till 25°C. Ligation reactions contained the three oligonucleotides, the respective DNA ligase4 (either WT or 580Q mutant), reaction buffer was the same as for the phosphorylation/annealing of the oligonucleotides but without PNK, incubation was performed at 37°C, the time of ligation is noted in the respective figure legend. Ligation reactions were quenched by the addition of equal volume of loading buffer (95% formamide, 18mM EDTA, 0.025% SDS) for 5min at 95°C and rapid cooling on ice. Products were characterized by denaturing PAGE (15% polyacrylamide TBE-Urea) in 0.5x TBE buffer and ligation was visualized and quantified with LICOR Odyssey imaging system.

RNA extraction and real-time PCR

Primers were designed with NCBI Primer-BLAST. RT-PCR primers should create amplicons spanning exon to exon junctions or designed within two separate exons with an > 1000bp intron in between. Cells were lysed by adding TRIzol (Thermo Fisher Scientific), RNA was extracted from immune cells using the QIAmp RNA Blood Mini Kit (Qiagen). RNA content was measured by NanoDrop Spectrophotometer (Thermo Fisher Scientific). Consequently, first-strand cDNA was generated via GoScript™ Reverse Transcription System (according to the manufacturer's protocol Promega) using random hexamer primers and 3mM $MgCl_2$. RT-PCR reactions contained variable amounts of template cDNA. PCR amplification was carried out in triplicates using the GoTaq qPCR reagents (Promega), amplification was done using Real Time PCR cycler Vii A7 (ThermoFisher Scientific). PCR conditions were 40 cycles à 30s - 95°C, 1 min - 61°C. Transcript levels were normalized for the house keepers *ACTB, GAPDH*, *PGK1* and *HRPT1* and subsequently to healthy controls, using the equation by W.Pfaffl[120].

Oligonucleotides

All oligonucleotides were ordered with Microsynth (Switzerland).

	Forward	Reverse	Source
Sanger			
LIG4 mutation at c.1739	TTGGCCAAGTATTGGAAGCCT	GCTGGCTATCTGTTCCACTCA	This paper
RT-PCR			
ACTB Actin β	CTCCTTAATGTCACGCACGAT	CATGTACGTTGCTATCCAGGC	121
PKG1 Phosphoglycerate kinase1	GTTGACCGAATCACCGACCT	GTCGACTCTCATAACGACCCG	121
HPRT Hypoxanthine-guanine phosphor-ribosyltransferase	ATGGACAGGACTGAACGTCT	TCCAGCAGGTCAGCAAAGAA	121
GAPDH Glyceraldehyde-3 phosphate dehydrogenase	TCTTCTTTTGCGTCGCCAGCC	CCCAATACGACCAAATCCGTTGA	121
LIG1	TTTGTACGCCTTCGACCTCA	TGCTCGATGTCCTTGGTGTC	This paper
LIG3 a+b	CTTTTCAGCCAAGCCCAACA	CGAAACTCCCGTAGCAGACA	This paper
LIG4	CACCTTGCGTTTTCCACGAA	CAGATGCCTTCCCCCTAAGTTG	This paper
Ligation assay			
Backbone	TAAGCGATGCTCTCACCGAGAATGGCAAGGGCCAGTTTTTCT		122
nicked substrate	Phos- CTCGGTGAGAGCATCGCTTA		122
nicked substrate with Dye	DYE-AGAAAAACTGGCCCTTGCCATT Dye = dyomics781		This paper

Serum autoantibody microarray

Autoantibodies testing was performed using a protein array at the Microarray Core Facility of the University of Texas Southwestern Medical Center (UTSW, USA). In short, sera was incubated with self-antigens printed on a micro-chip, autoantibodies were detected with Cy3 and Cy5 anti-human IgG resp. IgM. Chips were read with a Genepix scanner, the resulting images were analysed using Genepix Pro 6.0 software. The net fluorescence intensity (NFI) for each antigen was calculated by subtracting the PBS control. The signal-to-noise ratio (SNR) was computed to decipher the true signal

$$\frac{median_{(Signal)} - median_{(Background)}}{(SD_{Background})}$$

Next an antibody-score was calculated = ln(NFI*SNR+1). R studio was used for the visualization and statistical computation - heatmaps were generated using the gplot heatmap.2 package, the Limma package[107] was used to compute differential expression and empirical Bayes statistics.

Statistical analysis and illustrations

For statistical evaluation, the tests that were utilized are specified in each figure legend. Error bars show the standard deviation (SD) centered on the mean unless otherwise indicated. P values were compared to an α-threshold of 0.05, set as arbitrary significance level. Data was analyzed with GraphPad Prism software (v9), linear model for differential expression analysis were performed using the R platform. Specific tests are detailed in the respective figure legends.

Figures were generated with Affinity Designer (v.1.9), cartoons were generated using the free version of the illustration tool BioRender.

Supplementary material
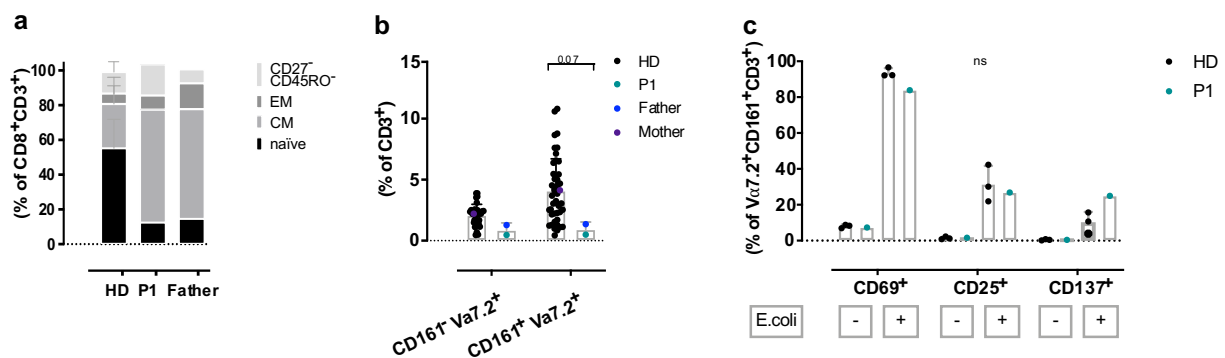
Supplementary figure S1| Multiple autoimmune manifestations and reduction of naïve T cells in the peripheral blood. a) Relative quantification of naïve, EM, CM and CD27⁻CD45RO⁻ T cell frequency with the CD8⁺CD3⁺ T cell population in P1 and her father compared to healthy controls. **b)** Subgrouping of the TCR Vα7.2⁺ T cells in mucosa-associated invariant T cells (MAIT Vα7.2⁺CD161⁺) and non-MAIT cells (Vα7.2⁺CD161⁻). **c)** MAIT cells from the peripheral blood were stimulated *in vitro* with E.coli extracts (10^8 CFU/mL) for two days, which was followed by an assessment of different surface activation markers. Statistical testing was performed in (b) using a 2wayANOVA and Šidàk correction. For the evaluation in (c) a non-parametric Kruskal-Wallis test with Dunn's correction was used.

Supplementary figure 2| P1 and her father display *TRA* and *TRB* repertoires which are preserved. a) The frequency of CDR3 lengths present within unique *TCRB* sequences was assessed. The CDR3 length distribution of the healthy mother is shown additionally as a grey shadow. **b)** A graphical representation of the *VJ* gene pairing within the total *TCRA* sequences is depicted as a heatmap. The pairing of distal *V* with distal *J* genes is highlighted by a box. **c)** Facet plot for unique clonotypes and the utilization of 62 individual *TRAV* gene segments is depicted for P1 and her father (LIG4) compared to three healthy controls (HD). Differential expression analysis was assessed with the linear modeling approach and an empirical Bayes method with a false discovery rate of 5%. Pseudogenes were filtered out beforehand.

Supplementary figure 3 |Evaluation of the *FAS* variant and transcription levels for *LIG1* and *LIG3*. a) Evaluation of naïve CD4+ T cells (CD27+CD45R0-, gray box) in P1, her father and the two healthy family members – the mother and brother. Depiction of representative flow cytometric plots, frequency within the plots indicates the percentage within the CD4+CD3+ cells. **b)** Representative flow cytometric investigation of double-negative (DNT, CD4-CD8- TCRαβ+TCRγδ- T cells) in PBMCs (gray box). The number next to the box indicates DNT frequency with the TCRαβ+TCRγδ- T cell population. **c)** Numeric evaluation of DNT in the father and P1, compared to HD and the mother. Gray shading indicates the ALPS-related threshold for DNTs[123]. **d)** Serum levels of vitamin B12 (left side) and soluble FAS ligand (FasL, right side) were measured at multiple time points in index patient P1. Gray shading indicates the ALPS-related threshold[123]. **e)** *In vitro* apoptosis assay with PHA T cell blasts. Annexin V binding (AnnV+) was measured by flow cytometry after four hours of stimulation with either anti-CD3 (OKT3), anti-FAS, recombinant FAS-ligand (rec FASL) and staurosporine (stauro). **f)** Reversed transcription qPCR was used to measure *LIG1* and *LIG3* mRNA levels in PBMCs of the patients and healthy controls including the mother. The relative quantity (RQ) was normalized to house-keeping genes and the mean of the HDs. Variance testing in (e) was performed using a Mann-Whitney test with multiple comparison correction.
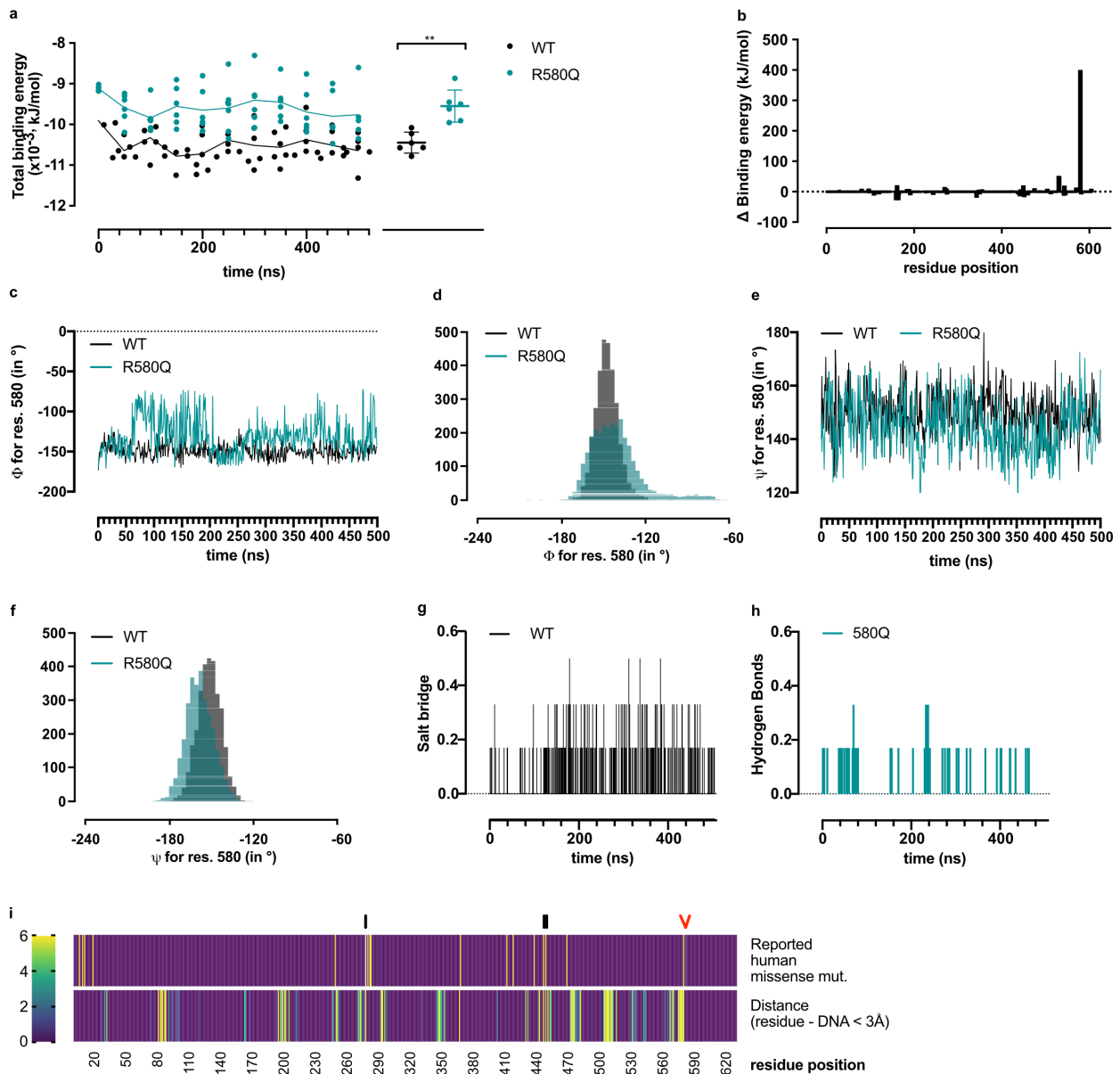
Supplementary figure 4|Decreased enzymatic activity and DNA binding capacity of the mutant LIG4 R580Q. a) Molecular dynamic simulations were used to quantify the total binding energy of the enzyme and the DNA. In each panel, the results from six independent trajectories (>500ns) for the WT and R580Q are shown. Right side, average for each time series analysis is depicted. **b)** The binding energy difference (mutant − WT) was investigated as a function of the residue position (x-axis). **c)** Representative time series of the torsion angle ϕ for the WT *vs*. R580Q residue side-chains. **d)** Distribution of the WT *vs*. R580Q residue side-chains ϕ angles. The dihedrals were extracted at an interval of 1ns over the last 100ns of simulation. **e)** Representative time series analysis regarding the torsion angle ψ for the WT *vs*. R580Q residue side-chains. **f)** Distribution of the WT *vs*. R580Q residue side-chains ψ angles. The dihedrals were extracted at intervals of 1ns over the last 100ns of simulation. Quantification of the number of g) salt bridges (for the WT) vs. **h)** hydrogen bonds (for the R580Q) formed over time between the residue and the DNA backbone. i) The human mutations found in suppl. table S2, were analyzed for their putative DNA binding capacity, utilizing the WT LIG4 trajectories and reporting if a residue was closer than <3Å to the DNA. Known missense mutations are highlighted in black on the top, the red arrow head indicates the R580Q. Significance testing in (a) was performed using the Mann-Whitney test. Mean ± SEM.

a

(% of CD25^{hi}FoxP3^{+}CD4^{+})

HD
Mother
P1
Father

CD137 PD-1 LAG3

αCD3/αCD28 − + − + − +

b

SSC-A / FSC-A

viability-NIR

CD8 / CD4

CD45RA / CCR6

FSC-H / FSC-A

SSC-A / CD3

CCR4 / CXCR3 — Th2, Th1 CCR6^{-}

CCR4 / CXCR3 — Th17, Th1* CCR6^{+}

c

TNFα^{+} (% of resp. CD3^{+})

HD
Mother
P1
Father

CD4^{+} CD8^{+}

d

(% of resp. CD4^{+}CD3^{+})

IL-2+ IL-17A+ IFNg+ TNFa+   IL-2+ IL-17A+ IFNg+ TNFa+

CD45R0^{-} CD45R0^{+}

e

(% of resp. CD8^{+}CD3^{+})

HD
P1

IL-2+ IFNg+ TNFa+   IL-2+ IFNg+ TNFa+

CD45R0^{-} CD45R0^{+}

f

CXCR5^{+}PD1^{+} (% of CD4^{+}CD3^{+})

P1

g

Cysteine index *TCRAD*   Cysteine index *TCRB*

unique sequences

total sequences

Mother
Father
P1

h

Hydrophobic index *TCRAD*   Hydrophobic index *TCRB*

unique sequences

total sequences

Mother
Father
P1

i

% of total V$_H$4-34 (×10^{-2})

HD
Father
P1

Naïve MD Memory IGHG IGHA

j

IgM autoantibody score

-log (adj.p)

log (fold change)

Supplementary figure S5| Qualitative changes in T$_{regs}$ and B cell tolerance in the two *LIG4* mutation carriers. a) Quantification of the expression of the indicated inhibitory and co-stimulatory molecules, which were measured on the surface of the CD4^{+}FoxP3^{+} T cell population, after in vitro activation with anti-CD3 and anti-CD28 mAb (αCD3αCD28). **b)** Flow cytometric gating strategy for the distinction of CD4^{+} T helper subsets in the peripheral blood based on chemokine receptor expression. T$_{H1}$ (CCR6^{-}CD45RA^{-}CXCR3^{+}CCR4^{-}), T$_{H1}$* (CCR6^{+}CD45RA^{-}CXCR3^{+}CCR4^{-}), T$_{H2}$ (CCR6^{-}CD45RA^{-}CXCR3^{-}CCR4^{+}) and T$_{H17}$ (CCR6^{+}CD45RA^{-}CXCR3^{-}CCR4^{+}). **c)** Enumeration of the intracellular TNFα production in CD4^{+} and CD8^{+} T cells after in vitro stimulation with PMA/ionomycin, by flow cytometry. **d + e)** Quantification of intracellularly produced IL-2, IFNγ, IL-17A and TNFα in CD45R0^{-} (naïve) and CD45R0^{+} (memory) CD4^{+} and CD8^{+} T cells after in vitro stimulation with PMA/ionomycin in P1 and healthy controls. **f)** Peripheral T follicular helper CD4^{+} T cells (PD-1^{+}CXCR5^{+}) were assessed in P1, gray shading indicates the in-house reference range of healthy individuals. **g)** The cysteine index and h) the hydrophobic index - two CDR3-related self-reactive indicators, were computed in the TCRAD and TCRB sequences, in total sequences

(x-axis) and unique sequences (y-axis). **i)** Relative quantification of B cell transcripts carrying the auto-reactive V_H4-34 gene segment without accounting for the mutational status. **j)** Fold change of auto-antigen specific serum IgM of the father vs. gender-matched healthy controls probed against a protein microarray chip coated with self-antigens. The horizontal dashed line indicates the significance level. In (a/c/d/e) we used a Mann-Whitney test with multiple comparison correction to estimate significance. In (d+e) the SD of the healthy controls was added manually to P1 to allow for variance testing. In (j) the significance was determined by fitting a linear model and empirical Bayes statistics.
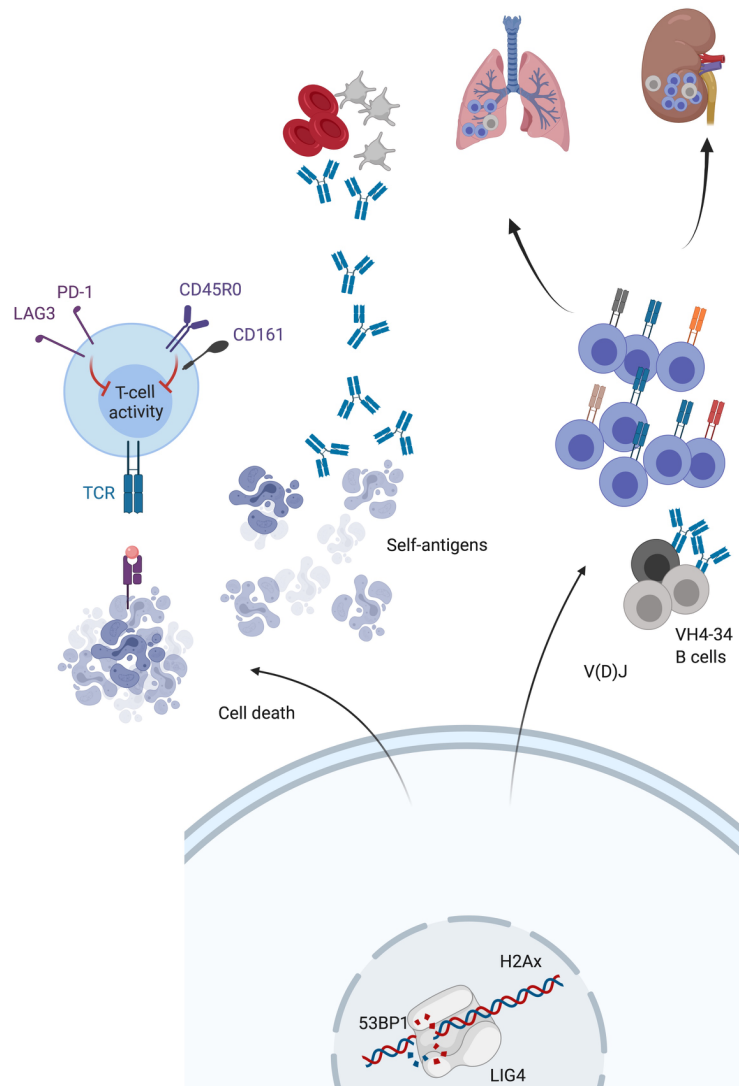
Supplementary figure S6 | Putative mechanism linking suboptimal DNA damage response and immune dysregulation. The suboptimal DNA break resolution could confer cell death susceptibility. This would be associated with a reduction of rapidly proliferating cells (ex. lymphocytes) and enhanced homeostatic proliferation. The later, could promoted the survival of autoreactive lymphocytes. T and B lymphocytes would *per se* display a subtle skewed antigen-receptor repertoire (due to an altered V(D)J recombination) and reduced selection against self-reactivity. In parallel, the cellular fragility and infection susceptibility is associated with the release of self-antigens – which would additionally fuel self-reactivity. The cartoon has been created with BioRender (using the free version).

Supplementary video V1 | Suboptimal interaction of R580Q with the DNA as compared to WT. The video shows a trajectory of the WT arginine, left, and R580Q mutant, right, and their respective interaction with the DNA duplex interaction. Residues p.580, are depicted as sticks with the carbon, nitrogen respectively, oxygen atoms colored in black, blue and red. Spheres appear when the residue side-chain is <4Å of oxygen (red) or phosphate atoms (brown) of the adenylated DNA strand. The experiment was repeated six times for each enzyme, resulting in similar observations.
Link to video: https://www.dropbox.com/sh/ryzawc7jsimae84/AABeCsJS3yVI56ExVJ5G4e6fa?dl=0

Supplementary table S1| Hematological and immunological parameters in patients with *LIG4* haploinsufficiency. Blue indicates values below the normal range, red indicates values above.

			Control Mother	Affected subjects P1	Father
Gender			F	F	M
Age			50y	19-23y	50y
Hemoglobin	g/L	(m 140-180, f 120-160)	144	113	140
MCV	fl	(81-100)	87.5	80.5	83.3
Thrombocytes	x10^9/L	150-450	235	116	159
WBC	Leukocytes	(3.5-10 x10^9/L)	5.56	3.22	5.23
	Neutrophils	40-74% (1.3-6.7 x10^9/L)	59.4% (3.37)	75.2% (2.42)	65.1% (3.47)
	Monocytes	3.4-9% (0.12- 0.62 x10^9/L)	5.7% (0.33)	6.5% (0.21)	8.9% (0.47)
	Eosinophils	0-7% (0-0.3 x10^9/L))	1.6% (0.04)	1.2% (0.04)	4.3% (0.23)
	Basophils	0-1.5% (0-0.09 x10^9/L)	0.9% (0.05)	0.9% (0.03)	0.8% (0.04)
	Total lymphocytes	20-45% (0.9-3.3 x10^9/L)	31.1% (1.77)	15.2% (0.49)	19.2% (1.02)
% of lymphocytes	Total T cells, CD3$^+$	55-86% [742-2750 /uL]	70.9%	74.5% [706/uL]	76%
	Total B cells, CD19$^+$CD20$^+$	5-22% [80-616 /uL]	4.83%	2.8% [134/uL]	7.8%
	Total NK cells, CD56$^+$CD16$^+$	5-26% [84-724 /uL]		9% [85/uL]	
	Surface markers	Range %/ [cells/uL]			
Total helper T	CD3$^+$ CD4$^+$	33-58% [404-1612/uL]	75.7%	60% [571/uL]	63.5%
Helper T, naïve	CD27$^+$ CD45RO$^-$	15.7-54.7%	40.2%	4.41%	9.75%
Helper T, central memory	CD27$^+$ CD45RO$^+$	8-28.8%	51%	85.9%	73.8%
Helper T, effector memory	CD27$^-$ CD45RO$^+$	16.8-57.4%	7.56%	9.46%	15.1%
Helper T, follicular	CXCR5$^+$PD1$^+$	6.9-19.1%		6.6%	
Helper T, regulatory	CD25^{hi}CD127^{low}	6.1-11%	6.23%	2.3%	3.41%
Helper, recent thymic emigrants	CD31$^+$CD27$^+$CD45R0$^-$	14.1-37.2% (% of naïve)		39.8%	
Total cytotoxic T	CD3$^+$ CD8$^+$	13-39% [220-1129/uL]	22.5%	15% [141/uL]	32.2%
Cytotoxic T, naïve	CD27$^+$ CD45RO$^-$	7-62.5%	32.7%	2%	12.7%
Cytotoxic T, central memory	CD27$^+$ CD45RO$^+$	0.6-4.4%	38.8%	0.4%	62.2%
Cytotoxic T, effector memory	CD27$^-$ CD45RO$^+$	4.3-64.5%	9.87%	68.3%	17.2%
Cytotoxic, TMRA	CD45RA$^+$CD62L$^-$	8.1-60.5%		29.2%	
Total double negative T	TCRαβ$^+$TCRγδ$^-$CD8$^-$CD4$^-$	≤ 2.5%	0.67%	0.66%	2.68%
Total, γδ T	TCRβ$^-$TCRγδ$^+$		2.87%	3.21%	2.45%
Total, Vα7.2$^+$	Vα7.2$^+$ CD3$^+$	3.26-8.85% (of CD3$^+$)	6.1%	1%	2.94%
Total, MAIT	Vα7.2$^+$ CD161$^+$CD3$^+$	1.55-6.77% (of CD3$^+$)	4.1%	0.47%	1.41%
Plasma blasts	CD27$^+$CD38$^+$CD20$^-$CD19$^+$	0.1-3% (of CD19$^+$ cells) [1-5/uL]	0.97%	<0.1% [<1/uL]	0.24%
Naïve B	IgD$^+$IgM$^+$CD27$^-$	25.1-92.4% [66-228/uL]	56.8%	60.5% [81/uL]	30.6%
Switched memory B	IgD$^-$IgM$^-$CD27$^+$	2.4-32.6% [8-102/uL]	5.88%	0.1% [<1/uL]	15%
Transitional B	IgD$^+$IgM$^+$CD38$^+$	0.3-2.9% [1-5uL]	0.77%	2.2% [<1/uL]	0.56%
Marginal zone like B	IgD$^+$IgM$^+$CD27$^+$	3.1-59.7% [8-172/uL]	14.3%	32.3% [43/uL]	26%
CD21^{low} B cells	CD21^{low}	0.5-4.7% [1-12/uL]	3.79%	16.6%	7.2%
Immunoglobulins					
IgG		7-16 g/L	9.73	<0.1	6.58
IgM		0.4-2.3 g/L	0.824	0.25	1.32
IgA		0.7-4 g/L	2.95	< 0.06	0.58

F female, Ig immunoglobulin, MAIT mucosa associated invariant T cell, m male, MCV mean corpuscular volume, NK Natural killer, WBC white blood count.

Supplementary table S2: Additional in silico predictor values for missense mutations in *LIG4* and *FAS*.

Genetic location	Nucleotide change (c.DNA)	Zygosity	Gene symbol	AA change	REVEL	SpliceAI	PrimateAI
13_108209530	c.A1739G	Het	*LIG4*	p.R580Q R[CGA] > Q[CAA]	0.82	0 (no consequence)	0.736
10_90768694	c.G383A	Homo (Father) Het (P1)	*FAS*	p.R128K R[AGA] > K[AAA]	na	na	na

AA amino acid, Chr. chromosome, Het heterozygous, Homo homozygous, REVEL, SpliceAI Splice artificial intelligence, PrimateAI primate artificial intelligence.

Supplementary table S3 | Clinical and genetic features of published patients with *LIG4* deficiency. Patients are ordered according to the 5' position of the first mutated allele.

Patient mutation	Genomic/biochemical / cellular characterization	Immunology and bone marrow	Infections	Autoimmunity/-inflammation and cutaneous phenotype	Malignancy and other observations	Sex & age at onset	Ref
p.A11G/ c.C32G p.N412K/ c.T1236T comp. het.		Ig normal, ⇓ CD4⁺ cells	Mild URTI	-	-	M, 1y	[95]
L19W p.L635fs*10X/ c.1904delT comp. het.		Ig normal ⇓ B cells, ⇓ CD4⁺ T cells, ⇑ % CD4⁺ CM ⇓ NK cells Pancytopenia			Growth restriction, microcephaly, neuro-developmental delay	M, 3y	[124]
p.S205Lfs*27X / c.613delT p.L635Rfs*10X/ c.1904delT comp. het.		⇓ IgG ⇓ B cells ⇓ Thrombocytes	Urinary tract infections sepsis, diarrhea	Erythema	icterus, tubulopathy	M, 0.25y	[125,126]
p.S205Lfs*27X /c.613delT p.R814X/c.C2440T comp. het.	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	Pancytopenia	-	-	Metastatic anal cancer Dubowitz syndrome	F, 34y	[125,128]
p.S205Lfs*29X / c.613delT p.H282L/ c.C845A comp. het. + balanced translocation (t(1;19)(q21;p13))	none	Leucopenia, anemia, thrombocytopenia Normal IgG/IgM/IgA, ⇓ IgG2 No B cells ⇓ CD3, ⇓ CD8, ⇓ NK	Multiple viral and bacterial infections, RTI, UTI, CMV, norovirus	Hepatomegaly	Failure to thrive, microcephaly, bloody diarrhea	M, 2.5y*	[97]
p.S205Lfs*29X / c.613delT p.H282L/ c.C845A comp. het. + balanced translocation (t(1;19)(q21;p13))	none	Leucopenia, anemia, thrombocytopenia ⇓ CD3, ⇓ CD19, ⇓ NK ⇓ IgA				M, 0.75y	[97]
p.M249V/ c.A745Gc p.K424fs*20X/c.1271_1275 delAAAGA comp. het		⇓ IgM, IgG T⁻B⁻NK⁺ CID Neutropenia	Pulmonary aspergillosis		Non-Hodgkin lymphoma (mouth)	F, 14y *	[129]
p.M249V/ c.A745G p.K424fs*20X/c.1271_1275 delAAAGA comp. het	Cellular bleomycin sensitivity normal	⇓ IgG, IgM, IgA, IgE ⇓ CD3, CD19, NK cells Progressive pancytopenia	Recurrent infections (respiratory, gut, urinary), bilateral bronchiectasis	Jaundice sclerosing cholangitis	Hepatosplenomegaly 1° amenorrhea, portal hypertension with esophageal varices −, gut bleeding	F, 1y	[95]
p.R278H/ c.G833A homo	R278H: Reduced adenylation, reduced ligation activity to 5-10% of wt protein. V(D)J impacts signal joint formation[130]. XRCC4 interaction intact[127]. R278H-mouse model[131].	⇑ IgG, normal IgM, ⇓ IgA ⇓ CD8⁺ T cells ⇓ B cells ⇑ NK cells	Chronic calicivirus	-	-	3y	[132]
p.R278H/ c.G833A homo	R278H: Reduced adenylation, reduced ligation activity to 5-10% of wt protein. V(D)J impacts signal joint formation[130]. XRCC4 interaction intact[127]. R278H-mouse model[131].	⇓ IgG, ⇓ IgA, ⇓ IgM B cell absent ⇓ naïve CD4⁺/CD8⁺ T cells NK normal	Perianal abscesses, suppurative hidradenitis, pneumonia (H. influenzae type B)	Hypopigmentation		M, 16y	[124]

Patient mutation	Genomic/biochemical / cellular characterization	Immunology and bone marrow	Infections	Autoimmunity/-inflammation and cutaneous phenotype	Malignancy and other observations	Sex & age at onset	Ref
p.R278H/ c.G833A homo	R278H: Reduced adenylation, reduced ligation activity to 5-10% of wt protein. V(D)J impacts signal joint formation[130]. XRCC4 interaction intact[127]. R278H-mouse model[131].	⇓ IgA, ⇓ IgM, IgG normal ⇓ B cells ⇓ naïve CD4$^+$/CD8$^+$ T cells NK normal	Asymptomatic	Hypopigmentation		M, 7y	[124]
p.R278H/ c.G833A p.K424fs*20X/c.1271_1275 delAAAGA comp. het	R278H: Reduced adenylation, reduced ligation activity to 5-10% of wt protein. V(D)J impacts signal joint formation[130]. XRCC4 interaction intact[127]. R278H-mouse model[131].	No IgM, IgA, IgG normal, ⇓ IgG3 ⇓ B cells ⇓ CD4$^+$, ⇓ naïve CD4/CD8 T cells (⇓) Hemoglobin TCR spectraltyping polyclonal, no TRECs	Recurrent sinopulmonal/ ear infections, Bronchiectasis			M, 6y	[133]
p.R278H/ c.G833A p.K424fs*20X/c.1271_1275 delAAAGA comp. het	R278H: Reduced adenylation, reduced ligation activity to 5-10% of wt protein. V(D)J impacts signal joint formation[130]. XRCC4 interaction intact[127]. R278H-mouse model[131].	IgG normal, ⇓ IgM and IgA ⇓ CD4$^+$ and CD8$^+$ ⇓ B cells Normal NK counts	Pneumonia, otitis	-	-	8y	[132]
p.R278H/ c.G833A + p.A3V + p.T9I/ c.C8T + c.C26T homo for all 3 mut	R278H+A3V+T9I: In vitro no adenylation and DNA ligation[30].	⇓Ig Pancytopenia (Ec, lymphocytes, Tc)	Warts, (bronchiectasis)		DLBCL (EBV neg)	M, 9y	[30]
p.R278H/ c.G833A homo	R278H: Reduced adenylation, reduced ligation activity to 5-10% of wt protein. V(D)J impacts signal joint formation[130]. XRCC4 interaction intact[127]. R278H-mouse model[131].	NA			AL leukemia Prophylactic Radiation: ulceration, lethargy, tetraparesis	F, 14y *	[39,130,134,135]
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA comp. het	R278H: Reduced adenylation, reduced ligation activity to 5-10% of wt protein. V(D)J impacts signal joint formation[130]. XRCC4 interaction intact[127]. R278H-mouse model[131].	Ig normal ⇓ CD4 cells			Non-Hodkin lymphoma (brain)	M, 1.5y	[95]
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA (probably) comp. het	R278H: Reduced adenylation, reduced ligation activity to 5-10% of wt protein. V(D)J impacts signal joint formation[130]. XRCC4 interaction intact[127]. R278H-mouse model[131].	Ig normal ⇓ CD4 and CD8 cells	Recurrent pneumonia, sepsis, vulvovaginal candidiasis, diaper area dermatitis		DLBCL (lung)	F, 0.75y	[95]
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275	R278H: Reduced adenylation, reduced ligation activity to 5-	Ig normal ⇓ CD4 and CD8 cells	Recurrent pneumonia, sepsis, CMV, viral		DLBCL	M, 0.75y	[95]

Patient mutation	Genomic/biochemical / cellular characterization	Immunology and bone marrow	Infections	Autoimmunity/-inflammation and cutaneous phenotype	Malignancy and other observations	Sex & age at onset	Ref
delAAAGA probably) comp. het	10% of wt protein. V(D)J impacts signal joint formation[130]. XRCC4 interaction intact[127]. R278H-mouse model[131].	CD19, NK cells normal ⇓ Hemoglobin	encephalitis, recurrent infections (gut, respiratory, urinary), perianal ulcers				
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA comp. het							10
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA (probably) comp. het	No evidence	Ig NA ⇓ Thrombocytes	Diarrhea, pneumonia, otitis media, oral candidiasis	Vitiligo		M, 11m	60
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA comp. het	No evidence	Ig NA ⇓ naïve CD4/CD8 T cells	Pneumonia, stomatitis, recurrent fever, diarrhea		Impaired liver function, intestinal ulceration	M, 1m	60
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA (probably) comp. het	No evidence	Ig NA ⇓ naïve CD4/CD8 T cells	Omphalitis, pneumonia, oral candidiasis, stomatitis, skin & soft tissue infection			F, 0.25m	60
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA comp. het	No evidence	Ig NA ⇓ Thrombocytes	Pneumonia, diarrhea	Erythroderma		M, 4m	60
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA (probably) comp. het	No evidence	⇓ IgG, IgM		Eczema	Generalized lymphadenopathy	M, 1m	60
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA comp. het	No evidence	Ig NA ⇓ naïve CD4/CD8 T cells Pancytopenia	Pneumonia, oral candidiasis, stomatitis, herpes simplex, diarrhea			F, 1.2y	60
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA (probably) comp. het	No evidence	⇓ IgG, IgM ⇓ naïve CD4/CD8 cells ⇓ Thrombocytes	Recurrent URTI, diarrhea, otitis media, pneumonia			F, 3y	60
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA comp. het	Sanger Radiosensitivity study	⇓ IgM, IgA ⇓ T cells, B cells ⇓ naïve T cells		Rush	Failure to thrive	F, 0.25y *	136
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA (probably) comp. het		⇓ IgM and IgG, no IgA T⁻B⁻NK⁺ SCID lymphopenia	Otitis, bronchiolitis, pneumonia, sepsis			F, 1.5y	137
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275		⇓ IgM, maternal IgG, no IgA small thymus	not reported EBV PTLD post HSCT			F, birth	137

Patient mutation	Genomic/biochemical / cellular characterization	Immunology and bone marrow	Infections	Autoimmunity/-inflammation and cutaneous phenotype	Malignancy and other observations	Sex & age at onset	Ref
delAAAGA comp. het		T⁻B⁻NK⁺ SCID lymphopenia					
p.R278H/ c.G833A p.K424fs*20X/ c.1271_1275 delAAAGA (probably) comp. het		⇓ IgM, no IgA Lymphopenia	NA	VOD post HSCT	-	F, birth	138
p.H282L/ c.A118T (iso?) p.D423fs422X/ c.1544-1548delAAAGA (iso?) comp. het		⇑ IgM and IgG, IgA normal Lymphopenia, neutropenia, thrombocytopenia	Recurrent infections, sepsis, chronic diarrhea	Autoimmune cytopenia	Non Hodgkin lymphoma (brain and lung)	F, 2y *	138
p.H282L/ c.C845T c.C26T (SNP) p.R582fsX/ c.1747_1751delAAGAT comp. het		⇓ IgM, IgA B cell lymphoproliferation Eosinophilia, lymphopenia	EBV after HSCT	Omenn syndrome Hepatosplenomegaly lymphadenopathy	GVHD	F, 0.5y	139
p.K283E / c.A847A p.K424fs*20X/ c.1271_1275 delAAAGA (probably) comp. het		Pancytopenia				F, 1.5y	140
p.K283E / c.A847A p.K424fs*20X/c.1271_1275 delAAAGA comp. het	Sanger	Ig NA Hemorrhagic syndrome Aplasia	HPV warts, recurrent respiratory infections, chronic diarrhea	-		M, 9y	140
p.Q229R/ c.A875G p.K436Rfs*20/ c.1307_1311del comp. het.							10
c.G907A c.1904delA comp. het	none	Leukopenia ⇓ T cells (CD4, CD8), (⇓) B cells, NK cells normal ⇓ IgG, IgG2, IgA IgM/E normal Thrombocytes normal Neutropenia			Microcephaly, microsomia, dysmorphic features	F, birth	141
p.D368V / c.A1103T p.W447C / c.G1341T comp. het		⇓ IgG, IgM, IgA ⇓ T cells, B cells	Bronchiectasis, long-term presence of vaccine-derived rubella virus	Granulomatous dermatitis	Villous atrophy	F, 4y	142
p.D368Y / c.G1102T homo	Sanger	⇓ IgG, IgM, IgA ⇓ T cells (CD4, CD8), ⇓B cells NK cell normal	Multilobar pneumonia Parainfluenza virus type III, P. aeruginosa, P. jirovecii	Eczema	Failure to thrive, microcephaly Thymus: Absence of Hassall's corpuscles, lack cortico-medullary demarcation, lack of T cells and Langerhans histiocytes in the thymic medulla	F, 0.3y	143
p.L418Mfs*3/ c.1245_1250dupGATGC p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	⇓ IgG ⇓ Thrombocytes	-	-	Growth failure	F, 1.75y	144

Patient mutation	Genomic/biochemical / cellular characterization	Immunology and bone marrow	Infections	Autoimmunity/- inflammation and cutaneous phenotype	Malignancy and other observations	Sex & age at onset	Ref
p.K424fs*20/ c.1271_1274del p.R814X/ c.C2440T comp. het							[10]
p.K424fs*20X/ c.1271_1275 delAAAGA p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	⇓ IgG ⇓ CD3, CD8 ⇓ CD19 ⇓ Ec, Tc	-	Psoriasis	-	F, 2.5y	[144]
p.K424fs*20X/ c.1271_1275 delAAAGA p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	⇓ IgG, ⇓ CD3, CD8, CD4 ⇓ CD19	-	-	Hip dysplasia, toe syndactyly	M, 2y	[144]
p.K424fs*20X/ c.1271_1275 delAAAGA p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	⇓ IgG ⇓ CD19 ⇓ Thrombocytes	Mild recurrent (gut, skin, respiratory)	-	-	F, 2y	[144]
p.K424fs*20X/ c.1271_1275 delAAAGA p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	IgG normal ⇓ CD19 ⇓ Thrombocytes	Mild recurrent (gut, skin, respiratory)	Hypopigmentation	Toe syndactyly	M, 5.5y	[144]
p.K424fs*20X/ c.1271_1275 delAAAGA p.R814X/ c.C2440T comp. het	UV-exposure of fibroblasts (normal). R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	⇓ IgM, IgG IgA normal Pancytopenia	Respiratory infections		Brachymetatarsy, syndactyly	F, 7y	[140]
p.K424fs*20X/ c.1271_1275 delAAAGA p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	⇓ IgG ⇓ CD4+, no CD19+ pancytopenia (self-resolved) ⇓ Thrombocytes	mild, influenza		Cerebral aneurysma 1° ovary failure	F, 17y	[144]
p.K424fs*20X/ c.1271_1275 delAAAGA p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	⇓ IgG ⇓ CD4+, ⇓ naïve T cells, no CD19+ pancytopenia (persistent) ⇓ Thrombocytes	mild recurrent (skin, gut, respiratory)		Atrial-ventricular septal defect, atrophic kidney, rip hypoplasia, fusion carpal bones, abnormal C1 vertebrae, platybasia, 1° ovary failure	F, 11y	[144]
p.Q433del/ c.1297-1299delCAA homo		⇓ IgM, IgG, IgA T-B-NK+ SCID ⇓ Ec	Respiratory infections Candidiasis, chronic diarrhea (failure to thrive)			F, 1.5y*	[145]
p.Y438H/ c.T1312c homo mutation in LRIG2 (homo)	Sanger	(⇓)IgG2, IgM IgG, IgG1, IgG3, IgG4 normal ⇑ IgA ⇓ CD4+, no CD19 Pancytopenia	Pneumonia, otitis media, sinusitis, oral candidiasis, urosepsis			F, 7y	[146]
p.K449Q/ c.A1345C p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	⇓ IgM, IgG, IgA, IgE ⇓ CD3, CD4, naïve CD4 and CD8	Recurrent pneumonia, sinusitis		-	F, 7y	[73]

Patient mutation	Genomic/biochemical / cellular characterization	Immunology and bone marrow	Infections	Autoimmunity/-inflammation and cutaneous phenotype	Malignancy and other observations	Sex & age at onset	Ref
		⇓ CD19 TRB mildly reduced diversity/ IGH preserved, shorter CDR3					
p.K449Q/ c.A1345C p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	IgG normal ⇓ CD3, CD4, naïve CD4 and CD8 ⇓ CD19 (⇓) Neutrophils TRB mildly reduced diversity/ IGH preserved, shorter CDR3	Walking pneumonia		-	M, 23y	[73]
p.K449Q/ c.A1345C p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	⇓ IgG, ⇑ IgA ⇓ CD3, CD4, naïve CD4 and CD8 ⇓ CD19 ⇓ Hemoglobin (⇓) Neutrophils TRB mildly reduced diversity/ IGH preserved, shorter CDR3	Asymptomatic	Asymptomatic	Asymptomatic	F, 12y	[73]
p.K449Q/ c.A1345C p.R814X/ c.C2440T comp. het		⇓ IgG, IgM, IgA ⇓ CD4+ T cells ⇓ B cells NK counts normal	Recurrent URTI/ LRTI, bronchiectasis, cutaneous ringworm infection	-	-	17y	[132]
p. G469E/ c.G1406A p.R814X/ c.C2440T comp. het	G469E, normal XRCC4 interaction, no adenylation, no dsDNA nor nick ligation[30]. R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127] IR/Bleomycin of PBMCs augmented cell death, increased chromosomal breakage	IgG normal ⇓ T cells, B cells ⇓ Hemoglobin ⇓ Thrombocytes Hypoplastic bone marrow	Recurrent sinopulmonal infections	Psoriasiform erythrodermic squamous skin patches	Hypergonadotrophic hypogonadism	F, 9y	[30,147]
p.R505Cfs*12X/ c.1512_1513delTC p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	Ig NA ⇓ Thrombocytes	-	-	-	F, 3.5y	[144]
p.K588del/ c.1762delAAG homo		⇓ IgM Pancytopenia	Recurrent infections (sinopulmonal)			F, 10y	[148]
p.K588del/ c.1762delAAG homo		⇓ IgM, IgG	URTI, urinary tract infections		spontaneous chromosomal breakage	M, 6y	[148]
p.R580X/c.C1738T p.R814X/ c.C2440T comp. het	p.R580X → truncated protein, lacks nuclear localization signal. R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	Ig NA Pancytopenia	Sinusitis	hypothyroidism, hypogonadism, DM II, chronic cutaneous affection, photosensitivity, telangiectasia	Myelodysplasia, pancytopenia	M, 46y	[127]

Patient mutation	Genomic/biochemical / cellular characterization	Immunology and bone marrow	Infections	Autoimmunity/-inflammation and cutaneous phenotype	Malignancy and other observations	Sex & age at onset	Ref
p.R580X/ c.C1738T p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	Ig NA	Respiratory tract infections	Hypothyroidism, amenorrhea Photosensitivity, psoriasis	-	F, 48y	[127]
p.Y698X/ c.C2094T p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	Ig NA Pancytopenia (self-resolved) ⇓ Thrombocytes	-		Anal atrasia (with rectovaginal fistula), esotropia	F, 8y	[125]
p.A797Dfs*3/ c.2386_2389dupATTG p.R814X/ c.C2440T comp. het	R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	⇓ IgG ⇓ CD3, CD8 ⇓ CD19 ⇓ Thrombocytes	Mild recurrent (skin, gut, respiratory)	Psoriasis		F, 2.5y	[125]
p.R814X/ c.C2440T homo	WB: severely reduced protein R814X reduced adenylation, impaired interaction with XRCC4, reduced ligation[127]	Ig NA (aplasia after chemo)	Sepsis (neutropenia after chemo)	Hypogonadism	ALL	M, 4y*	[149]
p.R871H/ c.G2612A homo	Sanger, augmented cell death after IR	Ig normal Lymphocytes normal Hemoglobin normal Thrombocytes normal		Behçet disease (recurrent meningitis, genital/oral ulcers), anterior uveitis	Dysembryoplastic neuroepithelial tumor	M, 10y	[150]
c.2736+3delC p.A3V + p.T9I/c.C8T + c.C26T	none	⇓ IgA ⇓ T cells	EBV stomatitis/ encephalitis		EBV+ DLBCL (lung & brain)	F, 2y	[151]
NA	NA	NA	NA	Anemia Cytopenia	AML Died after irradiation/ Chemotherapy Not Seckel Syndrome because irradiation sensitivity	F, 26y *	[148]

AML acute myeloid leukaemia, comp. het compound heterozygous, DLBCL diffuse large B cell lymphoma, Ec erythrocytes, F female, homo homozygous mutation, HSCT hematopoietic stem cell transplantation, Ig immunoglobulin, IR irradiation, LCL lymphoblastoid cell lines, M male, NA not available, URTI upper respiratory tract infection, y years. cDNA sequence refers to NM_001098268. Lymphocytes reference values were compared to Tosato F et al[152].
* died after conditioning or radio/chemotherapy.

References

1. Caron, P., van der Linden, J. & van Attikum, H. Bon voyage: A transcriptional journey around DNA breaks. *DNA Repair (Amst)*. **82**, (2019).
2. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).
3. Pascal, J. M. DNA and RNA ligases: structural variations and shared mechanisms. *Curr. Opin. Struct. Biol.* **18**, 96–105 (2008).
4. Notarangelo, L. D., Bacchetta, R., Casanova, J. L. & Su, H. C. Human inborn errors of immunity : An expanding universe. **1662**, (2020).
5. Notarangelo, L. D., Kim, M.-S., Walter, J. E. & Lee, Y. N. Human RAG mutations: biochemistry and clinical implications. *Nat. Rev. Immunol.* **16**, 234–46 (2016).
6. Tangye, S. G. *et al.* Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. *J. Clin. Immunol.* **40**, 24–64 (2020).
7. Frank, K. M. *et al.* Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature* **396**, 173–177 (1998).
8. Sharpless, N. E. *et al.* Impaired Nonhomologous End-Joining Provokes Soft Tissue Sarcomas Harboring Chromosomal Translocations, Amplifications, and Deletions. *Mol. Cell* **8**, 1187–1196 (2001).
9. Ruccia, F. *et al.* Homozygous DNA ligase IV R278H mutation in mice leads to leaky SCID and represents a model for human LIG4 syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 3024–3029 (2010).
10. Berland, A. *et al.* PROMIDISα: A T-cell receptor α signature associated with immunodeficiencies caused by V(D)J recombination defects. *J. Allergy Clin. Immunol.* **143**, 325-334.e2 (2019).
11. Chitty-Lopez, M. *et al.* Asymptomatic Infant With Atypical SCID and Novel Hypomorphic RAG Variant Identified by Newborn Screening: A Diagnostic and Treatment Dilemma. *Front. Immunol.* **11**, (2020).
12. Davis, M. M. *et al.* LIGAND RECOGNITION BY αβ T CELL RECEPTORS. *Annu. Rev. Immunol.* **16**, 523–544 (1998).
13. Davis, M. M. *et al.* T Cell Receptor Biochemistry, Repertoire Selection and General Features of TCR and Ig Structure. in *Ciba Found Symp.* 94–104 (1997). doi:10.1002/9780470515280.ch7
14. Lee, Y. N. *et al.* Characterization of T and B cell repertoire diversity in patients with RAG deficiency. *Sci. Immunol.* **1**, 1–13 (2016).
15. Shannon, C. E. The mathematical theory of communication. *MD. Comput.* **14**, 306–17 (1963).
16. Simpson, H. Measurement of Diversity. *Nature* **163**, 688–688 (1949).
17. Brandle, D., Muller, C., Rulicke, T., Hengartner, H. & Pircher, H. Engagement of the T-cell receptor during positive selection in the thymus down-regulates RAG-1 expression. *Proc. Natl. Acad. Sci.* **89**, 9529–9533 (1992).
18. Rowe, J. H. *et al.* Abnormalities of T-cell receptor repertoire in CD4+ regulatory and conventional T cells in patients with RAG mutations: Implications for autoimmunity. *J. Allergy Clin. Immunol.* **140**, 1739-1743.e7 (2017).
19. Wehr, C. *et al.* A new CD21low B cell population in the peripheral blood of patients with SLE. *Clin. Immunol.* **113**, 161–171 (2004).
20. De Villartay, J. P., Fischer, A. & Durandy, A. The mechanisms of immune diversification and their disorders. *Nat. Rev. Immunol.* **3**, 962–972 (2003).
21. Uduman, M., Shlomchik, M. J., Vigneault, F., Church, G. M. & Kleinstein, S. H. Integrating B Cell Lineage Information into Statistical Tests for Detecting Selection in Ig Sequences. *J. Immunol.* **192**, 867–874 (2014).
22. Ghraichy, M. *et al.* Maturation of naïve and antigen-experienced B-cell receptor repertoires with age. *bioRxiv* 609651 (2019). doi:10.1101/609651
23. Bashford-Rogers, R. J. M. *et al.* Analysis of the B cell receptor repertoire in six immune-mediated diseases. *Nature* **574**, 122–126 (2019).
24. Woodbine, L., Gennery, A. R. & Jeggo, P. A. Reprint of 'The clinical impact of deficiency in DNA non-homologous end-joining'. *DNA Repair (Amst)*. **17**, 9–20 (2014).
25. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).
26. Kaminski, A. M. *et al.* Structures of DNA-bound human ligase IV catalytic core reveal insights into substrate binding and catalysis. *Nat. Commun.* **9**, (2018).
27. Kircher, M., Witten, D., Jain, P., O'Roak, B. & Cooper, G. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310–5 (2014).
28. Adzhubei, I., Schmidt, S., Peshkin, L., Ramensky, V. & Gerasimova, A. A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248–9 (2017).
29. Kumar, P., Henikoff, S. & Ng, P. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073–81 (2009).
30. Girard, P. M., Kysela, B., Härer, C. J., Doherty, A. J. & Jeggo, P. A. Analysis of DNA ligase IV mutations found in LIG4 syndrome patients: The impact of two linked polymorphisms. *Hum. Mol. Genet.* **13**, 2369–2376 (2004).
31. Oliveira, J. B. *et al.* Revised diagnostic criteria and classification for the autoimmune lymphoproliferative syndrome (ALPS): Report from the 2009 NIH International Workshop. *Blood* **116**, 35–41 (2010).
32. Bignucolo, O., Leung, H. T. A., Grzesiek, S. & Bernèche, S. Backbone Hydration Determines the Folding Signature of Amino Acid Residues. *J. Am. Chem. Soc.* **137**, 4300–4303 (2015).
33. Bignucolo, O., Vullo, S., Ambrosio, N., Gautschi, I. & Kellenberger, S. Structural and Functional Analysis of Gly212 Mutants Reveals the Importance of Intersubunit Interactions in ASIC1a Channel Function. *Front. Mol. Biosci.* **7**, (2020).
34. Burgener, A. V. *et al.* SDHA gain-of-function engages inflammatory mitochondrial retrograde signaling via KEAP1–Nrf2. *Nat. Immunol.* **20**, 1311–1321 (2019).
35. Kaminski, A. M. *et al.* Structures of DNA-bound human ligase IV catalytic core reveal insights into substrate binding and catalysis. *Nat. Commun.* **9**, 1–12 (2018).
36. Srinivasan, J., Cheatham, T., Cieplak, P., Kollman, P. & Case, D. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices. *J. Am. Chem. Soc.* **120**, 9401–9409 (1998).
37. Kumari, R., Kumar, R. & Lynn, A. g_mmpbsa —A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **54**, 1951–1962 (2014).
38. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **98**, 10037–10041 (2001).
39. Riballo, E. *et al.* Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient. *Curr. Biol.* **9**, 699-S2 (1999).
40. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273**, 5858–5868 (1998).
41. Panier, S. & Boulton, S. J. Double-strand break repair: 53BP1 comes into focus. *Nat. Rev. Mol. Cell Biol.* **15**, 7–18 (2014).

42.   Arroyo Hornero, R. *et al.* CD70 expression determines the therapeutic efficacy of expanded human regulatory T cells. *Commun. Biol.* **3**, 375 (2020).

43.   Yi, H., Zhen, Y., Jiang, L., Zheng, J. & Zhao, Y. The phenotypic characterization of naturally occurring regulatory CD4+CD25+ T cells. *Cell. Mol. Immunol.* **3**, 189–195 (2006).

44.   Camisaschi, C. *et al.* LAG-3 Expression Defines a Subset of CD4 + CD25 high Foxp3 + Regulatory T Cells That Are Expanded at Tumor Sites. *J. Immunol.* **184**, 6545–6551 (2010).

45.   Raimondi, G., Shufesky, W. J., Tokita, D., Morelli, A. E. & Thomson, A. W. Regulated Compartmentalization of Programmed Cell Death-1 Discriminates CD4 + CD25 + Resting Regulatory T Cells from Activated T Cells. *J. Immunol.* **176**, 2808–2816 (2006).

46.   Pesenacker, A. M. *et al.* CD161 defines the subset of FoxP3 1 T cells capable of producing proinflammatory cytokines. **121**, 2647–2659 (2019).

47.   Afzali, B. *et al.* CD161 expression characterizes a subpopulation of human regulatory T cells that produces IL-17 in a STAT3-dependent manner. *Eur. J. Immunol.* **43**, 2043–2054 (2013).

48.   Acosta-Rodriguez, E. V *et al.* Surface phenotype and antigenic specificity of human interleukin 17–producing T helper memory cells. *Nat. Immunol.* **8**, 639–646 (2007).

49.   Stadinski, B. D. *et al.* Hydrophobic CDR3 residues promote the development of self-reactive T cells. *Nat. Immunol.* **17**, 946–955 (2016).

50.   Daley, S. R. *et al.* Cysteine and hydrophobic residues in CDR3 serve as distinct T-cell self-reactivity indices. *J. Allergy Clin. Immunol.* **144**, 333–336 (2019).

51.   Goodnow, C. C., Sprent, J., de St Groth, B. F. & Vinuesa, C. G. Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature* **435**, 590–597 (2005).

52.   Victora, G. D. & Nussenzweig, M. C. Germinal Centers. *Annu. Rev. Immunol.* **30**, 429–457 (2012).

53.   Meffre, E. & Wardemann, H. B-cell tolerance checkpoints in health and autoimmunity. *Curr. Opin. Immunol.* **20**, 632–638 (2008).

54.   Wardemann, H., Hammersen, J. & Nussenzweig, M. C. Human autoantibody silencing by immunoglobulin light chains. *J. Exp. Med.* **200**, 191–199 (2004).

55.   Larimore, K., McCormick, M. W., Robins, H. S. & Greenberg, P. D. Shaping of Human Germline IgH Repertoires Revealed by Deep Sequencing. *J. Immunol.* **189**, 3221–3230 (2012).

56.   Pugh-Bernard, A. E. *et al.* Regulation of inherently autoreactive VH4-34 B cells in the maintenance of human B cell tolerance. *J. Clin. Invest.* **108**, 1061–1070 (2001).

57.   Frank, K. M. *et al.* Late embryonic lethality and impaired V (D)J recombination in mice lacking DNA ligase IV. *Nature* **396**, 173–177 (1998).

58.   Altmann, T. & Gennery, A. R. DNA ligase IV syndrome; a review. *Orphanet J. Rare Dis.* **11**, 1–7 (2016).

59.   Rucci, F. *et al.* Homozygous DNA ligase IV R278H mutation in mice leads to leaky SCID and represents a model for human LIG4 syndrome. *Proc. Natl. Acad. Sci.* **107**, 3024–3029 (2010).

60.   Sun, B. *et al.* LIG4 syndrome : clinical and molecular characterization in a Chinese cohort. 1–9 (2020).

61.   Fairchild, C. R., Maybaum, J. & Kennedy, K. A. Concurrent unilateral chromatid damage and DNA strand breakage in response to 6-thioguanine treatment. *Biochem. Pharmacol.* **35**, 3533–3541 (1986).

62.   Lennard, L. The clinical pharmacology of 6-mercaptopurine. *Eur. J. Clin. Pharmacol.* **43**, 329–339 (1992).

63.   Hogarth, L. A. *et al.* The effect of thiopurine drugs on DNA methylation in relation to TPMT expression. *Biochem. Pharmacol.* **76**, 1024–1035 (2008).

64.   Wang, H. & Wang, Y. 6-Thioguanine Perturbs Cytosine Methylation at the CpG Dinucleotide Site by DNA Methyltransferases in Vitro and Acts as a DNA Demethylating Agent in Vivo †. *Biochemistry* **48**, 2290–2299 (2009).

65.   Allison, A. C. & Eugui, E. M. Mycophenolate mofetil and its mechanisms of action. *Immunopharmacology* **47**, 85–118 (2000).

66.   Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016).

67.   Lee, P. P. *et al.* The many faces of Artemis-deficient combined immunodeficiency — Two patients with DCLRE1C mutations and a systematic literature review of genotype–phenotype correlation. *Clin. Immunol.* **149**, 464–474 (2013).

68.   IJspeert, H. *et al.* Evaluation of the antigen-experienced B-cell receptor repertoire in healthy children and adults. *Front. Immunol.* **7**, 410 (2016).

69.   Halverson, R., Torres, R. M. & Pelanda, R. Receptor editing is the main mechanism of B cell tolerance toward membrane antigens. *Nat. Immunol.* **5**, 645–650 (2004).

70.   Ghia, P. *et al.* Ordering of human bone marrow B lymphocyte precursors by single-cell polymerase chain reaction analyses of the rearrangement status of the immunoglobulin H and L chain gene loci. *J. Exp. Med.* **184**, 2217–2229 (1996).

71.   Lang, J. *et al.* Receptor editing and genetic variability in human autoreactive B cells. *J. Exp. Med.* **213**, 93–108 (2016).

72.   Detanico, T. *et al.* Somatic mutagenesis in autoimmunity. *Autoimmunity* **46**, 102–114 (2013).

73.   Felgentreff, K. *et al.* Ligase-4 Deficiency Causes Distinctive Immune Abnormalities in Asymptomatic Individuals. *J. Clin. Immunol.* **36**, 341–353 (2016).

74.   Vera, G. *et al.* Cernunnos Deficiency Reduces Thymocyte Life Span and Alters the T Cell Repertoire in Mice and Humans. *Mol. Cell. Biol.* **33**, 701–711 (2013).

75.   Petersson, K., Pettersson, H., Skartved, N. J., Walse, B. & Forsberg, G. Staphylococcal Enterotoxin H Induces Vα-Specific Expansion of T Cells. *J. Immunol.* **170**, 4148–4154 (2003).

76.   Weng, Z., Pierce, B. & Baker, B. ATLAS Altered TCR Ligand Affinities and Structures database. (2020).

77.   Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).

78.   Simsek, D. & Jasin, M. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4–ligase IV during chromosomal translocation formation. *Nat. Struct. Mol. Biol.* **17**, 410–416 (2010).

79.   Goodnow, C. C. Multistep Pathogenesis of Autoimmune Disease. *Cell* **130**, 25–35 (2007).

80.   Bluestone, J. A. *et al.* The Immune Tolerance Network at 10 years: tolerance research at the bedside. *Nat. Rev. Immunol.* **10**, 797–803 (2010).

81.   Sakaguchi, S., Yamaguchi, T., Nomura, T. & Ono, M. Regulatory T Cells and Immune Tolerance. *Cell* **133**, 775–787 (2008).

82.   Chen, L. & Flies, D. B. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat. Rev. Immunol.* **13**, 227–242 (2013).

83.   Rieux-Laucat, F. What's up in the ALPS. *Curr. Opin. Immunol.* **49**, 79–86 (2017).

84.   Maccari, M. E. *et al.* A distinct CD38+CD45RA+ population of CD4+, CD8+, and double-negative T cells is controlled by FAS. *J. Exp. Med.* **218**, (2021).

85.   Grimbacher, B., Warnatz, K., Yong, P. F. K., Korganow, A. S. & Peter, H. H. The crossroads of autoimmunity and immunodeficiency: Lessons from polygenic traits and monogenic defects. *J. Allergy Clin. Immunol.* **137**, 3–17 (2016).

86.   Pascual, V. *et al.* Nucleotide sequence analysis of the V regions of two IgM cold agglutinins. Evidence that the VH4-21 gene segment is

responsible for the major cross-reactive idiotype. *J. Immunol.* **146**, 4385–91 (1991).

87. Grillot-Courvalin, C. *et al.* An anti-B cell autoantibody from Wiskott-Aldrich syndrome which recognizes i blood group specificity on normal human B cells. *Eur. J. Immunol.* **22**, 1781–1788 (1992).

88. Schickel, J. N. *et al.* Self-reactive VH4-34-expressing IgG B cells recognizecommensal bacteria. *J. Exp. Med.* **214**, 1991–2003 (2017).

89. Delmonte, O., Villa, A. & Notarangelo, L. Immune dysregulation in patients with RAG deficiency and other forms of combined immune deficiency immune deficiency. *Blood* (2020).

90. Anolik, J. & Sanz, I. B cells in human and murine systemic lupus erythematosus. *Curr. Opin. Rheumatol.* **16**, 505–512 (2004).

91. Hirabayashi, Y. Cell Death and Anti-DNA Antibodies. in *Apoptosis and Medicine* (InTech, 2012). doi:10.5772/48343

92. Sharabi, A. *et al.* Regulatory T cells in the treatment of disease. *Nat. Rev. Drug Discov.* **17**, 823–844 (2018).

93. Tesch, V. K. *et al.* Long-term outcome of LRBA deficiency in 76 patients after various treatment modalities as evaluated by the immune deficiency and dysregulation activity (IDDA) score. *J. Allergy Clin. Immunol.* **145**, 1452–1463 (2020).

94. Gardner, D. H. *et al.* 1,25(OH)2D3 Promotes the Efficacy of CD28 Costimulation Blockade by Abatacept. *J. Immunol.* **195**, 2657–65 (2015).

95. Staines Boone, A. *et al.* Failing to Make Ends Meet: The Broad Clinical Spectrum of DNA Ligase IV Deficiency. Case Series and Review of the Literature. *Front. Pediatr.* **6**, (2019).

96. Morris, E. C. Allogeneic hematopoietic stem cell transplantation in adults with primary immunodeficiency. *Hematology* **2020**, 649–660 (2020).

97. Schober, S. *et al.* Allogeneic hematopoietic stem cell transplantation in two brothers with DNA ligase IV deficiency: A case report and review of the literature. *BMC Pediatr.* **19**, 1–10 (2019).

98. Navarini, A. A. *et al.* Vedolizumab as a successful treatment of CTLA-4–associated autoimmune enterocolitis. *J. Allergy Clin. Immunol.* **139**, 1043-1046.e5 (2017).

99. Burgener, A. V. *et al.* SDHA gain-of-function engages inflammatory mitochondrial retrograde signaling via KEAP1–Nrf2. *Nat. Immunol.* **20**, 1311–1321 (2019).

100. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

101. Ghraichy, M. *et al.* Maturation of the Human Immunoglobulin Heavy Chain Repertoire With Age. *Front. Immunol.* **11**, 1–13 (2020).

102. Vander Heiden, J. A. *et al.* pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–1932 (2014).

103. Gupta, N. T. *et al.* Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data: Table 1. *Bioinformatics* **31**, 3356–3358 (2015).

104. Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–W40 (2013).

105. Lunter, G. & Goodson, M. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).

106. Yaari, G., Uduman, M. & Kleinstein, S. H. Quantifying selection in high-throughput Immunoglobulin sequencing data sets. *Nucleic Acids Res.* **40**, e134–e134 (2012).

107. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

108. Bienert, S. *et al.* The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* **45**, D313–D319 (2017).

109. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

110. Pettersen, E. F. *et al.* UCSF Chimera. A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

111. Bignucolo, O. & Bernèche, S. The Voltage-Dependent Deactivation of the KvAP Channel Involves the Breakage of Its S4 Helix. *Front. Mol. Biosci.* **7**, (2020).

112. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* NA-NA (2009). doi:10.1002/jcc.21367

113. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **52**, 1757–1768 (2012).

114. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

115. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).

116. MacKerell, A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

117. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

118. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

119. Jo, S., Lim, J. B., Klauda, J. B. & Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **97**, 50–58 (2009).

120. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, 45e – 45 (2001).

121. Bigler, M. B. *et al.* Stress-Induced In Vivo Recruitment of Human Cytotoxic Natural Killer Cells Favors Subsets with Distinct Receptor Profiles and Associates with Increased Epinephrine Levels. *PLoS One* **10**, e0145635 (2015).

122. Conlin, M. P. *et al.* DNA Ligase IV Guides End-Processing Choice during Nonhomologous End Joining. *Cell Rep.* **20**, 2810–2819 (2017).

123. Rao, V. K. & Oliveira, J. B. How I treat autoimmune lymphoproliferative syndrome. *Blood* **118**, 5741–5751 (2011).

124. Costa e Castro, A. *et al. Wide spectrum of manifestation of Ligase IV deficiency.* (2020).

125. Murray, J. E. *et al.* Extreme Growth Failure is a Common Presentation of Ligase IV Deficiency. *Hum. Mutat.* **35**, 76–85 (2014).

126. Ijspeert, H. *et al.* Clinical Spectrum of LIG4 Deficiency Is Broadened with Severe Dysmaturity, Primordial Dwarfism, and Neurological Abnormalities. *Hum. Mutat.* **34**, 1611–1614 (2013).

127. O'Driscoll, M. *et al.* DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. *Mol. Cell* **8**, 1175–1185 (2001).

128. Yue, J. *et al.* Identification of the DNA Repair Defects in a Case of Dubowitz Syndrome. *PLoS One* **8**, (2013).

129. Toita, N. *et al.* Epstein–Barr virus-associated B-cell lymphoma in a patient with DNA ligase IV (LIG4) syndrome. *Am. J. Med. Genet. Part A* **143A**, 742–745 (2007).

130. Riballo, E. *et al.* Cellular and Biochemical Impact of a Mutation in DNA Ligase IV Conferring Clinical Radiosensitivity. *J. Biol. Chem.* **276**, 31124–

31132 (2001).

131. Rucci, F. *et al.* Abnormalities of thymic stroma may contribute to immune dysregulation in murine models of leaky severe combined immunodeficiency. *Front. Immunol.* **2**, 1–13 (2011).

132. Dobbs, K. *et al.* Natural Killer Cells from Patients with Recombinase-Activating Gene and Non-Homologous End Joining Gene Defects Comprise a Higher Frequency of CD56bright NKG2A+++ Cells, and Yet Display Increased Degranulation and Higher Perforin Content. *Front. Immunol.* **8**, (2017).

133. Cifaldi, C. *et al.* Late-onset combined immune deficiency due to LIGIV mutations in a 12-year-old patient. *Pediatr. Allergy Immunol.* **28**, 203–206 (2017).

134. Plowman, P. N., Bridges, B. A., Arlett, C. F., Hinney, A. & Kingston, J. E. An instance of clinical radiation morbidity and cellular radiosensitivity, not associated with ataxia-telangiectasia. *Br. J. Radiol.* **63**, 624–628 (1990).

135. Riballo, E. *et al.* A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to γ-H2AX foci. *Mol. Cell* **16**, 715–724 (2004).

136. Slatter, M. A. & Gennery, A. R. Update on DNA-Double Strand Break Repair Defects in Combined Primary Immunodeficiency. *Curr. Allergy Asthma Rep.* **20**, (2020).

137. Buck, D. *et al.* Severe combined immunodeficiency and microcephaly in siblings with hypomorphic mutations in DNA ligase IV. *Eur. J. Immunol.* **36**, 224–235 (2006).

138. Enders, A. *et al.* A Severe Form of Human Combined Immunodeficiency Due to Mutations in DNA Ligase IV. *J. Immunol.* **176**, 5060–5068 (2006).

139. Grunebaum, E., Bates, A. & Roifman, C. M. Omenn syndrome is associated with mutations in DNA ligase IV. *J. Allergy Clin. Immunol.* **122**, 1219–1220 (2008).

140. Dard, R. *et al.* DNA ligase IV deficiency: Immunoglobulin class deficiency depends on the genotype. *Pediatr. Allergy Immunol.* **28**, 298–303 (2017).

141. Brunet, B. A. & Dave, N. Unique heterozygous presentation in an infant with DNA ligase IV syndrome. *Ann. Allergy, Asthma Immunol.* **119**, 379–380 (2017).

142. Buchbinder, D. *et al.* Rubella Virus-Associated Cutaneous Granulomatous Disease: a Unique Complication in Immune-Deficient Patients, Not Limited to DNA Repair Disorders. *J. Clin. Immunol.* **39**, 81–89 (2019).

143. Liao, W., Ngan, B. Y., Merico, D., Dadi, H. & Roifman, C. M. A novel mutation in LIG4 in an infant presenting with severe combined immunodeficiency with thymic medullary dysplasia. *LymphoSign J.* lymphosign-2017-0001 (2017). doi:10.14785/lymphosign-2017-0001

144. Murray, J. E. *et al.* Extreme Growth Failure is a Common Presentation of Ligase IV Deficiency. *Hum. Mutat.* **35**, 76–85 (2014).

145. Van Der Burg, M. *et al.* A new type of radiosensitive T-B-NK+ severe combined immunodeficiency caused by a LIG4 mutation. *J. Clin. Invest.* **116**, 137–145 (2006).

146. Fadda, A. *et al.* Two hits in one: Whole genome sequencing unveils LIG4 syndrome and urofacial syndrome in a case report of a child with complex phenotype. *BMC Med. Genet.* **17**, 1–7 (2016).

147. Gruhn, B. *et al.* Successful bone marrow transplantation in a patient with DNA ligase IV deficiency and bone marrow failure. *Orphanet J. Rare Dis.* **2**, 1–8 (2007).

148. Hayani, A., Suarez, C. R., Molnar, Z., LeBeau, M. & Godwin, J. Acute myeloid leukaemia in a patient with Seckel syndrome. *J. Med. Genet.* **31**, 148–149 (1994).

149. Ben-Omran, T. I., Cerosaletti, K., Concannon, P., Weitzman, S. & Nezarati, M. M. A patient with mutations in DNA Ligase IV: Clinical features and overlap with Nijmegen breakage syndrome. *Am. J. Med. Genet. Part A* **137A**, 283–287 (2005).

150. Taskiran, E. Z. *et al.* A Novel Missense LIG4 Mutation in a Patient With a Phenotype Mimicking Behçet's Disease. *J. Clin. Immunol.* **39**, 99–105 (2019).

151. Sharapova, S. O. *et al.* Next generation sequencing revealed DNA ligase IV deficiency in a 'developmentally normal' patient with massive brain Epstein-Barr virus-positive diffuse large B-cell lymphoma. *Clinical Immunology* **163**, 108–110 (2016).

152. Tosato, F. *et al.* Lymphocytes subsets reference values in childhood. *Cytom. Part A* **87**, 81–85 (2015).

The human adaptive immune system is an ensemble of heterogeneous cell populations, interconnected *via* dynamic pathways for host protection against foreign- and self-attacks. Genetic defects that lead to immune deficiency are summarized as inborn errors of immunity (IEI)[61]. Disentangling genetic from acquired immunity alterations have been challenging geneticists since 1931[62].

Traditionally, IEI were considered as rare diseases, found only in one in 10'000 – 50'000 births. The great leap of next-generation DNA sequencing corrected the estimation to be tenfold higher[61]. The recent pandemic might have taught us that IEI are even more abundant than expected. Severe COVID-19 occurred in 1% of previously healthy individuals[63]. Previous work demonstrated that infections with common viruses could unmask monogenetic susceptibilities [64], which was paralleled in severe Sars-CoV2 infections[65]. Infectious agents are recognized by antigen receptors on B and T lymphocytes. Three enzymes RAG1, RAG2 and LIG4 are crucial for B and T receptor generation[4]. For each of those enzymes immune deficiencies have been described with hypomorphic biallelic mutations[4,14]. Intriguingly, population analysis suggests that as many as 1 in 425 persons could be carrying a heterozygous *RAG* missense variant[66], the impact on immunity and immune-dysregulation are inadequately defined. We aimed at filling this gap and characterizing *RAG* and *LIG4* haploinsufficiency.

By using murine models with different *Rag1/Rag2* mutations we established a) models with titrated RAG activity and their impact on lymphocyte development, b) that double heterozygous *Rag* mutations were sufficient to reduce T cell repertoire clonality, c) that RAG activity reduction of up to 25% was associated with efficient anti-viral immunity, while reduction of RAG activity beyond 90% was associated with no measurable innate nor adaptive immune responses, d) that double heterozygous *Rag* mutations – reduction of RAG activity by 40% - lead to an impaired anti-viral immune response associated with augmented autoantibody production. We had the unique opportunity to investigate human patients, with novel dominantly inherited IEI, which were associated with heterozygous mutations in *RAG1/RAG2* respectively *LIG4.* Both genotypes presented with subtle changes in the TCR and BCR repertoires, whereas immune-dysregulation dominated the clinical manifestation. Incomplete penetrance is often observed in IEI's caused by haploinsufficiency[67]. The contribution of modifier genes, epigenetic modifications, environmental factors and somatic events are all putative contributors to the clinical phenotype[67].

We did not find evidence in the literature nor in GWAS metanalyses that relatives of biallelic mutated *RAG* or *LIG4* patients displayed more immune dysregulation than the general population. This observation could be attributed to the nature of the mutations *per se* or the cross-sectional design of many studies that did not allow for thorough investigation of the not obviously diseased family members. However, studies in hypomorphic *Lig4* murine models observed that in the heterozygous mutation carriers, the overall survival was reduced, lymphocyte development was altered, and that radiosensitivity and sarcoma formation was augmented[68–70]. In keeping, heterozygous carriers of *ATM* mutations, have been described to display abnormal apoptosis and cancer susceptibility[71,72]. Haploinsufficiency with incomplete penetrance was seen previously in IEI disorders associated with immune dysregulation such as CTLA-4 haploinsufficiency[73], autoimmune lymphoproliferative syndrome (ALPS, caused principally by *FAS* haploinsufficiency)[74], *IKZF1*[75,76], *SOCS1*[77] or *FOXN1*[78] haploinsufficiency. Understanding the core mechanisms underlying immune dysregulation in the *RAG* or *LIG4* mutation carriers might harbor the potential for later extrapolation to more common immune dysregulations like allergy and autoimmunity in general.

So far, allogeneic bone marrow transplantation is the only curative treatment option in *RAG* and *LIG4* deficiencies. The condition regimens are complex as reduces doses of cytoreductive agents are better tolerated but present the risk of mixed chimerism with re-expansion of residual host's T cells[79]. Thus, the exciting field of targeted precision medicine in IEI opens opportunities for therapy alternatives. Different avenues are exploited for individual IEIs, such as CTLA-4-Fc fusion proteins in CTLA-4 deficiency[80], selective PI3Kδ inhibitors in

activated PI3Kδ syndrome (APDS)[81] or gene editing platforms with engineered nucleases or base-editing enzymes combined with adenovirus-mediated guide RNA delivery in monogenetic blood disorders[82]. Some of these compounds may be used in the patients described here as they support immune-regulation.

A promising option is the *ex vivo* genetic modification of autologous CD34$^+$ hematopoietic stem cell precursor, approved recently for adenosine deaminase (ADA)-SCID[83]. Regarding the above described *RAG1/RAG2* and *LIG4* deficiency, future experiments need to clearly answer whether the heterozygous missense mutation(s) lead to haploinsufficiency *vs.* a negative dominance effect. The distinction is essential as gene therapies delivering a wild-type gene copy would not be curative in the context of a dominant-negative mutation. Independently, the constitutive expression of a wild-type *LIG4* transgene could have deleterious consequences, as uncontrolled ligation of DNA breaks incorporates the risk of malignancy development, which is *per se* elevated in LIG4 deficiency (suppl. table S3 in LIG4 manuscript). Pre-clinical data regarding therapy options in murine *Rag* deficient mice confirmed the complexity of *Rag* gene dosage. By delivering wild-type copies of the human *RAG1* gene with self-inactivating lentiviral vectors into *Rag* null mice, the authors achieved in half of the cohort little correction of T cell immunity with low-grade B cell antibody responses[84,85], while the other half developed a life-threatening Omenn syndrome-like phenotype[84]. This data is in line with our findings that low levels of RAG activity are insufficient to allow for an adaptive immune response, whereas intermediate levels of RAG activity are associated with the risk of immune dysregulation. Additionally, it has to be kept in mind that uncontrolled RAG activity is associated with genetic translocations[23].

We are excited to observe the development of different gene-editing approaches and their potential for personalized IEI therapy, such as base editing[86,87], insertion of an entire gene coding sequence at the endogenous locus *via* nuclease-directed DNA breaks and homologous recombination directed repair[88]. One of the latest DSB-free approaches is termed prime-editing and is based on a "search-and-replace" strategy[89]. The latter might be a promising choice in a host with defective DNA repair machinery.

Our experiments underpin the still underestimated clinical importance of heterozygous mutations in *RAG* and *LIG4* deficiency. A correct molecular diagnosis is essential to guide a personalized treatment.

References

1. Goodnow, C. C., Sprent, J., de St Groth, B. F. & Vinuesa, C. G. Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature* **435**, 590–597 (2005).
2. Notarangelo, L. D., Kim, M.-S., Walter, J. E. & Lee, Y. N. Human RAG mutations: biochemistry and clinical implications. *Nat. Rev. Immunol.* **16**, 234–46 (2016).
3. Backhaus, O. Generation of Antibody Diversity. in *Antibody Engineering* (InTech, 2018). doi:10.5772/intechopen.72818
4. Notarangelo, L. D., Kim, M.-S., Walter, J. E. & Lee, Y. N. Human RAG mutations: biochemistry and clinical implications. *Nat. Rev. Immunol.* **16**, 234–46 (2016).
5. Notarangelo, L. D., Bacchetta, R., Casanova, J. L. & Su, H. C. Human inborn errors of immunity: An expanding universe. *Sci. Immunol.* **5**, eabb1662 (2020).
6. Bruton, O. Agammaglobulinemia. *Pediatrics* **9**, 722–8 (1952).
7. Fischer, A. *et al.* Autoimmune and inflammatory manifestations occur frequently in patients with primary immunodeficiencies. *J. Allergy Clin. Immunol.* **140**, 1388-1393.e8 (2017).
8. Thaventhiran, J. E. D. *et al.* Whole-genome sequencing of a sporadic primary immunodeficiency cohort. *Nature* **583**, 90–95 (2020).
9. Zhang, S.-Y. *et al.* Human inborn errors of immunity to infection affecting cells other than leukocytes: from the immune system to the whole organism. *Curr. Opin. Immunol.* **59**, 88–100 (2019).
10. Burgener, A. V. *et al.* SDHA gain-of-function engages inflammatory mitochondrial retrograde signaling via KEAP1–Nrf2. *Nat. Immunol.* **20**, 1311–1321 (2019).
11. Tangye, S. G. *et al.* Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. *J. Clin. Immunol.* **40**, 24–64 (2020).
12. Casanova, J.-L. & Abel, L. Lethal Infectious Diseases as Inborn Errors of Immunity: Toward a Synthesis of the Germ and Genetic Theories. *Annu. Rev. Pathol. Mech. Dis.* **16**, annurev-pathol-031920-101429 (2021).
13. De Villartay, J. P., Fischer, A. & Durandy, A. The mechanisms of immune diversification and their disorders. *Nat. Rev. Immunol.* **3**, 962–972 (2003).
14. Altmann, T. & Gennery, A. R. DNA ligase IV syndrome; a review. *Orphanet J. Rare Dis.* **11**, 1–7 (2016).
15. Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426 (2018).
16. Nagasawa, T. Microenvironmental niches in the bone marrow required for B-cell development. *Nat. Rev. Immunol.* **6**, 107–116 (2006).
17. Kumar, B. V., Connors, T. J. & Farber, D. L. Human T Cell Development, Localization, and Function throughout Life. *Immunity* **48**, 202–213 (2018).
18. Nemazee, D. Receptor editing in lymphocyte development and central tolerance. *Nat. Rev. Immunol.* **6**, 728–740 (2006).
19. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575–581 (1983).
20. Kuo, T. C. & Schlissel, M. S. Mechanisms controlling expression of the RAG locus during lymphocyte development. *Curr. Opin. Immunol.* **21**, 173–178 (2009).
21. Wilson, A., Held, W. & MacDonald, H. R. Two waves of recombinase gene expression in developing thymocytes. *J. Exp. Med.* **179**, 1355–1360 (1994).
22. Grawunder, U. *et al.* Down-regulation of RAG1 and RAG2 gene expression in PreB cells after functional immunoglobulin heavy chain rearrangement. *Immunity* **3**, 601–608 (1995).
23. Teng, G. *et al.* RAG Represents a Widespread Threat to the Lymphocyte Genome. *Cell* **162**, 751–765 (2015).
24. Notarangelo, L. D., Kim, M. S., Walter, J. E. & Lee, Y. N. Human RAG mutations: Biochemistry and clinical implications. *Nat. Rev. Immunol.* **16**, 234–246 (2016).
25. Ji, Y. *et al.* The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* **141**, 419–431 (2010).
26. Moshous, D. *et al.* Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* **105**, 177–186 (2001).
27. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).
28. Schlissel, M. S. Regulating antigen-receptor gene assembly. *Nat. Rev. Immunol.* **3**, 890–899 (2003).
29. Dudley, E. C., Petrie, H. T., Shah, L. M., Owen, M. J. & Hayday, A. C. T cell receptor β chain gene rearrangement and selection during thymocyte development in adult mice. *Immunity* **1**, 83–93 (1994).
30. Krangel, M. S. Mechanics of T cell receptor gene rearrangement. *Curr. Opin. Immunol.* **21**, 133–139 (2009).
31. Palmer, E. & Naeher, D. Affinity threshold for thymic selection through a T-cell receptor–co-receptor zipper. *Nat. Rev. Immunol.* **9**, 207–213 (2009).
32. Singer, A., Adoro, S. & Park, J. H. Lineage fate and intense debate: Myths, models and mechanisms of CD4- versus CD8-lineage choice. *Nat. Rev. Immunol.* **8**, 788–801 (2008).
33. Geier, J. K. & Schlissel, M. S. Pre-BCR signals and the control of Ig gene rearrangements. *Semin. Immunol.* **18**, 31–39 (2006).
34. Bednarski, J. J. & Sleckman, B. P. At the intersection of DNA damage and immune responses. *Nat. Rev. Immunol.* **19**, 231–242 (2019).
35. Kinoshita, K. & Honjo, T. Unique and unprecedented recombination mechanisms in class switching. *Curr. Opin. Immunol.* **12**, 195–198 (2000).
36. Roco, J. A. *et al.* Class-Switch Recombination Occurs Infrequently in Germinal Centers. *Immunity* **51**, 337-350.e7 (2019).
37. Stratigopoulou, M., van Dam, T. P. & Guikema, J. E. J. Base Excision Repair in the Immune System: Small DNA Lesions With Big

Consequences. *Front. Immunol.* **11**, (2020).

38. Bétermier, M., Borde, V. & de Villartay, J. P. Coupling DNA Damage and Repair: an Essential Safeguard during Programmed DNA Double-Strand Breaks? *Trends Cell Biol.* **30**, 87–96 (2020).

39. O'Driscoll, M. & Jeggo, P. A. The role of double-strand break repair - Insights from human genetics. *Nat. Rev. Genet.* **7**, 45–54 (2006).

40. Lieber, M. R. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annu. Rev. Biochem.* **79**, 181–211 (2010).

41. Schatz, D. G. & Swanson, P. C. V(D)J Recombination: Mechanisms of Initiation. *Annu. Rev. Genet.* **45**, 167–202 (2011).

42. Celeste, A. *et al.* H2AX haploinsufficiency modifies genomic stability and tumor susceptibility. *Cell* **114**, 371–383 (2003).

43. Zimmermann, M. & De Lange, T. 53BP1: Pro choice in DNA repair. *Trends Cell Biol.* **24**, 108–117 (2014).

44. Reid, D. A. *et al.* Organization and dynamics of the nonhomologous end-joining machinery during DNA double-strand break repair. *Proc. Natl. Acad. Sci.* **112**, E2575–E2584 (2015).

45. Lescale, C. *et al.* RAG2 and XLF/Cernunnos interplay reveals a novel role for the RAG complex in DNA repair. *Nat. Commun.* **7**, 10529 (2016).

46. Zhao, B., Rothenberg, E., Ramsden, D. A. & Lieber, M. R. The molecular basis and disease relevance of non-homologous DNA end joining. *Nat. Rev. Mol. Cell Biol.* (2020). doi:10.1038/s41580-020-00297-8

47. Tadi, S. K. *et al.* PAXX Is an Accessory c-NHEJ Factor that Associates with Ku70 and Has Overlapping Functions with XLF. *Cell Rep.* **17**, 541–555 (2016).

48. Tumbale, P. P. *et al.* Two-tiered enforcement of high-fidelity DNA ligation. *Nat. Commun.* **10**, 1–13 (2019).

49. Waters, C. A. *et al.* The fidelity of the ligation step determines how ends are resolved during nonhomologous end joining. *Nat. Commun.* **5**, 4286 (2014).

50. Li, H., Vogel, H., Holcomb, V. B., Gu, Y. & Hasty, P. Deletion of Ku70, Ku80, or Both Causes Early Aging without Substantially Increased Cancer. *Mol. Cell. Biol.* **27**, 8205–8214 (2007).

51. Ege, M. *et al.* Omenn syndrome due to ARTEMIS mutations. *Blood* **105**, 4179–4186 (2005).

52. Espejel, S. *et al.* Shorter telomeres, accelerated ageing and increased lymphoma in DNA-PKcs-deficient mice. *EMBO Rep.* **5**, 503–509 (2004).

53. Gao, Y. *et al.* A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* **95**, 891–902 (1998).

54. Abramowski, V. *et al.* PAXX and Xlf interplay revealed by impaired CNS development and immunodeficiency of double KO mice. *Cell Death Differ.* **25**, 444–452 (2018).

55. Frank, K. M. *et al.* DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway. *Mol. Cell* **5**, 993–1002 (2000).

56. Nijnik, A. *et al.* Impaired lymphocyte development and antibody class switching and increased malignancy in a murine model of DNA ligase IV syndrome. *J. Clin. Invest.* **119**, 1696–1705 (2009).

57. Walter, J. E. *et al.* Expansion of immunoglobulin-secreting cells and defects in B cell tolerance in Rag-dependent immunodeficiency. *J. Exp. Med.* **207**, 1541–1554 (2010).

58. Rucci, F. *et al.* Homozygous DNA ligase IV R278H mutation in mice leads to leaky SCID and represents a model for human LIG4 syndrome. *Proc. Natl. Acad. Sci.* **107**, 3024–3029 (2010).

59. Barnes, D. E., Tomkinson, A. E., Lehmann, A. R., Webster, A. D. B. & Lindahl, T. Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. *Cell* **69**, 495–503 (1992).

60. Maffucci, P. *et al.* Biallelic mutations in DNA ligase 1 underlie a spectrum of immune deficiencies. *J. Clin. Invest.* **128**, 5505–5516 (2018).

61. Tangye, S. G. *et al.* Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. *J. Clin. Immunol.* **40**, 24–64 (2020).

62. Garrod, A. *The Inborn Factors in Disease*. (Clarendon Press, 1931).

63. Morens, D. M. & Fauci, A. S. Emerging Pandemic Diseases: How We Got to COVID-19. *Cell* **182**, 1077–1092 (2020).

64. Casanova, J.-L. & Abel, L. Lethal Infectious Diseases as Inborn Errors of Immunity: Toward a Synthesis of the Germ and Genetic Theories. *Annu. Rev. Pathol. Mech. Dis.* **16**, (2021).

65. Zhang, S. Y. *et al.* Severe COVID-19 in the young and healthy: monogenic inborn errors of immunity? *Nat. Rev. Immunol.* **20**, 455–456 (2020).

66. Kumánovics, A. *et al.* Estimated disease incidence of RAG1/2 mutations: A case report and querying the Exome Aggregation Consortium. *J. Allergy Clin. Immunol.* **139**, 690-692.e3 (2017).

67. Rieux-Laucat, F. & Casanova, J. L. Autoimmunity by haploinsufficiency. *Science (80-.).* **345**, 1560–1561 (2014).

68. Ruccia, F. *et al.* Homozygous DNA ligase IV R278H mutation in mice leads to leaky SCID and represents a model for human LIG4 syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 3024–3029 (2010).

69. Frank, K. M. *et al.* Late embryonic lethality and impaired V (D)J recombination in mice lacking DNA ligase IV. *Nature* **396**, 173–177 (1998).

70. Sharpless, N. E. *et al.* Impaired Nonhomologous End-Joining Provokes Soft Tissue Sarcomas Harboring Chromosomal Translocations, Amplifications, and Deletions. *Mol. Cell* **8**, 1187–1196 (2001).

71. Shigeta, T. *et al.* Defective control of apoptosis and mitotic spindle checkpoint in heterozygous carriers of ATM mutations. *Cancer Res.* **59**, 2602–7 (1999).

72. Bay, J.-O. *et al.* High incidence of cancer in a family segregating a mutation of the ATM gene: Possible role of ATM heterozygosity in cancer. *Hum. Mutat.* **14**, 485–492 (1999).

73. Schubert, D. *et al.* Autosomal dominant immune dysregulation syndrome in humans with CTLA4 mutations. *Nat. Med.* **20**,

1410–1416 (2014).

74. Oliveira, J. B. *et al.* Revised diagnostic criteria and classification for the autoimmune lymphoproliferative syndrome (ALPS): Report from the 2009 NIH International Workshop. *Blood* **116**, 35–41 (2010).

75. Bogaert, D. J. *et al.* A novel IKAROS haploinsufficiency kindred with unexpectedly late and variable B-cell maturation defects. *J. Allergy Clin. Immunol.* **141**, 432-435.e7 (2018).

76. Kuehn, H. S. *et al.* Loss of B Cells in Patients with Heterozygous Mutations in IKAROS. *N. Engl. J. Med.* **374**, 1032–1043 (2016).

77. Hadjadj, J. *et al.* Early-onset autoimmunity associated with SOCS1 haploinsufficiency. *Nat. Commun.* **11**, 5341 (2020).

78. Bosticardo, M. *et al.* Heterozygous FOXN1 Variants Cause Low TRECs and Severe T Cell Lymphopenia, Revealing a Crucial Role of FOXN1 in Supporting Early Thymopoiesis. *Am. J. Hum. Genet.* **105**, 549–561 (2019).

79. Morris, E. C. Allogeneic hematopoietic stem cell transplantation in adults with primary immunodeficiency. *Hematology* **2020**, 649–660 (2020).

80. Schwab, C. *et al.* Phenotype, penetrance, and treatment of 133 cytotoxic T-lymphocyte antigen 4–insufficient subjects. *J. Allergy Clin. Immunol.* **142**, 1932–1946 (2018).

81. Rao, V. K. *et al.* Effective "activated PI3Kδ syndrome"–targeted therapy with the PI3Kδ inhibitor leniolisib. *Blood* **130**, 2307–2316 (2017).

82. Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).

83. Aiuti, A., Roncarolo, M. G. & Naldini, L. Gene therapy for ADA-SCID, the first marketing approval of an ex vivo gene therapy in Europe: paving the road for the next generation of advanced therapy medicinal products. *EMBO Mol. Med.* **9**, 737–740 (2017).

84. van Til, N. P. *et al.* Recombination-activating gene 1 (Rag1)–deficient mice with severe combined immunodeficiency treated with lentiviral gene therapy demonstrate autoimmune Omenn-like syndrome. *J. Allergy Clin. Immunol.* **133**, 1116–1123 (2014).

85. Pike-Overzet, K. *et al.* Correction of murine Rag1 deficiency by self-inactivating lentiviral vector-mediated gene transfer. *Leukemia* **25**, 1471–1483 (2011).

86. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

87. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).

88. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–1389 (2013).

89. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).

Abbreviations

Abbreviation		Abbreviation	
°C	Degree Celsius	J gene	Joining gene
c1	Chi1 dihedral angle	kb	Kilo basepairs
AA	Amino acid	kJ	Kilojoule
a-EJ	Alternative end-joinig	Lag-3	Lymphocyte activation gene 3
ANOVA	Analysis of variance	LCMV	Lymphocytic choriomeningitis virus
APC	Antigen presenting cell	LIG1	DNA ligase 1
ATM	Ataxia Telengiectasia Mutated	LIG3	DNA ligase 3
ATP	Adenosine triphosphate	LIG4	DNA ligase 4
BCR	B cell receptor	Ly6C	Lymphocyte antigen 6 complex, locus C
BE	Binding energy	PAXX	Paralogue of XRCC4 and XLF
BHK-21	Baby Hamster Kidney cells	PBMCs	Peripheral blood mononucleated cells
BM	Bone marrow	PBS	Phosphate buffered saline
bp	Base pairs	PD-1	Programmed cell death-1
BSS	Balanced Salt Solution	pfu	plaque forming unit
CADD-PHRED	Combined Annotation Dependent Depletion	μg	Microgram
CD	Clusters of differentiation	MHC	Major histocompatibility complex
c.DNA	Complementary DNA	Min	Minute
CDR3	Complementary-determining region	μL	Microliter
CID	Combined immunodeficiency	mL	Millilitre
CID-G/AI	Combined immunodeficiency with granulomas and/or autoimmunity	MOI	Multiplicity of infection
CMV	Cytomegaly virus	mRNA	Messenger ribonucleic acid
CSR	Class switch recombination	mut	Mutation
CTL	Cytotoxic T cell	NHEJ	Non-homologous end joining
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats	NK	Natural killer cell
DDR	DNA damage response	ns	None significant
D gene	Diversity gene	nt	nucleotide
DMEM	Dulbecco's modified Eagle's medium	p53BP1	Phosphorylated 53-binding protein
DNA	Deoxyribonucleic acid	PCR	Polymerase chain reaction
DN	Double negative	SD	Standard deviation
DP	Double positive	RT-PCR	Reat time polymerase chain reaction
DSB	DNA double strand breaks	SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
DTT	Dithiothreitol	SEM	Standard error of the mean
EDTA	Ethylen-diamine-tetra-acetic acid	SHM	Somatic hypermutation
ESP	Exome Sequencing Project	SSC	Side scatter
FACS	Flow assisted cell sorting	SP	Single positive
FCS	Fetal calf serum	TBE	Tris-base, boric acid and EDTA
FDR	False discovery rate	TCR	T cell receptor
ffu	Focus forming unit	TCRAD	Human locus which encompasses the genes coding for the TCRδ and the TCRα chains
FoxP3	factor forkhead box protein 3	TCRB	Human locus which encompasses the genes coding for the TCRδ and the TCRα chains
FSC	Forward scatter	TCRG	Human locus which encompasses the genes coding for the TCRδ and the TCRβ chains
gDNA	Genomic DNA	T_{FH}	T follicular helper cells
GFP	Green fluorescent protein	Tg	Transgene
gnomAD	Genome Aggregation Database	T_H	CD4$^+$ T helper cell
GP	Glycoprotein	TRA	Human T cell receptor-α chain gene locus
γH2AX	Histone variant H2AX phosphorylated at serine 129	TRB	Human T cell receptor-β chain gene locus
HD	Healthy donor	Trb	Mouse T cell receptor-β chain gene locus
HDR	Homology directed repair	TNFα	Tumour necrosis factor alpha
HR	Homologous recombination	Treg	CD4$^+$ T regulatory T cell
HPRT	Hypoxanthine phosohorybosyl-transferase	RAG	Recombination activating gene
Ig	Immunoglobulin	RSS	Recombination signal sequence
IL	Interleukin	RT PCR	Reverse transcription polymerase chain reaction
IR	Ionizing radiation	V(D)J	Variable (Diversity) Joining
IFN-α	Interferon alpha	V gene	Variable gene
IFN-γ	Interferon gamma	WT	Wild type
i.v.	Intravenous	XLF	XRCC4-like factor (synonym Cernunnos, NHEJ-factor 1)
		XRCC4	X-ray repair cross-complementing protein 4

Contribution to the presented projects

I wrote both manuscripts presented and conceptualized main and supplementary figures as well as the tables.

Regarding the RAG manuscript: I designed, planned, performed and analyzed the big majority of experiments presented (exceptions are listed below). I established murine infection models in our laboratory and the corresponding read-outs. I planned/performed/analyzed the murine *in vivo* and *in vitro* experiments. As we could not observe alterations of the immune response of RAG1$^{mut/+}$ and RAG2$^{mut/mut}$ mice I decided to use different virus strains and to generate/investigate the double heterozygous *Rag* mutant mice, as well as following up on the reduction of virus-specific CD8$^+$ T cells and hepatic immunopathology. I established an analysis strategy for the *Trb* sequencing and autoantibodies microarray experiments. I conceived and wrote every mouse license, license renewals and end of the year reports. I genotyped all the mice and was in charge of the communication with the animal facility, veterinary service and I organized the import/rederivation of our strains.

Concerning the patients: I designed, performed and analyzed the immunophenotyping of PBMCS, establishment of a healthy control cohort, biobanking of PBMCs/protein lysates/DNA/RNA, genomic DNA isolation, Sanger sequencing and isolation of bone marrow cells. I did not perform the evaluations contributing to fig. 1a – d, 1f and 1n.

I obtained fellowships from the SNF, Novartis Foundation for Medical-Biological Research and Freiwillige Akademische Gesellschaft – which covered 34 months of my salary.

Regarding the LIG4 manuscript: I conceived, performed and analyzed the immunophenotyping of PBMCS and *in vitro* experiments, I established a healthy control cohort, biobanking of PBMCs/protein lysates/DNA/RNA, genomic DNA isolation, Sanger sequencing. I conceived and performed the *in vitro* ligation experiments and the analysis of the autoantibody microarrays. Intrigued by the punctual autoimmune manifestations, I went through the clinical documentations of the P1 and her family to establish a time line with the respective lab parameters. As I was curious draw parallels with other LIG4 deficient patients I generated the table with the published LIG4 deficient patients.

The molecular dynamics simulations incl. video was performed by Olivier Bignucolo. The BCR sequencing experiment and its analysis was performed by Marie Ghraichy/Valentin von Niederhäusern/ Johannes Trück. The human TCR experiment and its preliminary analysis was performed by Ottavia Delmonte, Luigi D. Notarangelo and Stephen Daley. The WES experiment and its analysis were performed Rebecca Higgins/Adhideb Gosh/Alexander Navarini. Stephan Ehl measured soluble FASL and vitamin B12, Helmut Hopfer organized and interpretated the biopsy specimens. Florian Marquardsen made T cell blasts and the apoptosis assay of the LIG4 family.

"Science is not only a disciple of reason but, also, one of romance and passion." Stephen Hawking

Personal information
Annaïse Julie Jauch
Nationality: Swiss
DOB: 1984/10/05
Linkedin:
https://www.linkedin.com/in/annaïse-
jauch-34b321b7/

Current address
in den Ziegelhöfen 87
Postfach 147
4054 Basel
Switzerland
+41 79 767 1188
annaise.jauch@unibas.ch

Curriculum Vitae

Profile summary

Detail-oriented with strong critical thinking skills and team-spirit
Results oriented, personal initiative, responsible, flexible, curious, self-management, enduring

Education & Career

March 2021 – March 2022	University Hospital Basel, FELIX PLATTER Spital Internal medicine assistant physician
Aug 2015 – March 2021	MD-Ph.D programme of the Swiss National Science Foundation DBM, University Hospital Basel, Prof. Mike Recher, Laboratory for Primary Immunodeficiencies. Thesis: "Dominantly inherited combined immunodeficiency as a consequence of heterozygous germline mutations in V(D)J recombination enzymes." - summa cum laude -
2013 – 2015	ISREC, EPFL Lausanne, Prof. Freddy Radtke Research intern Projects: "The role of TSLP in *in-vivo* melanoma/non-small cell lung cancer development and progression."
2007 – 2013	Swiss federal medical exams Master of Medicine, University of Basel, Switzerland. Grade point 6 Master theses: "High prevalence of infections and autoimmunity in patients with thymoma. The Role of PI3K/ Akt signalling pathways in EBV mediated B cell metabolic switch."
2004 - 2007	Bachelor of Science in Psychology, University of Basel, Switzerland. Grade point 5.5
2003- 2004	Volunteer English teacher in Thailand

Clinical and Research Experience

2008-2013	Project leader, Children and Youth Health Service Basel-Stadt, Unit for Health Promotion and Prevention
2013	Gynaecology / Obstetrics, University Hospital Stellenbosch, South Africa (1 month)
2012 -2013	Immunology, Prof. Ch. Hess laboratory, University Hospital of Basel (5 months)
2012	Neurology, Kantonsspital Aarau (1 month)
2012	Internal medicine, Ospidal Val Müstair (2 months)
2012	Dermatology, Hôpital Universitaire de Genève (1 month)
2012	Paediatrics, Hôpital Samaritain Vevey (1 month)
2012	Legal medicine, Kantonsspital St. Gallen (1 month)
2012	Surgery Department, Kantonsspital Lucerne (2 months)
2011	Pathology, Pathology Department University Hospital Basel (1 month)
2011	Ocular Genetics, McGill Montreal Children's Hospital, Canada (1 month)

2011	Immunology, Prof. Ch. Hess laboratory, University Hospital of Basel (1 months)
2009	Transplantation Immunology and Nephrology, Laboratory of Prof. Ed Palmer, University Hospital of Basel (2 months)
2008	Obstetrics and New-borns, Clinic Linde, Biel/Bienne (2 months)
2007	Research Assistant, Unit for Gynaecological Social and Psychosomatic Medicine, University Hospital Basel

Publications

Jauch A Kesten F, Daikeler T and Holbro A. A patient with fever of unknown origin: a common cause with an atypical presentation. Ther Umsch. 2012

Holbro A, **Jauch A**, Lardinois D, Tzankov A, Dirnhofer S and Hess C. High prevalence of infections and autoimmunity in patients with thymoma. Hum Immunol. 2012

Ghraichy M, Galson JD, Kovaltsuk A, von Niederhäusern V, Pachlopnik Schmid J, Recher M, **Jauch AJ,** Miho E, Kelly DF, Deane CM, Trück J. Maturation of the Human Immunoglobulin Heavy Chain Repertoire with Age. Frontiers in Immunology 2020.

Deigendesch N, Hirsiger JR, Bigler MB, Ghosh A, Harder D, **Jauch A**, Espinosa JM, Recher M, Navarini A, Daikeler T and Berger CT. Tumoural soft tissue calcification in Down Syndrome: association with heterozygous germline SAMD9 mutation and hyperactive type I interferon signaling. Rhumatology. 2020

Frick C, Dettinger P, Renkawitz J, J**auch A**, Berger CT, Recher M, Schroeder and Mehling M. Nano-scale microfluidics to study 3D chemotaxis at the single cell level. PLoSOne. 2018.

Bantug GR, Fischer M, Grähler J, Balmer ML, Unterstab G, Develioglu L, Steiner R, Zhang L, Costa ASH, Gubser P, Burgener AV, Sauder U, Löliger J, Belle R, Dimeloe S, Lötscher J, **Jauch A**, Recher M, Hönger G, Hall M, Romero P, Frezza C and Hess. Mitochondria- Endoplasmic Reticulum Contact Sites Function as Immunometabolic Hubs that Orchestrate the Rapid Recall Response of Memory CD8$^+$ T cells. Immunity 2018.

Marquardsen F, Baldin F, Wunderer F, Al-Herz W, Mikhael R, Baz Z, Rezaee F, Hanna R, Meyer B, **Jauch A**, Bigler M, Burgener AV, Higgins R, Navarini AA, Geha R, Notarangelo LD, Hess C, Berger CT, Bloch DB and Recher M. Detection of Sp110 deficiency by flow-cytometry and application to screening patients for veno-occlusive disease with immunodeficiency. JCI. 2017

Navarini AA, Hruz P, Berger CT, Hou TZ, Schwab C, Gabrysch A, Higgins R, Frede N, Padberg Sgier BC, Kämpe O, Burgener AV, Marqua F, Baldin F, Bigler M, Kistner A, **Jauch A,** Bignucolo O, Meyer B, Meienberg F, Mehling M, Jeker LT, Heijnen I, Daikeler TD, Gebbers JO, Grimbacher B, Sansom DM, Jeker R, Hess C, Recher M. Vedolizumab as a successful treatment of CTLA-4-associated autoimmune enterocolitis. J Allergy Clin Immunol. 2017

Balmer ML, Ma E, Bantug GR, Grählert J, Pfister S, Glatter T, **Jauch A,** Dimeloe S, Slack E, Dehio P, Krzyzaniak M, King C, Burgener AV Fischer M, Develioglu, Belle R, Recher M, Bonilla W, Macpherson AJ, Hapfelmeier S, Jones R and Hess C. Memory CD8$^+$ T Cells Require Increased Concentrations of Acetate Induced by Stress for Optimal Function. Immunity 2016

Gubser PM, Bantug GR, Razik L, Fischer M, Dimeloe S, Hoenger G, Durovic B, **Jauch A,** Hess C. Rapid effector function of memory CD8$^+$ T cells requires an immediate-early glycolytic switch. Nat Imm 2013

Kirchhofer, J, Steffen, T, Müller, B, and **Jauch, A**. Final report on the evaluation of the project Burzelbaum. Basel, Switzerland: Health promotion and prevention Basel-Stadt. 2007

References

| Prof.
Mike Recher | Laboratory and medical outpatient unit for Primary Immunodeficiencies University Hospital Basel.
mike.recher@usb.ch, +41 61 556 52 97 |
| Prof.
Christoph Hess | Immunobiology Laboratory and medical outpatient unit, University Hospital Basel & Department of Medicine, University of Cambridge, UK
chess@unibas.ch, +41 61 328 68 30 |

Language skills

German	Mother tongue
French	2nd mother tongue
English	Fluent - speaking, reading and writing
Italian	Basic proficiencies
Rumantsch	Basic proficiencies
Thai	Basic proficiencies

Fellowships and awards

2018	Antelope@Novartis mentee, University of Basel
2018	Novartis Medical-Biological Research grant, 60'000.-
2018	Fellowship Freiwillige Akademische Gesellschaft, 10'000.-
2017	Travel grant University Basel
2014 - 2016	MD PhD fellowship by the Swiss National Fond SNF and Krebsforschung Schweiz (Swiss Cancer research): 180'000.-