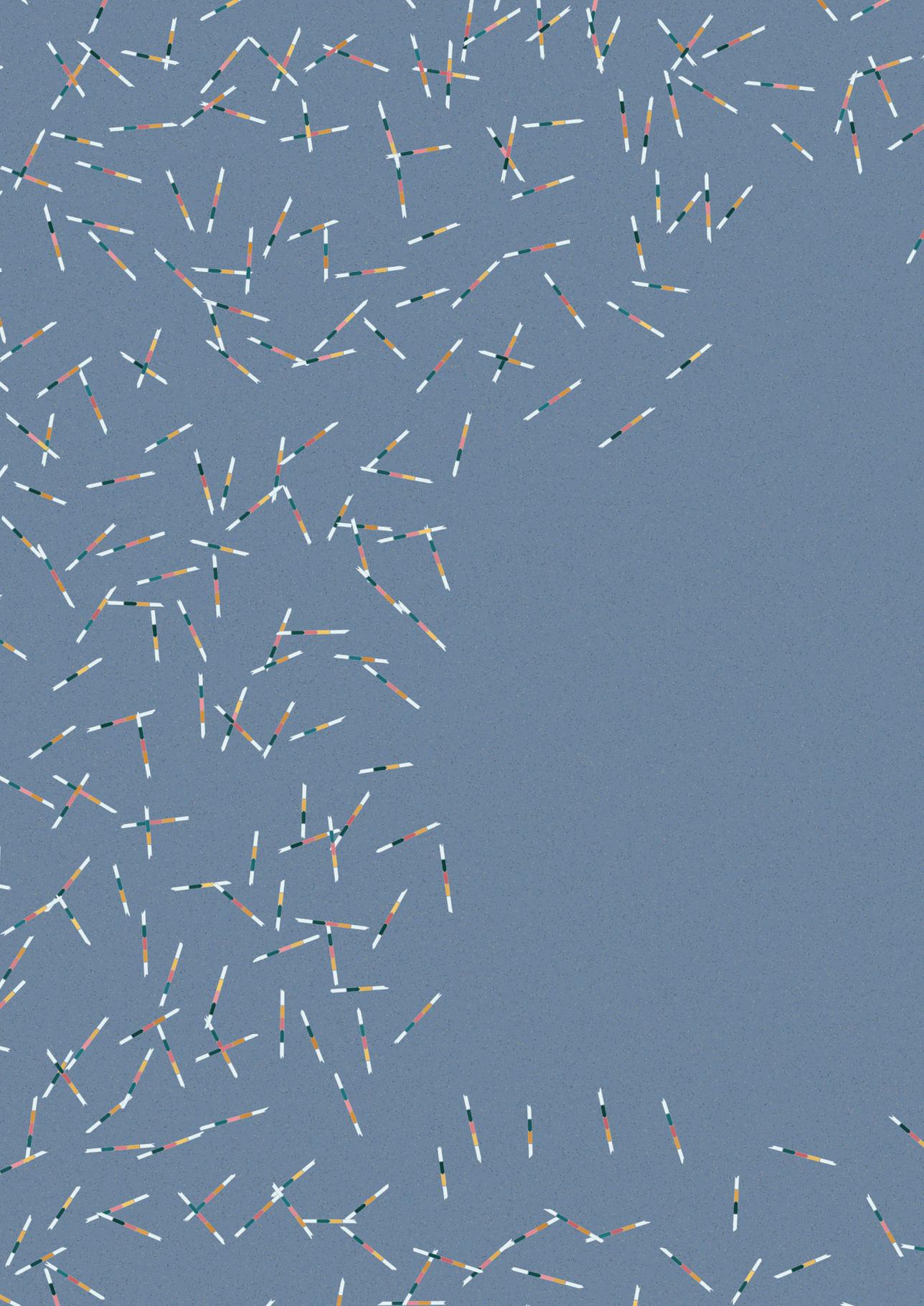


Lisa Ott de Bruin

Cut and paste

Using
CRISPR/Cas9
to model RAG1
deficiency





Cut and paste



Cut and paste: using CRISPR/Cas9 to model RAG1 deficiency

ISBN: 978-90-393-7057-5

Author: Lisa Ott de Bruin

Cover design: Jos de Bruin, artificialdesign.nl

Lay out: Jos de Bruin, artificialdesign.nl

Printed by: Ipskamp Printing

All rights reserved. No part of thesis may be reproduced or transmitted in any form by any means without permission in writing of the author. The copyright of published articles may have been transferred to the respective journals.

Copyright © 2018, Lisa Ott de Bruin, Rotterdam, The Netherlands.

CUT AND PASTE

using CRISPR/Cas9 to model RAG1 deficiency

KNIPPEN EN PLAKKEN:

*het gebruik van CRISPR/Cas9 om RAG1 deficiëntie te modelleren
(met een samenvatting in het Nederlands)*

Proefschrift

*ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties in het openbaar te
verdedigen op woensdag 14 november 2018 des middags te 12.45 uur*

door

Lisa Martina Ott de Bruin

geboren op 22 april 1987

te Beverly, MA, Verenigde Staten van Amerika

PROMOTOREN: Prof. N.M. Wulffraat
Prof. L.D. Notarangelo

COPROMOTOREN: Dr. M.L. Boes
Dr. J.P. Manis

Preface

This is a thesis about RAG, a crucial initiator of V(D)J recombination. V(D)J recombination is the DNA shuffling mechanism that makes it possible for our adaptive immune system to generate the enormous diversity of T- and B-cells needed to recognize the limitless variety of pathogens that attack our bodies. I remember being impressed by the ingenuity of it all when I took an introductory immunology course. I also found it the most difficult part of that course. Since then, my sense of awe for this ingenious system has only grown.

During medical school Prof. Wulffraat gave me the opportunity to explore the field of immunology further. He asked me to help with the literature search and organization of the EULAR task force, a group of pediatric immunologists and vaccine experts from Europe and Brazil. Their task was to critically appraise the current literature and to come up with evidence-based recommendations for vaccinating pediatric patients with immune diseases, in other words whether it was effective and safe to vaccinate these patients. This experience further contributed to my interest in the underlying mechanisms of the immune system.

Prof. Fasth, one of the EULAR task force experts, introduced me to Prof. Notarangelo, a world expert in the field of primary immunodeficiencies: a great researcher and an excellent clinician. He sees pediatric patients from all over the world to find out what immunodeficiency they suffer from and to develop novel cures for immunodeficiencies.

When I started in the Notarangelo lab as a medical student in 2011, I first learned to generate and to culture induced pluripotent stem cells (iPSCs), using Meganuclease and TALEN for genome-editing (**chapter 2**). These enzymes can be specifically designed to target the human *RAG1* locus to introduce a double strand DNA break that could lead to the correction of a mutation. Using this technology turned out to be quite challenging but I managed to show that our TALEN was able to target the *RAG1* locus in fibroblasts. A few months later, I had the opportunity to continue my work at Prof. Notarangelo's lab and do my PhD research there. At that point, a novel genome-editing tool had just been developed, CRISPR/Cas9, which quickly became very popular due to its efficiency and versatility. Using these new tools, I continued to work on genome-editing to model and correct primary immunodeficiencies, with a focus on RAG1 deficiency. As a result, my thesis is focused on CRISPR/Cas9 and RAG1. They actually have a lot in common: both are endonucleases, 'scissors' that can cut and paste DNA.

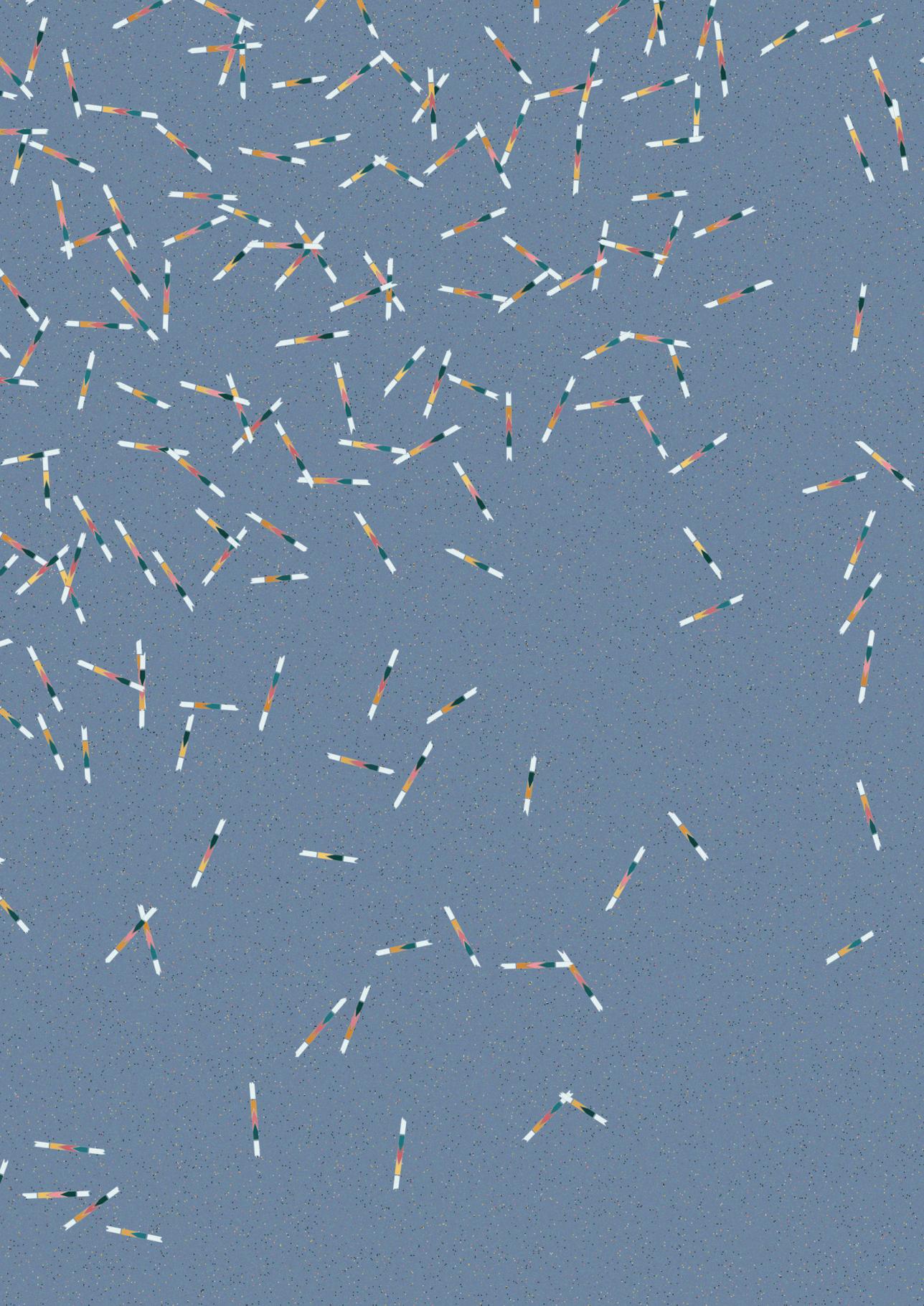


Table of contents

Chapter 1	General introduction to the immune system and RAG deficiency	1
Chapter 2	Novel Genome-Editing Tools to Model and Correct Primary Immunodeficiencies <i>Frontiers in Immunology</i> . 2015 May 21;6:250.	15
Chapter 3	Modeling altered T-cell development with human induced pluripotent stem cells from patients with <i>RAG1</i> mutations and distinct immunological phenotypes <i>Blood</i> . 2016 Aug 11;128(6):783-93.	37
Chapter 4	Rapid generation of novel models of RAG1 deficiency by CRISPR/Cas9-induced mutagenesis in murine zygotes <i>Oncotarget</i> . 2016 Mar 15;7(11):12962-74.	61
Chapter 5	Hypomorphic <i>Rag1</i> mutations affect composition of the pre-immune repertoire at early stages of lymphoid development <i>Blood</i> . 2018 Jul 19;132 (3):281-292	83
Chapter 6	General discussion	129
Chapter 7	Summary	139
	Nederlandse samenvatting	143
	Acknowledgments	149
	Curriculum Vitae	157
	Publications	161

1

2

3

4

5

6

7



1 General introduction to the immune system and RAG deficiency

Overview of the immune system

Our bodies are under permanent attack from pathogens that come in many different forms: viruses, bacteria, fungi, protozoan parasites. Still, our immune system is able to recognize and fight all these different pathogens, without attacking our own cells or tissues. To achieve this the immune system relies on two different systems: the innate immune system and the adaptive or 'acquired' immune system.

The innate immune system

A pathogen needs to cross the border between our bodies and its environment. The epithelial surface lining the gut, respiratory, urinary and reproductive tract thus forms a first line of defense. Pathogens that are able to cross this barrier will be confronted by the innate immune system. This system is able to recognize pathogens by the general molecular patterns associated with these pathogens. These pathogen-associated molecular patterns (PAMPs) can be recognized by a number of different components of the innate immune system, such as macrophages or neutrophils. When their receptors bind to a PAMP, this leads to a cascade of cytokines initiating the inflammatory response, which in turn activates the complement system, which contributes to the induction of phagocytosis (ingestion of pathogens), the recruitment of other inflammatory cells and, as discussed later, the triggering of adaptive immune responses.

A PAMP based line of defense is not sufficient to deal with all the different kinds of attack our bodies face. Viruses for instance, do not express these

molecular patterns. To deal with these, the immune system uses a different approach. Virus DNA or RNA entering a host cell can initiate a type I interferon response which blocks viral replication in cells or, if everything fails, can induce apoptosis of infected cells. Indirectly, type I interferons also activate Natural killer (NK) cells that can use their lower MHC-I expression to recognize and subsequently kill cancer cells or cells that have been infected by viruses. NK cells are derived from the same lymphoid progenitors as the adaptive system, but behave more like innate immune cells. Their receptors are not rearranged, they are mainly present in mucus and they respond early to infection.

The innate system does not require training, it comes pre-loaded with a general ability to recognize and react to pathogens. This means it can react quickly, but by necessity also in a rather generic manner. To develop a more specific and efficient response to a specific attack, it calls upon the adaptive system. Dendritic cells play an important role in this process. They act as a link between the innate and the adaptive system. Like macrophages and neutrophils, they recognize microbial parts through their pattern recognition receptors and can phagocytose the pathogen. Dendritic cells then cleave the pathogen into peptides and present these on MHC molecules. The activated dendritic cell then migrates to a nearby lymphoid organ such as a lymph node and presents the peptide to a T-cell to activate the adaptive immune system.

The adaptive immune system

The adaptive immune system relies on B- and T-cells to destroy pathogens. These lymphocytes develop out of the lymphoid progenitors. These lymphoid progenitors in turn grow out of hematopoietic stem cells. The progenitors either continue to develop in the bone marrow to become B-cells, or migrate to the thymus to become T-cells (**Fig. 1**). Central to the adaptive system is its ability to create an immense variety of unique T-cell receptors (TCR) and B-cell immunoglobulins. These immunoglobulins (Ig) can function as B-cell receptor (BCR) or as antibody when secreted into the plasma (IgM, IgG, IgA, IgE). In addition, certain antigen-stimulated lymphocytes will differentiate into memory cells. That way a more sensitive, rapid and effective response will be elicited the next time that antigen is encountered¹.

In the remainder of this thesis I will focus on the adaptive system and the role played by recombination-activating gene 1 (RAG1) in achieving its diversity.

B- and T-cell development

The great variety in lymphocytes used by the adaptive system relies on ingeniously rearranging the DNA of lymphoid progenitors. These lymphoid

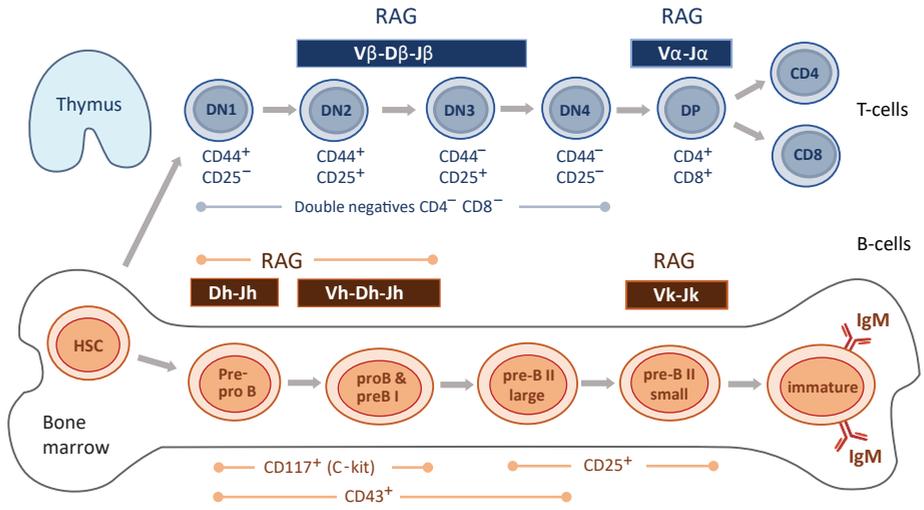


Figure 1: T- and B-cell development in thymus and bone marrow (cell surface markers for mice).

progenitors grow out of hematopoietic stem cells in the bone marrow. Common lymphoid progenitors then either continue to develop in the bone marrow to become B-cells, or migrate to the thymus to become T-cells (**Fig. 1**). Recombination-activating genes 1 (RAG1) and 2 (RAG2) are lymphoid specific proteins that are expressed during distinct stages of T- and B-cell development and are crucial for generating a diverse pool of mature T- and B-cells. RAG1 and RAG2 are key players in the rearrangement of the different chains that make up the TCR and B-cell immunoglobulins.

B-cell immunoglobulins consist of a heavy chain (Igh) and light chain (Igκ or Igλ), both with a variable antigen binding part. During the (pre-)pro-B and pre-B-I cell stage in the bone marrow, the heavy chain rearranges on one allele (**Fig. 1**). If this results in a productive allele, the cell continues to the pre-B-II cell stage where V to J light chain rearrangement takes place. Next, a productive heavy chain is paired with a productive light chain to form a functional BCR (IgM) and become an immature B-cell. At this stage the cell has to pass a final checkpoint for autoreactivity. If the BCR does not display high affinity for self-antigens (autoantigens) the cell will enter the circulation as a unique naïve B-cell. If the BCR does bind self-antigens with high affinity, the cell will undergo another round of receptor editing in a new attempt to get a non-autoreactive BCR. This mainly involves RAG-induced rearranging of the *light chain*. If this fails as well, the cell will undergo apoptosis.

T-cell development involves a similar process in the thymus. Some (5-15%)

T-cells consist of a γ (TR γ) and δ (TR δ) chain TCR, but the vast majority of TCRs consist of an α (TR α) and β chain (TR β) (85-95%)². For the purpose of this thesis, I will focus on $\alpha\beta$ T-cells. At the CD4⁻CD8⁻ double negative (DN) stage the *TRB* rearranges. During the CD4⁺CD8⁺ double positive (DP) stage the *TRA* is rearranged and a productive TR α is paired with a productive TR β to form a functional TCR. As for B-cells, a T-cell has to pass checkpoints for autoreactivity before it can exit the thymus as a naïve single positive T-cell (CD4⁺ or CD8⁺). Thymic epithelial cells express autoimmune regulator (AIRE) that enables expression of specific self-antigens (tissue-restricted antigens – TRAs), that normally are not expressed in thymic tissue. Depending on the strength of TCR binding to these self-antigens, a cell undergoes apoptosis or is positively selected to become either a regulatory T-cell (T_{Reg}) or an effector T-cell. The latter is either a cytotoxic T-cell (CD8⁺) or T helper cell (CD4⁺). T_{Reg} cells act as an important peripheral checkpoint for autoimmunity, as they suppress T effector cells.

The next section discusses how RAG is involved in the receptor rearrangement. Based on V(D)J recombination, this occurs at several distinct stages of T- and B-cell development.

V(D)J recombination

The variable antigen binding part of the TCR and the variable antigen binding part of the BCR are created by combinational joining of variable (*V*), diversity (*D*) and joining (*J*) genes (**Fig. 2**). During B- and T-cell development, precursor cells start off with a germline locus consisting of either many different *V*, *D* and *J* genes (for *TRB*, *TRD* and *heavy chain*) or just many different *V* and *J* genes (for *TRA*, *TRG* or *light chain*). V(D)J recombination is a process of rearrangement that relies on a cut and paste technique to select and join specific genes, one for each group. The constant domains of the TCR and BCR consist of constant genes, i.e. a sequence of genes that is not affected by rearrangement.

RAG1 and RAG2 play a key role in this V(D)J recombination. Two RAG1 and two RAG2 molecules form a heterotetramer that binds to the recombination signal sequences (RSS's) that flank the *V*, *D* and *J* coding elements (**Fig. 2**). RAG2 does not participate in direct binding to the RSS, but stabilizes the binding of the RAG complex to the RSS and interacts with the *V*, *D*, *J* coding elements³, as could recently be established using crystallographic and cryo-electron microscopy structures of the RAG1-RAG2 complex^{3,4}.

RSS's are composed of conserved heptamer (5'-CACAGTG-3') and nonamer (5'-ACAAAAACC-3') elements, which are separated by a degenerate spacer of either 12 or 23 nucleotides⁵. A 12-nucleotide RSS can only combine with

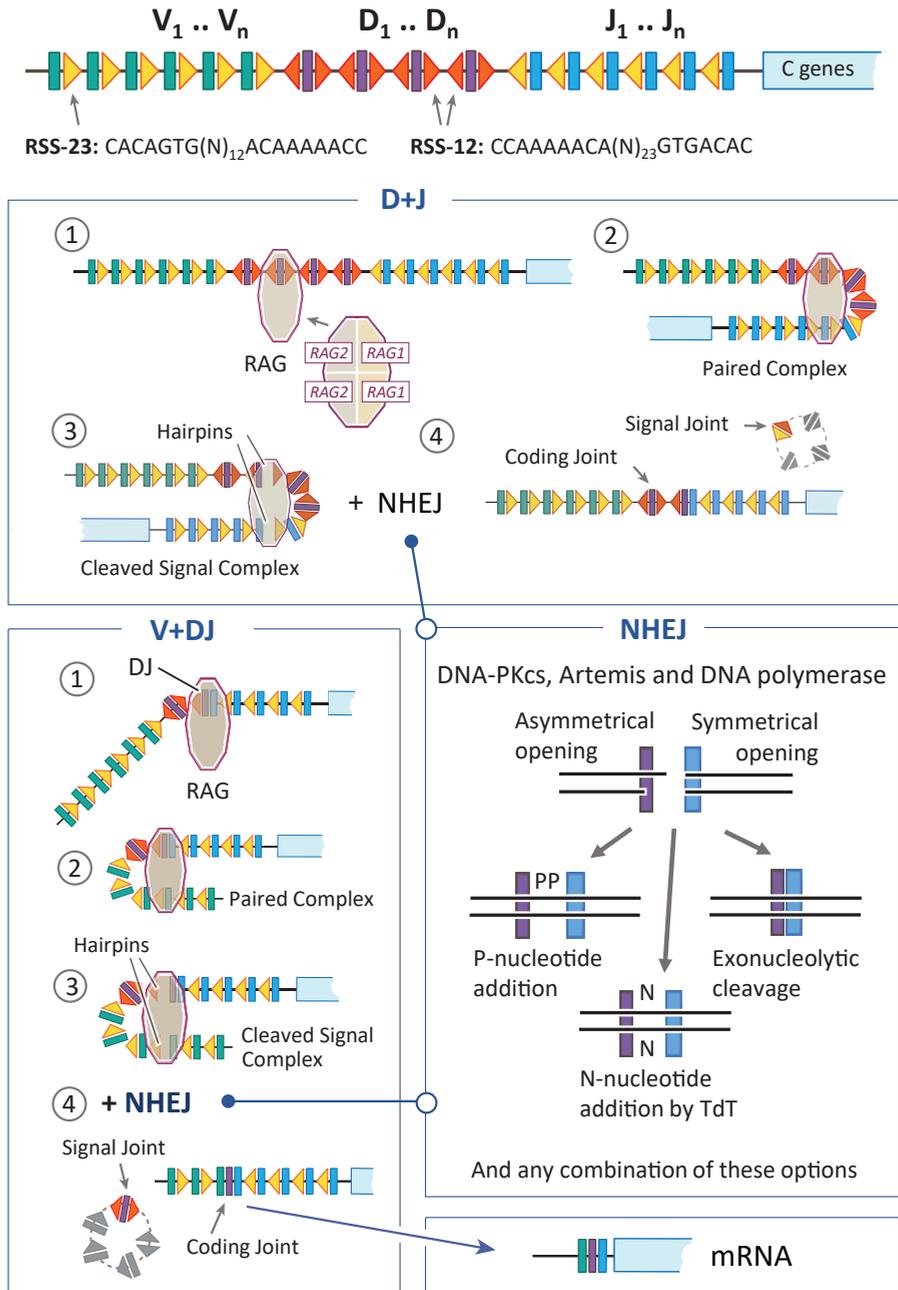


Figure 2: V(D)J recombination. Two RAG1 and two RAG2 molecules form a heterotetramer (shown as RAG) that binds to recombination signal sequences flanking the V, D and J genes. 12-RSS combines with 23-RSS to ensure sequential D to J (top) and V to DJ joining (bottom). Binding to a pair of RSSs (1) RAG1 introduces a nick on one strand of the DNA between RSS heptamer and flanking gene, generating a paired complex (2). Sealed hairpinned coding ends and blunted signal ends are formed, bound to RAG in a cleaved signal complex (3). Finally, the NHEJ pathway is initiated (4), as detailed in the NHEJ panel.

a 23-nucleotide RSS, the 12-23 rule. This ensures that for the *TRB*, *TRD* and *heavy* chain locus, a *V* can only rearrange to a *D* and not directly to a *J*, whereas for the *TRA*, *TRG* and *light chain*, a *V* rearranges directly to a *J*. Selection of the specific *V*, *D* and *J* genes is not stochastic but is based on the intrinsic quality of the RSS's and on accessibility and epigenetic modifications of the TCR and BCR.⁵⁻⁷

Upon binding to a pair of RSS's, RAG1 introduces a nick on one strand of the DNA between the RSS heptamer and the flanking coding element, generating a paired complex. The resulting hydroxyl group on the 3' end of the coding flank attacks the phosphodiester bond on the opposite DNA strand in a transesterification reaction, forming covalently closed hairpins at the coding ends and blunted signal ends. The RAG1-RAG2 heterotetramer remains bound to these in a cleaved signal complex. Subsequently, DNA dependent protein kinase catalytic subunit (DNAPKcs) activates Artemis, which opens the hairpins. Both signal ends and coding ends are then processed by the nonhomologous end joining (NHEJ) pathway to enable joining of broken ends.

The last step of non-homologous end joining introduces junctional diversity at the coding joint. Whereas signal ends containing the RSSs are precisely ligated, imprecise joining of coding ends may occur, resulting from the introduction or omission of nucleotides at the junction site. Asymmetrical opening of the hairpin coding ends allows incorporation of palindromic sequences (P-nucleotides) during the joining process. Terminal deoxynucleotidyl transferase (TdT) may introduce additional N-nucleotides in the junction. Finally, exonucleolytic cleavage may 'chew' nucleotides at the boundary between the two coding ends. This 'junctional diversity' contributes to the overall diversity of the TCR and BCR repertoires⁸. The signal joint is a circular by-product of this process, consisting of unused *V*, *D* and *J* gene segments. Measuring these circular by-products, called T-cell receptor excision circles (TRECs) or kappa-deleting recombination excision circles (KRECs) can be used to identify a defect in V(D)J recombination during newborn screening.

TCR and BCR diversity

The total diversity of the unique TCRs and BCRs that can be achieved is the sum of the combinatorial (different *V* (*D*) and *J* genes) and junctional diversity of each chain, and the combining of the different chains (TR α and TR β , TR γ and TR δ or heavy and light chains).¹⁵ Each antigen-receptor chain contributes three loops to the antigen-combining site, known as Complementarity-Determining Regions (CDR). CDR1 and CDR2 are encoded by *V* elements, whereas CDR3 is encoded by *V* genes, *D* genes and *J* genes and the junctions in between. Therefore, the CDR3 region contributes most to the diversity of the antigen-

binding part and is the region that is usually studied.

To study this region, one can use traditional methods such as CDR3 spectratyping or flow cytometric analysis of *V* genes. In CDR3 spectratyping, different CDR3s are distinguished based on length only, but length is influenced both by the usage of longer *V*, *D* or *J* genes and by nucleotide addition at the junction between the *V*, *D* and *J* genes, a result of junctional diversity. Flow cytometric analysis uses FACS antibodies that recognize certain *V* families, a tedious and time-consuming method that can only be used to study *V* families instead of individual *V* genes. We used next generation sequencing (NGS), also called high-throughput sequencing. NGS offers a much more in depth and detailed analysis of TCR and BCR repertoire. The specific NGS strategies used for this thesis are explained in the supplement (methods section, p. 121) of **chapter 5**.

Clinical spectrum of RAG deficiency

RAG deficiency is a primary immunodeficiency (PID). PIDs comprise a heterogeneous group of genetic diseases in which part of the immune system is missing or functions improperly. As next generation sequencing techniques such as whole exome and whole genome sequencing become more widely available, the number of genetic defects found in patients with primary immunodeficiencies is rapidly increasing. The latest number of known PID genes is close to 300¹⁵ (to compare: at the time of writing the review given in chapter 2, around 230 PID causing genes had been described¹⁶). These sequencing techniques not only have helped the discovery of novel genes, but also facilitated detecting novel mutations in known PID genes, characterized by phenotypes not previously known to be associated with those genes. Patients with mutations in *RAG1* or *RAG2* can have many different distinct clinical phenotypes. This spectrum is much broader than originally thought.⁸ Some of these mutations, so called null mutations, lead to a nonfunctional RAG and thus a complete developmental block in T-cells and B-cells. This results in a severe combined immunodeficiency (SCID) phenotype. Other mutations lead to some RAG activity and allow partial T- and B-cell development. These hypomorphic mutations can result in a wide range of phenotypes.⁸

Severe Combined Immunodeficiency (SCID)

SCID is characterized by a complete block in T-cell development. B-cells can be absent (T-B-SCID) or present (T-B+SCID), but are always impaired due to the lack of T-cells. Genes that may cause T-B+SCID when mutated are: *IL2RG*, *JAK3*, *IL7RA*, *PTPRC*, *CD3D*, *CD3E*, *CD3Z* and *CORO1A*.

Functional null mutations in genes involved in V(D)J recombination, such as *RAG1/RAG2* or genes encoding components of the NHEJ pathway can all cause T-B-SCID, characterized by a complete lack of functional T- and B-cells which leads to fatal early-onset infections¹⁸, unless immune reconstitution can be achieved, usually with hematopoietic stem cell transplantation (HSCT)¹⁹. NK cells are not dependent on V(D)J recombination for their development, therefore patients with null mutations in genes involved in V(D)J recombination have normal NK cell counts.

Mutations in genes involved in the NHEJ pathway (*DNA-PKcs*, *Artemis*, *Cernunnos/XLF*, *XRCC4* and *DNA ligase IV*), lead to DNA repair defects. This makes cells more susceptible to the damaging effects of ionizing radiation, resulting in genomic instability, tumor development or cell death. For this reason, SCID caused by defects in the NHEJ pathway is also called radiosensitive SCID. Since NHEJ is important for many cells in the body, patients with NHEJ defects will often display additional characteristics including microcephaly, dysmorphic facial features and growth delay²⁰. In contrast, NHEJ is not affected in RAG patients, therefore they do not show impaired DNA repair.

Besides *RAG1/RAG2* and NHEJ genes, other genes have also been shown to lead to T-B-SCID when mutated: e.g. *ADA* and *AK2*. Newborn screening results in the US have shown that *RAG1* and *RAG2* mutations account for approximately 17%-28,6% of all cases of SCID and SCID-related conditions (including Omenn Syndrome, atypical SCID and $\gamma\delta$ SCID described in the next section).^{9,21} In the Netherlands, a cohort of 43 SCID cases showed that 26% was caused by mutations in *RAG1* and *RAG2*, the majority of these being *RAG1* mutations.¹⁰

Omenn syndrome

Omenn syndrome was first described in 1965.²² It is characterized by very low residual T-cell development which results in early-onset generalized erythroderma, lymphadenopathy, hepatosplenomegaly, eosinophilia and severe hypogammaglobulinemia with increased IgE levels. Tissue-specific, (self) antigen driven oligoclonal and activated T-cells infiltrate multiple organs, while B-cells are typically absent.²³ Hypomorphic mutations, in particular in *RAG1*, that allow low recombination activity are the most common cause of Omenn syndrome.²⁴

Leaky SCID, Atypical SCID, $\gamma\delta$ SCID

When there is residual RAG activity, with some development of T-cells and without the clinical manifestations of Omenn syndrome, this is generally referred to as Leaky SCID or atypical SCID.²⁵ Expansion of specifically $\gamma\delta$

T-cells was described in patients with RAG deficiency and disseminated CMV infection.^{26,27}

Novel phenotypes of RAG deficiency

The phenotypes described above were the result of absent or very low residual RAG activity. Hypomorphic mutations with higher residual activity have been identified in patients with delayed-onset combined immunodeficiency associated with granulomas and/or autoimmunity (CID-G/AI). These patients present later in childhood or even in young adulthood. They have low to normal T- and B- cell numbers, low to normal total immunoglobulin levels, partially preserved immunity and autoimmunity or granulomatous disease.²⁸⁻³⁶ Specific autoimmune features that have been described in RAG patients are cytopenias, vitiligo, psoriasis, myasthenia gravis and Guillain-Barre. Other related phenotypes that have been described in hypomorphic RAG patients are CD4⁺ idiopathic T-cell lymphopenia³⁷, Common variable immunodeficiency³⁸, IgA deficiency³⁹, selective deficiency of polysaccharide-specific antibody responses⁴⁰, hyper-IgM syndrome⁴¹ and sterile chronic multifocal osteomyelitis⁴². In **chapter 5** we describe how we used CRISPR/Cas9 to generate 3 different mouse models with mutations corresponding to *RAG1* mutations found in patients with CID-G/AI.

Newborn screening

Newborn screening for severe combined immunodeficiency (SCID) using assays to detect T-cell receptor excision circles (TRECs) began in Wisconsin in 2008.⁹ Currently the quantification of TRECs has been included in newborn screening in all but three states in the US (Indiana, Alabama, Louisiana). In the Netherlands, a recent study showed that out of the 43 patients diagnosed with SCID, 9 died of infectious complications before treatment could be initiated, indicating that newborn screening would be crucial to catch these patients early, and treat them before they develop infections¹⁰. A pilot study suggested that indeed measuring TRECs is a suitable method for SCID newborn screening in the Netherlands¹¹. In addition, kappa-deleting recombination excision circles (KRECs) can be measured to show B-cell defects¹². Starting April 2018, newborns in the provinces of Gelderland, Utrecht and Zuid-Holland, will be screened for SCID using a combined TREC/KREC assay, as part of the ongoing SONNET-study. TREC/KREC assays have already been used in newborn screening pilot studies in Sweden and Spain^{13,14}. Several other European countries are currently developing pilot studies. Norway and Iceland have already fully implemented national newborn screening for SCID.

Models to study RAG1 deficiency

With the currently available sequencing tools, the number of genetic defects found in patients with primary immunodeficiencies is rapidly increasing. In order to confirm whether the reported mutation is indeed disease causing, to study the underlying mechanism and to test potential treatment options, there is a need for disease models with that specific mutation. Such models can be either *in vitro* human models or *in vivo* animal models. I used both approaches to study RAG1 deficiency. To create an *in vitro* model I used patient induced pluripotent stem cells (iPSCs). To study the *in vivo* effects of hypomorphic *RAG1* mutations I generated mouse models using CRISPR/Cas9 based genome-editing.

Induced Pluripotent stem cells (iPSCs)

To study the T-cell differentiation potential of RAG1 patients, we could take hematopoietic stem cells and differentiate these *in vitro*. However, patient hematopoietic stem cells are scarce and difficult to work with. In 2012 Yamanaka received the Nobel prize for his groundbreaking discovery that any adult cell can be turned into a stem cell by introducing four pluripotency genes, Oct3/4, Sox2, c-Myc and Klf4⁴³. Due to this discovery, it is possible to generate unlimited amounts of patient stem cells, derived from a piece of skin, such as the umbilical cord or foreskin. The promise of iPSCs is further discussed in **chapter 2**.

Chapter 3 describes how the differentiation potential of iPSCs derived from RAG patients was studied. While it is of clear advantage to be able to study human cells, there are only limited possibilities to study the phenotypic effects of these mutations in an *in vitro* setting.

Mouse models

Mouse models make it possible to study patient mutations *in vivo*. RAG1 deficiency is characterized by a very broad clinical and immunological phenotypic spectrum where both genetic and environmental factors play a role. To unravel the contribution of each of these factors *in vivo*, mouse models are indispensable. They make it possible to fully control environmental factors and study T- and B-cell development in detail. **Chapter 2** explains how CRISPR/Cas9 based genome-editing can be used to generate specific mouse models. **Chapters 4** and **5** describe how I generated and characterized several *Rag1* mouse models.

Thesis outline

The aims of this thesis were twofold: to better understand RAG1 deficiency using genome-editing and to help prepare the ground for using these techniques to correct mutations in *RAG1*. In **chapter 2** we review current treatment strategies for primary immunodeficiencies (PID) and describe how novel genome-editing tools can be used to model and correct PID. This chapter addresses the advantages of CRISPR/Cas9 over other endonucleases. This review also provides the background for the mouse models of **chapter 4** and **5** and explains how iPSCs can be used to model disease, as is later demonstrated in **chapter 3**.

For the study reported in **chapter 3**, the Notarangelo lab generated three induced pluripotent stem cell lines (iPSCs) of patients with mutations in *RAG1*. Two patients showed a SCID phenotype (P1 and P2), while one patient had an Omenn phenotype (P3). For this study I generated and characterized the iPSCs from fibroblasts of a SCID patient (P1) carrying a homozygous c.1428delC mutation, predicted to result in frameshift and premature termination (p.N476Kfs*16).

The Zúñiga-Pflücker lab used their *in vitro* assay to differentiate these iPSCs into T-cells and compare their immunological phenotype including TCR repertoire to T-cells of healthy donor iPSCs. iPSCs are a promising tool to model PID and potentially perform pre-clinical correction studies, as explained in **chapter 2**.

Chapter 4 describes how I used CRISPR/Cas9 to generate different *Rag1* mouse models in a single-step by targeting residue 838. The functional role of this region of RAG1 had not been studied previously. For the study reported in **chapter 5** I used CRISPR/Cas9 to generate 3 different hypomorphic *Rag1* mouse models with point mutations described in patients with CID-G/AI. These mouse models were used to characterize this immunological phenotype, including the TCR and BCR repertoire, in detail. In particular, we focused on early T- and B-cell development which is difficult to study in patients.

References

1. Alberts B. Molecular biology of the cell. [Hauptbd.], [Hauptbd.]. New York, NY: Garland Science; 2015.
2. van Dongen JJ, Comans-Bitter WM, Wolvers-Tettero IL, Borst J. Development of human T lymphocytes and their thymus-dependency. *Thymus*. 1990;16(3-4):207-234.
3. Ru H, Chambers MG, Fu TM, Tong AB, Liao M, Wu H. Molecular Mechanism of V(D) J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell*. 2015;163(5):1138-1152.
4. Kim MS, Lapkouski M, Yang W, Gellert M. Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature*. 2015;518(7540):507-511.
5. Feeney AJ, Goebel P, Espinoza CR. Many levels of control of V gene rearrangement frequency. *Immunol Rev*. 2004;200:44-56.
6. Teng G, Schatz DG. Regulation and Evolution of the RAG Recombinase. *Adv Immunol*. 2015;128:1-39.
7. Liu Y, Subrahmanyam R, Chakraborty T, Sen R, Desiderio S. A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity*. 2007;27(4):561-571.
8. Notarangelo LD, Kim MS, Walter JE, Lee YN. Human RAG mutations: biochemistry and clinical implications. *Nat Rev Immunol*. 2016;16(4):234-246.
9. Kwan A, Abraham RS, Currier R, et al. Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA*. 2014;312(7):729-738.
10. de Pagter AP, Bredius RG, Kuijpers TW, et al. Overview of 15-year severe combined immunodeficiency in the Netherlands: towards newborn blood spot screening. *Eur J Pediatr*. 2015;174(9):1183-1188.
11. Blom M, Pico-Knijnenburg I, Sijne-van Veen M, et al. An evaluation of the TREC assay with regard to the integration of SCID screening into the Dutch newborn screening program. *Clin Immunol*. 2017;180:106-110.
12. Nakagawa N, Imai K, Kanegane H, et al. Quantification of kappa-deleting recombination excision circles in Guthrie cards for the identification of early B-cell maturation defects. *J Allergy Clin Immunol*. 2011;128(1):223-225 e2.
13. Barbaro M, Ohlsson A, Borte S, et al. Newborn Screening for Severe Primary Immunodeficiency Diseases in Sweden-a 2-Year Pilot TREC and KREC Screening Study. *J Clin Immunol*. 2017;37(1):51-60.
14. de Felipe B, Olbrich P, Lucenas JM, et al. Prospective neonatal screening for severe T- and B-lymphocyte deficiencies in Seville. *Pediatr Allergy Immunol*. 2016;27(1):70-77.
15. Nemazee D. Mechanisms of central tolerance for B cells. *Nat Rev Immunol*. 2017;17(5):281-294.
16. Picard C, Al-Herz W, Bousfiha A, et al. Primary Immunodeficiency Diseases: an Update on the Classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. *J Clin Immunol*. 2015;35(8):696-726.
17. Al-Herz W, Bousfiha A, Casanova JL, et al. Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency. *Front Immunol*. 2014;5:162.
18. Fischer A, Notarangelo LD, Neven B, Cavazzana M, Puck JM. Severe combined immunodeficiencies and related disorders. *Nat Rev Dis Primers*. 2015;1:15061.
19. Pai SY, Logan BR, Griffith LM, et al. Transplantation outcomes for severe combined immunodeficiency, 2000-2009. *N Engl J Med*. 2014;371(5):434-446.
20. Woodbine L, Gennery AR, Jeggo PA. The clinical impact of deficiency in DNA non-homologous end-joining. *DNA Repair (Amst)*. 2014;16:84-96.
21. Kwan A, Church JA, Cowan MJ, et al. Newborn screening for severe combined immunodeficiency and T-cell lymphopenia in California: results of the first 2 years. *J Allergy Clin Immunol*. 2013;132(1):140-150.
22. Omenn GS. Familial Reticuloendotheliosis with Eosinophilia. *N Engl J Med*. 1965;273:427-432.
23. Rieux-Laucat F, Bahadoran P, Brousse N, et al. Highly restricted human T cell repertoire in peripheral blood and tissue-infiltrating lymphocytes in Omenn's syndrome. *J Clin Invest*. 1998;102(2):312-321.

24. Villa A, Santagata S, Bozzi F, et al. Partial V(D)J recombination activity leads to Omenn syndrome. *Cell*. 1998;93(5):885-896.
25. Shearer WT, Dunn E, Notarangelo LD, et al. Establishing diagnostic criteria for severe combined immunodeficiency disease (SCID), leaky SCID, and Omenn syndrome: the Primary Immune Deficiency Treatment Consortium experience. *J Allergy Clin Immunol*. 2014;133(4):1092-1098.
26. de Villartay JP, Lim A, Al-Mousa H, et al. A novel immunodeficiency associated with hypomorphic RAG1 mutations and CMV infection. *J Clin Invest*. 2005;115(11):3291-3299.
27. Ehl S, Schwarz K, Enders A, et al. A variant of SCID with specific immune responses and predominance of gamma delta T cells. *J Clin Invest*. 2005;115(11):3140-3148.
28. Schuetz C, Huck K, Gudowius S, et al. An immunodeficiency disease with RAG mutations and granulomas. *N Engl J Med*. 2008;358(19):2030-2038.
29. De Ravin SS, Cowen EW, Zarembek KA, et al. Hypomorphic Rag mutations can cause destructive midline granulomatous disease. *Blood*. 2010;116(8):1263-1271.
30. Avila EM, Uzel G, Hsu A, et al. Highly variable clinical phenotypes of hypomorphic RAG1 mutations. *Pediatrics*. 2010;126(5):e1248-1252.
31. Henderson LA, Frugoni F, Hopkins G, et al. Expanding the spectrum of recombination-activating gene 1 deficiency: a family with early-onset autoimmunity. *J Allergy Clin Immunol*. 2013;132(4):969-971 e961-962.
32. Walter JE, Rosen LB, Csomos K, et al. Broad-spectrum antibodies against self-antigens and cytokines in RAG deficiency. *J Clin Invest*. 2015;125(11):4135-4148.
33. Sharapova SO, Migas A, Guryanova I, et al. Late-onset combined immune deficiency associated to skin granuloma due to heterozygous compound mutations in RAG1 gene in a 14 years old male. *Hum Immunol*. 2013;74(1):18-22.
34. Patiroglu T, Akar HH, Gilmour K, et al. Atypical severe combined immunodeficiency caused by a novel homozygous mutation in Rag1 gene in a girl who presented with pyoderma gangrenosum: a case report and literature review. *J Clin Immunol*. 2014;34(7):792-795.
35. Chen K, Wu W, Mathew D, et al. Autoimmunity due to RAG deficiency and estimated disease incidence in RAG1/2 mutations. *J Allergy Clin Immunol*. 2014;133(3):880-882 e810.
36. Buchbinder D, Baker R, Lee YN, et al. Identification of patients with RAG mutations previously diagnosed with common variable immunodeficiency disorders. *J Clin Immunol*. 2015;35(2):119-124.
37. Kuijpers TW, Ijspeert H, van Leeuwen EM, et al. Idiopathic CD4⁺ T lymphopenia without autoimmunity or granulomatous disease in the slipstream of RAG mutations. *Blood*. 2011;117(22):5892-5896.
38. Abolhassani H, Wang N, Aghamohammadi A, et al. A hypomorphic recombination-activating gene 1 (RAG1) mutation resulting in a phenotype resembling common variable immunodeficiency. *J Allergy Clin Immunol*. 2014;134(6):1375-1380.
39. Kato T, Crestani E, Kamae C, et al. RAG1 deficiency may present clinically as selective IgA deficiency. *J Clin Immunol*. 2015;35(3):280-288.
40. Geier CB, Piller A, Linder A, Sauerwein KM, Eibl MM, Wolf HM. Leaky RAG Deficiency in Adult Patients with Impaired Antibody Production against Bacterial Polysaccharide Antigens. *PLoS One*. 2015;10(7):e0133220.
41. Chou J, Hanna-Wakim R, Tirosh I, et al. A novel homozygous mutation in recombination activating gene 2 in 2 relatives with different clinical phenotypes: Omenn syndrome and hyper-IgM syndrome. *J Allergy Clin Immunol*. 2012;130(6):1414-1416.
42. Reiff A, Bassuk AG, Church JA, Campbell E, Bing X, Ferguson PJ. Exome sequencing reveals RAG1 mutations in a child with autoimmunity and sterile chronic multifocal osteomyelitis evolving into disseminated granulomatous disease. *J Clin Immunol*. 2013;33(8):1289-1292.
43. Takahashi K, Okita K, Nakagawa M, Yamanaka S. Induction of pluripotent stem cells from fibroblast cultures. *Nat Protoc*. 2007;2(12):3081-3089.



2 Novel genome-editing tools to model and correct primary immunodeficiencies

Frontiers in Immunology, Vol. 6, art. 250, 2015



Lisa Ott de Bruin^{1,2}
Stefano Volpi^{3,4}
Kiran Musunuru⁵

¹ Division of Immunology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

² Department of Pediatric Immunology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands

³ UO Pediatria 2, Istituto Giannina Gaslini, University of Genoa, Genoa, Italy

⁴ Division of Immunology and Allergy, University Hospital of Lausanne, Laboratory Center of Epalinges (CLE), Epalinges, Switzerland

⁵ Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA

Abstract

Severe combined immunodeficiency (SCID) and other severe non-SCID primary immunodeficiencies (non-SCID PID) can be treated by allogeneic hematopoietic stem cell transplantation, but when HLA-matched donors are lacking, this can be a high-risk procedure. Correcting the patient's own hematopoietic stem cells with gene therapy offers an attractive alternative. Gene therapies currently being used in clinical settings insert a functional copy of the entire gene by means of a viral vector. With this treatment, severe complications may result due to integration within oncogenes.

A promising alternative is the use of endonucleases such as ZFNs, TALENs and CRISPR/Cas9 to introduce a double-stranded break in the DNA and thus induce homology-directed repair. With these genome-editing tools a correct copy can be inserted in a precisely targeted "safe harbor". They can also be used to correct pathogenic mutations in situ and to develop cellular or animal models needed to study the pathogenic effects of specific genetic defects found in immunodeficient patients. This review discusses the advantages and disadvantages of these endonucleases in gene correction and modeling with an emphasis on CRISPR/Cas9, which offers the most promise due to its efficacy and versatility.

Introduction

P RIMARY IMMUNODEFICIENCIES (PIDs) comprise a heterogeneous group of rare, chronic diseases in which part of the immune system is missing or functions improperly. PIDs are caused by a myriad of different genetic defects and their clinical manifestations may vary significantly. On the clinical spectrum of PID, severe combined immunodeficiency (SCID) is the most severe form of immunodeficiency. SCID can be caused by many different genetic mutations that result in a developmental block in the production of T-cells with an additional primary or secondary defect in B-cells. NK cells may be lacking as well. SCID is characterized by increased susceptibility to life-threatening infections, particularly early in life. Newborn screening for SCID has been implemented in many states in the USA, facilitating early detection and improving treatment outcomes.¹⁻³ In addition to newborn screening, the advances in gene identification techniques such as exome and genome sequencing have greatly enhanced diagnostic capabilities in the field of PID. Over 230 PID causing genes have been described and novel gene defects continue to be discovered⁴. In parallel, the field of genome editing has progressed rapidly in the past few years, and many new tools are now available. These greatly ease the generation of *in vitro* models and animal models needed to study Mendelian disorders such as PID. Genome editing techniques hold great promise for treatment by direct gene correction as well. This review will address these novel genome-editing methodologies and how these tools can be applied to model and correct PID.

Hematopoietic stem cell transplantation

The current treatment of choice for SCID and other severe forms of PID is allogeneic hematopoietic stem cell transplantation (HSCT), which replaces defective hematopoietic lineages with functional cells. If a histocompatibility leucocyte antigen (HLA)-matched donor is available, conditioning chemotherapy is usually not indicated, because the patient has no T-cells to cause rejection⁵. However, HLA-matched donors may not be available. In those cases, depending on donor source and SCID genotype/phenotype, conditioning chemotherapy may be needed to facilitate robust and sustained engraftment of donor cells and improve immune reconstitution⁶. Although results of HSCT have greatly improved over the years, when HLA-matched donors are lacking or when the recipients suffer from ongoing active infections or other serious complications, clinical outcomes are still suboptimal⁵⁻¹². This is due to risks of conditioning chemotherapy, graft rejection, graft-versus-host disease (GvHD), and delayed immune reconstitution. For these patients, gene therapy, in which gene-mutated autologous hematopoietic stem cells (HSCs) are complemented

with a correct version of the gene, may offer an attractive alternative.

Gene therapy using viral vectors

In order to complement autologous HSCs, CD34⁺ hematopoietic stem cells are harvested from the patient and then transduced with a viral vector containing a correct copy of the gene along with regulatory elements that control gene expression, such as promoters and enhancers. The viral vector allows integration of the therapeutic transgene into the HSC genome. HSCs transduced with the vector are then infused back into the patient. As in allogeneic HSTC, the number of successfully transduced HSCs required to obtain optimal reconstitution depends on the selective advantage of the corrected HSCs over the patient HSCs without the correct gene¹³. In the first trials of gene therapy for PIDs, retro-viral vectors were used in which expression of the normal transgene was driven by the retrovirus long terminal repeat (LTR). With this approach, successful and durable T-cell reconstitution was achieved in patients with X-linked SCID (X-SCID)^{14,15}, adenosine deaminase (ADA) deficiency¹⁶⁻²⁰, and Wiskott-Aldrich syndrome (WAS)^{21,22}. Unfortunately, several patients developed leukemia. These serious adverse events were caused by preferential integration of retro-viral vectors in proximity of transcription initiation sites of genes (including oncogenes) and by the strong enhancer activity of the viral LTR, leading to increased and deregulated expression of the targeted oncogenes.²³⁻²⁶

To counter these adverse effects, much effort has gone in the development of safer viral vectors. A gene therapy trial to correct X-linked SCID using a self-inactivating retro-viral (SIN-RV) vector, in which the U3 enhancer was deleted from the LTR and expression was driven by the weaker eukaryotic human elongation factor 1 α (EF1 α) short promoter, is currently underway in Europe and in the USA. Preliminary results from this trial show a similar kinetics of T-cell recovery compared to that of the previous trial, but with significantly less integration within proto-oncogenes. Long-term safety effects remain to be studied²⁷. In addition to the safer SIN-RV vectors, lentiviral vectors are promising. *In vitro*^{28, 29} and *in vivo* studies³⁰⁻³² have demonstrated that lentiviral vectors integrate randomly in actively transcribed genes, without any preference for the transcription initiation sites and regulatory elements; thus, they are potentially safer³³. Therefore several trials using self-inactivating lentiviral (SIN-LV) vectors have been initiated³⁴, including a treatment trial of Wiskott Aldrich Syndrome (WAS) that shows promising results³⁵.

Genome editing using endonucleases

While SIN-RV and SIN-LV vectors demonstrate a safer integration site profile, greater control over vector site integration is still desired. Engineered

endonucleases that introduce double-stranded breaks (DSB) at specific sequences in the genomic DNA offer much more control over the integration site of viral vectors. Cells repair a DSB either through the error-prone process of non-homologous end-joining (NHEJ) or through homology-directed repair (HDR) in which a highly homologous template, either a sister chromatid or an exogenous double stranded or single stranded DNA template, is copied accurately. HDR can be used to either insert a gene into a specific “safe harbor” or to replace a defective gene *in-situ* (**Fig. 1**). “Safe harbors” are regions in genomic DNA that do not contain oncogenes and that can be disrupted without adverse consequences. One such safe harbor is the adeno-associated virus integration site 1 (*AAVS1*) locus. In order to insert a functional copy of the PID-causing gene into a specific locus such as *AAVS1*, one can use an engineered endonuclease to introduce a DSB at the site and a DNA repair template containing the gene flanked by two homology arms that match the *AAVS1* sequence³⁶⁻³⁹.

Alternatively, instead of adding a normal copy of the gene in the “safe harbor”, one can correct the defective PID gene *in situ*. In this case, the DSB is introduced close to the mutation, after which a repair template, containing the correct sequence flanked by two homology arms matching the sequences surrounding the target site, is inserted. When correcting the actual mutation itself, the endogenous promoter, enhancer and other regulatory elements are

Genome editing: harnessing natural repair mechanisms to modify DNA

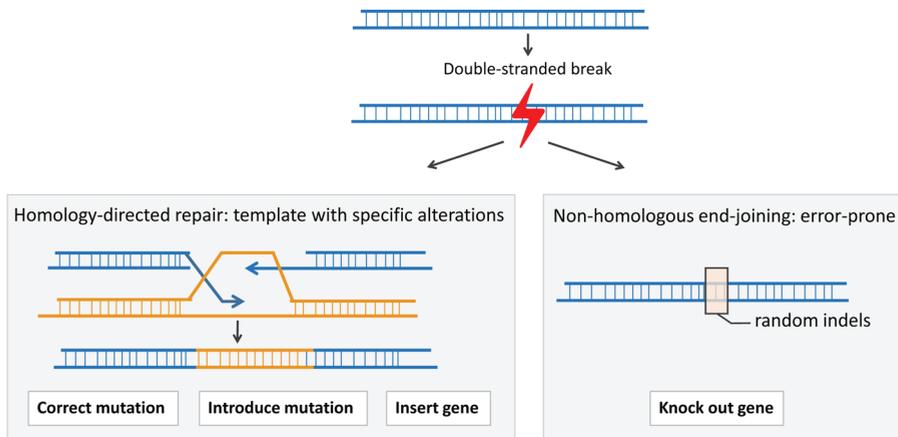


Figure 1: A double-stranded break (DSB) in the DNA can be repaired through the process of homologous recombination (HDR) or through the error-prone process of non-homologous end-joining (NHEJ). In HDR a template is used to correct the DSB. HDR can be used to precisely introduce a gene or part of a gene or even a point mutation, whereas NHEJ can result in insertions and/or deletions (indels) around the DSB. An indel can lead to a frameshift and an early stop codon.

used and, thus, physiological gene expression is preserved. This is beneficial when aiming to correct tightly regulated genes such as recombination-activating gene 1 (*RAG1*) and recombination-activating gene 2 (*RAG2*), the genes required for VDJ recombination during T-cell and B-cell development⁴⁰. In addition, in situ correction is ideal for dominant-negative mutations; in these cases, simple addition of the normal gene would be inadequate to rescue the phenotype, and a specific correction of the mutation is required. The endonucleases can be designed to only target the mutant sequence and spare the wild-type sequence. For the majority of PIDs, the disease phenotype is caused by recessive mutations. In these cases correcting one allele is sufficient to rescue the phenotype.

When using engineered endonucleases, several aspects need to be considered. First, the efficiency of introducing DSBs at the target site, the on-target efficiency, is important. When testing the endonuclease, on-target efficiency can be inferred from the proportion of alleles in a batch of cells showing deletions or insertions (“indels”) at the target site, because some of the introduced DSBs were repaired by the error-prone process of NHEJ. These indels can be easily captured using next-generation sequencing or alternatively, by studying the heteroduplex DNA hybridization of PCR products from the target site (e.g., with the Surveyor[®] assay). Second, a common concern with the use of endonucleases is off-target mutagenesis. This is the inadvertent introduction of mutations caused by DSBs at genomic sites other than the target site. Currently, several sequencing techniques are available to check for off-target mutagenesis. Some will be addressed in the next section when discussing different endonucleases.

A third aspect that needs to be considered when using engineered endonucleases is how to deliver the endonuclease and the DNA template to the cell. For ex vivo therapeutic applications, such as gene targeting in HSCs harvested from a patient, nucleofection is a non-viral method to introduce polynucleotides into the cells⁴¹. Its disadvantages include toxicity to the cells and low efficiency. Alternatively, viral vectors that do not integrate in the genome, such as integrase-deficient lentiviral vectors, adenoviral vectors, and vectors based on adeno-associated viruses can be used. These viruses enter the cell and express the endonuclease, without inserting it into the genome. These have been proven effective tools to deliver both the repair constructs and the endonucleases⁴⁴⁻⁴⁷. In vivo therapeutic applications present a greater challenge in that the delivery method must efficiently and specifically target the desired cells and spare the other cells within the whole body.

ZFNs, TALENs, and CRISPR/Cas9

Although a number of different genome-editing technologies are now in use, we describe three types of engineered endonucleases that have found broad use in the biomedical community, with a particular focus on the most recently developed nuclease system, CRISPR/Cas9, which has attracted widespread attention for its efficacy and versatility (**Fig. 2**, p. 23).

Zinc finger nucleases (ZFNs)

Zinc finger nucleases (ZFNs) consist of a pair of endonuclease domains of the bacterial FokI restriction enzyme flanked by two site-specific DNA-binding domains (**Fig. 2A**, p. 23). Upon binding of these domains, the FokI domains dimerize and introduce a DSB⁴². Even though academic consortia have developed open-source libraries for ZFN construction^{43,44}, engineering of site-specific ZFNs remains difficult for non-specialists. Depending on the ZFN architecture used, only limited sites in a genomic region can be targeted, which might be problematic if a specific mutation needs to be corrected⁴⁵⁻⁴⁸. One study showed off-target mutagenesis with ZFNs in one out of ten sites with a sequence similar to the target site³⁷. Two other studies also found that DSBs were introduced in off-target sites using ZFNs in a human tumor cell line^{49,50}. One of the ZFNs tested was designed to correct mutations causing X-SCID, i.e., mutations in the gene encoding the Interleukin 2 receptor gamma (*IL2RG*)^{49,51}. Variants of the endonucleases have been developed to reduce off-target mutagenesis. These consist of a mix of two distinct ZFNs with different FokI domains that are obligate heterodimers, such that a ZFN pair introduces a DSB only when the two distinct ZFNs are able to bind adjacent DNA regions⁵²⁻⁵⁴.

Transcription activator-like effector nucleases (TALENs)

Transcription activator-like (TAL) domains are tandem arrays with 10 to 30 repeats, each 33 to 35 aminoacids long, that bind and recognize extended DNA sequences⁵⁵. One domain of the TAL repeats is fused to a FokI endonuclease domain, creating a TAL effector nuclease (TALEN), similar to the ZFN. Upon binding of the two TAL effector nucleases (TALENs) to flanking DNA sequences, the FokI domains dimerize and introduce a DSB at the target site (**Fig. 2.B**). TALENs are easier to design and have fewer constraints on site selection than ZFNs. A disadvantage of using TALENs is their size. The DNA sequence encoding each TALEN is more than 3 kb in size, compared to only 1 kb for each ZFN. The larger size makes it harder to deliver TALEN-expressing plasmids into cells. Moreover, the highly repetitive nature of TALEN sequences due to their tandem arrays makes them more challenging to package into viral vectors for delivery⁵⁶. With respect to off-target mutagenesis, two studies

showed this phenomenon was minimal with TALENs^{57,58}. Unfortunately, not many studies have compared TALENs and ZFNs directly. However one study suggested that when targeting the *CCR5* gene, less off-target mutagenesis was seen with TALENs than with ZFNs⁵⁹.

CRISPR/Cas9

The latest in a series of new genome-editing tools is the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated 9 (Cas9) system. Cas9 is a protein used by bacteria to destroy foreign DNA. The foreign DNA is cleaved and incorporated as small sequences—called protospacer sequences—into the bacterial genome, and these sequences are then transcribed as short CRISPR RNAs (crRNAs). These crRNAs are then used to target and destroy any foreign DNA sequences that enter the cell and match those sequences. The bacterial Cas9 nuclease and a crRNA form a ternary complex with a second RNA component, the trans-activating crRNA (tracrRNA), which has a fixed sequence. This complex can engage double-stranded DNA, with the crRNA hybridizing the protospacer sequence and Cas9 binding a specific protospacer-adjacent motif (PAM).

Once the complex is engaged, Cas9 introduces a DSB 3bp upstream of the PAM⁶⁰. After the characterization of the CRISPR/Cas9 system in bacteria, investigators found that it could be used to introduce DSBs efficiently in mammalian DNA. The RNA components of the CRISPR/Cas9 system can be separate crRNA and tracrRNA molecules, or the two RNA molecules can be combined into a single guide RNA (gRNA)⁶¹⁻⁶⁴. The Cas9 protein from the species *Streptococcus pyogenes* is the most commonly used at present and uses a PAM with the sequence NGG. When Cas9 is used with a single gRNA (as is now usually the case), CRISPR/Cas9 represents a simple two-component system (**Fig. 2C**).

The advantages of CRISPR/Cas9 are the high efficiency of introducing DSBs into the genomes of mammalian cells and the ease of engineering. The Cas9 nuclease is always the same; to target a different region of the genome, only the protospacer region (20 nucleotides) of the gRNA needs to be altered. While theoretically the PAM requirement could be a limitation, on average the NGG PAM sequence needed by *S. pyogenes* Cas9 can be found every 8 bp in the genome, making it very likely to find a CRISPR/Cas9 target site near the mutation that needs to be corrected⁶³. In contrast, the PAM sequence used by *Streptococcus thermophilus* Cas9, NNAGAAW, occurs on average every 64 bp; the *N. meningitidis* Cas9 protein requires a NNNNGATT PAM, which occurs on average every 128 bp^{63,65,66}.

As with other engineered nucleases, the main concern regarding the use of

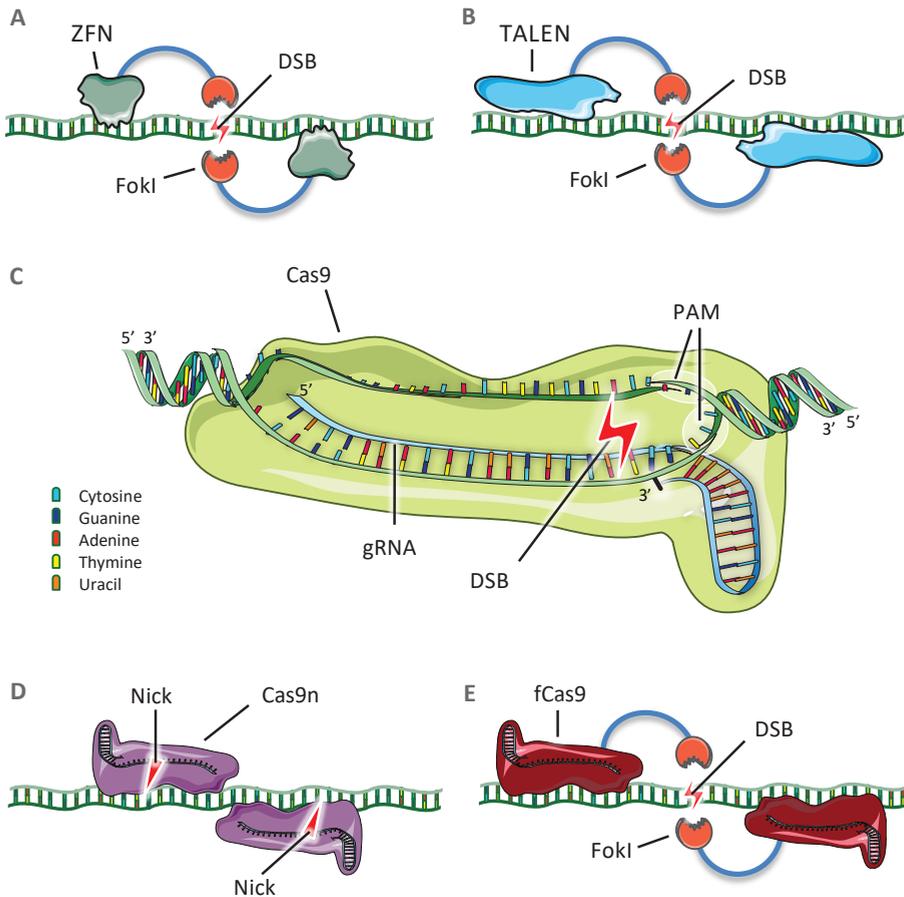


Figure 2: **Schematic representation of ZFNs, TALENs, and CRISPR/Cas9.**

[A] In the most commonly used CRISPR/Cas9 system, Cas9 forms a complex with a gRNA that recognizes and hybridizes a 20-bp protospacer in the genome. Cas9 binds the adjacent PAM sequence and introduces a double-stranded break (DSB) 3 bp upstream of the PAM sequence. [B] Cas9 nickases (Cas9n) are mutant variants that bind to flanking DNA sequences and generate single-strand nicks instead of DSBs. Two nicks are the equivalent of a DSB. [C] Another variant consists of catalytically inactive Cas9 fused to a FokI nuclease domain. When two FokI nucleases dimerize because the Cas9 proteins bind to flanking DNA sequences, a DSB is introduced between the binding sites.

CRISPR/Cas9 is off-target mutagenesis. Such mutagenesis appears to be highly gRNA dependent and most often occurs at sites with sequence similarity to the protospacer, with up to several mismatches tolerated; more mismatches are tolerated as the distance from the PAM increases^{61,62,67-70}. The number of potential off-target binding sites of Cas9 can vary widely, depending on the gRNA used^{71,72}.

Moreover, some studies suggest that even when the CRISPR/Cas9 is able to bind to an off-target site, the mismatches prevent actual DNA cleavage^{47,48}. A novel method to systematically assess off-target mutagenesis, the so-called genome-wide, unbiased identification of DSBs, enabled by sequencing (GUIDE-seq), was recently reported⁷³. This method is based on the detection of small synthetic double-stranded DNA oligodeoxynucleotides that are incorporated in the genomic DNA at the site of the DSBs through NHEJ.

To reduce off-target mutagenesis, many different variants of the CRISPR/Cas9 system with higher target specificity are being developed. In one strategy the length of the gRNA protospacer is reduced by up to three nucleotides, which appeared to make the gRNA less tolerant of mismatches and less likely to bind to off-target sites, thereby reducing the rate of off-target mutagenesis⁷⁴.

A different approach uses Cas9 “nickases,” which are mutated variants of Cas9 that each introduce a single-stranded break (called a nick) instead of a DSB (**Fig. 2D**, p. 23). When two distinct gRNAs matching to two distinct sequences flanking the target site are used, Cas9 will produce two separate nicks that together are the equivalent of a DSB and can thus induce repair by NHEJ or HDR. Off-target mutagenesis is substantially reduced with the double-nickase strategy because the two nicks occur in proximity only when Cas9 binds to two adjacent sequences that resemble the protospacers, which is very unlikely to occur elsewhere in the genome⁷⁵⁻⁷⁷.

Finally, investigators have combined the most desirable properties of CRISPR/Cas9 and ZFNs/TALENs by fusing a catalytically dead Cas9 to a FokI domain. Two Cas9-FokI fusion proteins are guided to flanking sequences around a target site by a pair of gRNAs. Upon DNA binding of the Cas9 domains, the FokI domains dimerize and generate a DSB (**Fig. 2E**, p. 23). As with the double-nickase strategy, binding of Cas9 to two separate nearby sequences is required for the generation of a DSB, and accordingly the off-target mutagenesis rate is greatly reduced^{78,79}.

Genome editing results in human stem cells

A number of studies have demonstrated the feasibility of genome editing in human stem cells with ZFNs, TALENs, and CRISPR/Cas9.^{36,37,51,62,77,80-84} In one such study, intestinal stem cells of cystic fibrosis patients with homozygous delta508 mutations in the *CFTR* gene were corrected using CRISPR/Cas9. Corrected and uncorrected stem cells were differentiated into organoids and compared. The corrected organoids showed swelling in response to forskolin treatment, as expected in the presence of a functional CFTR protein, whereas the mutant organoids failed to swell⁸⁵.

Another group transfected human HSCs with ZFNs that disrupted *CCR5*, the chemokine receptor used by human immunodeficiency virus (HIV) to infect cells. The targeted cells were transplanted into immunodeficient NOD/SCID/IL2 γ -null (NSG) mice, which then exhibited human hematopoiesis. Moreover, when the transplanted mice were infected with HIV, there was a selective survival advantage for the *CCR5* knockout cells, protection of the human-derived T-cell populations, and a reduction in HIV viral load⁴¹.

In SCID and most types of PID, multiple cell types of the immune system are defective. In order to restore the immune function one needs to target the patient's own HSCs⁸⁴. The most common form of SCID, X-SCID, is caused by mutations in the gene encoding interleukin 2 receptor gamma (*IL2RG*). Several groups have successfully used ZFNs to target and induce HDR in the *IL2RG* locus in various human cell types, including HSCs and embryonic stem cells^{51,84,86}. One of these studies corrected an *IL2RG* mutational hotspot in HSCs using ZFNs and showed multilineage hematopoietic differentiation upon transplantation of gene-corrected cells into NSG mice⁸⁴. Another study successfully used specific TALENs to target and induce HDR in the *IL2RG* locus of Jurkat cells⁸⁷.

Challenges when targeting HSCs are the low efficiency of HDR and the risk of losing multilineage potential when manipulating and expanding gene-corrected cells *in vitro*. Several cell types such as T-cells, HSCs, and fibroblasts can be reprogrammed into induced pluripotent stem cells (iPSCs) by transducing these cells with a SIN-lentiviral vector expressing four pluripotency genes, *OCT4*, *SOX2*, *KLF4* and c-Myc.⁸⁸⁻⁹⁰ Studies have shown the feasibility of iPSC gene targeting to correct hematopoietic diseases such as sickle cell disease *in vitro* using ZFNs^{82,91} or TALENs⁹². Many different iPSCs from a number of patients with distinct immunodeficiencies have now been generated^{38,39,93,94}. These patient iPSCs can be corrected, or, alternatively, an original patient somatic cell can be corrected before being reprogrammed into iPSCs. ZFNs have been used to correct chronic granulomatous disease (CGD) by introducing up to five different functional genes into the *AAVS1* safe harbor in iPSCs generated from peripheral HSCs. Using *in vitro* myeloid differentiation, normal granulocytes were generated from the corrected iPSCs^{38,39}.

Successful differentiation of human iPSCs into T-cells *in vitro* has been recently reported⁹⁵⁻⁹⁷, making it possible to test the ability of genome-editing to restore T-cell differentiation capacity of iPSCs from patients with SCID. However, *in vivo* use of gene-edited iPSCs for correction of human PIDs is not yet ready for the clinic because of safety concerns related to the tumorigenic potential of iPSCs and because of difficulty in generating definitive hematopoietic stem cells from human iPSCs. Nonetheless, these recent achievements with *in vitro*

targeted differentiation of human iPSCs have great value in that they enable preclinical efficacy and safety studies of genome-editing approaches that may eventually be applied to human HSCs.

Modeling PIDs using genome-editing

In vitro models

Even though gene targeting using engineered endonucleases is not ready to be applied in a clinical setting, it already offers a valuable tool to model diseases at the cellular level. CRISPR/Cas9 has been shown to be particularly efficient in the generation of knockout cell lines; as described above, after a DSB has been introduced into the genomic region matching the protospacer sequence of the gRNA, the cell uses either HDR or NHEJ to repair the defect. NHEJ can result in indels, which in turn can cause frameshifts and the occurrence of premature stop codons. A knockout cell generated in this way can be clonally expanded into a cell line that can be used for modeling studies. In addition, a useful property of the CRISPR/Cas9 system is that multiple genes can be knocked out simultaneously if several gRNAs are used together (“multiplexing”). A particular advantage compared to RNA interference is that CRISPR/Cas9 can be used to target regions in the non-coding genome (e.g. promoter and enhancer regions)⁹⁸⁻¹⁰¹.

The advent of next-generation sequencing has stimulated a new wave of discovery of novel inborn errors of immunity¹⁰². The ability to correct patient-specific iPSCs or, conversely, to introduce patient-specific mutations into a wild-type iPSC line using endonucleases represents an invaluable tool to prove the pathogenicity of newly discovered mutations and to gain insight into disease mechanisms in different cell types, depending on patients’ phenotypes.

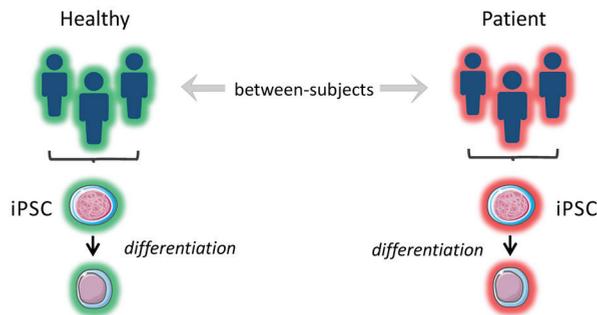
This approach also makes it possible to study the contribution of genetic background to the phenotypes arising from specific mutations by comparing patient-derived iPSCs with wild-type iPSCs into which the same mutations are introduced (**Fig. 3**). In a recent study, iPSCs were generated from patients with Parkinson disease caused by the G2019S mutation of the LRRK2 gene and from healthy controls. When comparing the whole-genome gene expression patterns, the investigators found a high degree of heterogeneity among the different iPSCs lines. However, when they used ZFNs to correct the mutation in three of the patient-derived iPSC lines and compared these lines to the original lines, and when they introduced the mutation into a healthy control line and compared this line to the original line, the lines were much more closely matched with respect to gene expression⁸³. This shows the importance of comparing isogenic lines, as confounding due to differences in genetic

background is minimized.

Animal models

Traditionally, animal models have been generated using homologous recombination: embryonic stem cells (ESCs) are electroporated with a highly homologous DNA template containing the sequence to be inserted but without using an engineered endonuclease to introduce a DSB. This approach results in a very low efficiency and requires the inclusion of an antibiotic resistance

A



B

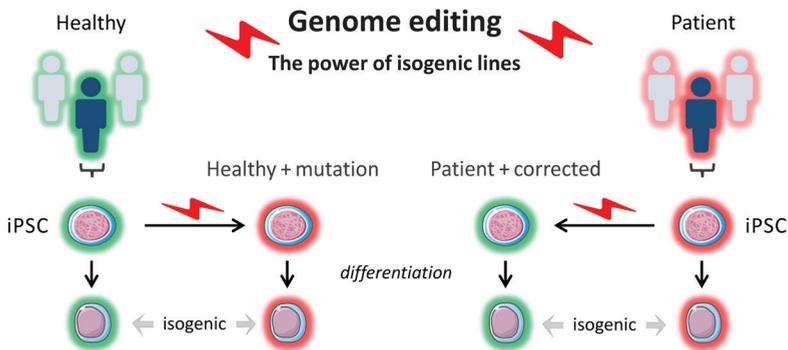


Figure 3 - In vitro modeling. [A] Induced pluripotent stem cells (iPSCs) are reprogrammed from a patient(s) and from a healthy control(s). The iPSCs are differentiated into a cell type of interest, and the phenotypes of the patient-derived cells are compared to the phenotypes of the healthy control cells. The cells that are compared do not have the exact same genetic background (genetically and epigenetically unmatched). This can lead to confounding. [B] Using genome editing with engineered nucleases like ZFNs, TALENs, and CRISPR/Cas9, a pathogenic mutation can be corrected in patient-derived cells or introduced into healthy control cells, and isogenic cell lines (i.e., identical genetic background) can be compared for relevant phenotypes.

gene in the inserted sequence for the selection of cells in which HDR has occurred. ESCs with the desired inserted sequence are then expanded, injected in blastocysts, and subsequently implanted in pseudogestant females. The resulting chimeric animals have to be further bred until the introduced mutation is transmitted through the germline. With the currently available genome-editing tools, this process can be greatly streamlined. Via the introduction of a DSB at the desired target site, a specific sequence can be efficiently introduced into ESCs without the need for an antibiotic resistance gene. Furthermore, to create a gene knockout, one can simply rely on NHEJ to produce indels leading to frameshifts and early stop codons.

In recent years, many animal models have been successfully generated using ZFNs, TALENs, or CRISPR/Cas9. TALENs and CRISPR/Cas9 have been used to generate knockout *C. elegans* models by injecting the endonucleases into the gonads¹⁰³⁻¹⁰⁵. Similarly, more complicated animal models can be generated by injecting the endonucleases in mRNA form directly in zygotes (**Fig. 4**). In the case of CRISPR/Cas9, this means that both the Cas9 and gRNA in RNA form are injected. Knock-in models can be generated by adding a DNA template to the injection mix, usually in the form of a single-stranded DNA oligonucleotide. Zebrafish models have been generated by injecting ZFNs or TALENs or CRISPR/Cas9 directly into the zygote¹⁰⁶⁻¹⁰⁹. This has been done in murine zygotes with ZFNs¹¹⁰⁻¹¹², TALENs^{113,114} and extremely efficiently with CRISPR/Cas9¹¹⁵⁻¹¹⁷.

New mouse models can be generated in just a few weeks, instead of taking one to two years as in the conventional strategy. With CRISPR/Cas9, the specific gRNA needed for the injections can be generated in a simple one-day procedure¹¹⁸. NSG mice have been efficiently generated in this way¹¹⁹. In other studies, the IgM locus has been successfully knocked out in rats via the injection of ZFNs and TALENs directly into the zygotes^{120,121}. Similarly, a rat model of X-SCID has been generated using ZFNs¹²². The multiplexing capacity of CRISPR/Cas9 has allowed for multiple genes being knocked out simultaneously¹²³. Endonucleases have been used to generate knockout models in animals not previously amenable to efficient genetic modification: rabbits with *IL2rg*, *Rag1*, or *Rag2* knockout¹²⁴⁻¹²⁷; hamsters with *Stat2* knockout¹²⁸; mutant pigs¹²⁹⁻¹³¹; and, most impressively, monkeys with *Rag1* knockout¹³². These kinds of animal models will enable disease studies of unprecedented sophistication.

Conclusion

During the last few years, the field of genome editing has shown tremendous progress. Currently, several endonucleases are available. We have described

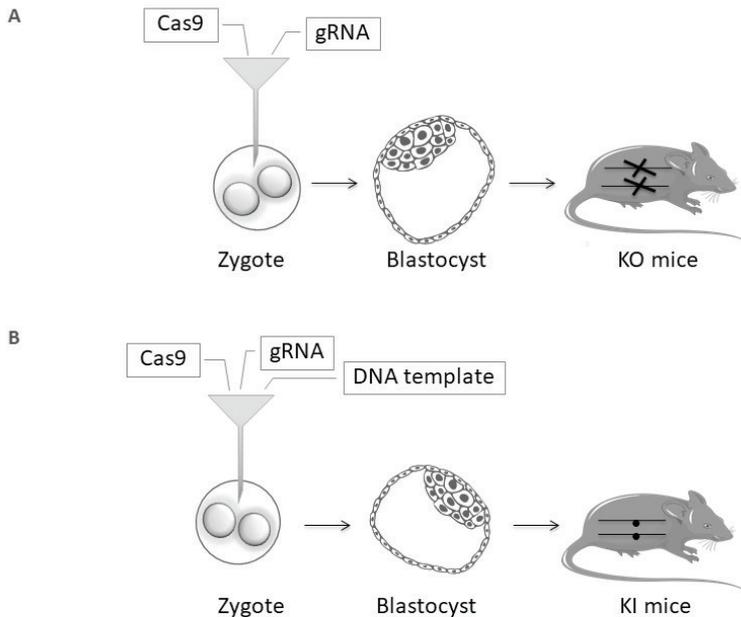


Figure 4: Schematic representation of zygote injection with CRISPR/Cas9.

[A] Injection of gRNA and Cas9 will lead to indels that can lead to a frameshift and an early stop codon thereby creating KO mice. [B] Addition of a highly homologous DNA template containing a specific mutation will result in KI mice, through the process of homology directed repair (HDR). Reagents are injected in the cytoplasm of the zygote. Alternatively, these can be injected in the pronucleus of the zygote, but cytoplasmic microinjection is simpler and less toxic. KI: knock-in, KO: knockout.

their advantages and disadvantages and how they can be used to model and correct PIDs. Efficiency and ease will continue to improve with further refinement of these tools, while endonuclease variants with increased specificity are being actively developed. To use genome editing in a clinical setting for the treatment of PID will require a further reduction in off-target mutagenesis and an improved yield of gene-corrected HSCs so that a sufficient number of cells for autologous transplantation and engraftment can be obtained. These issues notwithstanding, we expect the impact of genome editing on modern medicine to be revolutionary.

Acknowledgement

Lisa M Ott de Bruin is supported by T32 fellowship grant 5T32AI007512. Figures were produced using Servier Medical Art (www.servier.com). The authors thank Lauren Henderson MD, MMSc, Children's Hospital Boston, MA, USA, for her critical revision of this manuscript.

References

1. Puck JM. Laboratory technology for population-based screening for severe combined immunodeficiency in neonates: the winner is T-cell receptor excision circles. *J Allergy Clin Immunol*. 2012;129(3):607-616.
2. Borte S, von Döbeln U, Hammarström L. Guidelines for newborn screening of primary immunodeficiency diseases. *Curr Opin Hematol*. 2013;20(1):48-54.
3. Kwan A, Abraham RS, Currier R, et al. Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA*. 2014;312(7):729-738.
4. Al-Herz W, Bousfiha A, Casanova JL, et al. Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency. *Front Immunol*. 2014;5:162.
5. Buckley RH. Transplantation of hematopoietic stem cells in human severe combined immunodeficiency: long-term outcomes. *Immunol Res*. 2011;49(1-3):25-43.
6. Pai SY, Logan BR, Griffith LM, et al. Transplantation outcomes for severe combined immunodeficiency, 2000-2009. *N Engl J Med*. 2014;371(5):434-446.
7. Schuetz C, Neven B, Dvorak CC, et al. SCID patients with ARTEMIS vs RAG deficiencies following HCT: increased risk of late toxicity in ARTEMIS-deficient SCID. *Blood*. 2014;123(2):281-289.
8. Gennery AR, Slatter MA, Grandin L, et al. Transplantation of hematopoietic stem cells and long-term survival for primary immunodeficiencies in Europe: entering a new century, do we do better? *J Allergy Clin Immunol*. 2010;126(3):602-610 e601-611.
9. Antoine C, Müller S, Cant A, et al. Long-term survival and transplantation of haemopoietic stem cells for immunodeficiencies: report of the European experience 1968-99. *Lancet*. 2003;361(9357):553-560.
10. Buckley RH, Schiff SE, Schiff RI, et al. Hematopoietic stem-cell transplantation for the treatment of severe combined immunodeficiency. *N Engl J Med*. 1999;340(7):508-516.
11. Horn B, Cowan MJ. Unresolved issues in hematopoietic stem cell transplantation for severe combined immunodeficiency: need for safer conditioning and reduced late effects. *J Allergy Clin Immunol*. 2013;131(5):1306-1311.
12. Dvorak CC, Hassan A, Slatter MA, et al. Comparison of outcomes of hematopoietic stem cell transplantation without chemotherapy conditioning by using matched sibling and unrelated donors for treatment of severe combined immunodeficiency. *J Allergy Clin Immunol*. 2014;134(4):935-943 e915.
13. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, et al. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science*. 2000;288(5466):669-672.
14. Hacein-Bey-Abina S, Hauer J, Lim A, et al. Efficacy of gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med*. 2010;363(4):355-364.
15. Gaspar HB, Cooray S, Gilmour KC, et al. Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency. *Sci Transl Med*. 2011;3(97):97ra79.
16. Blaese RM, Culver KW, Miller AD, et al. T lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years. *Science*. 1995;270(5235):475-480.
17. Aiuti A, Cattaneo F, Galimberti S, et al. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N Engl J Med*. 2009;360(5):447-458.
18. Aiuti A, Slavin S, Aker M, et al. Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science*. 2002;296(5577):2410-2413.
19. Kohn DB, Weinberg KI, Nolta JA, et al. Engraftment of gene-modified umbilical cord blood cells in neonates with adenosine deaminase deficiency. *Nat Med*. 1995;1(10):1017-1023.
20. Gaspar HB, Cooray S, Gilmour KC, et al. Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction. *Sci Transl Med*. 2011;3(97):97ra80.

- 21.** Moratto D, Galiani S, Bonfim C, et al. Long-term outcome and lineage-specific chimerism in 194 patients with Wiskott-Aldrich syndrome treated by hematopoietic cell transplantation in the period 1980-2009: an international collaborative study. *Blood*. 2011;118(6):1675-1684.
- 22.** Boztug K, Schmidt M, Schwarzer A, et al. Stem-cell gene therapy for the Wiskott-Aldrich syndrome. *N Engl J Med*. 2010;363(20):1918-1927.
- 23.** Hacein-Bey-Abina S, Garrigue A, Wang GP, et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest*. 2008;118(9):3132-3142.
- 24.** Hacein-Bey-Abina S, Von Kalle C, Schmidt M, et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*. 2003;302(5644):415-419.
- 25.** Braun CJ, Boztug K, Paruzynski A, et al. Gene therapy for Wiskott-Aldrich syndrome—long-term efficacy and genotoxicity. *Sci Transl Med*. 2014;6(227):227ra233.
- 26.** Howe SJ, Mansour MR, Schwarzwaelder K, et al. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest*. 2008;118(9):3143-3150.
- 27.** Hacein-Bey-Abina S, Pai SY, Gaspar HB, et al. A modified gamma-retrovirus vector for X-linked severe combined immunodeficiency. *N Engl J Med*. 2014;371(15):1407-1417.
- 28.** Wu X, Li Y, Crise B, Burgess SM. Transcription start regions in the human genome are favored targets for MLV integration. *Science*. 2003;300(5626):1749-1751.
- 29.** Mitchell RS, Beitzel BF, Schroder AR, et al. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol*. 2004;2(8):E234.
- 30.** Montini E, Cesana D, Schmidt M, et al. Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat Biotechnol*. 2006;24(6):687-696.
- 31.** Montini E, Cesana D, Schmidt M, et al. The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. *J Clin Invest*. 2009;119(4):964-975.
- 32.** Biffi A, Bartolomea CC, Cesana D, et al. Lentiviral vector common integration sites in preclinical models and a clinical trial reflect a benign integration bias and not oncogenic selection. *Blood*. 2011;117(20):5332-5339.
- 33.** Naldini L. Ex vivo gene transfer and correction for cell-based therapies. *Nat Rev Genet*. 2011;12(5):301-315.
- 34.** Touzot F, Hacein-Bey-Abina S, Fischer A, Cavazzana M. Gene therapy for inherited immunodeficiency. *Expert Opin Biol Ther*. 2014;14(6):789-798.
- 35.** Aiuti A, Biasco L, Scaramuzza S, et al. Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science*. 2013;341(6148):1233151.
- 36.** Hockemeyer D, Wang H, Kiani S, et al. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol*. 2011;29(8):731-734.
- 37.** Hockemeyer D, Soldner F, Beard C, et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol*. 2009;27(9):851-857.
- 38.** Merling RK, Sweeney CL, Chu J, et al. An AAVS1-Targeted Minigene Platform for Correction of iPSCs From All Five Types of Chronic Granulomatous Disease. *Mol Ther*. 2014.
- 39.** Zou J, Sweeney CL, Chou BK, et al. Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPSC cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. *Blood*. 2011;117(21):5561-5572.
- 40.** van Til NP, Sarwari R, Visser TP, 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*. 2013.
- 41.** Holt N, Wang J, Kim K, et al. Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat Biotechnol*. 2010;28(8):839-847.
- 42.** Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*. 2010;11(9):636-646.
- 43.** Maeder ML, Thibodeau-Beganny S, Osiaik A, et al. Rapid “open-source” engineering

- of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell*. 2008;31(2):294-301.
- 44.** Maeder ML, Thibodeau-Beganny S, Sander JD, Voytas DF, Joung JK. Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nat Protoc*. 2009;4(10):1471-1501.
- 45.** Sander JD, Dahlborg EJ, Goodwin MJ, et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods*. 2011;8(1):67-69.
- 46.** Gupta A, Christensen RG, Rayla AL, Lakshmanan A, Stormo GD, Wolfe SA. An optimized two-finger archive for ZFN-mediated gene targeting. *Nat Methods*. 2012;9(6):588-590.
- 47.** Bhakta MS, Henry IM, Ousterout DG, et al. Highly active zinc-finger nucleases by extended modular assembly. *Genome Res*. 2013;23(3):530-538.
- 48.** Gaj T, Guo J, Kato Y, Sirk SJ, Barbas CF, 3rd. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat Methods*. 2012;9(8):805-807.
- 49.** Gabriel R, Lombardo A, Arens A, et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol*. 2011;29(9):816-823.
- 50.** Pattanayak V, Ramirez CL, Joung JK, Liu DR. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods*. 2011;8(9):765-770.
- 51.** Lombardo A, Genovese P, Beausejour CM, et al. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol*. 2007;25(11):1298-1306.
- 52.** Doyon Y, Vo TD, Mendel MC, et al. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods*. 2011;8(1):74-79.
- 53.** Miller JC, Holmes MC, Wang J, et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*. 2007;25(7):778-785.
- 54.** Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ, Cathomen T. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol*. 2007;25(7):786-793.
- 55.** Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. *Science*. 2011;333(6051):1843-1846.
- 56.** Holkers M, Maggio I, Liu J, et al. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res*. 2013;41(5):e63.
- 57.** Smith C, Gore A, Yan W, et al. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell*. 2014;15(1):12-13.
- 58.** Suzuki K, Yu C, Qu J, et al. Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. *Cell Stem Cell*. 2014;15(1):31-36.
- 59.** Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*. 2011;39(21):9283-9293.
- 60.** Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337(6096):816-821.
- 61.** Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol*. 2013;31(3):230-232.
- 62.** Mali P, Yang L, Esvelt KM, et al. RNA-guided human genome engineering via Cas9. *Science*. 2013;339(6121):823-826.
- 63.** Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339(6121):819-823.
- 64.** Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. *Elife*. 2013;2:e00471.
- 65.** Hou Z, Zhang Y, Propson NE, et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci U S A*. 2013;110(39):15644-15649.
- 66.** Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods*. 2013;10(11):1116-1121.
- 67.** Fu Y, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*. 2013;31(9):822-826.

- 68.** Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014;157(6):1262-1278.
- 69.** Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol*. 2013;31(9):839-843.
- 70.** Cradick TJ, Fine EJ, Antico CJ, Bao G. CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res*. 2013;41(20):9584-9592.
- 71.** Kescu C, Arslan S, Singh R, Thorpe J, Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol*. 2014;32(7):677-683.
- 72.** Wu X, Scott DA, Kriz AJ, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol*. 2014;32(7):670-676.
- 73.** Tsai SQ, Zheng Z, Nguyen NT, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*. 2014.
- 74.** Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*. 2014;32(3):279-284.
- 75.** Mali P, Aach J, Stranges PB, et al. Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*. 2013;31(9):833-838.
- 76.** Cho SW, Kim S, Kim Y, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res*. 2014;24(1):132-141.
- 77.** Ran FA, Hsu PD, Lin CY, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*. 2013;154(6):1380-1389.
- 78.** Tsai SQ, Wyvekens N, Khayter C, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*. 2014;32(6):569-576.
- 79.** Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol*. 2014;32(6):577-582.
- 80.** Zou J, Maeder ML, Mali P, et al. Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell*. 2009;5(1):97-110.
- 81.** Yusa K, Rad R, Takeda J, Bradley A. Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nat Methods*. 2009;6(5):363-369.
- 82.** Sebastiano V, Maeder ML, Angstman JF, et al. In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells*. 2011;29(11):1717-1726.
- 83.** Reinhardt P, Schmid B, Burbulla LF, et al. Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. *Cell Stem Cell*. 2013;12(3):354-367.
- 84.** Genovese P, Schirotti G, Escobar G, et al. Targeted genome editing in human repopulating haematopoietic stem cells. *Nature*. 2014;510(7504):235-240.
- 85.** Schwank G, Koo BK, Sasselli V, et al. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell*. 2013;13(6):653-658.
- 86.** Urnov FD, Miller JC, Lee YL, et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*. 2005;435(7042):646-651.
- 87.** Matsubara Y, Chiba T, Kashimada K, et al. Transcription activator-like effector nuclease-mediated transduction of exogenous gene into IL2RG locus. *Sci Rep*. 2014;4:5043.
- 88.** Yu J, Vodyanik MA, Smuga-Otto K, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007;318(5858):1917-1920.
- 89.** Takahashi K, Tanabe K, Ohnuki M, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*. 2007;131(5):861-872.
- 90.** Warlich E, Kuehle J, Cantz T, et al. Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming. *Mol Ther*. 2011;19(4):782-789.
- 91.** Zou J, Mali P, Huang X, Doweiy SN, Cheng L. Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood*. 2011;118(17):4599-4608.

- 92.** Sun N, Zhao H. Seamless correction of the sickle cell disease mutation of the HBB gene in human induced pluripotent stem cells using TALENs. *Biotechnol Bioeng.* 2014;111(5):1048-1053.
- 93.** Pessach IM, Ordovas-Montanes J, Zhang SY, et al. Induced pluripotent stem cells: a novel frontier in the study of human primary immunodeficiencies. *J Allergy Clin Immunol.* 2011;127(6):1400-1407 e1404.
- 94.** Park IH, Arora N, Huo H, et al. Disease-specific induced pluripotent stem cells. *Cell.* 2008;134(5):877-886.
- 95.** Kennedy M, Awong G, Sturgeon CM, et al. T lymphocyte potential marks the emergence of definitive hematopoietic progenitors in human pluripotent stem cell differentiation cultures. *Cell Rep.* 2012;2(6):1722-1735.
- 96.** Chang CW, Lai YS, Lamb LS, Jr., Townes TM. Broad T-cell receptor repertoire in T-lymphocytes derived from human induced pluripotent stem cells. *PLoS One.* 2014;9(5):e97335.
- 97.** Themeli M, Kloss CC, Ciriello G, et al. Generation of tumor-targeted human T lymphocytes from induced pluripotent stem cells for cancer therapy. *Nat Biotechnol.* 2013;31(10):928-933.
- 98.** Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science.* 2014;343(6166):80-84.
- 99.** Shalem O, Sanjana NE, Hartenian E, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science.* 2014;343(6166):84-87.
- 100.** Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera Mdel C, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol.* 2014;32(3):267-273.
- 101.** Zhou Y, Zhu S, Cai C, et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature.* 2014;509(7501):487-491.
- 102.** Casanova JL, Conley ME, Seligman SJ, Abel L, Notarangelo LD. Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies. *J Exp Med.* 2014;211(11):2137-2149.
- 103.** Waaijers S, Boxem M. Engineering the *Caenorhabditis elegans* genome with CRISPR/Cas9. *Methods.* 2014;68(3):381-388.
- 104.** Sugi T, Sakuma T, Ohtani Y, Yamamoto T. Versatile strategy for isolating transcription activator-like effector nuclease-mediated knockout mutants in *Caenorhabditis elegans*. *Dev Growth Differ.* 2014;56(1):78-85.
- 105.** Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat Methods.* 2013;10(8):741-743.
- 106.** Huang P, Xiao A, Zhou M, Zhu Z, Lin S, Zhang B. Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol.* 2011;29(8):699-700.
- 107.** Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol.* 2008;26(6):695-701.
- 108.** Auer TO, Duroure K, Concordet JP, Del Bene F. CRISPR/Cas9-mediated conversion of eGFP-into Gal4-transgenic lines in zebrafish. *Nat Protoc.* 2014;9(12):2823-2840.
- 109.** Gjini E, Mansour MR, Sander JD, et al. A Zebrafish Model of Myelodysplastic Syndrome Produced Through tet2 Genomic Editing. *Mol Cell Biol.* 2014.
- 110.** Meyer M, de Angelis MH, Wurst W, Kuhn R. Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. *Proc Natl Acad Sci U S A.* 2010;107(34):15022-15026.
- 111.** Carbery ID, Ji D, Harrington A, et al. Targeted genome modification in mice using zinc-finger nucleases. *Genetics.* 2010;186(2):451-459.
- 112.** Cui X, Ji D, Fisher DA, Wu Y, Briner DM, Weinstein EJ. Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nat Biotechnol.* 2011;29(1):64-67.
- 113.** Sung YH, Baek IJ, Kim DH, et al. Knockout mice created by TALEN-mediated gene targeting. *Nat Biotechnol.* 2013;31(1):23-24.
- 114.** Wefers B, Meyer M, Ortiz O, et al. Direct production of mouse disease models by embryo microinjection of TALENs and oligodeoxynucleotides. *Proc Natl Acad Sci U S A.* 2013;110(10):3782-3787.
- 115.** Wang H, Yang H, Shivalila CS, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell.* 2013;153(4):910-918.
- 116.** Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. One-step generation of mice

carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell*. 2013;154(6):1370-1379.

117. Shen B, Zhang J, Wu H, et al. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res*. 2013;23(5):720-723.

118. Yang H, Wang H, Jaenisch R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat Protoc*. 2014;9(8):1956-1968.

119. Li F, Cowley DO, Banner D, Holle E, Zhang L, Su L. Efficient genetic manipulation of the NOD-Rag1-/-IL2RgammaC-null mouse by combining in vitro fertilization and CRISPR/Cas9 technology. *Sci Rep*. 2014;4:5290.

120. Geurts AM, Cost GJ, Freyvert Y, et al. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science*. 2009;325(5939):433.

121. Tesson L, Usal C, Menoret S, et al. Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol*. 2011;29(8):695-696.

122. Mashimo T, Takizawa A, Voigt B, et al. Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. *PLoS One*. 2010;5(1):e8870.

123. Li W, Teng F, Li T, Zhou Q. Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat Biotechnol*. 2013;31(8):684-686.

124. Yan Q, Zhang Q, Yang H, et al. Generation of multi-gene knockout rabbits using the Cas9/gRNA system. *Cell Regen (Lond)*. 2014;3(1):12.

125. Wang Y, Fan N, Song J, et al. Generation of knockout rabbits using transcription activator-like effector nucleases. *Cell Regen (Lond)*. 2014;3(1):3.

126. Flisikowska T, Thorey IS, Offner S, et al. Efficient immunoglobulin gene disruption and targeted replacement in rabbit using zinc finger nucleases. *PLoS One*. 2011;6(6):e21045.

127. Song J, Zhong J, Guo X, et al. Generation of RAG 1- and 2-deficient rabbits by embryo microinjection of TALENs. *Cell Res*. 2013;23(8):1059-1062.

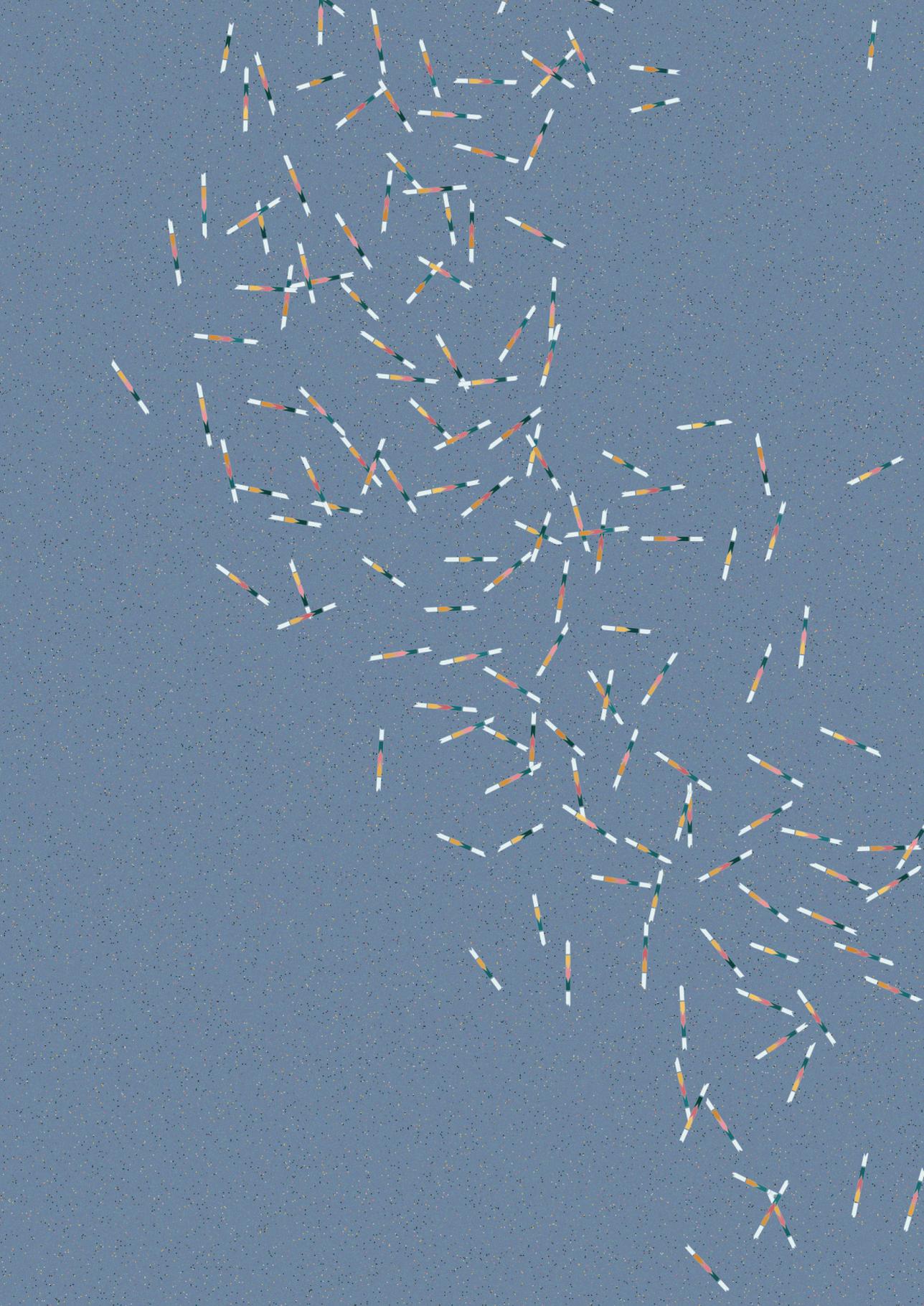
128. Fan Z, Li W, Lee SR, et al. Efficient gene targeting in golden Syrian hamsters by the CRISPR/Cas9 system. *PLoS One*. 2014;9(10):e109755.

129. Kurome M, Dahlhoff M, Bultmann S, et al. 361 growth hormone receptor mutant pigs produced by using the clustered regularly interspaced short palindromic repeats (crispr) and crispr-associated systems in in vitro-produced zygotes. *Reprod Fertil Dev*. 2014;27(1):269.

130. Zhou X, Xin J, Fan N, et al. Generation of CRISPR/Cas9-mediated gene-targeted pigs via somatic cell nuclear transfer. *Cell Mol Life Sci*. 2014.

131. Hai T, Teng F, Guo R, Li W, Zhou Q. One-step generation of knockout pigs by zygote injection of CRISPR/Cas system. *Cell Res*. 2014;24(3):372-375.

132. Niu Y, Shen B, Cui Y, et al. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell*. 2014;156(4):836-843.



3 Modeling altered T-cell development with human induced pluripotent stem cells from patients with *RAG1* mutations and distinct immunological phenotypes

Blood. 2016 Aug 11;128(6):783-93.



Patrick M. Brauer¹

Itai M. Pessach^{2,3}

Erik Clarke⁴

Jared H. Rowe²

Lisa Ott de Bruin²

Yu Nee Lee²

Carmen Dominguez-Brauer⁵

Anne M. Comeau⁶

Geneve Awong^{1,7}

Kerstin Felgentreff²

Yuhang H. Zhang⁸

Andrea Bredemeyer⁹

Waleed Al-Herz¹⁰

Likun Du²

Francesca Ververs²

Marion Kennedy⁷

Silvia Giliani¹¹

Gordon Keller⁷

Barry P. Sleckman⁹

David G. Schatz^{8,12}

Frederic D. Bushman⁴

Luigi D. Notarangelo^{2,13}

Juan Carlos Zúñiga-Pflücker¹

¹ Sunnybrook Research Institute, and Dept. of Immunology, University of Toronto, Toronto, ON, Canada

² Division of Immunology, Boston Children's Hospital, Boston, MA

³ Dept. of Pediatric Critical Care, Sheba Medical Center, Tel-Aviv University, Tel-Aviv, Israel

⁴ Dept. of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

⁵ Ontario Cancer Institute, University Health Network, Toronto, ON, Canada

⁶ Dept. of Pediatrics, University of Massachusetts Medical School, Worcester, MA

⁷ McEwen Centre for Regenerative Medicine, University Health Network, Toronto, ON, Canada

⁸ Dept. of Immunobiology and Molecular Biophysics and Biochemistry, Yale School of Medicine, New Haven, CT

⁹ Dept. of Pathology and Immunology, Washington University in St. Louis, St. Louis, MO

¹⁰ Dept. of Pediatrics, Faculty of Medicine, Kuwait University, Kuwait City, Kuwait

¹¹ Angelo Nocivelli Institute for Molecular Medicine, University of Brescia, Brescia, Italy

¹² Howard Hughes Medical Institute, New Haven, CT

¹³ Harvard Stem Cell Institute, Harvard University, Cambridge, MA

The online version and supplemental data can be found at:

<https://doi.org/10.1182/blood-2015-10-676304>.

Abstract

Primary immunodeficiency diseases comprise a group of heterogeneous genetic defects that affect immune system development and/or function. Here we use *in vitro* differentiation of human induced pluripotent stem cells (iPSCs) generated from patients with different recombination-activating gene RAG1 (*RAG1*) mutations to assess T-cell development and T-cell receptor (TCR) V(D)J recombination. RAG1-mutants from severe combined immunodeficient (SCID) patient cells showed a failure to sustain progression beyond the CD3⁻CD4⁻CD8⁻CD7⁺CD5⁺CD38⁻CD31^{-/lo}CD45RA⁺ stage of T-cell development to reach the CD3^{-/+}CD4⁺CD8⁺CD7⁺CD5⁺CD38⁺CD31⁺CD45RA⁻ stage. Despite residual mutant RAG1 recombination activity from an Omenn syndrome (OS) patient, similar impaired T-cell differentiation was observed, due to increased single-strand DNA breaks that likely occur due to heterodimers consisting of both an N-terminal truncated and a catalytically dead RAG1. Furthermore, deep-sequencing analysis of TCR- β (TRB) and TCR- α (TRA) rearrangements of CD3⁻CD4⁺CD8⁻ immature single-positive and CD3⁺CD4⁺CD8⁺ double-positive cells showed severe restriction of repertoire diversity with preferential usage of few Variable, Diversity, and Joining genes, and skewed length distribution of the TRB and TRA complementary determining region 3 sequences from SCID and OS iPSC-derived cells, whereas control iPSCs yielded T-cell progenitors with a broadly diversified repertoire. Finally, no TRA/ δ excision circles (TRECs), a marker of TRA/ δ locus rearrangements, were detected in SCID and OS-derived T-lineage cells, consistent with a pre-TCR block in T-cell development. This study compares human T-cell development of SCID vs. OS patients, and elucidates important differences that help to explain the wide range of immunologic phenotypes that result from different mutations within the same gene of various patients.

Introduction

Primarily immunodeficiency diseases (PIDs) include over 200 distinct disorders that variably influence immune system development and function, thereby causing increased susceptibility to infections, autoimmunity, and malignancies.¹ In particular, severe combined immunodeficiency (SCID) disorders are characterized by a block in T-cell development, variably associated with impaired B- and/or natural killer (NK) cell differentiation. Mechanisms that account for SCID include impaired cell survival (adenosine deaminase deficiency, reticular dysgenesis), defective response to cytokines (*IL7R*, *IL2RG*, and *JAK3* gene defects), impaired assembly of the antigen receptor (recombination-activating gene 1/2 (*RAG1/RAG2*), *DCLRE1C*, *PRKDC*, and *LIG4* defects), and defective signaling through the pre-T-cell receptor (TCR) and/or the TCR (defects of *CD3 δ* , *CD3 ϵ* , *CD3 ξ* , *ZAP70*, and *CD45*). Severe mutations of these genes block T-cell development prior to differentiation of single-positive thymocytes.

The *RAG1* and *RAG2* proteins initiate the process of Variable (V), Diversity (D), and Joining (J) gene recombination by recognizing recombination signal sequences (RSS) that flank the V-D-J gene elements within the TCR and immunoglobulin (Ig) gene loci.² RAGs first nick a single DNA strand, which allows them to introduce DNA double-strand breaks (DNA-DSBs) that are initially covalently sealed by a hairpin (coding ends [CEs]).³ Subsequently, the DNA-PK catalytic subunit (DNA-PKcs) protein activates DNA cross-link repair 1C (*DCLRE1C*; also known as Artemis), allowing opening of the hairpin. The DNA-DSBs are then repaired by proteins of the nonhomologous end-joining pathway, thereby permitting the juxtaposition of non-adjacent V-D-J genes.⁴ *RAG* mutations in humans are associated with a variety of clinical and immunologic phenotypes that reflect the biochemical consequences of the mutation and the effect of environmental factors.⁵ In patients with null mutations, complete failure of V(D)J recombination is associated with complete lack of circulating T and B lymphocytes, hence resulting in the T⁻ B⁻ NK⁺ form of SCID. We and others have shown that hypomorphic mutations that affect, but do not abolish, V(D)J recombination, are often associated with distinct immunologic and clinical phenotypes with residual presence of T, and in some cases B, lymphocytes.⁶⁻⁹ The presence of autologous, auto-reactive, activated, and oligoclonal T-cells that infiltrate and damage peripheral organs is a hallmark of Omenn syndrome (OS). In other cases, hypomorphic mutations may cause delayed disease onset, granuloma formation, autoimmunity, and/or dysgammaglobulinemia.⁵

Using an *in vitro* cellular platform in which RAG activity can be measured by analyzing recombination at an inverted green fluorescent protein (GFP) cassette flanked by RSS, we have shown that the phenotypic diversity of

human RAG deficiency correlates with the residual function of the mutant RAG protein.¹⁰ We found that mutations associated with OS have residual, yet markedly decreased, recombination activity. The observation that OS and T⁻B⁻NK⁺SCID may occur in affected members of the same family suggests that RAG mutations associated with these phenotypes can only support, at best, limited repertoire diversity. However, no studies have compared T-cell development in patients with RAG mutations associated with OS vs SCID.

Mouse models have been used to elucidate the functions of genes involved in PID, and SCID in particular. A mouse model for SCID was first reported by Bosma et al,¹¹ the result of a naturally occurring mutation in the *Prkdc* gene.¹² Although the *scid* mouse is initially deficient in functional T and B-cells, some young adult mice generate a low number of functional lymphocytes, and a leaky SCID phenotype is observed in most *scid* mice by 1 year.¹³ In contrast, the *Rag1* or *Rag2* null mice result in a non-leaky SCID, with a stringent block at the CD4⁻CD8⁻CD44⁻CD25⁺ double negative 3 stage of intrathymic T-cell development, resulting in absence of B or T lymphocytes.^{14,15} Mouse models of OS and of leaky SCID have been generated, such as Rag1 R972Q¹⁶, the Rag1 S723C⁷ and Rag2 R229Q¹⁷ mice. In addition to the mouse, SCID and SCID variants have also been modeled in the dog and horse.^{18,19}

Although animal models serve as an important tool for elucidating gene functions, and for determining how certain mutations result in PIDs, there is a clear need to study PIDs in a human context. There are differences in T-lymphocyte development between humans and mice,²⁰ and disease mechanisms likely differ as well. However, several obstacles exist that make it difficult to study the developmental pathophysiology of human SCID at the cellular and molecular level, including rarity of the disease, the urgency of treatment, and difficulties in obtaining appropriate tissue samples.

Recent work has demonstrated that T-cells can be generated from human induced pluripotent stem cells (iPSCs) *in vitro*.²¹⁻²³ This *in vitro* approach can reduce the need for using animal models in place of a more ethical, rapid, and more cost-effective means to conduct research within a human context, validating treatment or the repair of a patient's defective gene in the context of thymocyte differentiation. A first report that defective T-cell differentiation associated with SCID can be modeled using patient-derived iPSCs has been provided by demonstrating an early arrest of T-cell development of cells carrying an *IL2RG* mutation, responsible for X-linked SCID, and rescue of T-cell differentiation by means of transcription activator-like effector nucleases-mediated gene editing.²⁴ Although recent work in the field adds to the understanding of PIDs in a human context, these studies were not aimed at elucidating the causality of the broad spectrum of clinical and immunologic

phenotypes resulting from different mutations within the same gene. We reprogrammed dermal fibroblasts to generate iPSCs from patients with *RAG1* mutations that resulted in different clinical and immunologic phenotypes (SCID and OS). When control and patient-derived iPSCs were differentiated toward the T-cell lineage *in vitro*, a primary block in T-cell development at an early stage (CD7⁺CD5⁻ or CD7⁺CD5⁺ stage) was observed both in OS and in SCID cells, respectively. In both instances, developmental progression was briefly observed up to the CD4⁺CD8⁻ immature single-positive (ISP) and CD4⁺CD8⁺ double-positive (DP) stage, but was not sustained, which correlated with the lack of intracellular TCR- β (TRB) for SCID and OS cells, and increased detection of single-strand DNA breaks unique to OS cells. Additionally, assessment of TRB and TCR- α (TRA) recombination within the SCID and OS-derived ISP and DP populations showed limited repertoire diversity, with skewed usage of individual V-D-J genes, abnormal distribution, and length of the TRB and TRA complementary determining region3 (CDR3), and an absence of TRA/ δ excision circles (TRECs). Overall, these results confirm that *RAG* mutations associated with OS and SCID are similarly severe, with key differences unique to OS cells that help to explain the broad phenotypic diversity of *RAG1*-associated PIDs, and demonstrate the applicability of using iPSCs to model PID and to study mechanisms of altered T-cell development in humans.

Methods

Patients

Three patients with bi-allelic *RAG1* mutations were included in the study. Patient 1 (P1) was born to first cousin parents, and suffered from recurrent lower respiratory tract infections since the first month of life. Laboratory tests performed at 5 months of age demonstrated severe lymphopenia (140 cells/*mL*). Lymphocyte subsets were as follows: CD3: 1.4%; CD4: 0.3%; CD8: 0.1%; CD19: 1.1%; and CD16: 74.2%. Serum IgA (1 mg/dL) and IgM (2 mg/dL) were markedly decreased, whereas IgG serum levels (385 mg/dL) likely reflected maternally transferred IgG. Molecular analysis revealed a homozygous *RAG1* mutation (c.1428delC), predicted to result in frameshift and premature termination (p.N476Kfs*16).

P2 was a Kuwaiti infant born to consanguineous (first cousin) parents. She presented with diarrhea, growth arrest, recurrent respiratory tract infections, and oral candidiasis since the first months of life. At 11 months of age, the absolute lymphocyte count was 1848 cells/*mL*, with abnormal distribution of lymphocyte subsets (CD3: 15%; CD4: 14%; CD8: 0%; CD19: 0.5%; and CD16: 79%). *In vitro* proliferation to phytohemagglutinin was markedly decreased (2700 cpm vs 166 000 in a healthy control). Molecular analysis revealed a homozygous *RAG1*

mutation (c.1180C>T), resulting in the R394W amino acid substitution.

P3, of Slavic origin, was born to nonconsanguineous parents. An older sibling had been diagnosed with OS and had died of complications after hematopoietic stem cell transplantation. The patient developed generalized skin erythema, lymphadenopathy, and hepatomegaly at 2 months of age. Laboratory tests demonstrated leukocytosis (white blood cells, 46 100/mL), lymphocytosis (18 300 cells/mL), and eosinophilia (7039 cells/mL). Lymphocyte subsets were as follows: CD3: 91.9% and CD4: 62.4% (98.8% of which were CD45R0⁺); CD8: 27.6%; CD19: 0.2%; and CD16: 3.9%. Hypogammaglobulinemia (IgG: 245 mg/dL) with markedly elevated IgE (.1000 kU/L) were demonstrated. In vitro proliferation to phytohemagglutinin was nearly abolished (2000 cpm vs 172 000 in a healthy control). Maternal T-cell engraftment was ruled out by microsatellite analysis. Molecular analysis demonstrated compound heterozygous *RAG1* mutations (c.256_257del; c.2164G.A), predicted to cause the K86Vfs*33 and E722K protein changes, respectively.

On the basis of the Primary Immune Deficiency Treatment Consortium criteria²⁵, P1 and P2 met the definition of SCID, whereas P3 was diagnosed with OS.

Online materials

See supplemental Methods, available on the Blood web site. Online material includes methods for: (1) analysis of recombination activity of the mutant RAG1 proteins; (2) generation and characterization of RAG1-deficient iPSCs; (3) iPSC maintenance and differentiation; 4) CD34⁺ isolation; (5) OP9-DL-4 coculture and differentiation; (6) flow cytometry; (7) next-generation sequencing of TRB and TRA repertoires; (8) Quantification of TRECs; (9) single cell gel electrophoresis (Comet assay); (10) Southern blot analysis; (11) protein expression and purification for *in vitro* analysis; (12) *in vitro* DNA nicking assays; (13) *in vitro* binding assays; and (14) statistical analysis.

Results

Low recombination activity and cleavage defects

In order to investigate the effects of RAG1 deficiency on human T-cell development, we generated iPSCs from dermal fibroblasts obtained from 2 patients with SCID and 1 patient with OS. P1 and P2 with SCID were both homozygous for distinct *RAG1* mutations, whereas compound heterozygosity was demonstrated in P3 with OS (**Fig. 1**). The SCID P1 single nucleotide deletion at residue 1428 is predicted to result in frameshift and early truncation of the RAG1 protein with only the nonamer-binding region (NBR) domain of the catalytic core intact. The SCID P2 and OS P3 substitution

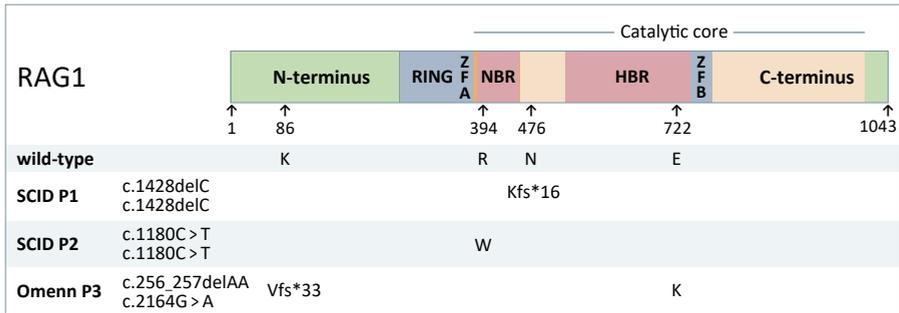


Figure 1: Scheme of RAG1 protein and mutations. WT human RAG1 consists of 1043 amino acids and includes an N-terminus domain, really interesting new gene (RING) finger sequence, zinc finger sequences zinc finger A (ZFA) and zinc finger B (ZFB), the catalytic core, which contains the nonamer- and heptamer-binding regions, and a C-terminus domain. The amino acid positions affected by mutations identified in patients with SCID (P1 and P2) and with OS (P3), and the respective consequences on amino acid sequence are shown. P1 and P2 were homozygous for a frameshift (N476Kfs*16) and a missense (R394W) mutation, respectively. P3 was compound heterozygous for a missense (E722K) and a frameshift (K86Vfs*33). For the latter, an alternative start codon can be used resulting in an N-terminus truncated protein with normal sequence from Met183 onward with cytoplasmic localization. HBR, heptamer-binding region; NBR, nonamer-binding region.

mutations within the NBR or heptamer-binding region respectively, are predicted to severely limit recognition and/or binding of the RSS. The OS P3 deletion (c.256_257delAA) was previously reported to produce an N-terminal truncated RAG1 protein with an intact catalytic domain, due to use of an alternative start codon; this mutant is severely blocked from entering the nucleus.²⁶ To characterize the levels of recombination activity of each of these *RAG1* mutants, Abelson-transformed *Rag1*^{-/-} pro-B-cells containing an inverted GFP cassette flanked by RSS (pMX-INV cassette) were transduced with retroviral vectors encoding for the various RAG1 mutant proteins or wild-type (WT) RAG1 and GFP expression was used as a read-out of the recombination activity.¹⁰ Total cell lysates were obtained for immunoblot analysis in order to assess RAG1 expression levels (**Fig. 2A**, p. 44). The antibody used (D36B3; Cell Signaling) is N-terminus specific and therefore does not recognize the product of the c.256_257delAA mutation.

As shown in **Fig. 2B-C**, p. 44, the RAG1 mutations associated with SCID supported extremely low recombination activity when compared with WT RAG1 (P1, c.1428delC: 0.10 ± 0.05%; P2, c.1180C.T: 0.10 ± 0.04%). One of the 2 mutant alleles associated with OS in P3 (c.2164G.A) had virtually undetectable recombination activity (0.04% ± 0.03), whereas the other allele (256_257delAA) had residual, though markedly decreased, function (3.48 ± 0.35%). These data are consistent with the previous demonstration that the 256_257delAA frameshift mutation results in a prematurely truncated, catalytically inactive RAG1 mutant protein, but use of an alternative

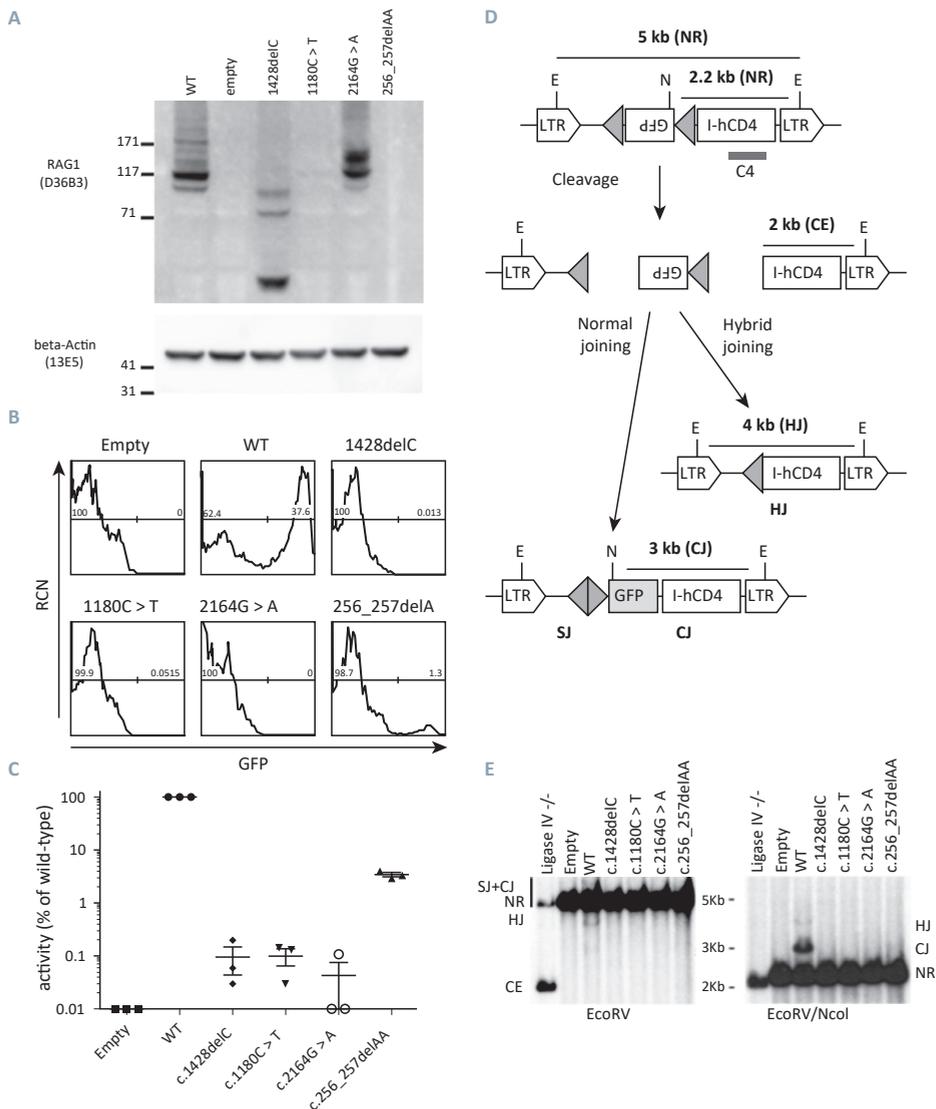


Figure 2: Recombination activity of mutant RAG1. [A] Immunoblot analysis of Rag12/2 pro-B Abelson cells expressing WT human RAG1, empty vector, or hypomorphic RAG1-mutant alleles from P1 (1428delC), P2 (1180C.T), and P3 (256_257delAA and 2164G.A). N-terminus-specific anti-RAG1 antibody (D36B3) cannot detect the N-terminus-truncated 256_257delAA. b-Actin serves as a loading control, and molecular weight markers (in kDa) are marked. [B] Flow cytometric analysis of Rag12/2 pro-B Abelson cells containing a pMX-INV cassette with an inverted GFP sequence flanked by RSS (shown as gray triangles in [D]), and transduced with retroviral vectors expressing WT, empty, or hypomorphic RAG1 mutants. GFP1 events indicate rearrangement. [C] Recombination activity in transduced Rag12/2 pro-B-cells, expressed as a percentage of activity detected in cells expressing WT RAG1. For each construct, 3 independent experiments were performed, and mean value \pm 6 standard deviation is shown. [D] Schematic representation showing nonrearranged (NR) products flanked by Escherichia coli RV (EcoRV) (E) sites (5 kb), or EcoRV (E) and Nocardia corallina I (NcoI) (N) sites (2.2 kb). Recombination results in

downstream translation start site permits expression of an N-terminally truncated RAG1 protein with low recombination activity.²⁶

RAG1 catalyzes recombination of the V-D-J genes through cleaving and rejoining of the TCR locus. To establish whether the RAG1 mutations identified in patients lead to unresolved, unjoined DNA breaks, we analyzed the Abelson pro-B-cells expressing mutant RAG1 (**Fig. 2D-E**). As expected, Southern-blot analysis of the pro-B Abelson cells reconstituted with WT RAG1 revealed a band of the expected size for the rearranged pMX-INV cassette (**Fig. 2E**). Conversely, the RAG1 mutations did not support the generation of detectable levels of pMX-INV rearrangement, with the exception of the c.256_257delAA mutant, for which a very faint band corresponding to rearranged pMX-INV was observed (**Fig. 2E**), reflecting very low levels of recombination observed in OS patients. These results are consistent with the residual recombination activity of this mutant demonstrated by flow cytometry analysis of GFP expression (**Fig. 2B-C**). Additionally, none of the RAG1-reconstituted samples showed accumulation of unrepaired CEs, which are easily detected in a ligase IV^{-/-} cell line, thus confirming that the RAG1 mutants have defects in DNA cleavage.

Generation of CD34⁺ hematopoietic progenitor cells from RAG1-mutant iPSCs

In order to study the developmental block in T-cell development associated with different RAG1 mutants, we took advantage of iPSCs that we had previously generated, starting from dermal fibroblasts of SCID P2 and OS P3²⁷, as well as of a control iPSC line that had been derived from a healthy control.²⁸ To generate iPSCs from SCID P1, we transduced fibroblasts with a lentiviral vector expressing the 4 Yamanaka factors (OCT4, SOX2, KLF4, and c-MYC). The resulting iPSC colonies were manually picked and expanded. Mutation analysis at the *RAG1* locus confirmed their patient-specific origin (**Fig. 3A**, p. 46). Immunohistochemistry (**Fig. 3B**) and real-time polymerase chain

[Fig. 2 cont.]

normal joining, producing a 3 kb fragment (coding joining or CJ) flanked by N and E sites, or hybrid joining (HJ) flanked by V sites and producing a 4 kb fragment. Successful cleavage but unresolved CEs produce a 2 kb fragment. C4 represents the location of the probe to visualize the various fragments for Southern blotting. [E] Southern blot showing products of rearrangement of the pMX-INV vector in pro-B Abelson cells expressing ligase IV null, empty vector, WT RAG1, or one of the four hypomorphic RAG1 mutants. Hybridization with the C4 probe with EcoRV digest reveals NR, HJ, or CE events, and EcoRV1NcoI digest reveals HJ, CJ, or NR events. LTR, long terminal repeat; RCN, relative cell number; SJ, signal joining.

reaction (PCR) (**Fig. 3C**) demonstrated expression of pluripotency and stemness markers, as well as silencing of the KLF4 and c-MYC transgenes. Finally, cytogenetic analysis demonstrated a lack of karyotypic abnormalities (**Fig. 3D**).

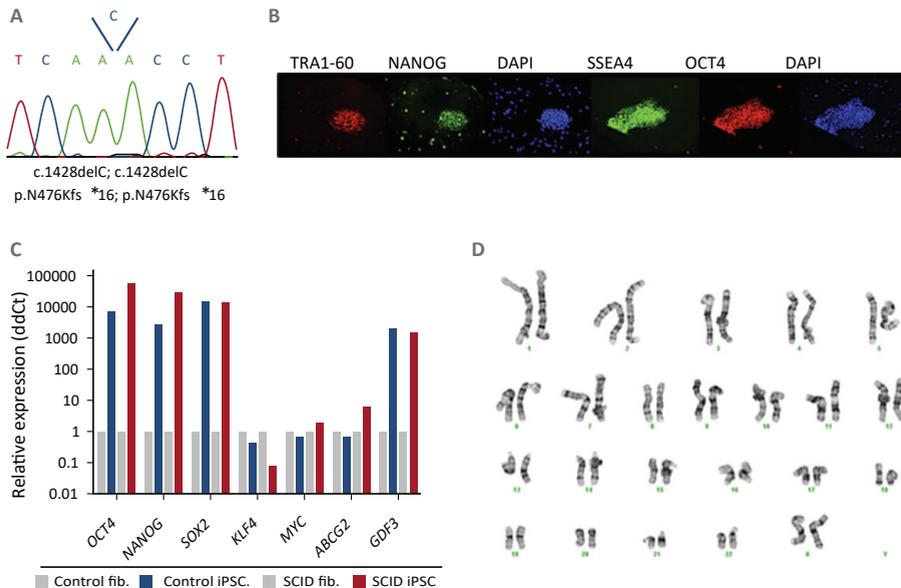


Figure 3: Characterization of an iPSC line reprogrammed from fibroblasts from P1.

[A] Sequencing of genomic DNA from the iPSC line, revealing homozygosity for the same single nucleotide deletion at position 1428 identified in parental fibroblasts, and resulting in the p.N476Kfs*16 mutation. [B] Immunofluorescent staining and [C] quantitative PCR analysis of pluripotency markers expressed by the iPSCs. [D] Karyotypic integrity of P1 iPSC line.

To generate CD34⁺ hematopoietic progenitor cells, control- and *RAG1*-mutated iPSCs were allowed to differentiate into embryoid bodies,²¹ and CD34⁺ progenitor cells were isolated by magnetic-activated cell separation after 8 to 13 days (**Fig. S1**, online, p. 38).

T-cell development analysis from control- and patient-derived iPSC lines

To analyze T-cell developmental progression of control and *RAG1*-mutated iPSCs, embryoid body-derived CD34⁺ cells were transferred onto OP9-DL-4 cells in the presence of appropriate cytokines and growth factors to induce T-lineage differentiation. Flow cytometric analyses after 3 to 4 weeks of coculture showed that control, SCID, and OS-derived cells had gained CD7 expression, indicative of T-lineage differentiation (**Fig. 4A**). Additionally, nearly all the cells had downregulated CD34 expression, whereas a majority of the cells also expressed CD5 and CD38. However, a larger proportion of SCID and OS cells were CD7⁺ CD5⁻ or CD7⁺ CD38⁻ as compared with control cells.

Further analysis showed that a larger proportion of *RAG1*-mutated cells remained CD45RA⁺ as compared with what was observed in control cells. Although both control and *RAG1*-mutated cells reached a CD4⁺ stage of



Figure 4: In vitro T-lineage differentiation of control and SCID iPSC lines. Flow cytometric analysis of T-lineage developmental progression of control- and patient-derived cells. iPSCs were allowed to differentiate for 8 days into embryoid bodies, and magnetic bead-purified CD34⁺ cells were cocultured with OP9-DL-4 cells. **[A]** Cells from P1 and P2 with SCID, and from P3 with OS attained normal expression of early markers of T-lineage differentiation (CD7, CD5, and CD38) upon 3 to 4 weeks of coculture with OP9-DL-4 cells. **[B]** After 4 to 5 weeks of coculture, cells from a healthy control progress to the CD4⁺ CD8αβ⁺ DP stage of differentiation, with the appearance of CD3⁺ TRA/TRB⁺ cells. By contrast, SCID- and OS-derived cells were mostly blocked at the CD7⁺ CD31^{+/+} CD45RA⁺ stage of differentiation, with a virtual absence of CD4 and CD8a/b expression, and lack of CD3⁺ cells. In (A-B), cells were pre-gated for lymphocytes (SSCxFSC), DAPI-, and CD45⁺. DAPI, 4',6-diamidino-2-phenylindole; FSC, forward scatter; SSC, side scatter.

differentiation, a larger proportion of control cells expressed CD4, with nearly half also expressing CD8a and CD8b. In contrast, fewer CD4⁺ cells were obtained in culture of SCID P1 and OS cells, with only ~10% to 30% also expressing CD8 α and CD8 β (**Fig. 4A**, p. 47). At this time point, neither control nor RAG1-mutated cells showed CD3 expression at the cell surface.

Within 4 to 5 weeks of coculture, as expected, both control and RAG1-mutated cells expressed CD7 and CD5 (**Fig. 4B**, p. 47). Control cells showed a sustained developmental progression, as illustrated by expression of CD38 and by reaching the CD31⁺ CD45RA⁻ stage of T-cell differentiation, which also includes expression of CD4 and CD8 α/β . By contrast, SCID- or OS-derived cells did not co-express CD38 and CD7, and were mostly blocked at the CD31^{-/+} CD45RA⁺ stage of differentiation, with a virtual absence of CD4 and CD8 α/β expression. Cell surface expression of CD3 was observed in control cells, but not in SCID- or OS-derived cells. Additionally, intracellular staining for TRB expression showed that 31% of control cells had a productively rearranged TCR, whereas virtually no TCR expression could be detected in SCID or OS cells (**Fig. S2**, online, p. 38). These results suggest that although an initial progression to a CD45RA⁻ CD31⁺ CD4⁺ CD8 α/β stage of differentiation can be observed in RAG1-mutated CD34⁺ cells cultured on OP9-DL-4 cells, this is not sustained at later time points, with the cells showing the more expected block in T-cell differentiation. Similar results were obtained in at least 3 independent experiments for each iPSC line (**Fig. S3A-K**, online, p. 38). To assess the extent of apoptosis/cell death, cells were stained with Annexin V and SYTOX Green (**Fig. S3L** online, p. 38), and showed a similar amount of Annexin V-positive staining for each cell line.

OS cells have a higher propensity for DNA breaks

In order to investigate the potential cause for the lack of sustained developmental progression observed in OS cells, an alkaline single cell electrophoresis (Comet assay) was performed,^{6,29} that would allow to detect DNA nicks or breaks, which relax and open regions of the affected supercoiled DNA resulting in trailing tails (resembling comets) of DNA. Comets were consistently observed in OS cells sorted from coculture day 25 (immediately prior to the block observed for SCID or OS-derived cells), whereas these were absent in equivalent control- or SCID-derived cells (**Fig. 5A-B**). To determine whether SCID- or OS-associated mutations within the catalytic core of RAG1 allow the production of mutant protein capable of DNA nicking, we performed an *in vitro* oligonucleotide cleavage assay (**Fig. S4A**, online, p. 38). No nicking activity was observed for either R394W (SCID) or E722K (OS) mutants (**Fig. S4B**, online, p. 38). The p.N476Kfs*16 mutant of SCID P1 lacks the catalytic core, and therefore it was not possible (nor necessary) to assay its

activity. By contrast, the c.256_257delAA mutant allele of OS P3 is predicted to produce an N-terminus truncated protein (due to usage of a downstream inframe translation initiation site), with an intact catalytic domain.²⁶ Consistent with this, this mutant retained nicking activity, as revealed by the Comet assay (Fig. 5A-B).

V(D)J gene recombination analysis of early T-cells from OS- and SCID-derived iPSC lines

In spite of severe lymphoid depletion in the thymus,³⁰ patients with OS have a variable number of circulating T-cells, which however, lack a diverse TCR

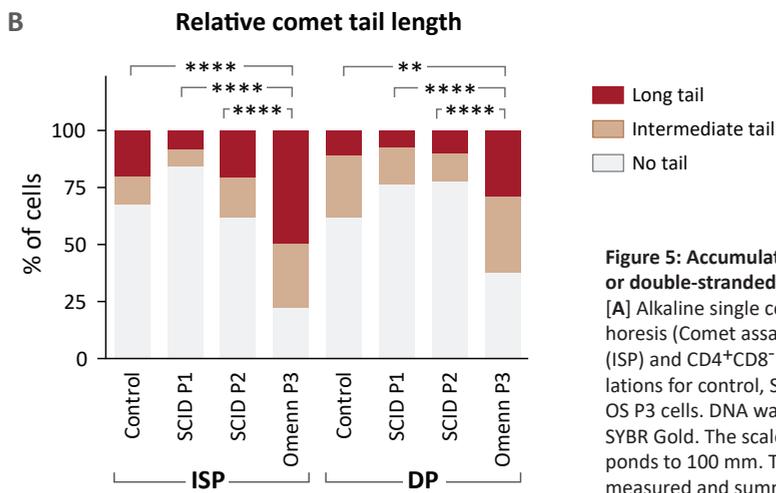
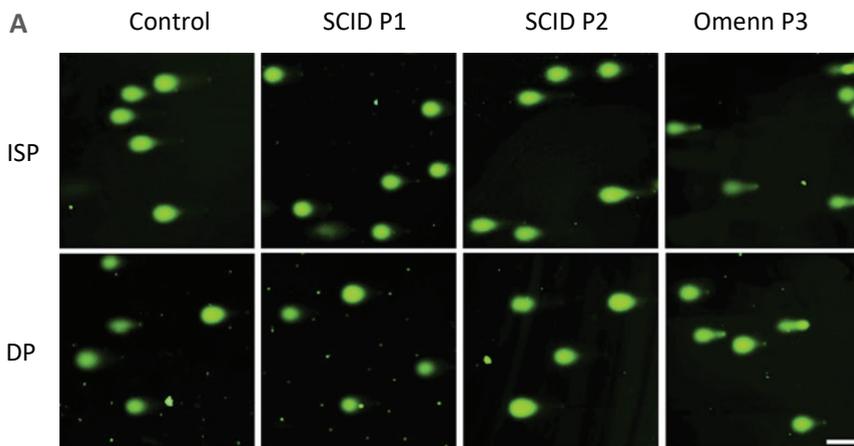


Figure 5: Accumulation of single- or double-stranded DNA breaks. [A] Alkaline single cell gel electrophoresis (Comet assay) of CD4⁺CD8⁻ (ISP) and CD4⁺CD8⁺ (DP) subpopulations for control, SCID P1, P2, and OS P3 cells. DNA was visualized by SYBR Gold. The scale bar corresponds to 100 μm. Tail lengths were measured and summarized in [B]. ****p, .0001; **p, .01.

repertoire. To determine to what extent restriction of the TCR repertoire is already present at early stages of T-cell development in patients with OS, deep sequencing analysis of the TRB and TRA V(D)J rearrangements was performed on genomic DNA extracted from sorted CD4⁺ ISP and CD4⁺ CD8⁺ DP cells generated *in vitro* from control-, OS-, and SCID-derived iPSCs (**Table 1**). Due to low cell numbers, the SCID P1 DP population was not analyzed for TRA rearrangement. A reduction in the proportion of unique and total TRB and TRA rearranged sequences was evident in OS and SCID samples as compared with control (**Fig. S5A**, online, p. 38). Broad diversity of *TRBV* and *TRAV* gene

Cell line	Population	Total	Unique	Productive		Clonality
				total	unique	
Control	CD4 ⁺ CD8 ⁻	451 195	3 300	428 110	2512	0.52
		603 206	6 949	478 446	2875	0.52
Control	CD4 ⁺ CD8 ⁺	734 020	8 775	666 299	6930	0.30
		900 284	14 151	724 386	8414	0.28
SCID P1	CD4 ⁺ CD8 ⁻	1 026 133	366	1 021 305	219	0.60
Omenn P3	CD4 ⁺ CD8 ⁻	132 386	120	131 048	76	0.49
Omenn P3	CD4 ⁺ CD8 ⁺	784 927	301	782 376	217	0.65

Table 1. Summary of deep sequencing analysis of the TRB V(D)J rearrangements

usage and of V-J pairing was demonstrated among both unique (**Fig. 6A-B**) and total (**Fig. S5B-C**) sequences from sorted ISP and DP cells derived from control iPSCs, whereas a dramatic restriction of TRB and TRA repertoires was observed in equivalent OS and SCID cell subsets (**Fig. 6A-B; Fig. S5B-C**, online, p. 38). To further analyze the ability of control- and RAG1-mutated cells to accomplish rearrangements at the TRA/locus, we quantified the levels of TRECs in sorted DP cells. TRECs were readily detectable in control, but not in SCID or OS cells (**Fig. 6C**). Virtual spectratyping analysis of the distribution of TRB CDR3 length among unique sequences showed a normal distribution for ISP and DP cells derived from control iPSCs, whereas OS and SCID cells were characterized by a skewed distribution of CDR3 lengths, with presence of longer sequences (**Fig. 6D**, top panel, p. 52). Similarly, a skewed distribution of CDR3 lengths was also observed when analyzing the most abundant V-J TRB rearrangements (**Fig. 6E**, p. 52).

Interestingly, TRA CDR3 length distribution revealed two peaks in SCID and OS cells, suggesting that some clones have normal CDR3 length whereas others exhibit an abnormally reduced length (**Fig. 6D**, lower panel, p. 52). Indeed, when analyzing some of the most abundant V-J TRA rearrangements, both

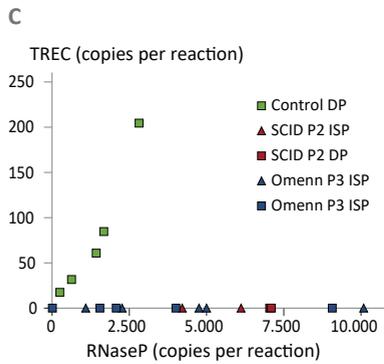
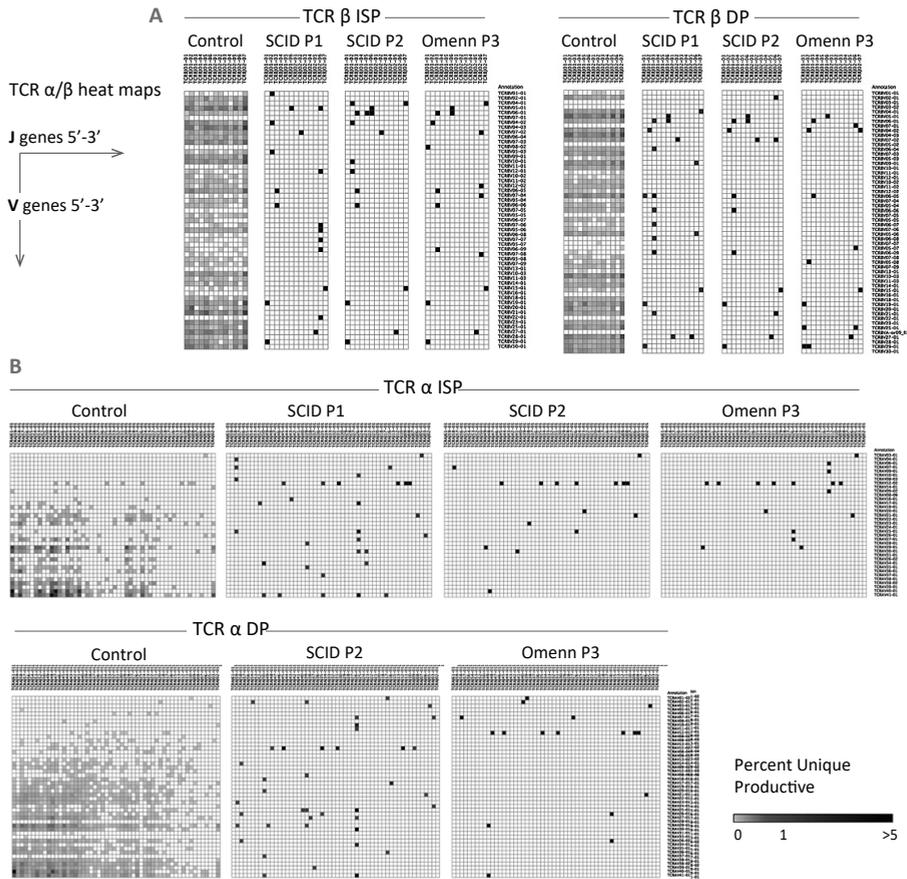
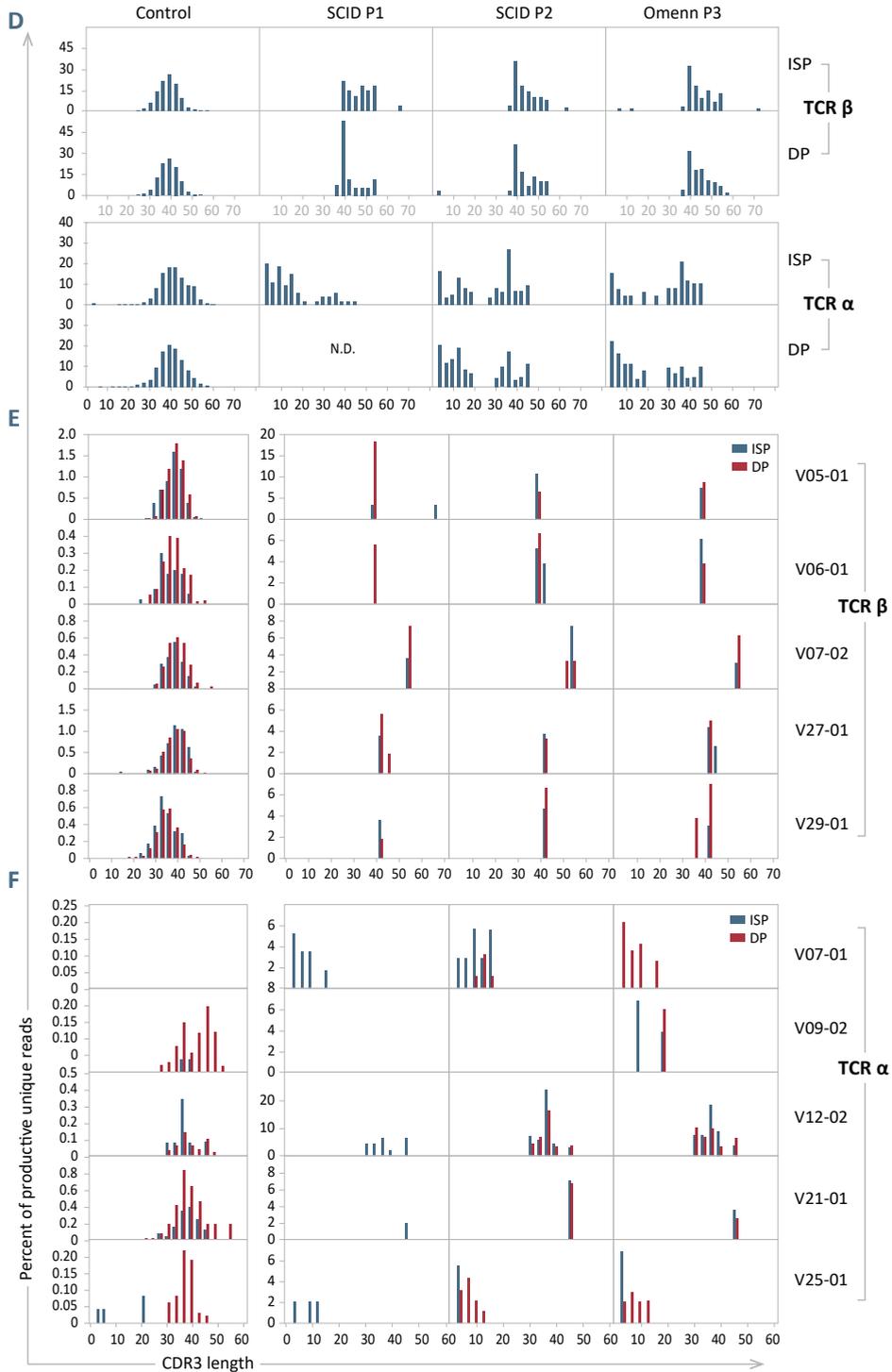


Figure 6: Next generation sequencing analysis of TCR repertoire upon *in vitro* T-lineage differentiation of control, SCID (P1 and P2), and OS (P3)-derived iPSCs. [A] Heat map representation of percentage of TRB VJ (orientated via chromosomal 59 to 39 distribution) pairings among unique sequences in ISP (left) and DP (right) T-lineage cells derived from the indicated patient iP lines. Results demonstrate 1 representative sample from 2 experiments with similar results. [B] Heat map representation of percentage of TRA VJ (orientated via chromosomal 59 to 39 distribution) pairings among total sequences (all d rearrangements excluded from analysis) in ISP (left) and DP (right) T-lineage cells derived from the indicated patients' iP lines among unique sequences. Results demonstrate 1

representative sample from 2 experiments with similar results. [C] Quantitative PCR analysis of TRECs in control, SCID P2, and OS P3 cells. RNase P was used as an internal control for quality of genomic DNA amplification. **Continued on next page:** [D] Virtual spectratyping, showing skewing in the distribution of CDR3 lengths among unique TRB or TRA sequences expressed by ISP (top) and DP (bottom) cells in SCID P1, P2, and Omenn P3 compared with control. [E] Distribution of CDR3 length for 5 more commonly expressed V genes in each sample for unique TRB sequences. [F] Distribution of CDR3 length for 5 more commonly expressed V genes in each sample for unique TRA sequences. Results demonstrate 1 representative sample from 2 experiments with similar results. RNase P, ribonuclease P.

Figure 6, continued from previous page.



shortened (V07-01 and V25-01) or normal (V12-02 and V21-01) CDR3 lengths can be seen (**Fig. 6F**, p. 52).

Finally, tree map representation of CDR3 amino acid sequences showed increased diversity, and reduced clonotypic expansion, in ISP and DP cells from control as compared with SCID and OS cells for both TRB (**Fig. S5E**, online, p. 38) and TRA (**Fig. S5F**, online, p. 38) repertoires.

Discussion

Combined immunodeficiencies include a heterogeneous group of conditions, whose broad clinical and immunologic phenotype depends on the nature of the gene affected, and the specific mutations involved.^{31,32} Here, we initially observed normal expression of markers corresponding to early stages of T-cell development (CD7, CD5, and CD38), which was also seen in cells with *RAG1* mutations associated with SCID and OS. However, these cells showed delayed expression of CD5 and CD38, resulting in a partial block at the CD7⁺ CD5⁻CD38⁻ stage. Nonetheless, both SCID- and OS-derived cells maintained some ability to initially express CD4, CD8 α , and CD8 β . Prior observations in the mouse indicated that *RAG1/2* expression occurs at the CD4⁻CD8⁻ double-negative stage, but is downregulated by pre-TCR signal to allow maturation to the CD4⁺ CD8⁺ DP stage, followed by a second wave of *RAG* expression and TRA recombination.³³ The emergence of a low number of DP cells during culture of iPSC-derived CD34⁻ cells from patients with *RAG1*-deficient SCID suggests that the first wave of *RAG1* expression and V(D)J recombination are not strictly required for the development of DP cells in humans, or that even these seemingly null *RAG1* mutations support minimal levels of recombination activity that suffice for partial progression to the DP stage. Control cells continued to mature, as demonstrated by expression of CD3 and TRA/TRB, whereas the presence of DP cells was rapidly lost in culture of SCID- and OS-derived cells, where only CD5⁺ CD7^{hi} cells were maintained. At later time points, SCID- and OS-derived cells remained as CD38⁻, CD31⁻ and CD45RA⁺, which is consistent with a failure to effectively traverse the pre-TCR-dependent β -selection checkpoint of T-cell differentiation.

Although the *RAG1* mutation associated with OS supported residual levels of recombination activity when tested with the Abelson pro-B-cell system, iPSC-derived CD34⁺ cells from the patient with OS did not show an enhanced ability to support T-cell differentiation *in vitro* when compared with SCID-derived cells. It is possible that the residual *RAG1* function in OS-derived cells is sufficient to initiate V(D)J recombination, but that these cells fail to survive and cannot progress to later stages. The *RAG1* protein plays an indispensable role in DNA nicking and cutting, and is also involved in the joining process.

Southern blot analysis revealed that introduction of the hypomorphic c.256_257delAA *RAG1* mutant into Abelson-transformed pro-B-cells did not result in accumulation of cells with unresolved DNA-DSBs, as observed for ligase IV-deficient cells. These data are consistent with the notion that this *RAG1* mutation associated with OS is defective in one of the steps prior to double-strand DNA cleavage.

The assay is limited by the requirement of a specific probe, therefore is unable to detect promiscuous, off-target DNA-DSBs that could conceivably occur in cells with mutant *RAG1*, nor does it show single-strand breaks. To address both possibilities, we further analyzed the efficiency of DNA cleavage of *RAG1* mutants associated with OS by employing the single cell gel electrophoresis (Comet assay), which was performed under alkaline conditions in order to reveal single- and double-strand DNA breaks. Surprisingly, CD4⁺ CD8⁻ and CD4⁺ CD8⁺ derived from the patient with OS consistently yielded comets.

One possible explanation is that the hypomorphic *RAG1* mutant may inappropriately nick DNA, while failing to introduce appropriate DNA-DSBs. According to this idea, the initial hydrolysis step of cleavage would be functional, but the reaction would be selectively defective for the subsequent transesterification step of hairpin formation, in which the 3' hydroxyl freed by single-strand cleavage attacks the backbone of the other strand. Single-strand cuts (or off-target DNA-DSBs) could accumulate and eventually lead to cell death, whereas on-site DNA-DSBs would yield the observed low-level and skewed recombination outcomes.

To test the SCID and OS mutants directly for nicking activity, we used the catalytic domain of SCID P2 and OS P3 to conduct an oligonucleotide cleavage assay, which revealed that these mutants have no nicking activity nor can they bind DNA (**Fig. S4A-B**, online, p. 38). Although the N-terminus truncated mutant of OS P3 retains an intact catalytic core, it is predominantly retained in the cytoplasm, and therefore has severely restricted activity.²⁶ However, OS P3 was a compound heterozygote for the c.256_257delAA and the c.2164G>A (p.E722K) mutations.

To explain the presence of comets and the inability to complete T-cell differentiation in OS P3, we have developed a model (**Fig. S6**, online, p. 38) that takes into account that dimerization of *RAG1* precedes nuclear translocation of the *RAG* complex. Swanson et al have shown that a heterodimer containing cleavage-competent and cleavage-incompetent components can nick DNA efficiently.³⁴ Similarly, a heterodimer of the *RAG1* mutants in OS P3 cells could enter the nucleus via E722K and bind DNA through the N-terminus truncated protein, which can also nick one RSS. However, the second site cannot be nicked by E722K, which has no activity,

thereby preventing hairpin formation.^{35,36} The RAG1 remains tightly bound to the site of the nick,³⁷ and cells with single-strand DNA breaks accumulate and are readily observed by the Comet assay. However, persistence of single-strand DNA breaks in immature T-lineage cells would culminate in impaired cell survival, and thus failure to sustain T-cell differentiation.

A recent publication by Teng et al demonstrated that, although RAG1 can bind ≈ 1800 sites in human thymocytes, off-target activity is prevented by reducing cryptic recombination signals near RAG1 binding sites,³⁸ thereby protecting the genome against oncogenesis. In light of this, our data indicate that DNA nicks or breaks are largely resolved when WT RAG1 is present, or cannot occur in SCID P1 or P3 cells with catalytically inactive RAG1, whereas these breaks persist in OS P3 cells (**Fig. 5**, p. 49). Assessing whether some *RAG1* mutations result in recognition of different off-target sites would be important, as the inherent protective mechanisms would not be in place, thereby increasing the risk of oncogenic genome rearrangements.

The observation that *in vitro* T-cell differentiation was similarly inefficient, both for SCID and for the OS patient, is consistent with the notion that OS-associated RAG mutations are very severe and support only minimal levels of DNA recombination. This was also reflected by the severe restriction of TRB and TRA rearrangements that was documented in sorted ISP and DP cells from patients with SCID and OS as compared with control cells. These data also suggest that the oligoclonality of peripheral T-cells that has been reported in patients with OS^{39,40} is not solely due to peripheral selection, but also reflects an intrinsic and severe defect of V(D)J recombination that occurs in the thymus, as also indicated by previous studies from our group in post-mortem thymic samples from patients with OS.³⁰ Control cells showed a clear bias for proximal Va (3'-end) and Ja (5'-end) usage, whereas remarkably this was not the case for RAG1-mutant cells. To date, this is the first comparison of both TRB and TRA rearrangement of *in-vitro*-derived SCID and OS T-lineage cells, and further studies are required to elucidate the causality for the observed lack of proximal bias observed for TRA rearrangements in RAG1-deficient cells.

In addition to the overall restriction of TRB and TRA rearrangements, *in vitro* differentiated ISP and DP cells from patients with SCID or OS revealed a series of additional abnormalities, including preferential usage of certain V and J genes, and skewing of CDR3 length, with presence of unusually long or short CDR3 fragments. Similar abnormalities have been recently reported in peripheral T-cells from patients with OS.⁴¹ Variability of the length of the CDR3 loop is determined by V(D)J recombination.⁴² Although CDR3 lengths are typically highly constrained for the TRA/TRB chains,⁴³ a slight decrease in CDR3 length is observed as T lymphocytes mature from the ISP to the DP

stage.⁴⁴ To this end, the increased TRB CDR3 length that we have observed in sorted ISP and DP cells from patients with SCID and OS may indicate that these cells express a less mature TCR that may escape central tolerance, thus allowing autoreactive T-cells to exit the thymus and contribute to the OS phenotype.

In summary, we have demonstrated that iPSCs represent a valuable tool to study mechanisms of altered T-cell development in patients with SCID and OS, and may offer unique insights into genotype-phenotype correlation in patients carrying distinct mutations in the same gene.

Acknowledgments

The authors thank their patients, their families, and physicians; and the Courtney McIntosh and Vincent Cheng (Centre for Flow Cytometry and Microscopy, Sunnybrook Research Institute, Toronto, ON, Canada) for their expert cell sorting support, and Andrew Elias (Princess Margaret Cancer Centre, Toronto, ON, Canada) for his assistance with fluorescence microscopy.

This work was supported by grants from the National Institutes of Health, National Institute of Allergy and Infectious Diseases (1R01AI100887 [L.D.N. and J.C.Z.-P.] and 2U54AI082973 [L.D.N.]), the March of Dimes (1-FY13-500) (L.D.N.), and by The Krembil Foundation (J.C.Z.-P). J.C.Z.-P. is also supported by a Canada Research Chair in Developmental Immunology. P.M.B. was supported by a Canadian Institutes of Health Research Postdoctoral Fellowship award. I.M.P. is supported by the Israeli Centers of Research Excellence, Gene Regulation in Complex Human Disease, Center No 41/11. F.D.B. and E.C. were supported in part by R01AI 082020.

References

1. Picard C, Al-Herz W, Bousfiha A, et al. Primary immunodeficiency diseases: an update on the classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. *J Clin Immunol*. 2015;35(8):696-726.
2. Bassing CH, Swat W, Alt FW. The mechanism and regulation of chromosomal V(D)J recombination. *Cell*. 2002;109(suppl):S45-S55.
3. Roth DB, Menetski JP, Nakajima PB, Bosma MJ, Gellert M. V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. *Cell*. 1992;70(6):983-991.
4. Fugmann SD. RAG1 and RAG2 in V(D)J recombination and transposition. *Immunol Res*. 2001;23(1):23-39.
5. Notarangelo LD, Kim MS, Walter JE, Lee YN. Human RAG mutations: biochemistry and clinical implications. *Nat Rev Immunol*. 2016;16(4):234-246.
6. Villa A, Santagata S, Bozzi F, et al. Partial V(D)J recombination activity leads to Omenn syndrome. *Cell*. 1998;93(5):885-896.
7. Giblin W, Chatterji M, Westfield G, et al. Leaky severe combined immunodeficiency and aberrant DNA rearrangements due to a hypomorphic RAG1 mutation. *Blood*. 2009;113(13):2965-2975.
8. Schuetz C, Huck K, Gudowius S, et al. An immunodeficiency disease with RAG mutations and granulomas. *N Engl J Med*. 2008;358(19):2030-2038.
9. Abolhassani H, Wang N, Aghamohammadi A, et al. A hypomorphic recombination-activating gene 1 (RAG1) mutation resulting in a phenotype resembling common variable immunodeficiency. *J Allergy Clin Immunol*. 2014;134(6):1375-1380.
10. Lee YN, Frugoni F, Dobbs K, et al. A systematic analysis of recombination activity and genotype-phenotype correlation in human recombination-activating gene 1 deficiency. *J Allergy Clin Immunol*. 2014;133(4):1099-1108.
11. Bosma GC, Custer RP, Bosma MJ. A severe combined immunodeficiency mutation in the mouse. *Nature*. 1983;301(5900):527-530.
12. Blunt T, Finnie NJ, Taccioli GE, et al. Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell*. 1995;80(5):813-823.
13. Bosma MJ, Carroll AM. The SCID mouse mutant: definition, characterization, and potential uses. *Annu Rev Immunol*. 1991;9:323-350.
14. Mombaerts P, Iacomini J, Johnson RS, Herrup K, Tonegawa S, Papaioannou VE. RAG-1-deficient mice have no mature B and T lymphocytes. *Cell*. 1992;68(5):869-877.
15. Shinkai Y, Rathbun G, Lam KP, et al. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell*. 1992;68(5):855-867.
16. Khiong K, Murakami M, Kitabayashi C, et al. Homeostatically proliferating CD4 T cells are involved in the pathogenesis of an Omenn syndrome murine model. *J Clin Invest*. 2007;117(5):1270-1281.
17. Marrella V, Poliani PL, Casati A, et al. A hypomorphic R229Q Rag2 mouse mutant recapitulates human Omenn syndrome. *J Clin Invest*. 2007;117(5):1260-1269.
18. Perryman LE. Molecular pathology of severe combined immunodeficiency in mice, horses, and dogs. *Vet Pathol*. 2004;41(2):95-100.
19. Wiler R, Leber R, Moore BB, VanDyk LF, Perryman LE, Meek K. Equine severe combined immunodeficiency: a defect in V(D)J recombination and DNA-dependent protein kinase activity. *Proc Natl Acad Sci USA*. 1995;92(25):11485-11489.
20. Spits H. Development of alphabeta T cells in the human thymus. *Nat Rev Immunol*. 2002;2(10):760-772.
21. Kennedy M, Awong G, Sturgeon CM, et al. T lymphocyte potential marks the emergence of definitive hematopoietic progenitors in human pluripotent stem cell differentiation cultures. *Cell Reports*. 2012;2(6):1722-1735.
22. Vizcardo R, Masuda K, Yamada D, et al. Regeneration of human tumor antigen-specific T cells from iPSCs derived from mature CD8(1) T cells. *Cell Stem Cell*. 2013;12(1):31-36.
23. Chang CW, Lai YS, Lamb LS Jr, Townes TM. Broad T-cell receptor repertoire in T-lymphocytes derived from human induced pluripotent stem cells. *PLoS One*. 2014;9(5):e97335.

- 24.** Menon T, Firth AL, Scripture-Adams DD, et al. Lymphoid regeneration from gene-corrected SCID-X1 subject-derived iPSCs. *Cell Stem Cell*. 2015;16(4):367-372.
- 25.** Shearer WT, Dunn E, Notarangelo LD, et al. Establishing diagnostic criteria for severe combined immunodeficiency disease (SCID), leaky SCID, and Omenn syndrome: the Primary Immune Deficiency Treatment Consortium experience. *J Allergy Clin Immunol*. 2014;133(4):1092-1098.
- 26.** Santagata S, Gomez CA, Sobacchi C, et al. N-terminal RAG1 frameshift mutations in Omenn's syndrome: internal methionine usage leads to partial V(D)J recombination activity and reveals a fundamental role in vivo for the N-terminal domains. *Proc Natl Acad Sci USA*. 2000;97(26):14572-14577.
- 27.** Pessach IM, Ordovas-Montanes J, Zhang SY, et al. Induced pluripotent stem cells: a novel frontier in the study of human primary immunodeficiencies. *J Allergy Clin Immunol*. 2011;127(6):1400-1407.
- 28.** Weinacht KG, Brauer PM, Felgentreff K, et al. The role of induced pluripotent stem cells in research and therapy of primary immunodeficiencies. *Curr Opin Immunol*. 2012;24(5):617-624.
- 29.** Ostling O, Johanson KJ. Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells. *Biochem Biophys Res Commun*. 1984;123(1):291-298.
- 30.** Signorini S, Imberti L, Pirovano S, et al. Intrathymic restriction and peripheral expansion of the T-cell repertoire in Omenn syndrome. *Blood*. 1999;94(10):3468-3478.
- 31.** Milner JD, Holland SM. The cup runneth over: lessons from the ever-expanding pool of primary immunodeficiency diseases. *Nat Rev Immunol*. 2013;13(9):635-648.
- 32.** Villa A, Sobacchi C, Notarangelo LD, et al. V(D)J recombination defects in lymphocytes due to RAG mutations: severe immunodeficiency with a spectrum of clinical presentations. *Blood*. 2001;97(1):81-88.
- 33.** Wilson A, Held W, MacDonald HR. Two waves of recombinase gene expression in developing thymocytes. *J Exp Med*. 1994;179(4):1355-1360.
- 34.** Swanson PC. The DDE motif in RAG-1 is contributed in trans to a single active site that catalyzes the nicking and transesterification steps of V(D)J recombination. *Mol Cell Biol*. 2001;21(2):449-458.
- 35.** Yu K, Lieber MR. Mechanistic basis for coding end sequence effects in the initiation of V(D)J recombination. *Mol Cell Biol*. 1999;19(12):8094-8102.
- 36.** Bischerour J, Lu C, Roth DB, Chalmers R. Base flipping in V(D)J recombination: insights into the mechanism of hairpin formation, the 12/23 rule, and the coordination of double-strand breaks. *Mol Cell Biol*. 2009;29(21):5889-5899.
- 37.** Grawunder U, Lieber MR. A complex of RAG-1 and RAG-2 proteins persists on DNA after single-strand cleavage at V(D)J recombination signal sequences. *Nucleic Acids Res*. 1997;25(7):1375-1382.
- 38.** Teng G, Maman Y, Resch W, et al. RAG represents a widespread threat to the lymphocyte genome. *Cell*. 2015;162(4):751-765.
- 39.** Rieux-Laucat F, Bahadoran P, Brousse N, et al. Highly restricted human T cell repertoire in peripheral blood and tissue-infiltrating lymphocytes in Omenn's syndrome. *J Clin Invest*. 1998;102(2):312-321.
- 40.** Brooks EG, Filipovich AH, Padgett JW, Mamlock R, Goldblum RM. T-cell receptor analysis in Omenn's syndrome: evidence for defects in gene rearrangement and assembly. *Blood*. 1999;93(1):242-250.
- 41.** Yu X, Almeida JR, Darko S, et al. Human syndromes of immunodeficiency and dysregulation are characterized by distinct defects in T-cell receptor repertoire development. *J Allergy Clin Immunol*. 2014;133(4):1109-1115.
- 42.** Hughes MM, Yassai M, Sedy JR, et al. T cell receptor CDR3 loop length repertoire is determined primarily by features of the V(D)J recombination reaction. *Eur J Immunol*. 2003;33(6):1568-1575.
- 43.** Rock EP, Sibbald PR, Davis MM, Chien YH. CDR3 length in antigen-specific immune receptors. *J Exp Med*. 1994;179(1):323-328.
- 44.** Nishio J, Suzuki M, Nanki T, Miyasaka N, Kohsaka H. Development of TCRB CDR3 length repertoire of human T lymphocytes. *Int Immunol*. 2004;16(3):423-431.

4 Rapid generation of novel models of RAG1 deficiency by CRISPR/Cas9-induced mutagenesis in murine zygotes

Oncotarget. 2016 Mar 15;7(11):12962-74.



Lisa Ott de Bruin^{1,2}

Wei Yang³

Kelly Capuder¹

Yu Nee Lee¹

Maddalena Antolini¹

Robin Meyers⁴

Martin Gellert³

Kiran Musunuru^{5,6}

John Manis⁷

Luigi Notarangelo^{1,6}

¹ Division of Immunology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

² Pediatric Immunology, Wilhelmina Children's Hospital, Utrecht University Medical Center, Utrecht, The Netherlands

³ Laboratory of Molecular Biology, NIDDK, NIH, Bethesda, MD, USA

⁴ Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA, USA

⁵ Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA

⁶ Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA

⁷ Division of Transfusion Medicine, Department of Laboratory Medicine, Boston Children's hospital, Harvard Medical School, Boston, MA, USA

Abstract

Mutations in the recombination-activating gene RAG1 (*RAG1*) can cause a wide variety of clinical and immunological phenotypes in humans, ranging from absence of T and B lymphocytes to occurrence of autoimmune manifestations associated with expansion of oligoclonal T-cells and production of autoantibodies. Although the mechanisms underlying this phenotypic heterogeneity remain poorly understood, some genotype-phenotype correlations can be made. Currently, mouse models of Rag deficiency are restricted to *Rag1*^{-/-} mice and to knock-in models carrying severe missense mutations. The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/Cas9 system is a novel and powerful gene-editing strategy that permits targeted introduction of DNA double strand breaks with high efficiency through simultaneous delivery of the Cas9 endonuclease and a guide RNA (gRNA). Here, we report on CRISPR-based, single-step generation and characterization of mutant mouse models in which gene editing was attempted around residue 838 of *RAG1*, a region whose functional role had not been studied previously.

Introduction

The recombination-activating gene 1 (*RAG1*) and *RAG2* are critical for T and B-cell development. The *RAG1* and *RAG2* proteins form a complex that introduces DNA double strand breaks (DSBs) at the recombination signal sequences (RSSs) that flank the Variable (V), Diversity (D) and Joining (J) gene segments of the immunoglobulin (Ig) and T-cell receptor (TCR) genes, thereby initiating the process of V(D) J recombination that permits expression of Ig and TCR molecules¹. Mutations in *RAG1* or *RAG2* can lead to a wide variety of clinical and immunological phenotypes in humans, including complete absence of T and B-cells (T- B- Severe Combined Immunodeficiency (SCID))²; Omenn syndrome (OS) with lymphadenopathy, increased serum IgE, eosinophilia, erythroderma and detectable autologous, oligoclonal and activated T lymphocytes³⁻⁹; *RAG* deficiency with expansion of TCR $\gamma\delta$ ⁺ T-cells⁹; atypical/leaky SCID (LS) with some T and B-cells but no typical OS features^{10,11}; Combined Immunodeficiency with granuloma and/or autoimmunity (CID-G/AI)¹²⁻¹⁴, and CD4 lymphopenia¹⁵. The mechanisms underlying such phenotypic heterogeneity remain poorly defined, but in the past years some genotype-phenotype correlation has emerged^{16,17}.

The *RAG1* protein is highly conserved between humans and mice. Mouse *RAG1* consists of a RING finger (ZFA), a nonamer binding region (NBR), a dimerization and DNA-binding domain (DDBD), an RNase H-like catalytic domain containing the metal-chelating carboxylates D600, D708 and E962, and a large insertion between residues D708 and E962, which includes two Zinc binding regions (one is formed by C727 and C730, the other by H937 and H942) that together form one zinc finger binding domain (ZFB)¹⁸ (Fig. 1, p. 65). The stretch in between these residues is of unknown function, as also shown by two recent structural studies^{17,18}. The C-terminal domain (CTD) starts immediately after the catalytic residue E962 and interacts extensively with the DDBD domain. Based on recent crystallography data, mutations causing SCID and OS can be grouped in four classes. The first class of mutations destabilizes the tertiary structure, as is the case for mutations involving the zinc binding sites. The second class of mutations involves domains important for DNA binding. The third class of mutations involves the catalytic RNase H-like domain. Lastly, the fourth class involves the *RAG1/RAG2* interface¹⁸.

In addition to the original knock-out models, *Rag1*^{-/-}¹⁹ and *Rag2*^{-/-}²⁰, characterized by complete absence of T and B-cells, three mouse knock-in models of OS and LS have been reported: the hypomorphic *Rag2* mutation R229Q²¹ (involving the *RAG1/RAG2* interface), the hypomorphic *Rag1* mutation S723C²² (close to one of the zinc binding regions) and the R972Q²³

mutation (affecting the CTD). However, missense mutations in regions other than the NBR, DDBD, catalytic domain or zinc binding domain often show higher residual V(D)J recombination activity and are frequently seen in patients with less severe and delayed-onset disease, often associated with autoimmunity, as was the case for the human mutation R841W (mouse R838W)¹⁶. Therefore, we decided to target the region around residue 838 of the RAG1 locus, which falls within the catalytic residues 708 and 962 and does not involve any zinc binding regions.

Traditionally, in order to generate mouse models of human diseases, gene targeted embryonic stem cells (ESCs) are electroporated with a DNA template containing the desired mutation in the gene of interest flanked by homology arms. Usually, an excisable antibiotic resistance gene is also introduced in one of the homology arms to facilitate identification and selection of targeted clones. Homology-directed repair (HDR) is a low efficiency process that permits to replace the endogenous target ESC genomic sequence with that provided by the DNA template. Upon *in vitro* culture under antibiotic pressure and screening, by polymerase chain reaction (PCR), ESC clones that have been successfully targeted with the desired sequence are initially selected and expanded, and are then injected into blastocysts, and implanted in pseudo-gestating females. The resulting chimeric offspring animals have to be further bred until the introduced mutation is transmitted through the germline. Overall, this is a lengthy, rather inefficient, and expensive process.

Recently, the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR associated 9 (Cas9) system has emerged as a novel and powerful gene editing platform. The Cas9 nuclease is directed to a specific DNA site by a guide RNA (gRNA), allowing targeted introduction of DNA double strand breaks (DSBs), that can be repaired through the error-prone non-homologous end-joining (NHEJ) machinery, which often introduces insertions and/or deletions (indels) at the DNA DSB site, leading to frameshift and early termination and hence gene disruption. Alternatively, if a homologous DNA template sequence with a desired mutation is provided along with Cas9 and the gRNA, homology-mediated repair permits replacement of the endogenous target DNA sequence with the mutated sequence of interest. Cas9-based gene editing has therefore emerged as a novel and powerful strategy to generate knock-out (KO) and knock-in (KI) mouse models. Importantly, Cas9 and gRNA with or without a homologous single strand DNA oligonucleotide (ssODN) can be directly injected into multiple mouse zygotes that are then implanted into foster mothers, with no need for further selection. This method has significantly improved the efficiency of generating genetically modified mouse models, with significant containment of costs and time²⁴⁻²⁶.

Here we describe multiple *Rag1* mouse models generated with high efficiency in a single step by targeting zygotes with Cas9 and a *Rag1*-specific gRNA. Several mice with various mutations at and around residue 838 were studied, providing insights into the functional role of the region. This study also offers proof of principle that a similar strategy can be applied to develop novel knock-in mouse models that may help interrogate the specific role of individual amino acid residues within this region.

Materials and methods

Design of *Rag1*-specific gRNAs

In order to target the region around residue 838 of RAG1, a series of gRNAs were designed. In order to select gRNA protospacer sequences with limited off-site targeting, NCBI blast was used to search the mouse genome for sequence similarity. gRNA protospacer sequences were selected that had no perfect matches, 1-base mismatches, or 2-base mismatches, and as few 3-base mismatches as possible, in other regions of the mouse genome than the target sequence. As a result of this process, two different gRNAs (gRNA A and B) were selected which would introduce a DSB between nucleotide c. 2508 and c. 2509 (gRNA-A) and between nucleotide c. 2514 and c. 2515 (gRNA-B), respectively (**Fig. 1**). Each protospacer region was cloned in a gRNA cloning plasmid (Addgene)²⁷.

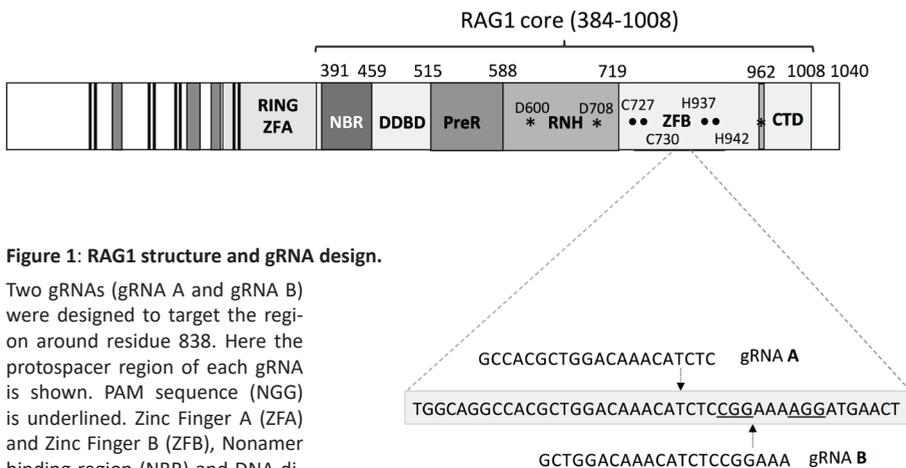


Figure 1: RAG1 structure and gRNA design.

Two gRNAs (gRNA A and gRNA B) were designed to target the region around residue 838. Here the protospacer region of each gRNA is shown. PAM sequence (NGG) is underlined. Zinc Finger A (ZFA) and Zinc Finger B (ZFB), Nonamer binding region (NBR) and DNA dimerization and binding domain (DDBD), pre-RNase (preR), the catalytic RNase H-like (RNH) domain and C-terminal domain (CTD). Residue numbers are given for the boundaries of the different domains. Catalytic residues D600, D708 and E962 are denoted by (*). ZFB (one domain) consists of two binding regions (residues 727/730 and residues 937/942), denoted by (●). The region in between the two zinc binding regions (that form one domain) was targeted.

Testing the efficiency of gRNA-mediated introduction of targeted DSBs

The efficiency of gRNA-A and gRNA-B in mediating targeting of the *Rag1* gene was compared in murine ESCs. ESCs were grown cultured in media consisting of DMEM high glucose supplemented with 15% FCS, 1% Pen/strep, 1% L-glutamine, 1% Non-essential amino acids and 10^{-4} M Beta mercapto-ethanol and 103U/ml LIF. Cells were transfected with 2.5 μ g of the pCas9-GFP plasmid (Addgene) and either 2.5 μ g gRNA-A or 2.5 μ g gRNA-B using the 4D Amaxa nucleofector (Lonza), using kit P3 and the protocol provided by the manufacturer. Forty-eight hours post-nucleofection, GFP-positive ESCs expressing Cas9-GFP were selected by FACS sorting. Genomic DNA was isolated and subjected to high throughput sequencing (HTS) using barcoded primers and the 454 GS Junior apparatus (Roche). For each gRNA, the efficiency of on-site targeting was defined as the percentage of reads containing an indel spanning the target site.

Zygote injection and analysis of *Rag1* gene targeting in the offspring

Two hundred and fifty C57BL/6 zygotes were injected with 100 ng/ μ l Cas9 mRNA (purchased from System Biosciences Cat#CAS500A-1) and 50 ng/ μ l gRNA A (generated from PCR product using Megashortscript T7 kit Life Technologies) into the cytoplasm, as previously described²⁴. A PCR of the gRNA was performed to add a T7 promoter to the gRNA sequence as previously described²⁵. Two hundred and thirty live embryos were implanted into seven recipient foster mothers. Pups were born two and a half weeks later. Tail gDNA of three-week-old pups was analyzed by PCR and Sanger sequencing for the presence of mutations in the *Rag1* gene. To this purpose, the genomic region spanning 2290 - 2723 (433 bp) was amplified using the following primers:

Forward	5' -CGTATCATGAGTCCGTGGAA-3'
Reverse	5' -TGACAGAGGGACTCTGACAC-3'

PCR was performed using a 20 μ l reaction with Phusion Hot start II (Lifetechnologies Cat# F-549L). The manufacturer's PCR settings were followed, using GC buffer 5X, a final concentration of 3% DMSO and an annealing temperature of 60C. For sixteen mice, in which no amplification products were obtained with the primers reported above, a PCR reaction was performed with primers that span a 1500 bp region of *Rag1* around the Cas9-induced cleavage site:

Forward	5' -TCTGGACTTGCCCTCTGT-3'
Reverse	5' -CCATTGAATCTTGGCTTTCC-3'

Breeding and control mice

The five mice with in-frame deletions (ID 25, 37, 48, 49, 51) and one with a frameshift deletion (ID 41), collectively referred to as $\Delta Rag1$ mice, were sacrificed and analyzed without breeding (**Table 1**, p. 69). To segregate the three alleles, female mouse 31 was bred with a wild type (*wt*) C57BL/6 male resulting in offspring with the different mutations. F1 mice with heterozygous H836Q mutation (*Rag1*^{+/H836Q} mice) were bred to get F2 mice homozygous for the H836Q mutation (*Rag1* H836Q mice). *Rag1*^{-/-} mice were purchased from the Jackson Laboratory. All mice described in this manuscript were on a C57BL/6 background.

FACS analysis of spleen, thymus and bone marrow cells

Total cellularity of thymus, spleen, bone marrow and the distribution of the various T and B-cell subsets were analyzed in 8-12 week-old $\Delta Rag1$ mice and *Rag1* H836Q mice. Both groups were compared to age-matched wild type (*wt*) and *Rag1*^{-/-} mice (Jackson Laboratory)¹⁹.

Splenocytes were stained with a T-cell panel consisting of CD4-PB (Ebioscience), CD8-PE/Cy5 (BD), CD3 PE-Cy7 (Biolegend) and a B-cell panel consisting of B220-APC (Biolegend), IgM-PE-cy5 (Ebioscience), CD43-PE (Ebioscience). Bone marrow cells were stained with B220-APC (Biolegend), IgM-Pe-Cy5 (Ebioscience), and CD43-PE (Ebioscience) antibodies. Thymocytes were stained with CD3-PE-Cy7 (Biolegend), CD4-PB (Ebioscience), CD8-PE-Cy5 (BD), CD44-FITC (BD), and CD25-PE (Ebioscience) antibodies, upon excluding B220 (Biolegend), Ter119 (Biolegend) and MAC1 (Biolegend) positive cells. FACS gating strategies are shown in **Fig. S1-S4**, p. 78. Standard FSC/SSC live gates and FSC/SSC lymphocyte gates were used.

Statistical analysis

Two-tailed unpaired *t*-test was used to compare results in $\Delta Rag1$ vs. *wt* mice and to compare *Rag1* H836Q to *wt* mice.

Results

In order to compare the ability of gRNA-A and gRNA-B to induce Cas9-mediated introduction of on-site DSBs at the *Rag1* locus, we analyzed the frequency of indels as detected by high-throughput sequencing (HTS) in nucleofected ESCs. Any deletion and/or insertion spanning the cutting site was counted. A higher frequency of on-site targeting was detected with gRNA-A (57% vs 15%). Therefore, gRNA-A was selected for injection into the murine zygotes.

ID	Nucleotide	Aminoacid
14	c.[2105_3014del]	p.[Phe702Phefs*4]
30	~700 bp del	
35	~600 bp del	
10	c.[2503_3021del]; c.[2508_2999del]	p.[Lys835_Asn1007]; p.[His836Glnfs*41]
50	c.[2505_2920del]	p.[Lys835Asnfs*9]
2	c.[2500_2811del]	p.[Asp834_His937del]
18	c.[2497_2785del]	p.[Leu833Argfs*45]
34	c.[2501_2763del]	p.[Asp834Aspfs*5]
40	c.[2508_2739del]	p.[His836Hisfs*61]
17	c.[2503_2671del]	p.[Lys835Profs*112]
23	c.[2507_2671del]	p.[His836Profs*150]
8	c.[2363_2514del]; c.[2502_2524del]	p.[Val788Glufs*29]; p.[Asp834Glufs*26]
21	c.[2494_2642del]	p.[Thr832Alafs*18]
4	c.[2503_265del]	p.[Lys835Profs*15]
*25	c.[2508_2627del]	p.[Leu837_His876del]
36	c.[2507_2516del]; c.[2507_2585del]	p.[His836Glnfs*2]; p.[His836Glnfs*2]
42	c.[2464_2512del]; c.[2508del]	p.[Ser822Glyfs*3]; p.[His836Hisfs*7]
**31	c.[2508T>A]; c.[2462_2509del]; c.[2508del]	p.[His836Gln]; p.[Ala821Valfs*5]; p.[His836Hisfs*6]
20	c.[2509_2540del]	p.[Leu837Glufs*20]
44	c.[2501_2515delins CGCTTTGGTAGATGCCTCCCATCCACTCGGTT]	p.[Asp834Alafs*12]
*37	c.[2507_2533del]; c.[2510del]	p.[His836_Lys844del]; p.[Leu837Profs*4]
28	c.[2500_2510del]	p.[Leu833Leufsf*31]
43	c.[2508_2512delinsAAA]	p.[His836Glnfs*31]
33	c.[2503_2516del]	p.[Lys835Lysfs*29]
9	c.[2503_2516del]	p.[Lys835Lysfs*29]
16	c.[2503_2516del]	p.[Lys835Lysfs*29]
27	c.[2472_2505delinsAAAGTT;2508_2509insTC]; c.[2503_2516del]	p.[Glu824Glufs*35]; p.[Lys835Lysfs*29]
29	c.[2502_2515del]; c.[2509_2510delinsGTCCTTT]	p.[Asp834Glufs*29]; p.[Leu837Valfs*6]
39	c.[2502_2515del]; c.[2508insC]	p.[Asp834Glufs*29]; p.[His836Hisfs*32]
52	c.[2502_2515del]	p.[Asp834Glufs*29]
13	c.[2508delinsAA]; c.[2502_2515del]	p.[His836Glnfs*32]; p.[Asp834Glufs*29]
5	c.[2502_2515del]	p.[Asp834Glufs*29]
22	c.[2502_2515del]	p.[Asp834Glufs*29]
38	c.[2499_2512del]	p.[Leu833Leufs*30]
45	c.[2499_2512del]; c.[2508_2509del]	p.[Leu833Leufs*30]; p.[His836Hisfs*31]
*41	c.[2496_2509del]	p.[Thr832Thrfs*31]
3	c.[2505_2508del]; c.[2507_2556del]	p.[Lys835Asnfs*5]; p.[His836Leufs*15]
32	c.[2506_2508delinsTG]	p.[His836Cysfs*5]
*48	c.[2493_2495del; 2512C>T]; c.[2493_2495del]	p.[Thr832del; Trp838del]; p.[Thr832del]
*49	c.[2507_25093del]; c.[2505-2573delinsG]	p.[His836del]; p.[His846Aspfs*9]
*51	c.[2504_2509del]; c.[2505del; 508_2509del]	p.[Lys835Ile; His836_Leu837del]; p.[Lys835_Leu837delinsAsn835_Ile836]
6	c.[2509_2510del]	p.[Leu837Profs*30]
19	c.[2507_2508del]	p.[His836Profs*31]
47	c.[2507_2508del]	p.[His836Profs*31]
53	c.[2508_2509del]	p.[His836Hisfs*31]
26	c.[2509_2510del]; c.[2508_2509insT]	p.[Leu837Profs*30]; p.[Leu837Serfs*31]
15	c.[2508_2509insA]	p.[Leu837Thrfs*31]

Upon injection of Cas9 mRNA and gRNA-A into 250 C57BL/6 zygotes, 230 live embryos were obtained and transferred into seven female recipient mice. After 18-20 days, 53 live pups were born. At 3 weeks of life, Sanger sequencing of the *Rag1* gene was performed using the tail gDNA as a template. Results of genotyping are shown in **Table 1**. For sixteen mice (ID 1, 2, 7, 10, 11, 12, 14, 18, 21, 24, 30, 34, 35, 40, 46 and 50), the initial amplification of 433 bp around the DNA cleavage site did not yield any product. Upon performing PCR with primers that span a 1.5 kb region around the cleavage site, smaller fragments were detected in ten of these mice (ID 2, 10, 14, 18, 21, 30, 34, 35, 40 and 50), indicating the occurrence of intragenic deletions, whose boundaries were defined by Sanger sequencing (**Table 1**). For six mice (ID 1, 7, 11, 12, 24, 46) no products were obtained even with this strategy, suggesting occurrence of either an even larger deletion or chromosomal translocation²⁸. These six mice with undefined mutations were omitted from any further analysis. In all other cases, Sanger sequencing showed either two distinct sequences, corresponding to different indels on different alleles, or one single sequence with an indel spanning the cutting site. The latter may reflect either homozygosity for that indel, or compound heterozygosity with the second allele failing to amplify due to a large deletion or translocation.

In summary, all 53 mice showed bi-allelic targeting of the *Rag1* locus. Forty-two of them showed deletions, 1 showed only an insertion, 9 showed a combination of insertions and deletions, and one mouse (ID 31, indicated by two asterisks in **Table 1**) showed somatic chimerism, with three distinct alleles. Upon subcloning, it was established that these mutant alleles corresponded to a c.2508T>A nucleotide substitution (p.His836Gln) a c. 2462_2509del (p. Ala821Valfs*5), and a 2508del (p.His836Hisfs*6). Five mice (ID 25, 37, 48, 49, 51; indicated by an asterisk in **Table 1**) showed in-frame deletions or insertions on one allele, and either frameshift deletions or large deletions on the other allele.

The sizes of the in-frame deletions corresponded to one (residue 836 in mouse 49), two (residue 832 and 838 in mouse 48), nine (residue 836-844 in mouse 37) or forty amino acids (residue 837-876 in mouse 25). For one mouse (ID 51), a 6 bp deletion on one allele resulted in substitution of Lysine to Isoleucine at position 835 (Lys835Ile) and deletion of residues 836 and

Table 1. When for each allele a different sequence is given (heterozygous), two different sequences were seen by Sanger sequencing. When only one sequence is given, this means only one sequence was amplified and detected by Sanger sequencing. This could either be due to an identical indel on both alleles (homozygous) or the second allele was not amplified by PCR due to a large deletion or translocation. * 5 mice with in-frame deletions and one with frameshift (as one example), were further analyzed by FACS without further breeding. **Mouse 31 showed 3 sequences, due to mosaicism and was further bred to separate the alleles.

837 (p.Lys835Ile; His836_Leu837del). On the other allele, a 3 bp deletion resulted in deletion of residue 837 with a substitution of Lysine to Asparagine in position 835 and substitution of Histidine to Isoleucine in position 836 (p. Lys835_Leu837delinsAsn835_Ile836).

The human RAG1 H839Q mutation, corresponding to H836Q in, has not been reported in patients, nor in public databases of genomic variants in the general population (dbSNP, 1000 genomes, ExAC). The human H839Q variant was however predicted to be damaging by SIFT, probably pathogenic by Polyphen-2, but neutral by Snap and SNP&Go program analyses. To predict the structural consequences of this mutation, a model was developed based on crystal structure of mouse RAG1¹⁸. Mouse RAG1 residues R838 - N852 are predicted to have extended interactions with the heptameric region of RSS DNA (**Fig. 2A** and **2B**). Therefore, any deletion in this region is predicted to have a deleterious effect on RSS binding. Residue 836 in particular is important for interaction between two alpha helices in RAG1. However, Histidine and Glutamine have similar hydrogen bonding potential. Glutamine has the hydrogen bond donor (NH₂ group) and acceptor (carboxyl oxygen) sites at positions equivalent to those in the imidazole group of Histidine (one ring NH as donor and the other N as acceptor). Therefore, substituting Histidine with Glutamine should have little effect on protein interactions and none on DNA binding (**Fig. 2C**).

Altogether, structural modeling predicted the H836Q mutation in mouse RAG1 to have little or no effect on protein expression and function. In contrast, **Fig. 2D** shows the possible consequence of deletion of H836 (observed in mouse 49) within an alpha helix. Assuming the alpha helix remains stable, the residues after H836 all shift by one register, and L837 thus would take the position of H836 in the deletion mutant, which would place a hydrophobic residue in a hydrophilic environment. Moreover, the R838 (Arginine) is shifted into a hydrophobic pocket normally occupied by L837 (Leucine) in wild type RAG1. Arginine is a longer amino acid than Leucine and has a more positive charge. Therefore, deletion of H836 is predicted to wreck the protein structure and interfere with RAG1 functions ranging from DNA binding to cleavage. The effect of deleting T832 (observed in mouse 48) is similar to deletion of 836, shifting the register of this alpha helix by one residue and resulting in charged residues in the hydrophobic core. Since T832 is an earlier residue in the same helix as H836, deleting T832 would be expected to have an even more severe effect, placing D834 in the position of L833, in addition to R838 being placed in the former position of L837 (**Fig. 2E**).

Next, we assessed the functional consequences of the various in-frame deletions or insertions to T and B-cell development, in the gene-targeted

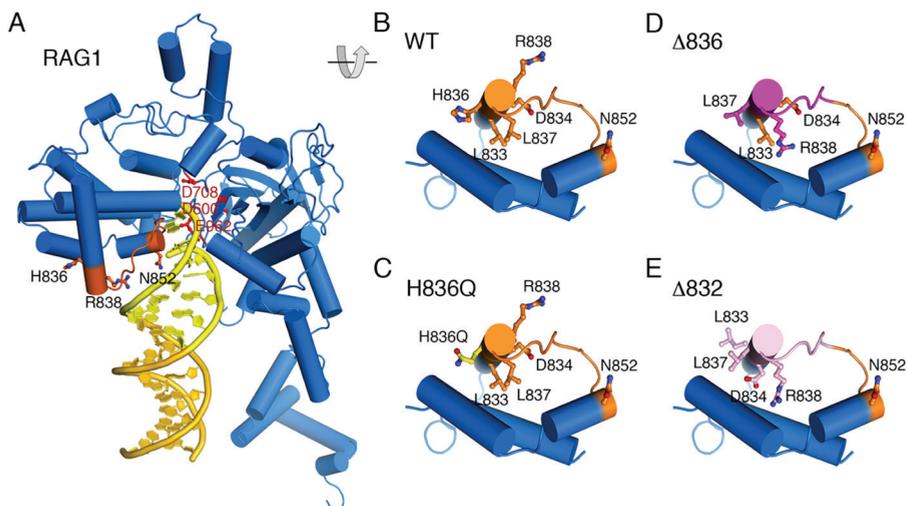


Figure 2: Structural model of RAG1-RSS complex.

[A] Wild type RAG1. RAG1 core is shown in blue. The yellow portion of the DNA is 7bp (heptamer), and the orange colored DNA represents the spacer region. The DNA contact region (residues 838-852) is highlighted in brown. [B] wild type RAG1 but a zoom in view of the protein helices surrounding the DNA binding region and rotated $\sim 90^\circ$ indicated by the grey diagram. [C] The H836Q mutation (Histidine substituted by Glutamine). Glutamine has the hydrogen bond donor (NH₂ group, shown in blue) and acceptor (carboxyl oxygen, shown in red) at positions equivalent to those in the imidazole group of Histidine (one ring NH as donor and the other ring N as acceptor). [D] The possible consequence of deletion of H836 in the middle of an alpha helix. The residues after H836 all shift by one register and L837 (Leucine) would occupy the H836 position, meaning a hydrophobic residue in a hydrophilic environment. R838 (Arginine) shifts into a hydrophobic pocket occupied by L837 in WT RAG1, thus destabilizing the protein structure and interfering with all RAG1 functions from DNA binding to cleavage. [E] The predicted effect of deleting residue 832 is similar to deleting residue 836, which is shifting the register of an alpha helix by one residue and resulting in charged residues in the hydrophobic core. Since this deletion occurs one helical turn before H836, the effect is even more severe. D834 is placed in the position of L833 in addition to R838 being placed in the L837 position.

mice. Since bi-allelic out-of-frame deletions in this region uniformly result in a complete T- B- SCID phenotype in humans, we focused our attention on mice carrying in-frame deletions or insertions. For mouse 31, which was found to be a somatic chimera, we segregated the 3 distinct mutant alleles by sequential breeding with *wt* mice. Intercrossing of heterozygous *Rag1*^{+/^{H836Q} mice yielded mice that were homozygous for the p.H836Q mutation, which were kept for further analyses. Finally, one mouse (ID 41), with a frameshift deletion leading to a stop codon at position 863, was included as a positive control of severe *Rag1* locus disruption.}

Peripheral blood of the five RAG1 targeted mouse strains with in-frame deletions and of the mouse with the frameshift mutation (collectively referred to as Δ *Rag1* mice), all showed complete absence of peripheral T- and B-cells by FACS analysis (Fig. S1, p. 78).

Total thymic cellularity in $\Delta Rag1$ mice was markedly decreased compared to that in *wt* mice, and consistent with that in *Rag1*^{-/-} mice (**Fig. 3A**). Immunophenotypic analysis of thymocyte subsets in these mice showed an arrest at the CD4⁻ CD8⁻ double negative stage (**Fig. 3B and 3C**), and in particular at the DN3 (CD44⁻ CD25⁺) stage of differentiation (**Fig. 3D and 3E**), which marks initiation of *Rag* gene expression. Importantly, a similar severe phenotype was observed also in mouse 49 with deletion of three nucleotides corresponding to deletion of H836 (**Table 1**, p. 69). By contrast, thymic

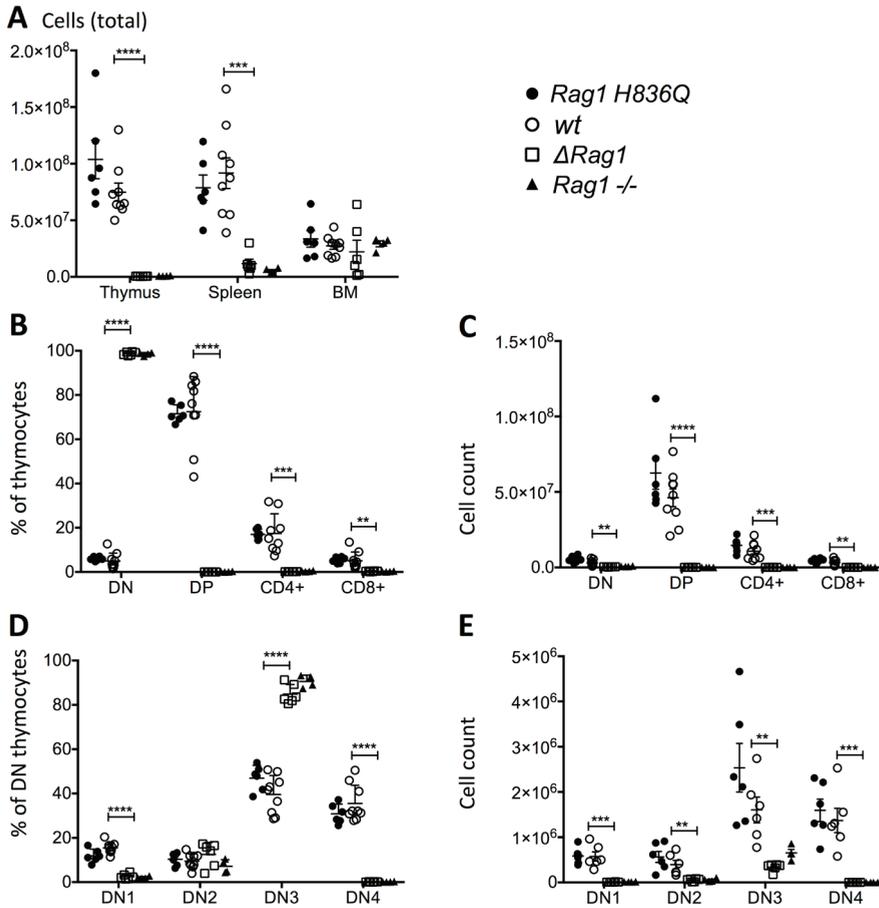


Figure 3: Total Cell counts and FACS Thymus.

[A] Total cell counts in Thymus, bone marrow (BM) and spleen. [B] FACS of thymocyte subsets (%) (excluding B220⁺, Ter119⁺ MAC1⁺). [C] Absolute cell counts for panel 3B. [D] Subsets of DN (%). [E] Absolute cell counts for panel 3D. *Rag1* H836Q (n=6), *wt* (n=9), *Rag1* knock-out (*Rag1*^{-/-}) (n=4). $\Delta Rag1$ (n=6) are mouse 25, 37, 41, 48, 49 and 51 with each different indels (**Table 1**, p. 69). Two tailed, unpaired T-tests between $\Delta Rag1$ and *wt* (** p ≤ 0.01, *** p ≤ 0.001, **** p ≤ 0.0001). There was no significant difference between *Rag1* H836Q and *wt*.

cellularity and thymocyte subset distribution were comparable in mice homozygous for the H836Q mutation (*Rag1 H836Q*) and in *wt* mice.

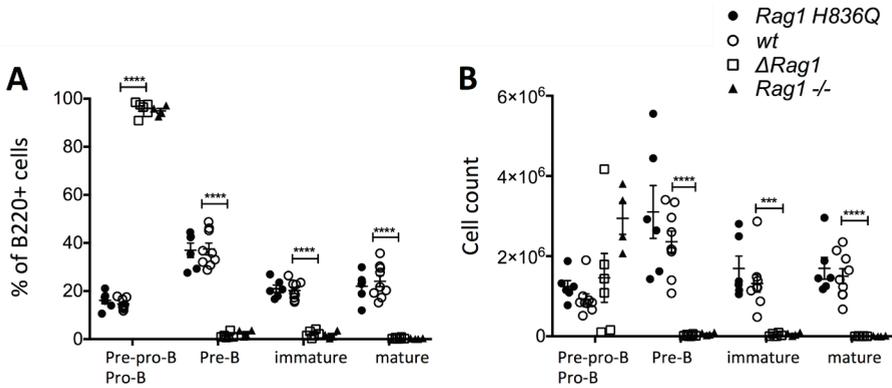


Figure 4: FACS Bone Marrow.

[A] Subsets of B220⁺ cells (%). Pre-pro- and pro-B-cell = B220⁺ CD43⁺ IgM⁻, Pre-B-cells = B220⁺ CD43⁻ IgM⁻, immature B-cells = B220^{low} CD43⁻ IgM⁺, mature B-cells = B220^{hi} CD43⁻ IgM⁺. [B] absolute cell counts of 4. A. *Rag1 H836Q* (n=6), *wt* (n=9), *Rag1* knock-out (*Rag1*^{-/-}) (n=4). Δ *Rag1* (n=6) are mouse 25, 37, 41, 48, 49 and 51, each with different indels (Table 1). Two tailed, unpaired T-tests between Δ *Rag1* and *wt* (***) p < 0.001, **** p < 0.0001). There was no significant difference between *Rag1 H836Q* and *wt*.

In the bone marrow the Δ *Rag1* showed a complete lack of pre-B-cells (B220⁺ CD43⁻ IgM⁻), immature B-cells (B220^{low} CD43⁻ IgM⁺) and mature B-cells (B220^{hi} CD43⁻ IgM⁺) and an arrest at pre-pro and pro-B-cell (B220⁺ CD43⁺ IgM⁻) stage of B-cell differentiation (**Fig. 4A** and **4B**). A similar pattern was observed in *Rag1*^{-/-} mice. In contrast, B-cell counts and cell subset distribution were comparable in *Rag1 H836Q* and *wt* mice.

In the spleen, Δ *Rag1* mice showed marked reduction of total cellularity, associated with lack of CD4⁺ and CD8⁺ cells (**Fig. 5A** and **5B**, p. 74). Furthermore, only B220⁺ IgM⁻ B-cells, but no mature B220⁺ IgM⁺ B-cells were detected (**Fig. 5C** and **5D**). The IgM⁻ B-cells were all CD43⁺, corresponding to pro-B-cells, whereas in *wt* mice a proportion of the IgM⁻ B-cells were CD43⁻, representing pre-B-cells (**Fig. 5E** and **5F**, p. 74).

By contrast, the percentages and absolute counts of CD4⁺ and CD8⁺ T-cells were comparable in *Rag1 H836Q* and *wt* mice (**Fig. 5A** and **5B**, p. 74), and so were the percentages and absolute counts of mature B220⁺ IgM⁺ B-cells (**Fig. 5C** and **5D**), marginal zone B-cells (MZ) and follicular B-cells (Fo) (**Fig. 6A** and **6B**, p. 75). Finally, CD4⁺ and CD8⁺ cells had a normal distribution of activated (CD44⁺CD62L⁻), memory (CD44⁺CD62L⁺) and naïve (CD4⁻CD62L⁺) cells in *Rag1 H836Q* mice compared to *wt* mice (**Fig. 6C-F**, p. 75).

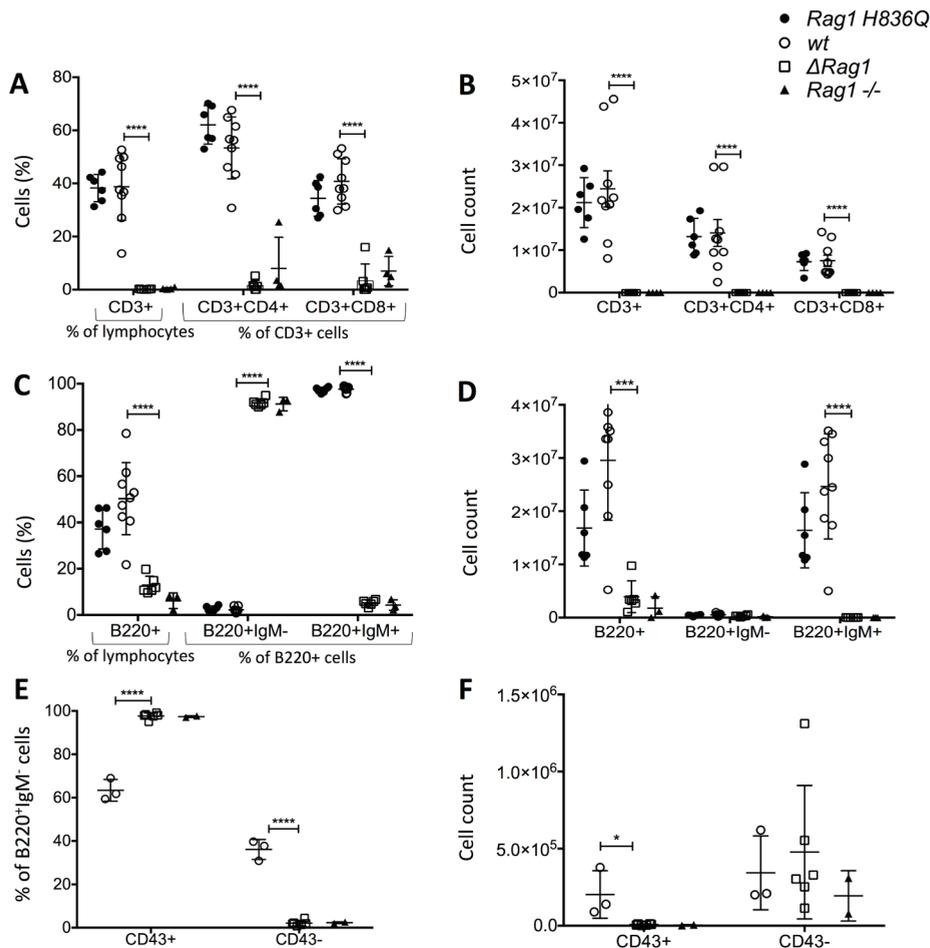


Figure 5: FACS Spleen. [A] CD3⁺ % of lymphocytes (SSC, FSC gate) and CD4⁺ and CD8⁺ as % of CD3⁺ cells. [B] CD3, CD4, CD8 absolute cell counts of 5.A. [C] B220⁺ % of lymphocytes (SSC, FSC gate) and precursors and immature (B220⁺IgM⁻) vs mature B-cells (B220⁺IgM⁺) (% of B220⁺ cells). [D] Absolute cell counts of 5.C. [E] Gated on B220⁺IgM⁻ cells to distinguish pre-B (CD43⁻) cells from pro-B-cells (CD43⁺). [F] Absolute cell counts of 5.E. Δ *Rag1* compared to *wt* and *Rag1*^{-/-}. Two tailed, unpaired t-tests between Δ *Rag1* and *wt* (* = $p \leq 0.05$, *** = $p \leq 0.001$, **** = $p \leq 0.0001$). No significant difference between *Rag1* H836Q and *wt* (5.A-D, not done for 5.E and 5.F).

Discussion

CRISPR/Cas9 is a highly efficient and versatile tool for performing genome-editing in mammalian cells *in vitro* and for generating animal models for *in vivo* studies. Here we showed that CRISPR/Cas9 can be efficiently used to generate several unique *Rag1* murine models in six weeks counting from the start of zygote injection to an F0 generation of weaning age, circumventing months of breeding as required for the traditional ES cell blastocyst injection. In addition,

no complicated cloning of constructs is required, and several models can be obtained at once. We showed that when testing different gRNAs *in vitro* and selecting the gRNA with the highest DNA cleavage efficiency, zygote injection can lead to a targeting efficiency of 100%, as was seen previously for other genes using the same injection conditions²⁴. Using this strategy, we were able to evaluate the effects of in-frame deletions in residues 832-877 of *RAG1*, a region of unclear function that does not contain any catalytic residues nor is involved in zinc binding. The mice we studied had deletions of one (residue 836 in mouse 49 and residue 832 in mouse 48), two (residue 832 and 838 in

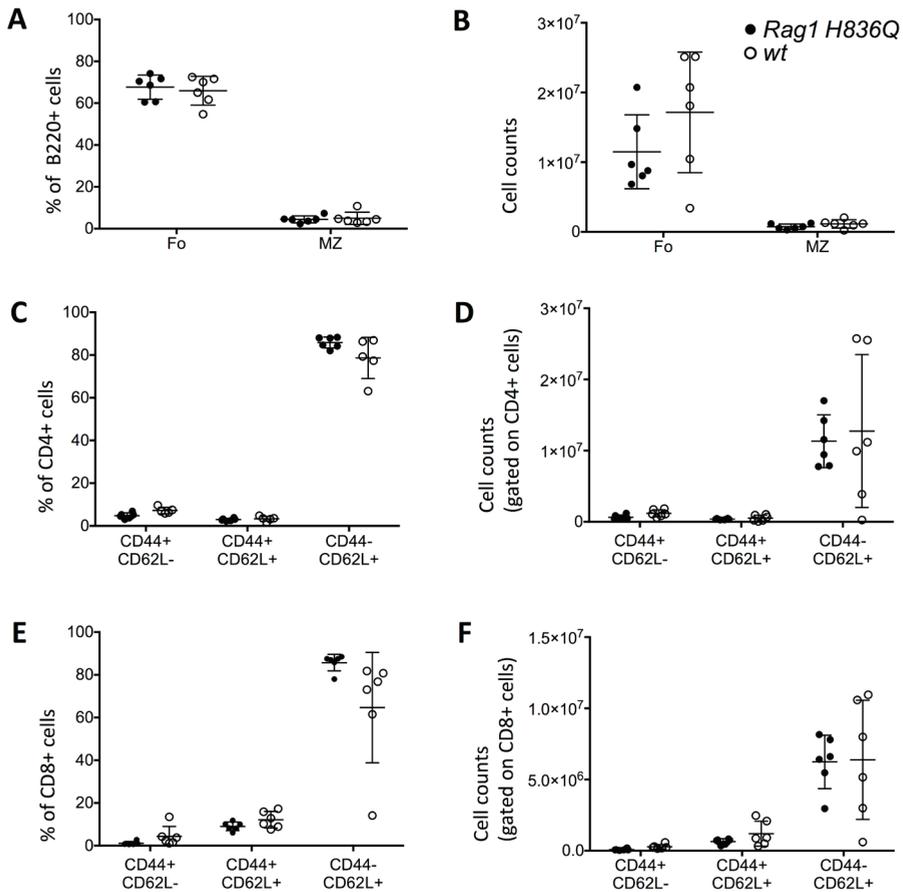


Figure 6: FACS Spleen. [A] Follicular (Fo) (B220⁺, CD93⁻, CD23⁺CD21⁻), and marginal zone (MZ) (B220⁺, CD93⁻, CD23⁻CD21⁻) B-cells in the spleen (% of B-cells). [B] Absolute cell counts of 6.A. [C] Activated (CD44⁺CD62L⁻), memory (CD44⁺CD62L⁺) and naïve (CD4⁻CD62L⁺) CD4⁺ T-cell subsets in the spleen (% of CD4⁺). [D] Absolute cell counts of 6.C. [E] Activated (CD44⁺CD62L⁻), memory (CD44⁺CD62L⁺) and naïve (CD4⁻CD62L⁺) CD8⁺ T-cell subsets in the spleen (% of CD8⁺). [F] Absolute cell counts of 6.E. Two tailed, unpaired t-tests between *Rag1 H836Q* (n=6) and *wt* (n=6). No significant difference between *Rag1 H836Q* and *wt*.

Target	Microhomology Score	Out-of-frame Score	
AGAGGAAAGGAAGAGATGGCAGGCCACGCTGGACAAACATCTCCGGAAAAGGATGAACTTAAAACCAATCATGAGGATGA	5312.1	69.3680465353	
Predicted Patterns	Microhomology	Length	Score
AGAGGAAAGGAAGAGATGGCAGGCCACGCTGGACAAACATCTCCGGAAAAGGATGAACTTAAAACCAATCATGAGGATGA	Wild Type		
AGAGGAAAGGAAGAGATGGCAGGCCACGCTGGA-----AAAGGATGAACTTAAAACCAATCATGAGGATGA	GGA	14	248.5
AGAGGAAAGGAAGAGATGGCAGGCCACGCTGGA-----TGAACTTAAAACCAATCATGAGGATGA	GGA	20	184.0
AGAGGAAAGGAAGAGATGGCAGGCCACG-----GAAAAGGATGAACTTAAAACCAATCATGAGGATGA	CG	17	170.8
AGAGGAAAGGAAGAGATGGCAGGCCACGCTGGACAAA-----AGGATGAACTTAAAACCAATCATGAGGATGA	AAA	12	164.7
AGAGGAAAGGAAGAGATGGCAGGCCACGCT-----CCGGAAAAGGATGAACTTAAAACCAATCATGAGGATGA	CT	12	164.7
AGAGGAAAGGAAGAGATGGCAGGCCACGCTGGACAAA-----GGATGAACTTAAAACCAATCATGAGGATGA	AAA	13	156.6
AGAGGAAAGGAAGAGATGGCAGGCC-----GGAAAAGGATGAACTTAAAACCAATCATGAGGATGA	CC	19	154.8
AGAGGAAAGGAAGAGATGGCAGGCCACGCTGGACAAAC-----TAAAACCAATCATGAGGATGA	AAC	20	147.2
AGAGGAAAGGAAGAGATGGCAGGCCACGCTGGACAAAC-----CAATCATGAGGATGA	AAAC	27	129.5
AGAGGAAAGGAAGAGATGGCAGG-----AAAAGGATGAACTTAAAACCAATCATGAGGATGA	GG	23	126.8

Figure 7: Microhomology prediction. Screenshot of output generated by online tool to predict microhomology mediated repair after cleavage by Cas9 with gRNA A. Input: 80bp around cutting site of gRNA A. (<http://www.rgenome.net/mich-calculator/>).

mouse 48), nine (residue 836-844 in mouse 37) or forty amino acids (residue 837-876 in mouse 25) (**Table 1**, p. 69). All of these in-frame deletions led to a severe phenotype, with a B-cell development block at the stage of pre-pro and pro-B-cells and a T-cell development block at DN3.

The use of various algorithms had yielded controversial results with regard to pathogenicity of the H836Q mutation. However, structural modeling studies suggested that this would be a neutral change. Indeed, analysis of the *in vivo* immunological phenotype of mice with a homozygous H836Q mutation showed intact T and B-cell development. Altogether, these data indicate the superiority of structural modeling over *in silico* predictive pathogenicity tools, and reinforce the importance of *in vivo* validation to assess the functional effects of missense mutations, especially in this locus that was not tested *in vivo* in mice.

Although CRISPR/Cas9 has been shown to be a successful and highly efficient genome editing tool for generating *in vivo* animal models of different species^{24,29-31}, somatic mosaicism may arise with this approach, because after pro-nucleus injection, CRISPR/Cas9-mediated cleavage can occur multiple times at various cell number stages in the morula²⁹, as was observed in mouse 31 in this study. In such cases, sequential breeding is required to segregate the various alleles.

One concern of the CRISPR/Cas9 technique is off-site targeting. We selected gRNAs with as little predicted off-site-targeting activity as possible. At the time of gRNA selection, NCBI blast was used to select the best gRNAs, as described in the Methods section. When novel sequence analysis tools became available (<http://crispr.mit.edu/>)³², the predicted off-site targeting activity was also investigated. For gRNA A, there were only 16 possible off-site targets, of which one had 3 mismatches (in position 1, 3 and 19 of the protospacer starting from the PAM site), and all of the others all had at least 4 mismatches, with at least 2 falling in the first 10 bp of the protospacer sequence. It has been reported that mismatches in the first 10 bp of the protospacer are most stringent and make off-site mutagenesis very unlikely³³. None of the possible off-site targets were in genes known to cause immunodeficiency.

The different indels that were generated by CRISPR/ Cas9 induced cleavage were not completely random. For example, there were 13 mice with 14bp frameshift indels, and only 4 different unique sequences (**Table 1**, p. 69). All but one of these sequences showed microhomology of 2 or 3 bp. When using the online tool to predict microhomology for the 80bp around the cutting site, the 14bp sequence that was found in 6 mice, was indeed on top of the list with the highest score for microhomology-mediated repair (<http://www.rgenome.net/mich-calculator/>)³⁴ (**Fig. 7**). This is important to realize when selecting a target site and a gRNA.

In summary, we have shown that CRISPR/Cas9 is a highly efficient tool to rapidly generate different mouse models of Rag1 deficiency and to study the functional consequences of such mutations. If a similar strategy is coupled with injection of single-stranded homology DNA template into the zygote, novel knock-in models can be generated rapidly and with high efficiency by means of HDR^{24,26}. This approach would be particularly relevant for modeling the phenotypic heterogeneity that is associated with RAG deficiency in humans by generating multiple mouse models in a short period of time.

ACKNOWLEDGMENTS

The authors thank Marianne Boes, PhD, University Medical Center Utrecht, Utrecht, The Netherlands, for her critical revision of this manuscript. This work was supported by grant R01AI100887 from the National Institute of Health, National Institute of Allergy and Infectious Diseases (to LN) and March of Dimes grant MOD - 6FY10-282 (to). LO is supported by T32 fellowship grant 5T32AI007512. The research of WY and MG is supported by the intramural program of the National Institute of Diabetes and Digestive and Kidney Diseases, NIH. JM is supported by grant R21A1088510 from the National Institute of Health, National Institute of Allergy and Infectious Diseases.

Supplemental figures

Representative dot plots with FACS gating strategies for each group.

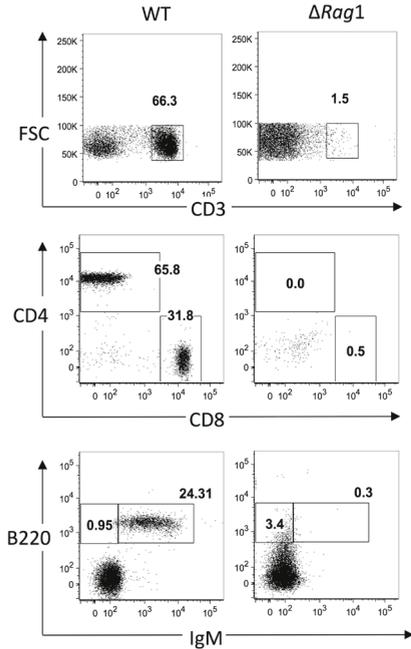
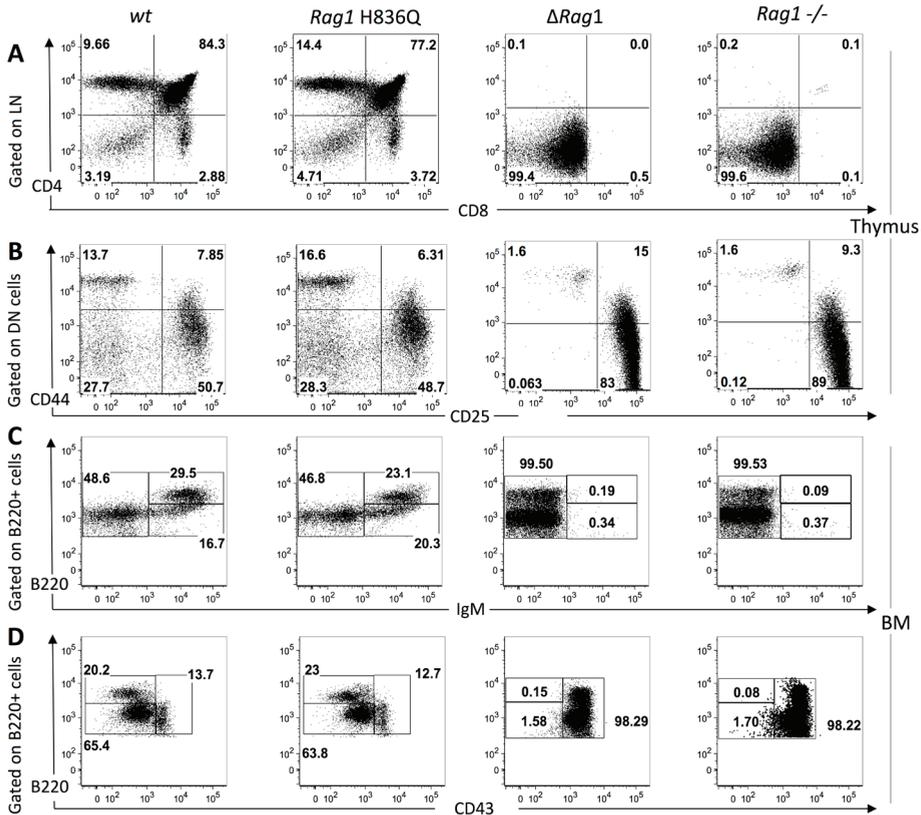


Figure S1 (←): FACS peripheral blood. FACS gating strategy to assess T and B cells in peripheral blood of deltaRag1 mice.

Figure S2 (↓): FACS Thymus and Bone Marrow. [A] FACS gating strategy for thymocyte subsets in Fig. 3B-C. CD4⁺, CD8⁺, CD4⁺CD8⁺ (DP: Double Positive) and CD4⁻CD8⁻ (DN: Double Negative). [B] FACS gating strategy for subsets of DN in Fig. 3D-E. DN1 (CD44⁺CD25⁻), DN2 (CD44⁺CD25⁺), DN3 (CD44⁻CD25⁺), DN4 (CD44⁻CD25⁻). [C] and [D] FACS gating strategy for B220⁺ cells in the bone marrow as described in Fig. 4A-B.



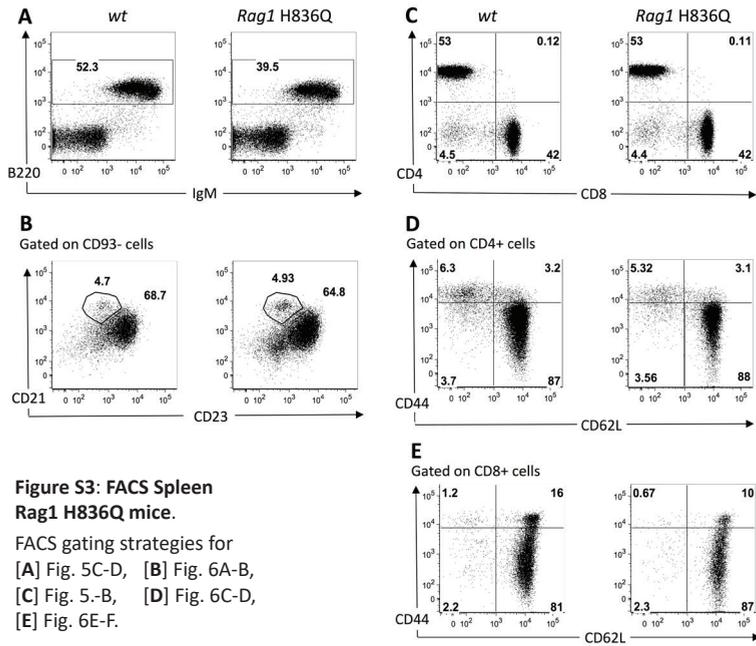


Figure S3: FACS Spleen Rag1 H836Q mice.

FACS gating strategies for [A] Fig. 5C-D, [B] Fig. 6A-B, [C] Fig. 5-B, [D] Fig. 6C-D, [E] Fig. 6E-F.

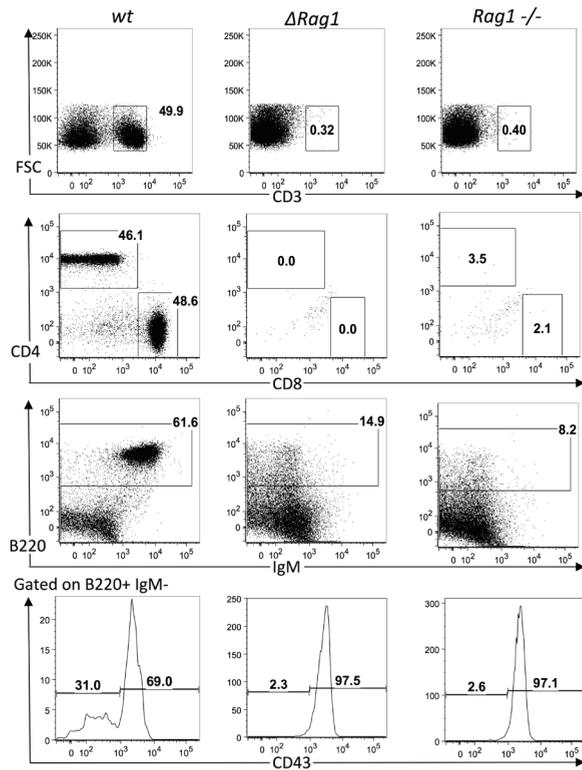


Figure S4: FACS spleen Δ Rag1 mice. FACS gating strategies for Fig. 5.

References

1. Schatz DG. V(D)J recombination. *Immunol Rev.* 2004;200:5-11.
2. Schwarz K, Gauss GH, Ludwig L, et al. RAG mutations in human B cell-negative SCID. *Science.* 1996;274(5284):97-99.
3. Villa A, Santagata S, Bozzi F, et al. Partial V(D)J recombination activity leads to Omenn syndrome. *Cell.* 1998;93(5):885-896.
4. Corneo B, Moshous D, Gungor T, et al. Identical mutations in RAG1 or RAG2 genes leading to defective V(D)J recombinase activity can cause either T-B-severe combined immune deficiency or Omenn syndrome. *Blood.* 2001;97(9):2772-2776.
5. Sobacchi C, Marrella V, Rucci F, Vezzoni P, Villa A. RAG-dependent primary immunodeficiencies. *Hum Mutat.* 2006;27(12):1174-1184.
6. Pasic S, Djuricic S, Ristic G, Slavkovic B. Recombinase-activating gene 1 immunodeficiency: different immunological phenotypes in three siblings. *Acta Paediatr.* 2009;98(6):1062-1064.
7. Dalal I, Tasher D, Somech R, et al. Novel mutations in RAG1/2 and ADA genes in Israeli patients presenting with T-B-SCID or Omenn syndrome. *Clin Immunol.* 2011;140(3):284-290.
8. Cassani B, Poliani PL, Moratto D, et al. Defect of regulatory T-cells in patients with Omenn syndrome. *J Allergy Clin Immunol.* 2010;125(1):209-216.
9. Villa A, Sobacchi C, Notarangelo LD, et al. V(D)J recombination defects in lymphocytes due to RAG mutations: severe immunodeficiency with a spectrum of clinical presentations. *Blood.* 2001;97(1):81-88.
10. de Villartay JP, Lim A, Al-Mousa H, et al. A novel immunodeficiency associated with hypomorphic RAG1 mutations and CMV infection. *J Clin Invest.* 2005;115(11):3291-3299.
11. Ehl S, Schwarz K, Enders A, et al. A variant of SCID with specific immune responses and predominance of gamma delta T cells. *J Clin Invest.* 2005;115(11):3140-3148.
12. Schuetz C, Huck K, Gudowius S, et al. An immunodeficiency disease with RAG mutations and granulomas. *N Engl J Med.* 2008;358(19):2030-2038.
13. De Ravin SS, Cowen EW, Zarembek KA, et al. Hypomorphic Rag mutations can cause destructive midline granulomatous disease. *Blood.* 2010;116(8):1263-1271.
14. Henderson LA, Frugoni F, Hopkins G, et al. Expanding the spectrum of recombination-activating gene 1 deficiency: a family with early-onset autoimmunity. *J Allergy Clin Immunol.* 2013;132(4):969-971 e961-962.
15. Kuijpers TW, Ijspeert H, van Leeuwen EM, et al. Idiopathic CD4+ T lymphopenia without autoimmunity or granulomatous disease in the slipstream of RAG mutations. *Blood.* 2011;117(22):5892-5896.
16. Lee YN, Frugoni F, Dobbs K, et al. A systematic analysis of recombination activity and genotype-phenotype correlation in human recombination-activating gene 1 deficiency. *J Allergy Clin Immunol.* 2013.
17. Zhang YH, Shetty K, Surleac MD, Petrescu AJ, Schatz DG. Mapping and Quantitation of the Interaction between the Recombination Activating Gene Proteins RAG1 and RAG2. *J Biol Chem.* 2015;290(19):11802-11817.
18. Kim MS, Lapkouski M, Yang W, Gellert M. Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature.* 2015;518(7540):507-511.
19. Mombaerts P, Iacomini J, Johnson RS, Herrup K, Tonegawa S, Papaioannou VE. RAG-1-deficient mice have no mature B and T lymphocytes. *Cell.* 1992;68(5):869-877.
20. Shinkai Y, Rathbun G, Lam KP, et al. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell.* 1992;68(5):855-867.
21. Marrella V, Poliani PL, Casati A, et al. A hypomorphic R229Q Rag2 mouse mutant recapitulates human Omenn syndrome. *J Clin Invest.* 2007;117(5):1260-1269.
22. Giblin W, Chatterji M, Westfield G, et al. Leaky severe combined immunodeficiency and aberrant DNA rearrangements due to a hypomorphic RAG1 mutation. *Blood.* 2009;113(13):2965-2975.
23. Khiong K, Murakami M, Kitabayashi C, et al. Homeostatically proliferating CD4 T cells are involved in the pathogenesis of an Omenn syndrome murine model. *J Clin Invest.* 2007;117(5):1270-1281.
24. Wang H, Yang H, Shivalila CS, et al. One-step generation of mice carrying mutations in

multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*. 2013;153(4):910-918.

25. Yang H, Wang H, Jaenisch R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat Protoc*. 2014;9(8):1956-1968.

26. Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell*. 2013;154(6):1370-1379.

27. Mali P, Yang L, Esvelt KM, et al. RNA-guided human genome engineering via Cas9. *Science*. 2013;339(6121):823-826.

28. Frock RL, Hu J, Meyers RM, Ho YJ, Kii E, Alt FW. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol*. 2014.

29. Niu Y, Shen B, Cui Y, et al. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell*. 2014;156(4):836-843.

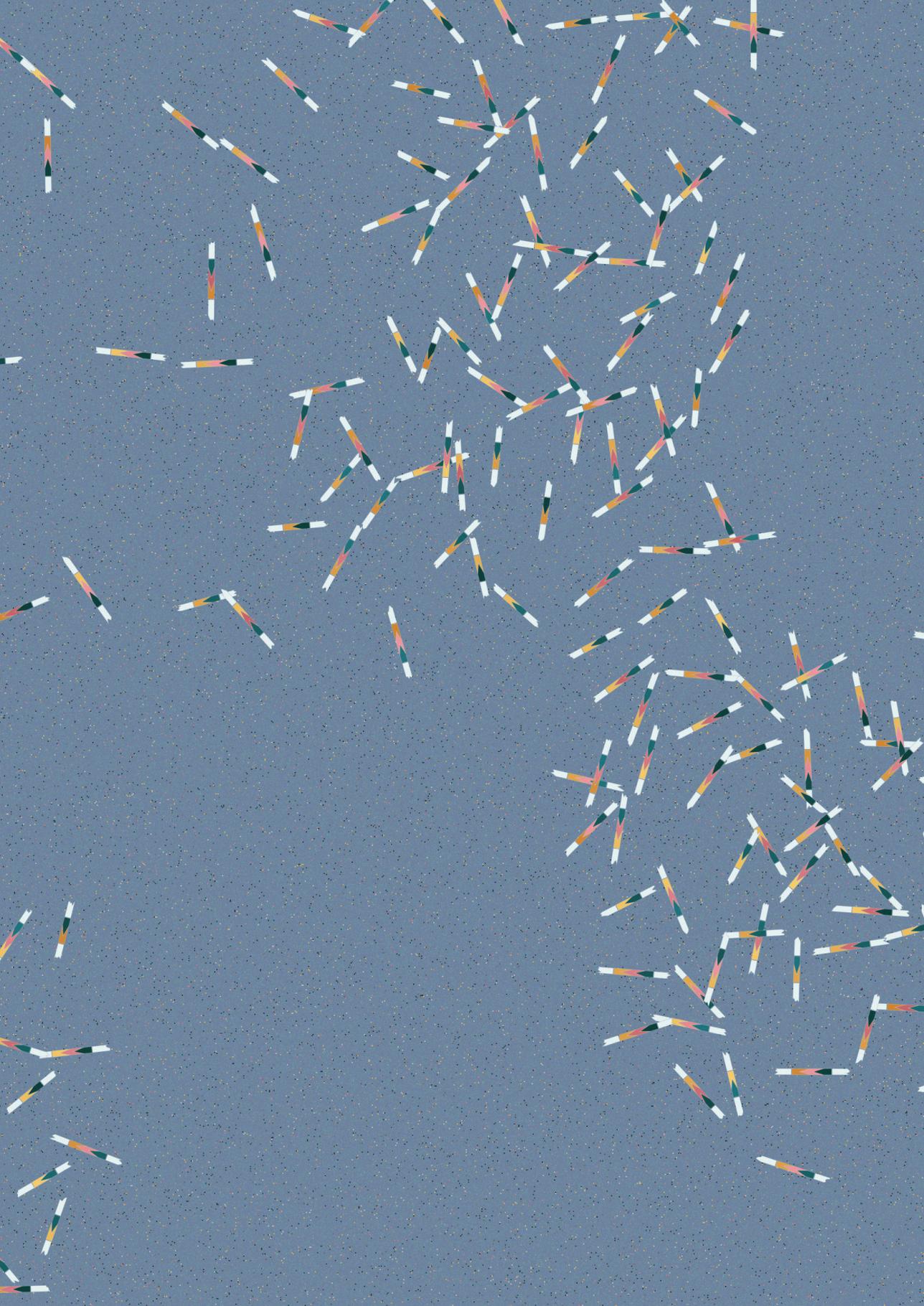
30. Shen B, Zhang J, Wu H, et al. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res*. 2013;23(5):720-723.

31. Li W, Teng F, Li T, Zhou Q. Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat Biotechnol*. 2013;31(8):684-686.

32. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281-2308.

33. Fu Y, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*. 2013;31(9):822-826.

34. Bae S, Kweon J, Kim HS, Kim JS. Microhomology-based choice of Cas9 nuclease target sites. *Nat Methods*. 2014;11(7):705-706.



5 Hypomorphic *Rag1* mutations alter the preimmune repertoire at early stages of lymphoid development

Blood. 2018 Jul 19;132 (3):281-292



L. M. Ott de Bruin^{1,2}
M. Bosticardo³
A. Barbieri⁴
S. G. Lin⁵
J. H. Rowe¹
P. L. Poliani⁶

K. Ching⁴
D. Eriksson⁷
N. Landegren⁷
O. Kämpe⁷
J. P. Manis⁴
L. D. Notarangelo³

¹ Division of Immunology, Boston Children's Hospital, Harvard Medical School, Boston, MA

² Department of Pediatric Immunology, Wilhelmina Children's Hospital, Utrecht University Medical Center, Utrecht, The Netherlands

³ Laboratory of Clinical Immunology and Microbiology, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

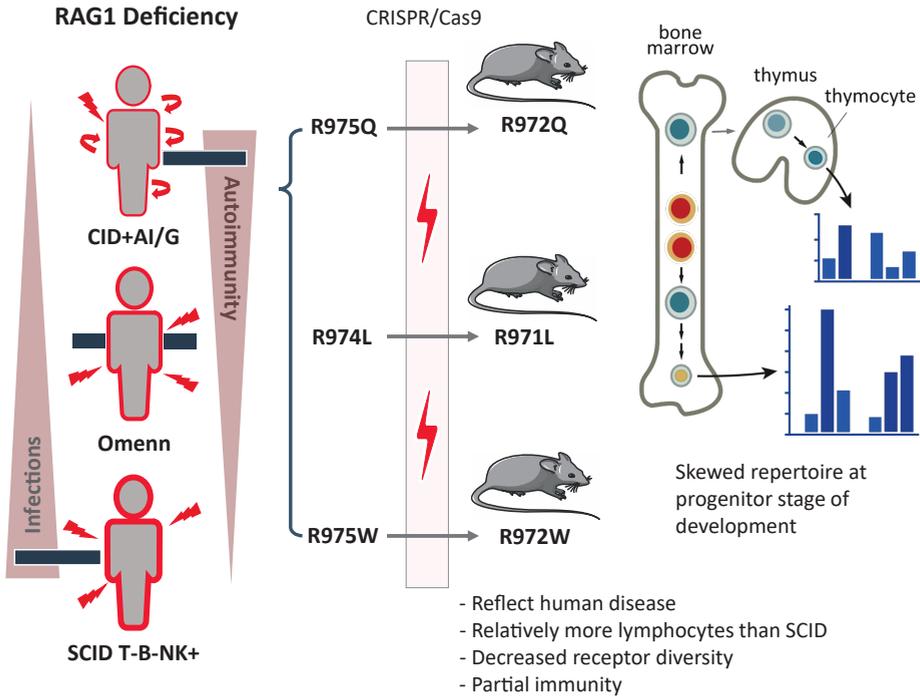
⁴ Department of Laboratory Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA

⁵ Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy

⁶ Center for Molecular Medicine, Department of Medicine (Solna), Karolinska Institutet, Stockholm, Sweden

Abstract

Hypomorphic *RAG1* mutations allowing residual T- and B-cell development have been found in patients presenting with delayed-onset combined immune deficiency with granulomas and/or autoimmunity (CID-G/AI) and abnormalities of the peripheral T- and B-cell repertoire. To examine how hypomorphic *Rag1* mutations affect the earliest stages of lymphocyte development, we used CRISPR/Cas9 to generate mouse models with mutations equivalent to those found in patients with CID-G/AI. Immunological characterization showed partial development of T and B lymphocytes, with persistence of naïve cells and preserved serum immunoglobulin but impaired antibody responses and presence of autoantibodies, thereby recapitulating the phenotype seen in patients with CID-G/AI. By using high-throughput sequencing, we identified marked skewing of *IghV* and *TrbV* gene usage in early progenitors, with a bias for productive *Igh* and *Trb* rearrangements after selection occurred and increased apoptosis of B-cell progenitors. Rearrangement at the *Igk* locus was impaired, and polyreactive immunoglobulin M antibodies were detected. This study provides novel insights into how hypomorphic *Rag1* mutations alter the primary repertoire of T and B-cells, setting the stage for immune dysregulation frequently seen in patients.



Introduction

Adaptive immunity relies on the dynamic response of lymphocytes to generate specific antigen receptors to fight pathogens. Recombination-activating gene 1 (RAG1) and 2 (RAG2) are crucial for effective combinatorial joining of variable (*V*), diversity (*D*), and joining (*J*) genes that encode the antigen-binding regions of T-cell receptor (TCR) and B-cell immunoglobulin molecules.¹ More than 200 disease-causing mutations in the *RAG1* and *RAG2* genes have been identified that can cause a wide spectrum of clinical and immunological phenotypes.² In particular, functionally null mutations cause a complete arrest of T- and B-cell development, resulting in T-B- severe combined immunodeficiency.³⁻⁵ Hypomorphic mutations allowing minimal residual function of RAG can lead to Omenn syndrome, with presence of a variable number of activated, oligoclonal T-cells that infiltrate and damage target tissues.⁶ By contrast, hypomorphic mutations with higher residual activity have been identified in patients with delayed-onset combined immunodeficiency associated with granulomas and/or autoimmunity (CID-G/AI).⁷

A significant proportion of patients with CID-G/AI carry missense mutations in the coding flank-sensitive region of the carboxy-terminal domain (CTD) of RAG1 (human amino acid 892-977; mouse amino acid 889-974; see **Fig. S1A**, p. 108). These mutations have been postulated to favor targeting of certain coding elements.⁸ Although abnormalities of the peripheral T- and B-cell repertoire have been observed in patients with CID-G/AI and *Rag1*-mutant mice,^{9,10} it is unknown whether similar abnormalities are present in the pre-immune repertoire of progenitor lymphocytes or whether they are secondary to antigen-mediated pressure in the periphery. We used gene editing to generate 3 mouse models carrying homozygous *Rag1* mutations (F971L, R972Q, and R972W), corresponding to human mutations (F974L, R975Q, and R975W) described in patients with CID-G/AI,^{7,11-13} to understand how these mutations affect repertoire composition, cell selection, and survival during T- and B-cell development.

Methods

Mice

Rag1-mutant mice were generated by gene editing as previously described.^{14,15} For the analysis of regulatory T-cells (T_{reg}) and conventional T-cells (T_{conv}), *Rag1*-mutant mice were crossed with Foxp3.EGFP-Cre mice. Animal work was conducted in specific pathogen-free conditions, in accordance with the US Public Health Service Policy on Humane Care and Use of Laboratory Animals, with protocols approved by Boston Children's Hospital (protocols 13-08-2472R and 16-05-3176R) and by the National Institute of Allergy and Infectious Diseases Animal Care and Use Committee (protocol LCIM 16E).

Flow cytometry, cell isolation, and histochemistry

Detailed methods for flow cytometry and cell isolation are provided in the Methods section, p. 121. Immunohistochemistry of thymus and spleen was performed as previously described.^{16,17}

Measurement of serum BAFF, immunoglobulin levels, specific antibody levels and autoantibodies

BAFF, immunoglobulin, and specific antibody serum levels were measured by enzyme-linked immunosorbent assay as described in the supplemental Methods. Serum immunoglobulin M (IgM) autoantibodies and anticytokine antibodies were measured using microarrays, as described in the Methods section, p. 121.

T- and B-cell repertoire analysis

High-throughput sequencing of TCR β (*Trb*) and immunoglobulin heavy chain (*Igh*) rearrangements was performed as described in the supplemental Methods. Analysis of V κ -J κ rearrangements in pre-B-cells was performed by polymerase chain reaction amplification as described.¹⁸

Statistical analysis

One-way analysis of variance with multiple comparisons was used when comparing wild-type (WT) with *Rag1*-mutant mice or comparing different *Rag1*-mutant mice. Two-way analysis of variance was used to compare antibody responses in WT and mutant mice. The Mann-Whitney *U* test was used when only 2 groups of mice were compared. Distribution of *V* and *J* gene usage was compared using the Kolmogorov-Smirnov test. Individual *V* and *J* gene usage was analyzed by the χ^2 test.

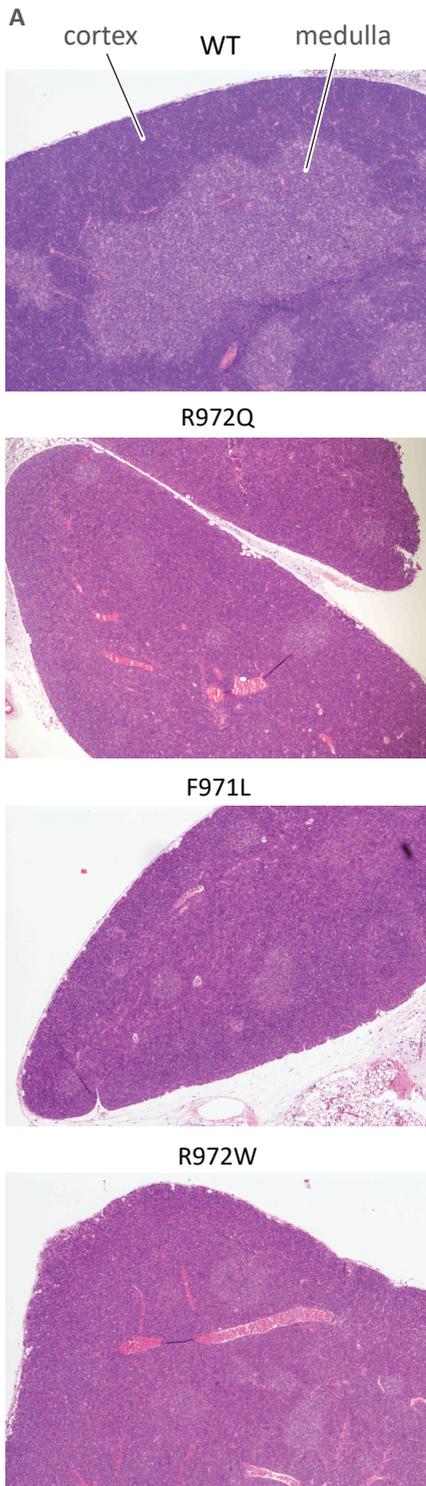
Results

Generation of mice with targeted mutations in RAG1 CTD

We selected 3 mutations (F971L, R972Q, and R972W) corresponding to human mutations (F974L, R975Q, and R975W) that have been previously described in patients with CID-G/AI. All 3 fall in the coding flank-sensitive region of RAG1 CTD⁸ (**Fig. S1A**, p. 108). Crystallography predicted that the R972 residue located near the catalytic amino acid E962 (**Fig. S1B**, p. 108) may participate in the recognition sequence specificity of the DNA coding flank that is directly adjacent to the recombination signal sequence.¹⁹ On the basis of amino acid properties and *in vitro* studies,¹⁰ we predicted that the R972Q and the F971L mutations would have a moderate effect on RAG1 protein stability. To extend our analyses, we included a mutation (R972W) that protein structure and *in vitro* activity predicted to be highly disruptive.⁷

Incomplete block of T- and B-cell development in *Rag1*-mutant mice

All 3 *Rag1*-mutant mouse models showed abnormalities of thymic architecture that were most prominent in R972W mice, with a marked decrease of the thymic medulla but partial preservation of the corticomedullary demarcation (**Fig. 1A**, p. 88). The total number of thymocytes was reduced 50- to 100-fold in all 3 models (**Fig. 1B**). Flow cytometric analysis revealed a near-complete block of thymocyte development at the CD4⁻CD8⁻ DN stage in R972W mice, similar to what was observed in *Rag1*^{-/-} (knockout) mice, whereas a significant fraction of CD4⁺CD8⁺ DP cells and CD4⁺CD8⁻ and CD4⁻CD8⁺ single-positive cells were detected in R972Q and F971L mice (**Fig. 1C**; **Fig. S2A**, p. 109).



In all mutant mice, DN cells were largely blocked at $CD25^+CD44^-$ DN3 stage (**Fig. 1D**; **Fig. S2B**, p. 109). An increase in the relative proportion of $\gamma\delta$ T-cells was noted in the 2 leakiest models (R972Q and F971L; **Fig. 1E**; **Fig. S2C**, p. 109).

In the bone marrow, a significant increase in the proportion of $B220^{lo}IgM^-CD43^+$ cells (pro-B-cells and pre-BI, here collectively called pro-B) was seen in *Rag1*-mutant mice as compared with WT mice (**Fig. 2A-B**, p. 90). By contrast, the proportion and absolute number of $B220^{lo}IgM^-CD43^-$ cells (small pre-B, including pre-BII), $B220^{lo}IgM^+CD43^-$ cells (immature), and $B220^{hi}IgM^+CD43^-$ cells (mature recirculating B-cells) were all decreased in *Rag1*-mutant mice (**Fig. 2A-C**, p. 90). This block in

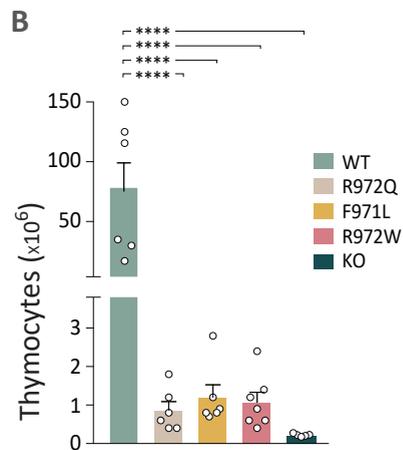


Figure 1: Thymic T-cell development in *Rag1* mutant mice

[A] Tissue sections of thymus stained with H&E for medulla (lighter staining) and cortex (darker staining). **[B]** Live cell counts from individual thymuses of wild-type (WT), hypomorphic *Rag1* mutant, and *Rag1* knock-out (KO) mice.

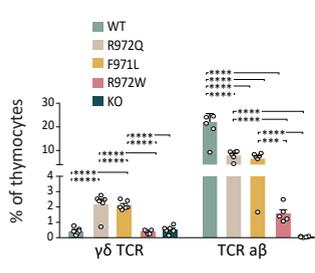
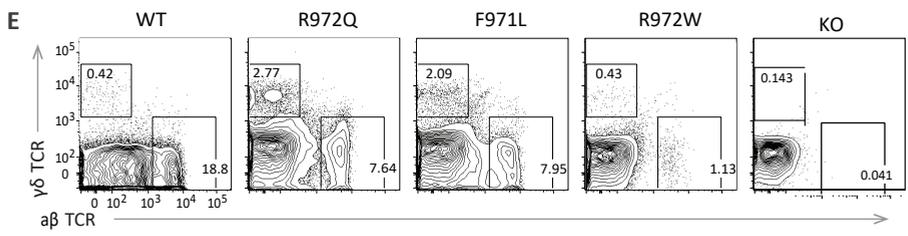
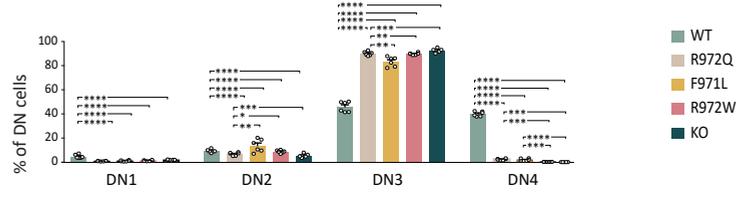
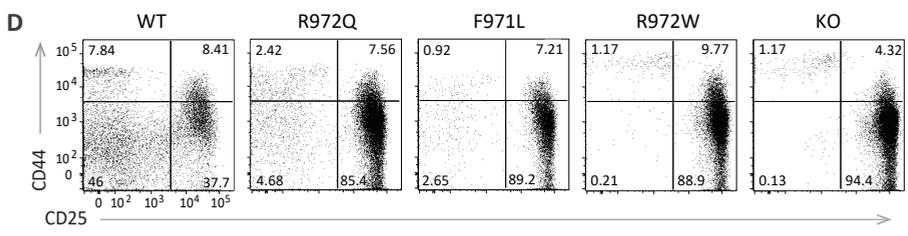
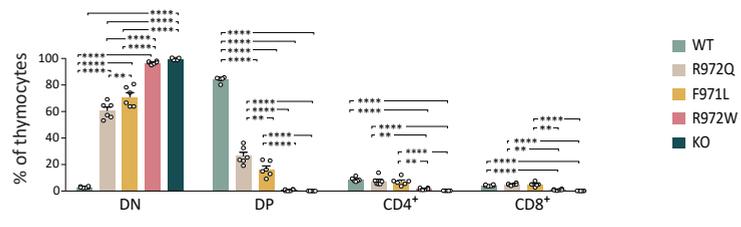
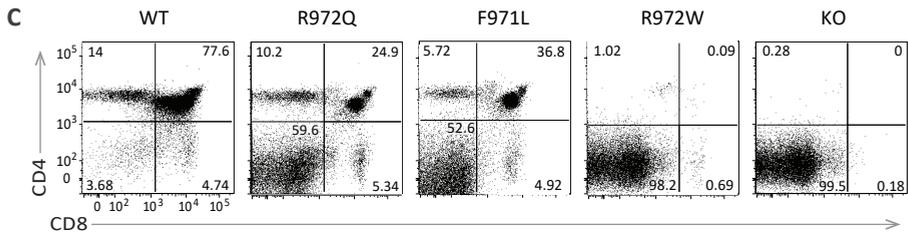
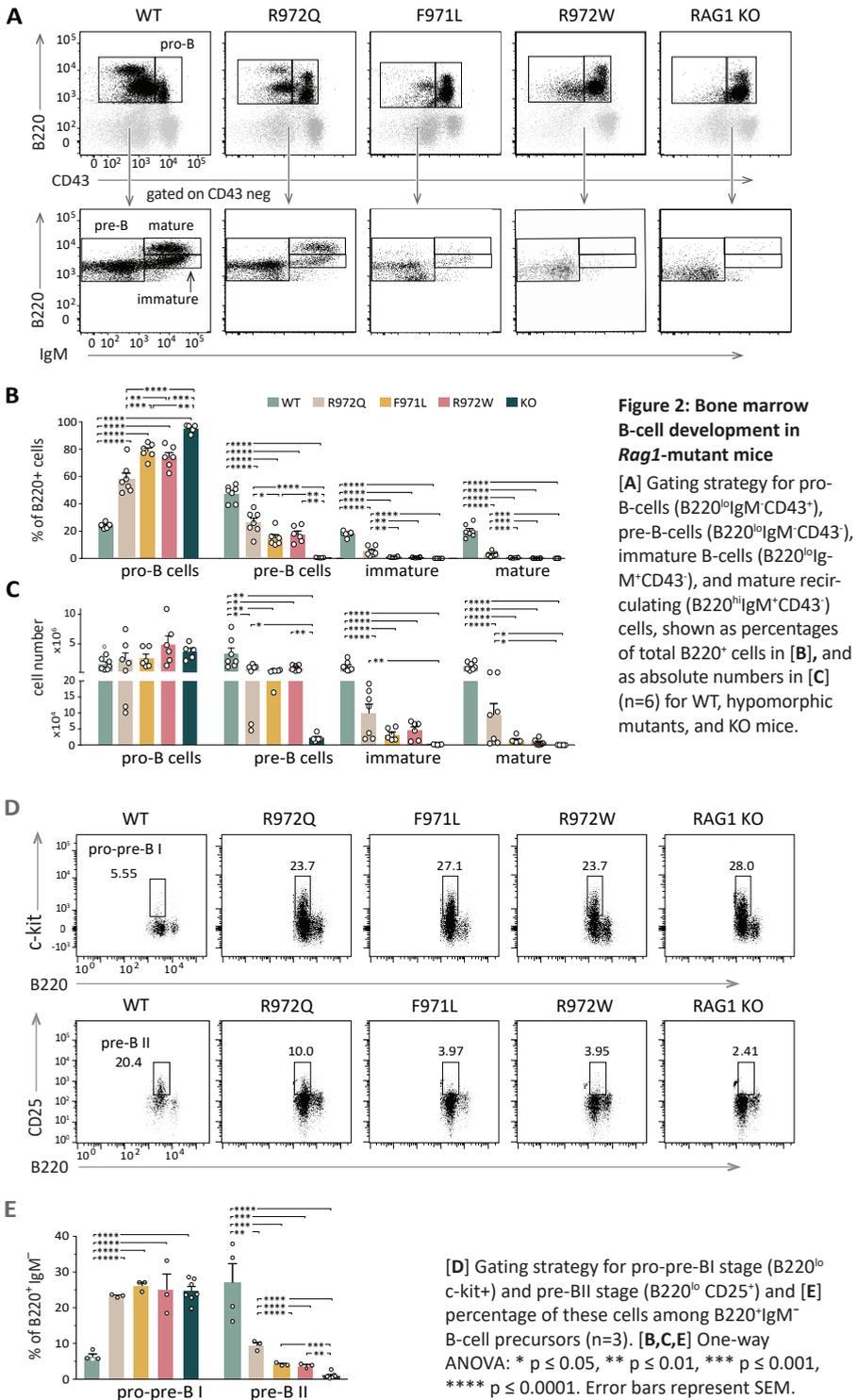


Figure 1 cont. Thymocyte developmental stages were analyzed by flow cytometry for [C] DN (CD4⁺CD8⁻), DP (CD4⁺CD8⁺) and the SP (CD4⁺ or CD8⁺) [D] Lineage negative DN populations, DN1 (CD44⁺CD25⁻), DN2 (CD44⁺CD25⁺), DN3 (CD44⁺CD25⁺), DN4 (CD44⁺CD25⁺) and [E] thymocytes expressing the αβ or γδ form of the TCR. Representative flow cytometry panels with n=6 thymuses per group (open circles). Error bars represent SEM. Statistical analysis with one-way ANOVA: * p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001, **** p ≤ 0.0001.



B-cell development was less severe in R972Q than in F971L and R972W mice. During B-cell development, RAG expression occurs at the pro-B stage for *Igh*²⁰ rearrangement and at the pre-BII stage for light chain (LC) rearrangement. To characterize these specific transitions, we performed flow cytometric analysis to identify the proportions of B220^{lo}c-kit⁺ pro/pre-BI and of B220^{lo}CD25⁺ pre-BII bone marrow cells. Among B220⁺ IgM⁻ B-cell precursors, a significant increase in the proportion of pro/pre-BI cells was demonstrated in all 3 hypomorphic *Rag1*-mutant mouse strains (**Fig. 2D-E**), indicating a major block at the assembly of LC. Although the proportion of B220^{lo}CD25⁺ pre-BII cells was significantly reduced in all 3 *Rag1* hypomorphic mutants compared with WT mice, this difference was less severe in R972Q mice (**Fig. 2E**), indicating a more pronounced leakiness of defective lymphocyte development in this model.

Rag1-mutant mice have a reduced number of mature T and B lymphocytes in the periphery

In the spleen, the number of CD3⁺, CD4⁺ and CD8⁺ cells was diminished in all 3 *Rag1*-mutant models, and this defect was more pronounced in F971L and R972W mice (**Fig. 3A-B**). In T-cell lymphopenic hosts, compensatory homeostatic proliferation can result in an increased proportion of peripheral T-cells with an activated phenotype and virtual absence of peripheral naïve T-cells.^{9,21,22} Although all 3 mutants showed depletion of naïve T-cells, a

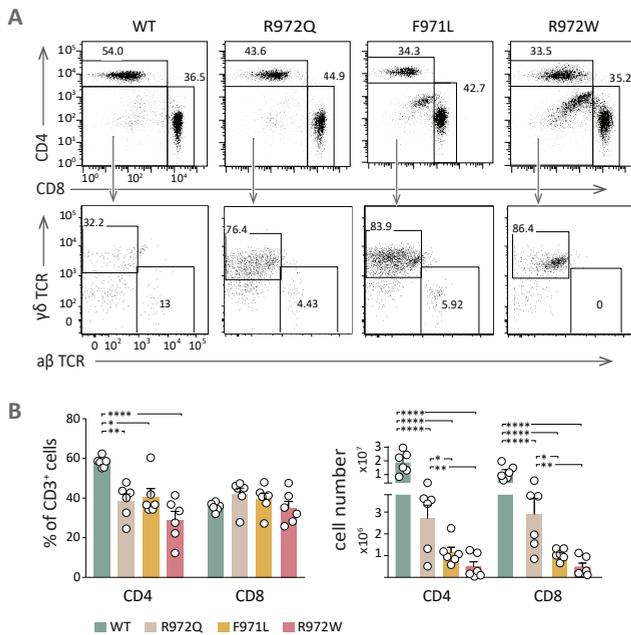


Figure 3. Distribution and phenotype of splenic T-cell subsets in *Rag1* mutant mice

A. Representative example of CD4⁺ and CD8⁺ T-cells (top), and of CD4⁻CD8⁻ cells expressing the $\gamma\delta$ form of the TCR (bottom) among splenic CD3⁺ cells. **B.** Percentages (among CD3⁺ cells), and, **C)** absolute count of CD4⁺ and CD8⁺ cells. One-way ANOVA: * p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001, **** p ≤ 0.0001.

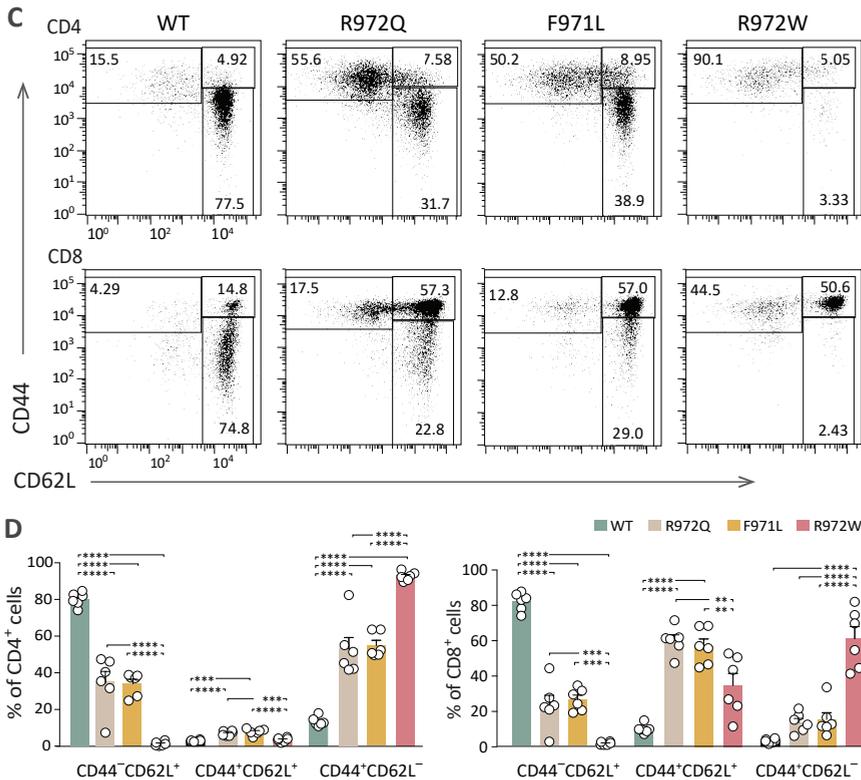
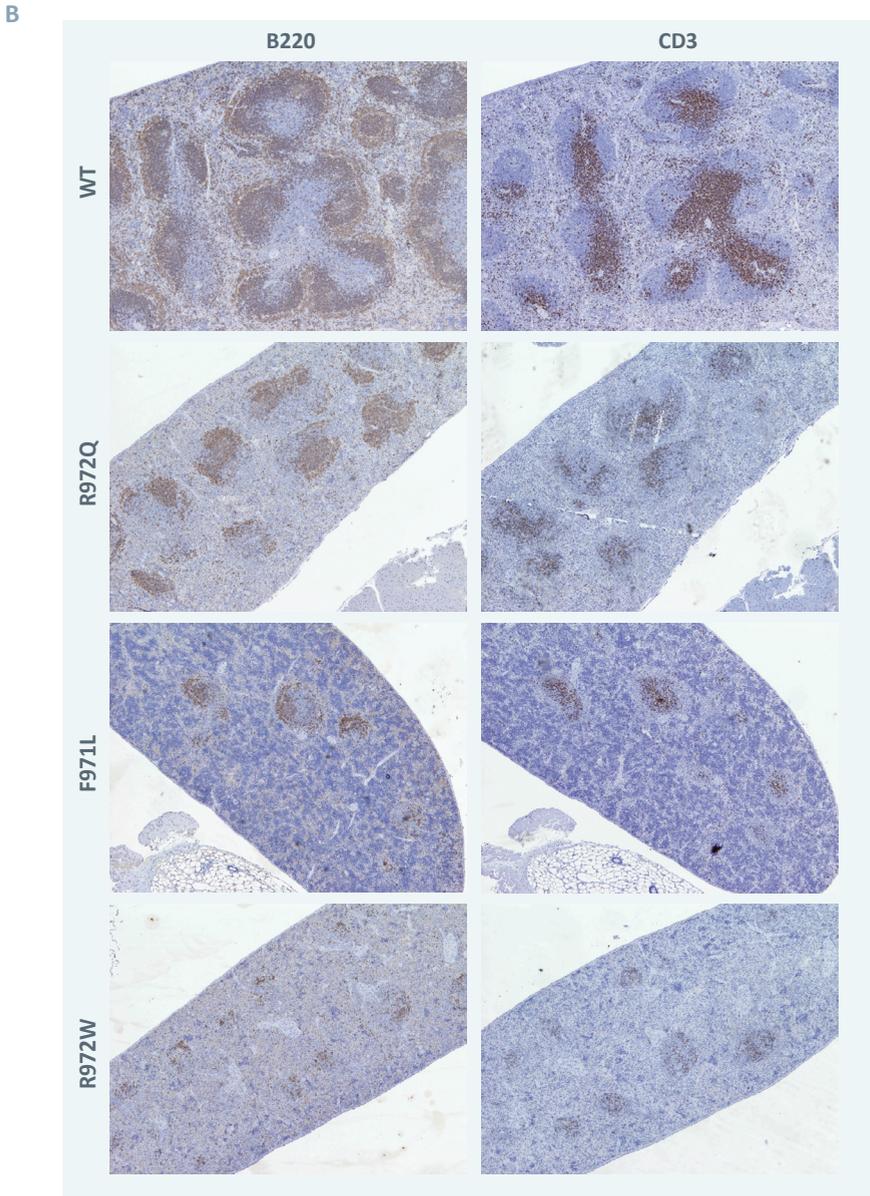
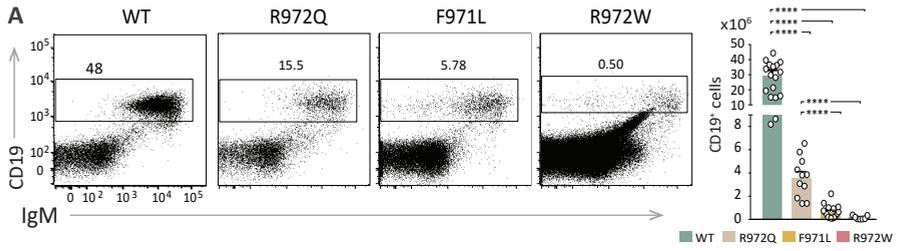


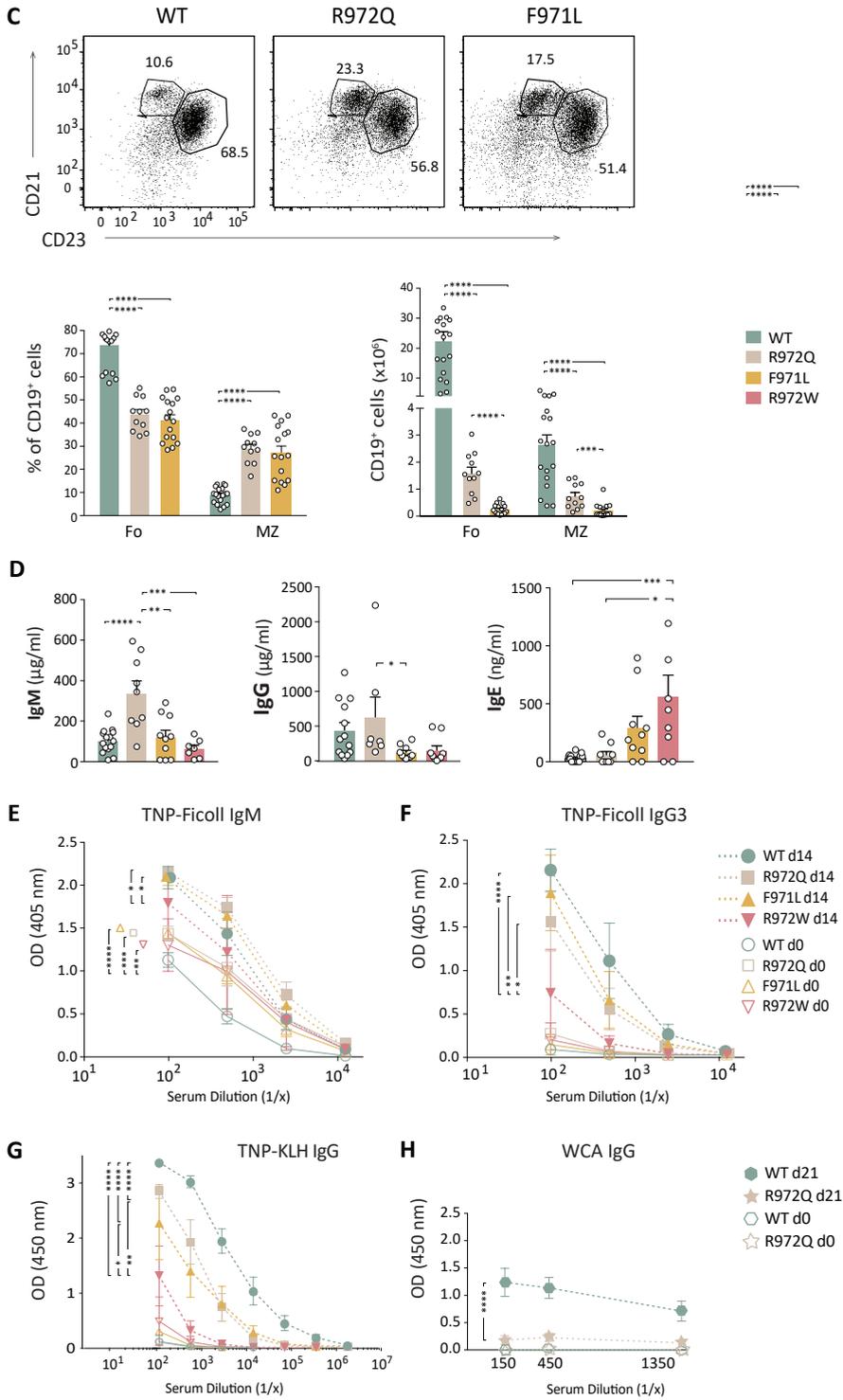
Figure 3: Distribution and phenotype of splenic T-cell subsets in *Rag1* mutant mice (cont.) [C] Gating strategy for naïve (CD62L⁺CD44⁻), central memory (CD44⁺CD62L⁺) and effector memory T-cells (CD44⁺CD62L⁻) among CD4⁺ (top) and CD8⁺ (bottom) T-cells. [D] Distribution of naïve, central memory and effector memory T-cells among total splenic CD4⁺ (top) and CD8⁺ (bottom) cells (n=6 per group). Error bars represent SEM. One-way ANOVA: * p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001, **** p ≤ 0.0001.

significant fraction of CD62L⁺CD44⁻ naïve cells was present in both F971L and R972Q mice (**Fig. 3C-D**). As compared with WT mice, the absolute number of B-cell splenocytes was reduced in all *Rag1*-mutant models and especially in R972W and F971L compared with R972Q mice (**Fig. 4A**). The number and size of B-cell follicles in the spleen correlated with the degree of B-cell lymphopenia (**Fig. 4B**). Analysis of the proportion and absolute number of follicular (CD19⁺CD93⁻CD21⁺CD23⁺) and MZ (CD19⁺CD93⁻CD21⁺CD23⁻) B-cells showed that B-cell depletion was more prominent for follicular than MZ B-cells and more severe in F971L than in R972Q mice (**Fig. 4C**).

Figure 4: Distribution and phenotype of splenic B-cells and humoral immunity in *Rag1* mutant mice [A] Representative examples (left) and absolute count (right) of mature CD19⁺ IgM positive splenic B-cells (error bars of SEM). [B] Sections of spleen stained for B220 and CD3 concentrated within follicles (dark staining).



[Figure continued on next page]



Dysregulation of humoral immunity in *Rag1*-mutant mice

In contrast to patients with Omenn syndrome, who are profoundly hypogammaglobulinemic but show increased levels of serum IgE, variable levels of serum IgG and IgM and of antigen-specific antibody responses have been reported in patients with CID-G/AI.^{11,13,23} R972Q mice had normal levels of IgG with increased serum IgM; increased serum IgE was detected in R972W mice (**Fig. 4D**). Naïve *Rag1*-mutant mice had higher serum titers of TNP-binding IgM (**Fig. 4E**). Upon immunization with the T-independent antigen TNP-Ficoll, equivalent amounts of anti-TNP IgM and IgG3 antibodies were detected in WT, R972Q, and F971L mice, whereas a defective response was observed in R972W mutants (**Fig. 4E-F**). The antibody response to the T-dependent antigen TNP-KLH was impaired in all 3 *Rag1*-mutant mice, although it was greatest for R972W mice (**Fig. 4G**). Furthermore, R972Q mice mounted an impaired antibody response to the Th17-dependent pneumococcal whole-cell antigen (**Fig. 4H**). Altogether, these data indicate that antibody responses to T-independent responses, potentially stemming from the demonstrable MZ B-cell population, are preserved in R972Q and F971L mice, but their T-dependent antibody responses are compromised, despite their relative leakiness of T- and B-cell development.

One of the characteristics of patients with hypomorphic mutations in the RAG1 CTD domain is the frequent association with autoimmune disease.²⁴ We hypothesized that increased levels of TNP-binding IgM antibodies in *Rag1*-mutant mice might reflect the presence of low-affinity, polyreactive antibodies. High levels of antiphosphorylcholine (PC) antibody were detected in the serum of 8- to 12-week-old R972Q-mutant mice (**Fig. 5A**, p. 96). Because B1 B-cells are thought to be a principal source of anti-PC antibodies,²⁵ we characterized the peritoneal compartment and observed a significant reduction of B1a cells in R972Q and F971L mice, whereas the B1b compartment was not significantly compromised (**Fig. 5B**, p. 96). These results suggest that elevated levels of anti-PC antibodies in R972Q mice cannot be attributed to expansion of B1a cells.

Figure 4: Distribution and phenotype of splenic B-cells and humoral immunity in *Rag1* mutant mice [A] Representative examples (left) and absolute count (right) of mature CD19⁺/IgM positive splenic B-cells (error bars of SEM). [B] Sections of spleen stained for B220 and CD3 concentrated within follicles (dark staining). [C] Spleens were analyzed by flow cytometry for Follicular (Fo) (CD19⁺CD93⁻CD21⁺CD23⁺) and marginal zone (MZ) B-cells (CD19⁺CD93⁻CD21⁺CD23⁻) with percentage and absolute number of CD19⁺ cells shown on the right [D] Immunoglobulin serum concentration (mean ± SEM) in naïve mice (n=6-10 per group). Mice were immunized with TNP-Ficoll. [E] TNP-specific IgM, and [F] IgG3 were measured by ELISA (n = 4). [G] Mice were immunized with T-dependent antigen (TNP-KLH) and TNP-specific IgG were measured by ELISA (n=4). [H] Th17 responses to immunization with Pneumococcal WCA antigen were measured by antigen specific total IgG. [A, C, D] one-way ANOVA, [E, F, G, H] two-way ANOVA. * p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001, **** p ≤ 0.0001.

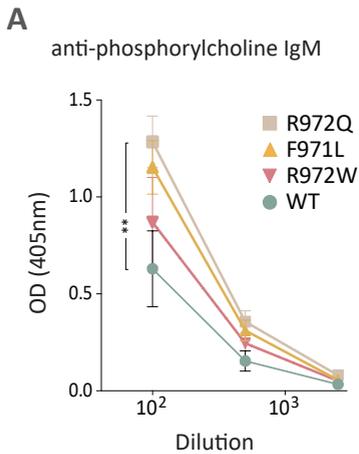


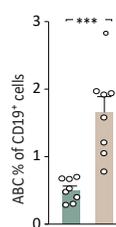
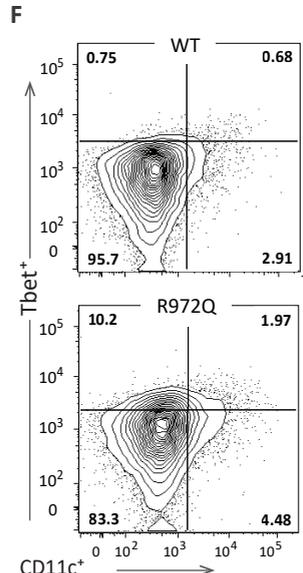
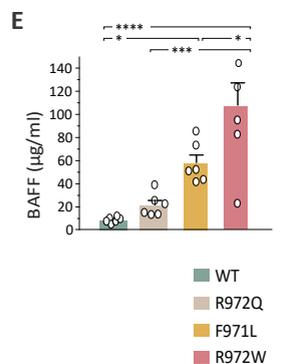
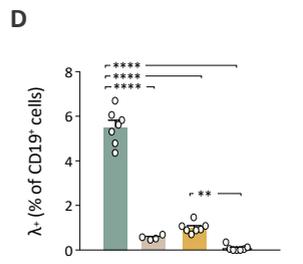
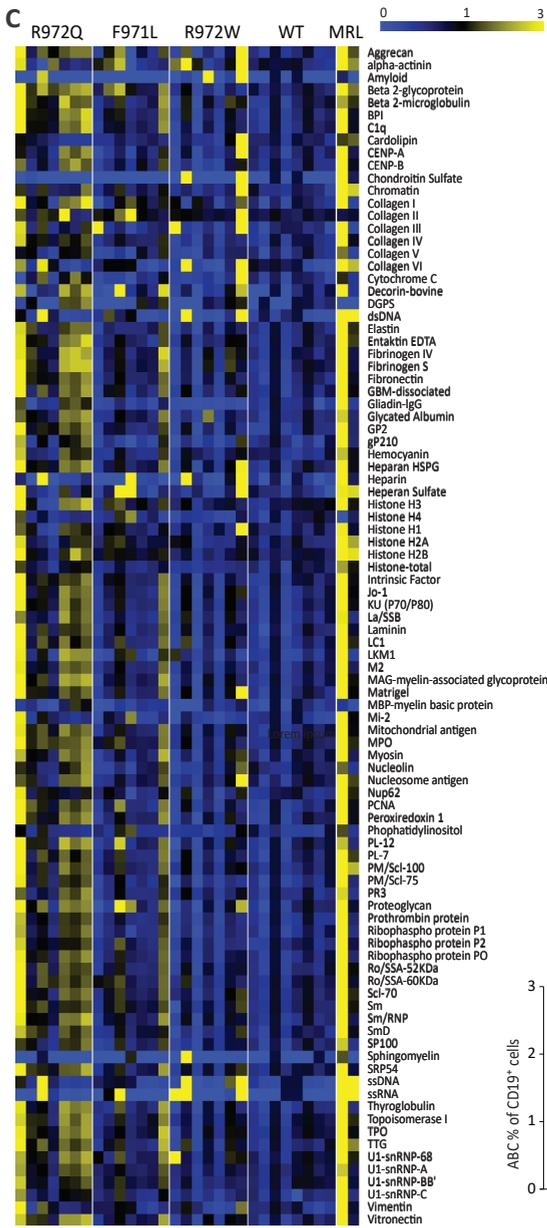
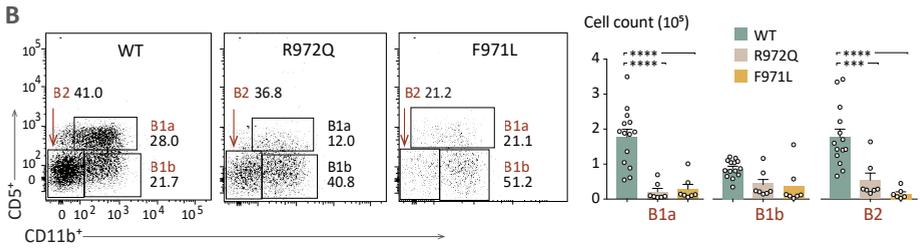
Figure 5: Serum Autoantibodies and mechanisms of altered B-cell tolerance in *Rag1* mutant mice (cont. on next page). [A] Anti-Phosphorylcholine (PC) IgM measured by ELISA. [B] Peritoneal lavage was analyzed by flow cytometry for B1 (CD19⁺CD11b⁻) and B2 (CD19⁺CD11b⁺) cells and B1a (CD19⁺CD11b⁻CD5⁺) and B1b (CD19⁺CD11b⁺CD5⁻) cells. [C] Heat map of serum IgM binding to a protein autoantigen microarray (n=7-8 mice per strain), normalized to WT reactivity. [D] Percentage of IgM⁺ B-cells in spleen expressing Igλ (n=4-7 per group). [E] BAFF serum concentrations measured by ELISA (n=4 per group). [F] Age-associated B-cells (ABCs) (Tbet⁺CD11c⁺) as percentage of CD19⁺ cells in spleen (n=7-8 per group). [A, B, D, E] One-way ANOVA, [F] Mann-Whitney: * p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001, **** p ≤ 0.0001. Error bars represent SEM.

When probing a proteomic microarray that contained a panel of 95 common self-antigens, IgM auto-antibodies were detected in the serum of all 3 mutant strains, most abundantly in the leakiest mutant, R972Q (**Fig. 5C**). Because the proteomic microarray was normalized for total IgM concentration, the increase seen in R972Q mice was not simply a result of the higher IgM levels in these mice. Specificity of the autoantibody repertoire was further interrogated for anticytokine antibodies, which have been seen in patients with CID-G/AI²⁴; however, no distinctive anticytokine antibody signature was observed in *Rag1* hypomorphic mutant mice (**Fig. S3**, p. 110).

We then looked for mechanisms that could account for autoantibody formation both centrally and in the periphery. RAG-mediated receptor editing and revision are important checkpoints of central B-cell tolerance to prevent autoimmunity.^{26,27} This process results in further *k*, and eventually *l*, LC rearrangement. The proportion of IgI expressing splenic B-cells was significantly decreased in all 3 mutant strains, indicating impairment of receptor editing (**Fig. 5D**). Elevated BAFF levels, which may allow rescue of self-reactive immature B-cells,^{28,29} were found in all 3 models and correlated with the severity of B-cell lymphopenia (**Fig. 5E**). Finally, R972Q mice had an increased proportion of splenic age-associated B-cells, which express CD11c and T-bet and have been reported to contribute to autoantibody production³⁰ (**Fig. 5F**). Altogether, these data suggest that multiple checkpoints of B-cell tolerance are affected in *Rag1*-mutant mice.

Abnormalities of B-cell repertoire at early stages of B-cell development

To investigate how *Rag1* mutations affect composition of the B-cell repertoire, we adopted a protocol using specific primers for each of the 4 *J* genes to detect in an unbiased manner *V(D)J* and *DJ* rearrangements, as well as productive



and nonproductive *Igh* recombination products.^{18,31} First, we analyzed B220⁺IgM⁻ CD43⁺ bone marrow B-cells, a population containing pro- and pre-BI cells, collectively termed pro-B-cells^{18,31} and in which both *DJ* and *V(D)J* rearrangements have occurred. At this stage, if neither allele undergoes productive *V(D)J* rearrangement, the cell undergoes apoptosis.²⁰ Although equal numbers of pro-B-cells and DNA input were sequenced, too few *Igh* rearrangements were detected in R972W mice, preventing further analysis. We recovered significantly fewer *DJ* and *V(D)J* rearrangement products in *Rag1* R972Q and F971L than in WT mice. To take this into account, absolute numbers instead of percentages were used to test for statistical differences (supplemental Tables 2 and 3). *DJ* rearrangements accounted for a higher percentage of total rearrangements in pro-B-cells from R972Q and F971L mice than in WT mice (**Fig. 6A**), consistent with decreased Rag1 activity and corresponding impaired sequential recombination.

In the periphery, 40% of mature B-cells are expected to harbor *V(D)J* rearrangements on both alleles (only 1 productive), and 60% are anticipated to carry a productive *V(D)J* rearrangement on 1 allele and a *DJ* rearrangement on the other allele.²⁰ Consistent with this, *DJ* rearrangements accounted for 31% of total rearrangements in splenic B-cells from WT mice. However, *DJ* rearrangements represented, 5% of rearrangements in splenic B-cells from R972Q mice (**Fig. 6A**). The low proportion of *DJ* rearrangements in mutant splenic B-cells and the lower absolute number of *Igh DJ* and *V(D)J* rearrangements in pro-B-cells from mutant mice suggest that 1 allele remains in germ line configuration. This implies that productive *V(D)J* rearrangement must occur on the other allele for the cell to survive. To test this hypothesis, we evaluated the proportion of productive and non-productive *Igh V(D)J* rearrangements at different stages of B-cell development. Unique sequences from pro-B-cells revealed a similar frequency of nonproductive *Igh* rearrangements in *Rag1*-mutant and in WT mice (**Fig. 6B**). By contrast, the proportion of nonproductive rearrangements in pre-B-cells was significantly lower in R972Q and F971L mice (6.5% and 6.8%, respectively) than in WT mice (27%), a pattern that was maintained in splenic B-cells (**Fig. 6B**). To further examine the efficiency of hypomorphic RAG-mediated recombination *in vivo*, we assayed for *Igk* gene rearrangements in pre-B-cells using a semiquantitative polymerase chain reaction strategy²⁰ and observed a fivefold reduction of *Vκ-Jκ* products in R972Q when compared with WT mice (**Fig. 6C**).

Reduced capacity of hypomorphic *Rag1* mutations to develop productive *V(D)J* rearrangements on at least 1 allele for both the *Igh* and *Igk* loci and ineffective pairing of the 2 could result in increased cell death. To test this hypothesis, we stained bone marrow B220^{lo}IgM⁻CD43⁺ (pro-B-cells), B220^{lo}IgM⁻CD43⁻ cells (pre-B-cells that have rearranged the LC and include

pre-BII cells), immature B-cells (B220^{lo}IgM⁺CD43⁻), and mature recirculating B-cells (B220^{hi}IgM⁺CD43⁻) from WT and R972Q mice. A significantly higher proportion of apoptotic cells was detected in pre-B- and immature B-cells from *Rag1*-mutant mice, with an increased trend seen for pro-B-cells (**Fig. 6D**).

Subsequently, we analyzed individual *V* and *D* gene usage at distinct stages of B lymphocyte development. Analysis of *D* gene usage in *DJ* rearrangements from sorted pro-B-cells did not reveal any differences between WT and R972Q or F971L mice (**Fig. 6E**). *V* gene usage is tightly regulated during early B-cell development and requires the interplay of transcription factors to alter accessibility of the *Igh* locus.^{32,33} The overall distribution of *IghV* gene usage among total unique *V(D)J* sequences was significantly different between pro-B-cells of WT and *Rag1*-mutant mice ($p=1.99e-12$ for R972Q, $p=4.27e-22$ for F971L; **Fig. 6F-G** and **Table S3**, p. 115). In particular, *Rag1*-mutant pro-B-cells revealed fewer rearrangements to the normally highly used proximal V_H2 and

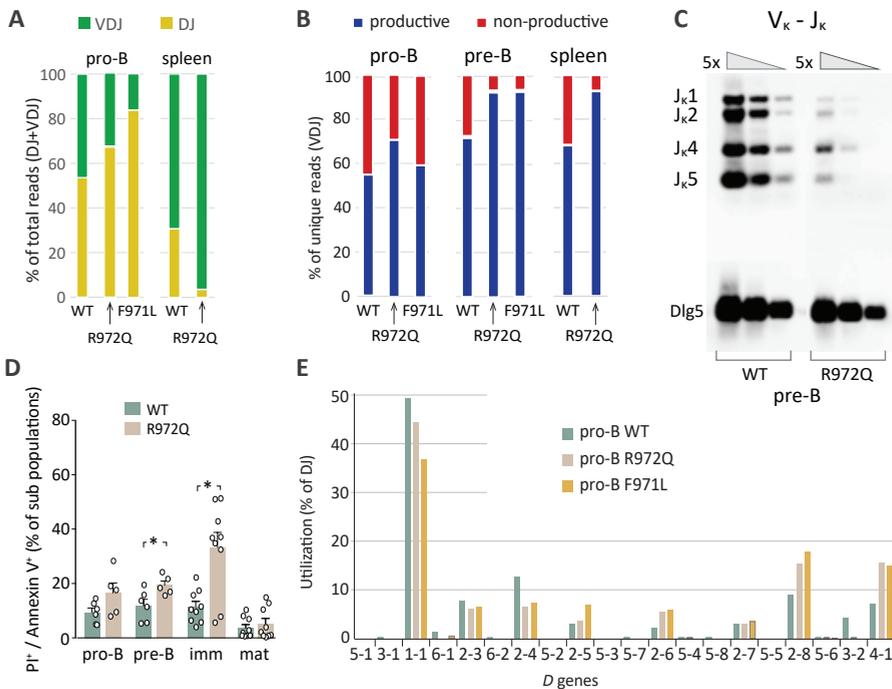


Figure 6: Abnormalities of *Igh* repertoire in *Rag1* mutant mice

A) Bar graph of *Igh* *VDJ* (green) and *DJ* (yellow) rearrangements as a percentage of total rearrangements recovered in pro-B-cells and IgM⁺ splenocytes. **B**) Bar graph of productive (blue) vs non-productive (red) reads of total unique *Igh* *VDJ* rearrangements in pro-B-cells, pre-B and IgM⁺ splenic B-cells. **C**) Fivefold dilution of input DNA was PCR amplified for *Vκ-Jκ* light chain rearrangements in pre-B-cells with loading control (*Dlg5*). **D**) Apoptotic B-cells through development were analyzed by flow cytometry using PI and Annexin. Mann-Whitney test, * $p \leq 0.05$. **E**) Frequency of *D* gene utilization of *Igh* *DJ* rearrangements in pro-B-cells. [Figure continued on next page]

V_H5 families, including the very proximal V_H5-2 .^{18,31} The 10 most proximal V genes were involved in 14.3% of total unique $V(D)J$ rearrangements in WT pro-B-cells, compared with 4.6% in R972Q pro-B-cells and 4.5% in F971L pro-B-cells (**Fig. 6F-G; Table S4**, p. 115). Conversely, there was a trend for more frequent usage of distal $IghV$ genes in Rag1-mutant pro-B-cells (**Fig. 6F**).

These findings indicate that the V gene repertoire is significantly skewed at the pro-B-cell stage, independent of classical selection after LC rearrangement. Similar trends in proximal and distal $IghV$ gene usage between WT and mutant mice were observed in pre-B-cells (**Fig. S4A-B**, p. 111), splenic B-cells (**Fig. 6H**), and peritoneal B1 cell (**Fig. S5**, p. 112), although less prominent than at the pro-B-cell level. Overall distribution of V genes was significantly different for all B-cell populations between WT and mutants. The difference was more significant for pro-B-cells than for the more mature B-cell populations and more significant for F971L than R972Q mice (supplemental Table 3). No differences were observed in CDR3 length or P or N nucleotide modifications between WT and R972Q pro-B and mature B-cells (**Fig. S4C**, p. 111).

Finally, despite altered usage of $IghV$ genes, we observed that B-cell splenocytes from R972Q mice had a diversified repertoire, as indicated by the D50 index, a measure for clonotypic expansion corresponding to the percentage of unique CDR3 sequences that account for 50% of the total number of sequences observed¹⁰ (**Fig. S4D**, p. 111).

T-cell repertoire of *Rag1* hypomorphic mutant mice is skewed at early stages of V(D)J recombination

Next-generation sequencing of the *Trb* repertoire revealed many similarities to what was observed at the *Igh* locus during B-cell development. In particular, the proportion of nonproductive *Trb* rearrangements was similar in DN4 cells of WT and mutant mice, but an increased proportion of productive rearrangements was detected in DP and peripheral T-cells from R972Q- and F971L-mutant compared with WT mice (**Fig. 7A**). Furthermore, important differences in *Trb V* and *J* gene usage were detected in thymocytes from *Rag1*-mutant and WT mice (**Fig. S6A-B**, p. 113), including DN4 thymocytes ($p=3.340e-08$; **Fig. 7B; Tables S3**, p. 115 and **S5**, p. 118), indicating that they had occurred before selection. The same pattern was documented in DP thymocytes (**Fig. 7C-D; Table S6**, p. 119), splenic $CD4^+CD4^- T_{conv}$ (**Fig. 7E-F; Tables S5**, p. 118 and **S6**, p. 119), and T_{reg} (**Fig. 7G-H; Tables S5** and **S6**). In the periphery, a mildly reduced diversity of the T-cell repertoire was observed in both T_{conv} and T_{reg} from R972Q and F971L mice as shown by modestly lower Shannon entropy index (**Fig. S6C**, p. 113); however, the capacity of T_{reg} to suppress proliferation of T effector cells in response to polyclonal mitogenic stimulation was preserved (**Fig. S7**, p. 113).

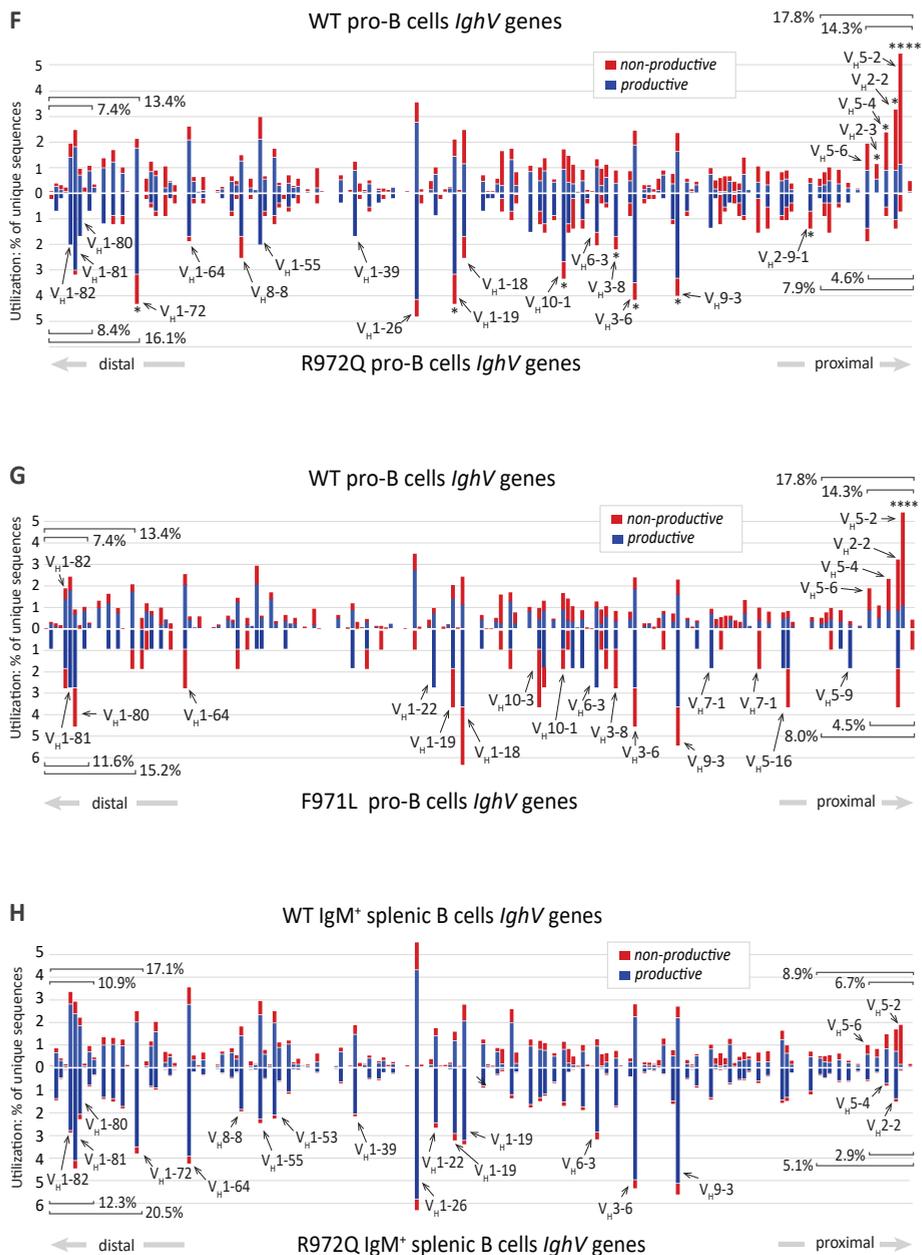


Figure 6 (continued): Abnormalities of *Igh* repertoire in *Rag1* mutant mice

[F] Frequency of *V* gene utilization in unique *Igh* VDJ rearrangements in pro-B-cells. Top shows WT, bottom shows R972Q mice. Individual *V* gene usage was compared with Chi-square test, * $p \leq 0.05$, **** $p \leq 0.0001$. [G] Frequency of *V* gene utilization in unique *Igh* VDJ rearrangements in pro-B-cells. Top shows WT (data as in F), bottom shows F971L mice. Individual *V* gene usage was compared with Chi-square test, * $p \leq 0.05$, **** $p \leq 0.0001$. [H] Utilization of *V* genes in unique *Igh* VDJ rearrangements in splenic IgM⁺ B-cells. Top shows WT, bottom shows R972Q mice.

Discussion

We have generated mouse models with missense mutations in the RAG1 CTD corresponding to those identified in patients with CID-G/AI and shown that 2 models (R972Q and F971L mice) mimic the human phenotype, with partial preservation of T- and B-cell development and immunoglobulin levels but abnormalities of T- and B-cell repertoire and autoantibody production. Alterations of T- and B-cell development did not affect all lymphocyte subsets equally. Generation of TCR $\gamma\delta^+$ T-cells was relatively preserved, perhaps because of the distinct developmental origin, which requires fewer RAG-dependent checkpoints.³⁴ In the spleen, partial preservation of MZ B-cell development was observed in R972Q and F971L mice, whereas follicular B-cells were severely depleted. The B-cell compartment seems most vulnerable to impaired RAG1 activity, with proportionately more depletion in F971L compared with R972Q mice. Similar findings in RAG1-deficient patients have been observed, where B-cell defects appear in greater magnitude than those in T-cells.² Additional studies of the constraints of RAG1 activity at the Igh versus the TCR locus may uncover mechanisms for this difference in severity of lymphoid development.

This study has uncovered novel effects of hypomorphic *Rag1* mutations on the composition of the pre-immune repertoire, cell survival, and selection at early stages of T and B lymphocytes. In particular, we observed a strong bias for productive *V(D)J* rearrangements in peripheral B- and T- cells from *Rag1*-mutant mice, previously recognized only in peripheral B-cells of patients with more severe hypomorphic mutations.^{35,36} Before LC rearrangement, both *DJ* and *V(D)J* rearrangements are detectable in pro-B-cells, but only productive *V(D)J* rearrangements subsequently pair with the $\lambda 5/V$ pre-B-cell surrogate LC complex enabling selection.³⁷ We saw no significant bias for productive versus nonproductive rearrangements at the pro-B-cell stage; however, a significantly higher proportion of productive *V(D)J* rearrangements were detected in pre-B-cells from *Rag1*-mutant than WT mice.

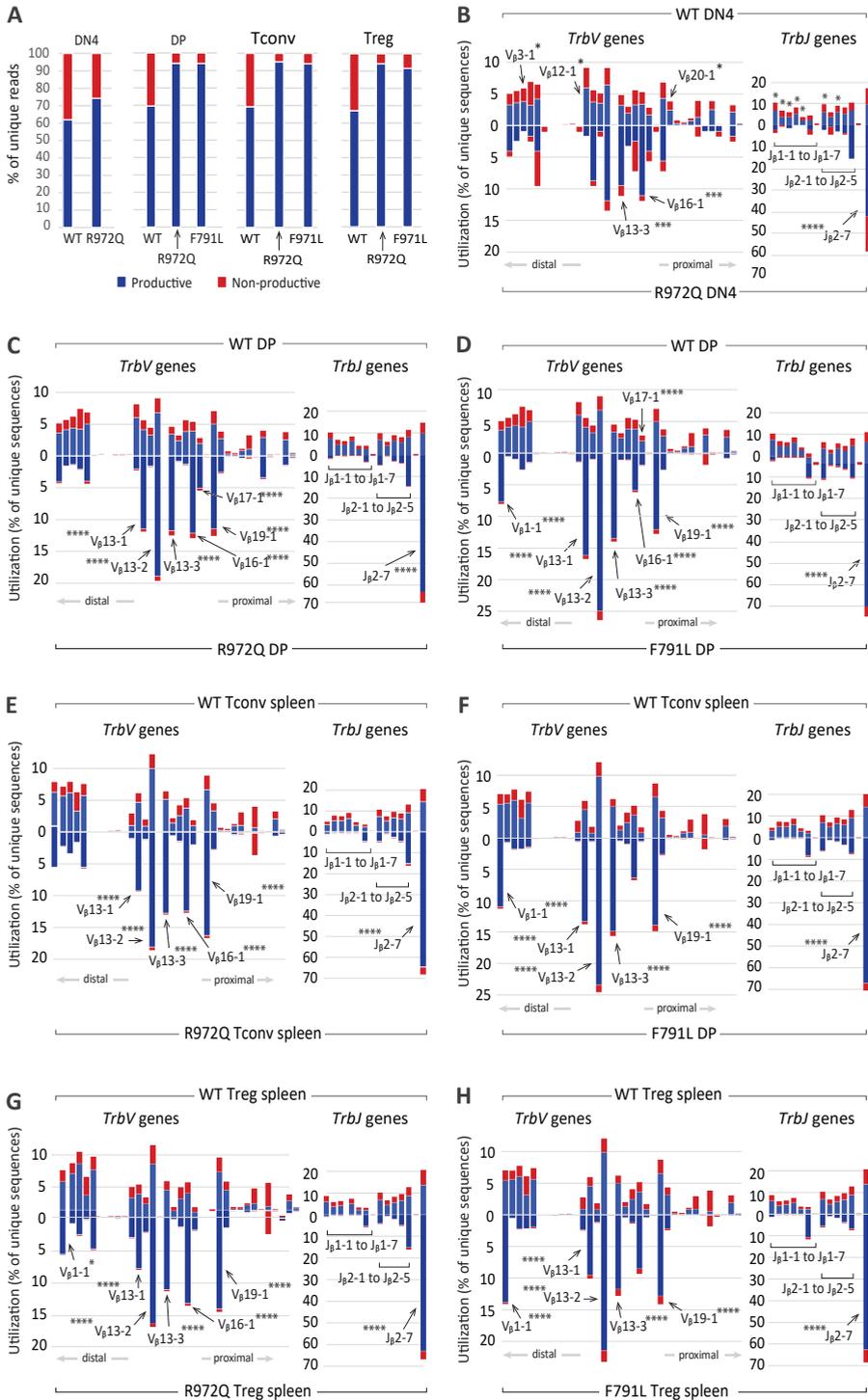
Figure 7: (→) Abnormalities of *Trb* repertoire in *Rag1* mutant mice

[A] Percentage of productive and non-productive reads of total unique *Trb* joins from WT and R972Q DN4 thymocytes (CD4⁺CD8⁺CD44⁺CD25⁻); WT, R972Q and F971L CD4⁺CD8⁺ double-positive (DP) thymocytes; WT, R972Q and F971L T conv (CD4⁺EGFP⁺) cells; and WT, R972Q and F971L Treg (CD4⁺EGFP⁺) cells.

[B] Utilization of *V* (left) and *J* (right) genes in *Trb* rearrangements in DN4 thymocytes (CD4⁺CD8⁺CD44⁺CD25⁻). Top shows WT, and bottom shows R972Q mice. [C] *Trb* unique sequences in CD4⁺CD8⁺ double-positive (DP) thymocytes. Top shows WT, and bottom shows R972Q mice. [D] *Trb* unique sequences in CD4⁺CD8⁺ double-positive (DP) thymocytes. Top shows WT, and bottom shows F971L mice. [E] *Trb* unique sequences in splenic T_{conv} (CD4⁺EGFP⁺) cells from WT (top) and R972Q (bottom) mice.

[Fig.7 continued] [F] *Trb* unique sequences in splenic T_{conv} (CD4⁺EGFP⁺) cells from WT (top) and F971L (bottom) mice. [G] Utilization of *V* (left) and *J* (right) genes in *Trb* unique sequences in splenic T_{reg} (CD4⁺EGFP⁺) cells from WT (top) and R972Q (bottom) mice. [H] Utilization of *V* (left) and *J* (right) genes in *Trb* unique sequences in splenic T_{reg} (CD4⁺EGFP⁺) cells from WT (top) and F971L (bottom) mice.

Chi square test, ***p ≤ 0.001, **** p ≤ 0.0001.



Our data are consistent with a model where in *Rag1*-mutant mice, fewer alleles accomplish *DJ* rearrangement, and only those cells with in-frame *V(D)J* rearrangements subsequently progress further along B-cell development, with a high proportion of cells maintaining the second allele in germ line configuration. A similar selective pressure was observed in T-cell development, with no significant bias for productive *Trb V(D)J* rearrangements at the DN4 stage of T-cell development, before selection and pairing with TCR α ,³⁴ but an increased proportion of productive rearrangements in DP thymocytes (where pairing with TCR α enables positive selection) and in the periphery. These findings indicate that despite the relatively leaky block in lymphoid development observed in R972Q and F971L mice, a significant impairment of gene rearrangement at the *Igh*, *Igk*, and *Trb* loci limits antigen receptor diversity from a very early stage in development.

Signaling at the pre-BI and pre-BII stages represents an important checkpoint during B-cell development. Failure to correctly signal via the newly generated antigen receptor results in cell death, usually at the immature B-cell stage of development, where LC rearrangement can potentially rescue mispaired or autoimmune-prone B-cells.^{1,27} Correspondingly, *Rag1*-mutant mice manifested increased apoptosis that was more prominent in immature B-cells, with a reduced proportion of Ig λ ⁺ B-cells in the spleen. However, enhanced apoptosis was also observed earlier in development (pre-B-cells), potentially because of failed surrogate LC pairing or failed LC rearrangement. These findings suggest that RAG activity unexpectedly dictates survival very early in lymphocyte development, possibly by constraining the timeframe that is responsive to a productive recombination.

Pro-B-cells from *Rag1*-mutant mice displayed a significantly decreased usage of proximal *IghV* genes. This is in marked contrast to what has been previously described in normal development, where proximal *V* genes are preferentially rearranged.^{31,38} Targeting of the *Igh* locus is regulated by several transcription factors that regulate pro- to pre-B-cell transition and recombination of the distal *IghV* genes and by cohesin and CTCF factors that affect *V* gene usage by controlling proximal to distal interactions.^{18,33,39-42} Impaired RAG activity may alter the timing dynamics and magnify the regulation of recombination by CTCF versus the transcription factors.^{40,43}

Using next-generation sequencing, we demonstrated that the skewing of *Trb* repertoire previously reported in the periphery of R972Q mice⁹ is already present at the DN4 stage. Remarkably, similar abnormalities of *Trb V* and *J* gene usage were observed in F971L and R972Q mice. One of the proposed functions of the CTD of RAG1 is to ensure recognition of specific antigen receptor genes, with *in vitro* data suggesting that this region interacts with the terminal 2 base pair coding ends of *V*, *D*, and *J* genes.^{8,44,45} However, no difference in *D* gene

usage in *DJ* rearrangements was observed in pro-B-cells of mutant mice, and no differences in the composition of the 2 base pair coding ends were detected when comparing *IghV* and *TrbV* genes that were differentially used in mutant mice versus WT mice (**Tables S7** and **S8**, p. 120).

Finally, *Rag1*-mutant mice displayed signs of immune dysregulation, with the presence of broad-spectrum autoantibodies, recapitulating similar features observed in patients with CID-G/AI.²⁴ This phenotype was especially evident in the leakiest model (R972Q mouse), where anti-PC antibodies were detected at higher titer. An important difference between patients with CID-G/AI and the mouse models described here is that the former have a characteristic anticytokine antibody signature (predominantly directed against type 1 interferon),²⁴ especially in patients with a history of severe viral infections, whereas such an anticytokine signature was not observed in *Rag1*-mutant mice. Future studies may help address whether infectious triggers elicit anticytokine antibodies in *Rag1*-mutant mice.

Various mechanisms may contribute to immune dysregulation both in patients and in mice with hypomorphic mutations. In particular, we have confirmed previous findings of altered receptor editing.^{17,46} Abnormalities in the composition of the preimmune repertoire may affect subsequent survival and selection of T and B-cells. In the periphery, relative expansion of age-associated B-cells and increased serum BAFF levels may also lead to breakage of B-cell tolerance. Furthermore, although T_{reg} function was apparently preserved, as measured by suppression of proliferation of T effector cells in response to polyclonal stimulation via CD3/CD28/CD2, abnormalities of T_{reg} repertoire in *Rag1*-mutant mice and patients⁴⁷ may limit their capacity to effectively suppress immune responses against defined sets of self-antigens.

Acknowledgments

The authors thank K. Musunuru for his guidance in designing the gRNA to generate the *Rag1*-mutant mice and W. Yang and M. Gellert for the structural prediction model of the RAG mutants (supplemental Figure 1); F. W. Alt for providing the HTGTS-Rep-seq platform and L. Charbonnier for help with T regulatory studies; and M. L. Boes for reading the manuscript and discussing the study.

This work was supported by the National Institutes of Health, National Institute of Allergy and Infectious Diseases grant #R01 AI100887 (J.P.M.) and by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health.

References

1. Schatz DG, Swanson PC. V(D)J recombination: mechanisms of initiation. *Annu Rev Genet.* 2011;45:167-202.
2. Notarangelo LD, Kim MS, Walter JE, Lee YN. Human RAG mutations: biochemistry and clinical implications. *Nat Rev Immunol.* 2016; 16(4):234-246.
3. Mombaerts P, Iacomini J, Johnson RS, Herrup K, Tonegawa S, Papaioannou VE. RAG-1-deficient mice have no mature B and T lymphocytes. *Cell.* 1992;68(5):869-877.
4. Schwarz K, Gauss GH, Ludwig L, et al. RAG mutations in human B cell-negative SCID. *Science.* 1996;274(5284):97-99.
5. Shinkai Y, Rathbun G, Lam KP, et al. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell.* 1992;68(5):855-867.
6. Villa A, Santagata S, Bozzi F, et al. Partial V(D)J recombination activity leads to Omenn syndrome. *Cell.* 1998;93(5):885-896.
7. Schuetz C, Huck K, Gudowius S, et al. An immunodeficiency disease with RAG mutations and granulomas. *N Engl J Med.* 2008; 358(19):2030-2038.
8. Wong SY, Lu CP, Roth DBA. A RAG1 mutation found in Omenn syndrome causes coding flank hypersensitivity: a novel mechanism for antigen receptor repertoire restriction. *J Immunol.* 2008;181(6):4124-4130.
9. Khiong K, Murakami M, Kitabayashi C, et al. Homeostatically proliferating CD4 T cells are involved in the pathogenesis of an Omenn syndrome murine model. *J Clin Invest.* 2007; 117(5):1270-1281.
10. Lee YN, Frugoni F, Dobbs K, et al. Characterization of T and B cell repertoire diversity in patients with RAG deficiency. *Sci Immunol.* 2016;1(6):eaah6109.
11. Henderson LA, Frugoni F, Hopkins G, et al. Expanding the spectrum of recombination-activating gene 1 deficiency: a family with early-onset autoimmunity. *J Allergy Clin Immunol.* 2013;132(4):969-971.e1-2.
12. Villa A, Sobacchi C, Notarangelo LD, et al. V(D)J recombination defects in lymphocytes due to RAG mutations: severe immunodeficiency with a spectrum of clinical presentations. *Blood.* 2001;97(1):81-88.
13. Avila EM, Uzel G, Hsu A, et al. Highly variable clinical phenotypes of hypomorphic RAG1 mutations. *Pediatrics.* 2010;126(5):e1248-e1252.
14. Ott de Bruin L, Yang W, Capuder K, et al. Rapid generation of novel models of RAG1 deficiency by CRISPR/Cas9-induced mutagenesis in murine zygotes. *Oncotarget.* 2016; 7(11):12962-12974.
15. Yang H, Wang H, Jaenisch R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat Protoc.* 2014;9(8):1956-1968.
16. Rucci F, Poliani PL, Caraffi S, et al. Abnormalities of thymic stroma may contribute to immune dysregulation in murine models of leaky severe combined immunodeficiency. *Front Immunol.* 2011;2(15):00015.
17. Cassani B, Poliani PL, Marrella V, et al. Homeostatic expansion of autoreactive immunoglobulin-secreting cells in the Rag2 mouse model of Omenn syndrome. *J Exp Med.* 2010;207(7):1525-1540.
18. Guo C, Yoon HS, Franklin A, et al. CTCF-binding elements mediate control of V(D)J recombination. *Nature.* 2011;477(7365): 424-430.
19. Kim MS, Lapkouski M, Yang W, Gellert M. Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature.* 2015;518(7540): 507-511.
20. Jung D, Giallourakis C, Mostoslavsky R, Alt FW. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol.* 2006;24: 541-570.
21. Marrella V, Poliani PL, Casati A, et al. A hypomorphic R229Q Rag2 mouse mutant recapitulates human Omenn syndrome. *J Clin Invest.* 2007;117(5):1260-1269.
22. Giblin W, Chatterji M, Westfield G, et al. Leaky severe combined immunodeficiency and aberrant DNA rearrangements due to a hypomorphic RAG1 mutation. *Blood.* 2009; 113(13):2965-2975.
23. De Ravin SS, Cowen EW, Zarembek KA, et al. Hypomorphic Rag mutations can cause destructive midline granulomatous disease. *Blood.* 2010;116(8):1263-1271.
24. Walter JE, Rosen LB, Csomos K, et al. Broad-spectrum antibodies against self-antigens

and cytokines in RAG deficiency. *J Clin Invest*. 2015;125(11):4135-4148.

25. Baumgarth N. The double life of a B-1 cell: self-reactivity selects for protective effector functions. *Nat Rev Immunol*. 2011;11(1):34-46.

26. Nemazee D. Receptor editing in lymphocyte development and central tolerance. *Nat Rev Immunol*. 2006;6(10):728-740.

27. Nemazee D. Mechanisms of central tolerance for B cells. *Nat Rev Immunol*. 2017;17(5): 281-294.

28. Lesley R, Xu Y, Kalled SL, et al. Reduced competitiveness of autoantigen-engaged B cells due to increased dependence on BAFF. *Immunity*. 2004;20(4):441-453.

29. Mackay F, Woodcock SA, Lawton P, et al. Mice transgenic for BAFF develop lymphocytic disorders along with autoimmune manifestations. *J Exp Med*. 1999;190(11):1697-1710.

30. Rubtsov AV, Rubtsova K, Fischer A, et al. Toll-like receptor 7 (TLR7)-driven accumulation of a novel CD11c1 B-cell population is important for the development of autoimmunity. *Blood*. 2011;118(5):1305-1315.

31. Lin SG, Ba Z, Du Z, Zhang Y, Hu J, Alt FW. Highly sensitive and unbiased approach for elucidating antibody repertoires. *Proc Natl Acad Sci USA*. 2016;113(28):7846-7851.

32. Teng G, Schatz DG. Regulation and evolution of the RAG recombinase. *Adv Immunol*. 2015; 128:1-39.

33. Proudhon C, Hao B, Raviram R, Chaumeil J, Skok JA. Long-range regulation of V(D)J recombination. *Adv Immunol*. 2015;128:123-182.

34. Krangel MS. Mechanics of T cell receptor gene rearrangement. *Curr Opin Immunol*. 2009;21(2):133-139.

35. Ohm-Laursen L, Nielsen C, Fisker N, Lillevang ST, 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*. 2008;28(5):588-592.

36. IJspeert H, Driessen GJ, Moorhouse MJ, et al. Similar recombination-activating gene (RAG) mutations result in similar immunobiological effects but in different clinical phenotypes. *J Allergy Clin Immunol*. 2014;133(4): 1124-1133.

37. Melchers F. The pre-B-cell receptor: selector of fitting immunoglobulin heavy chains for the B-cell repertoire. *Nat Rev Immunol*. 2005;5(7): 578-584.

38. Yancopoulos GD, Desiderio SV, Paskind M, Kearney JF, Baltimore D, Alt FW. Preferential utilization of the most JH-proximal VH gene segments in pre-B-cell lines. *Nature*. 1984; 311(5988):727-733.

39. Bertolino E, Reddy K, Medina KL, Parganas E, Ihle J, Singh H. Regulation of interleukin7-dependent immunoglobulin heavy-chain variable gene rearrangements by transcription factor STAT5. *Nat Immunol*. 2005;6(8): 836-843.

40. Ebert A, Hill L, Busslinger M. Spatial regulation of V(D)J recombination at antigen receptor loci. *Adv Immunol*. 2015;128:93-121.

41. Busslinger M. Transcriptional control of early B cell development. *Annu Rev Immunol*. 2004; 22:55-79.

42. Reynaud D, Demarco IA, Reddy KL, et al. Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nat Immunol*. 2008; 9(8):927-936.

43. Kumari G, Sen R. Chromatin interactions in the control of immunoglobulin heavy chain gene assembly. *Adv Immunol*. 2015;128:41-92.

44. Gerstein RM, Lieber MR. Coding end sequence can markedly affect the initiation of V(D)J recombination. *Genes Dev*. 1993;7(7B): 1459-1469.

45. Mo X, Bailin T, Sadofsky MJ. A C-terminal region of RAG1 contacts the coding DNA during V(D)J recombination. *Mol Cell Biol*. 2001;21(6):2038-2047.

46. Walter JE, Rucci F, Patrizi L, et al. Expansion of immunoglobulin-secreting cells and defects in B cell tolerance in Rag-dependent immunodeficiency. *J Exp Med*. 2010;207(7):1541-1554.

47. Rowe JH, Stadinski BD, Henderson LA, et al. Abnormalities of T-cell receptor repertoire in CD4¹ regulatory and conventional T cells in patients with RAG mutations: implications for autoimmunity. *J Allergy Clin Immunol*. 2017; 140(6):1739-1743.e7.

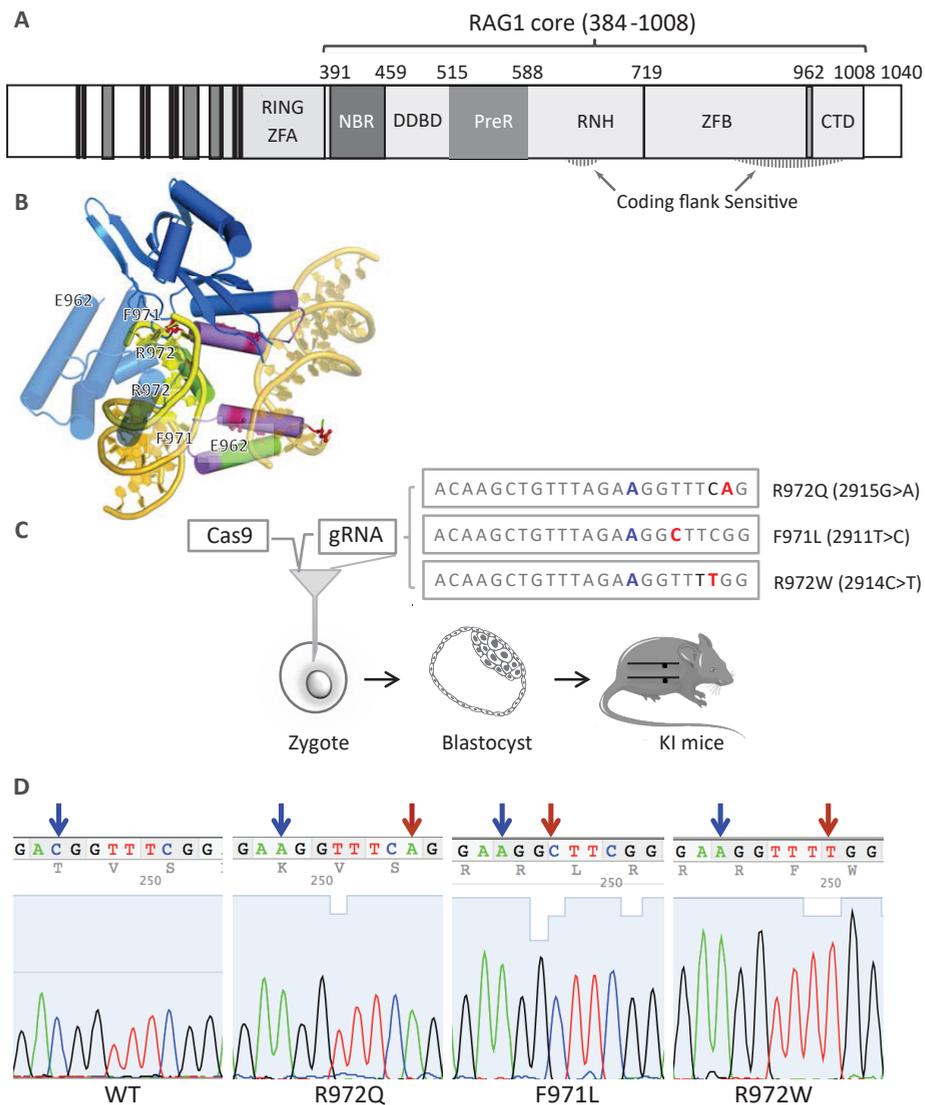


Figure S1: Predicted effect of Rag1 mutations based on crystallography data, and strategy for the generation of Rag1 mutant mice. [A] RAG1 protein structure, showing Zinc Finger A (ZFA) and Zinc Finger B (ZFB), Nonamer binding region (NBR) and DNA dimerization and binding domain (DDBD), pre-RNase (preR), the catalytic RNase H-like (RNH) domain and the C-terminal domain (CTD). Residue numbers are given for the boundaries of the different domains. Two coding flank sensitive regions run from amino acid 606-611 and from 889-974. **[B]** Structural model of RAG1-RSS complex. Both F971 and R972 are located on the bridging α -helix (colored in purple), which is immediately downstream from the catalytic residue E962 and is predicted to link the two RSS DNAs. **[C]** Zygote injection with a mixture of 3 different ssODNs each containing specific point mutations (red) and a common silent variant (blue). **[D]** Sanger sequences of resulting genotypes showing the silent variant (blue arrow) and the specific c.2915G>A, c.2911T>C and α c.2914C>T mutations (red arrows).

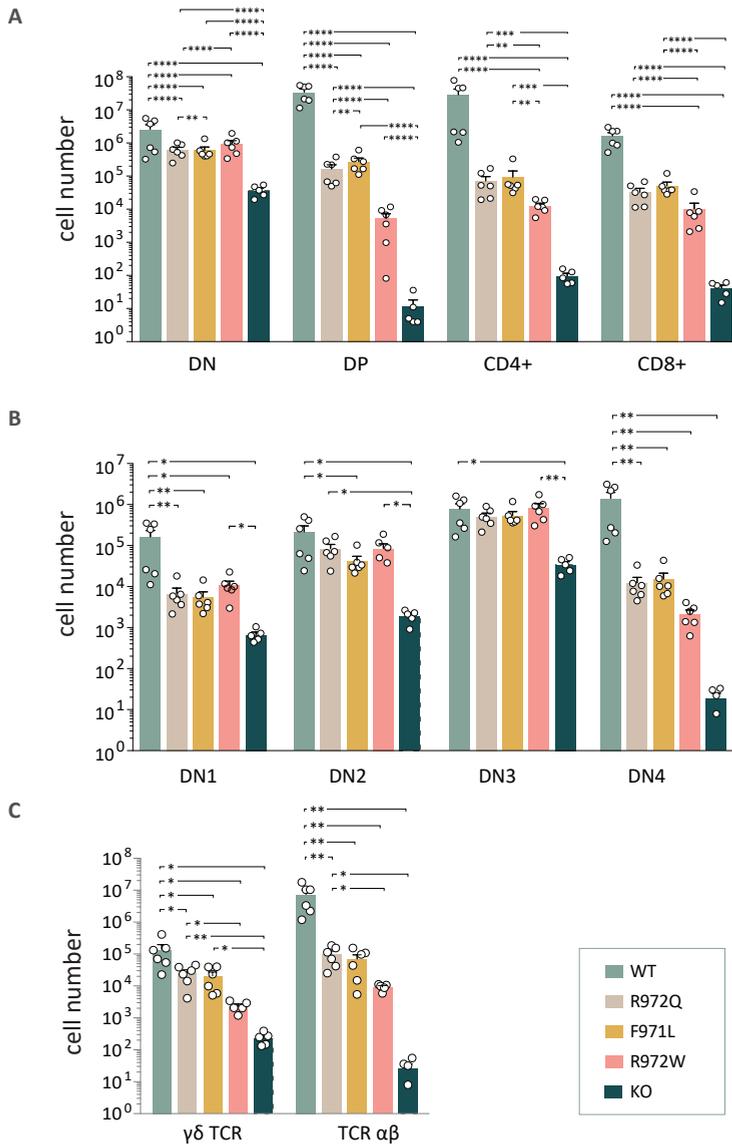


Figure S2: Absolute number of cells at various stages of lymphocyte development in the thymus [A] Absolute count of CD4⁻CD8⁻ double negative (DN), CD4⁺CD8⁺ double positive (DP) and single positive (CD4⁺ and CD8⁺) cells in the thymus of wild-type (WT), Rag1^{-/-} (KO) and indicated mutant mice. [B] Absolute count of DN thymocytes at various stages of development (DN1-DN4) based on CD44 and CD25 expression: CD44⁺CD25⁻ (DN1), CD44⁺CD25⁺ (DN2), CD44⁻CD25⁺ (DN3), and CD44⁻CD25⁻ (DN4). [C] Absolute count of $\gamma\delta$ TCR and $\alpha\beta$ TCR thymocytes. In all panels, 6 mice per group were analyzed, and statistical analysis was done with one-way ANOVA (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$)

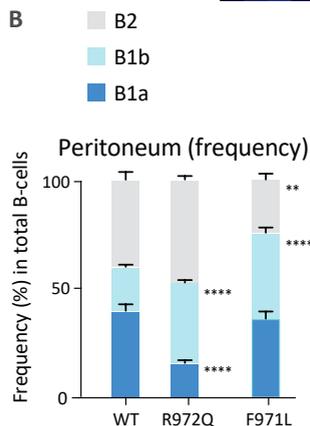
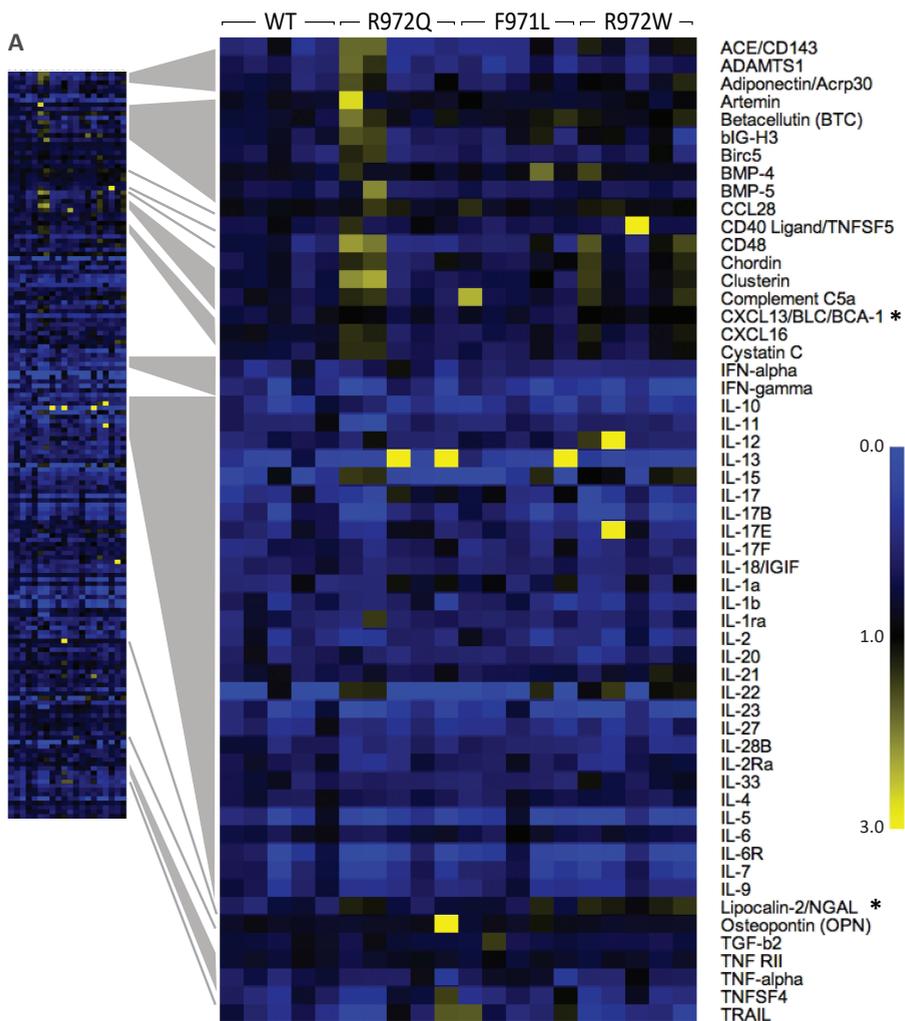


Figure S3: Anti-cytokine antibodies and distribution of peritoneal B1 and B2 subsets

A An array of full-length murine proteins was used to screen for cytokine autoantibodies in mutant ($n=5$ per strain) and WT ($n=5$) mice. The heatmap displays the ratio of signal intensities to the second SD over the WT mean, and therefore highlights signals that are stronger in the mutant mice. A selection of cytokines, for which autoantibodies had been found in patients previously, or that tested positive in our mice, are shown on the right. SEM test was used to test for statistical significance between groups (* FDR $\leq 1\%$). This was only the case for Lipocalin-2/NGAL and CXCL13/BCL/BCA-1 in the R972W model compared to WT. **[B]** Distribution of B2 ($CD19^+ CD11b^-$) cells, and B1a ($CD19^+ CD11b^+ CD5^+$) and B1b ($CD19^+ CD11b^+ CD5^-$) cells in peritoneal lavage. Statistical analysis was done with one way ANOVA (** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$). Error bars represent SEM.

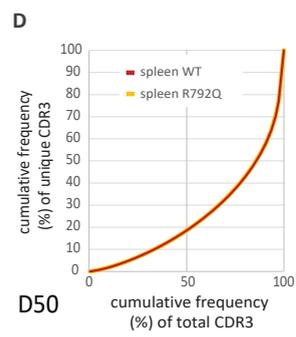
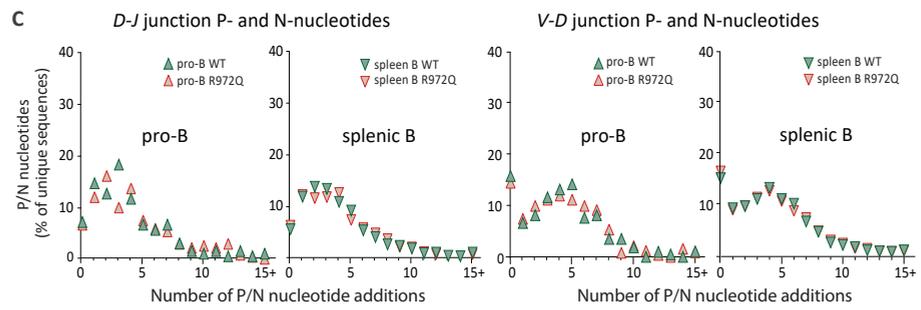
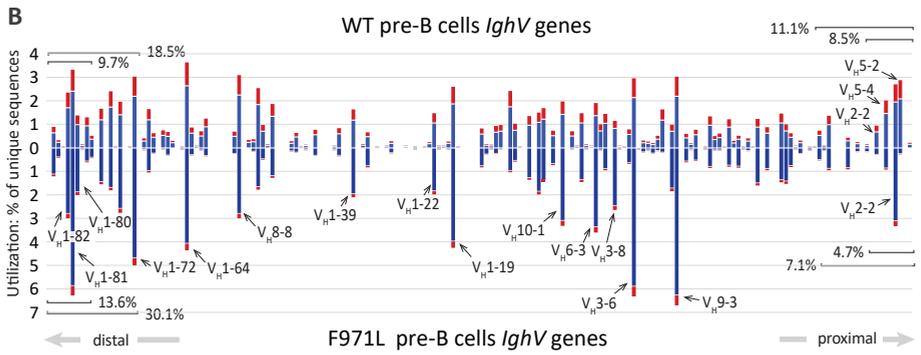
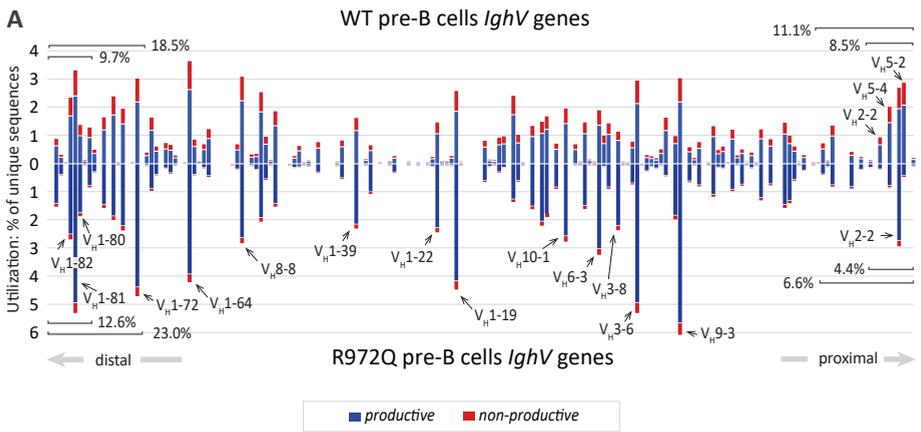


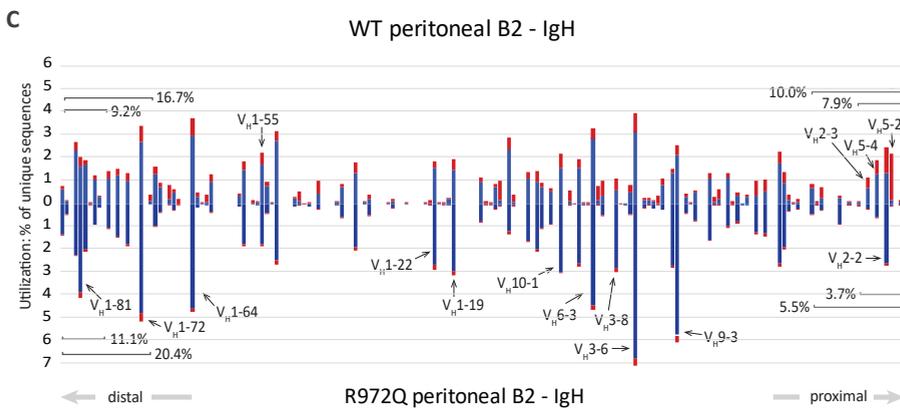
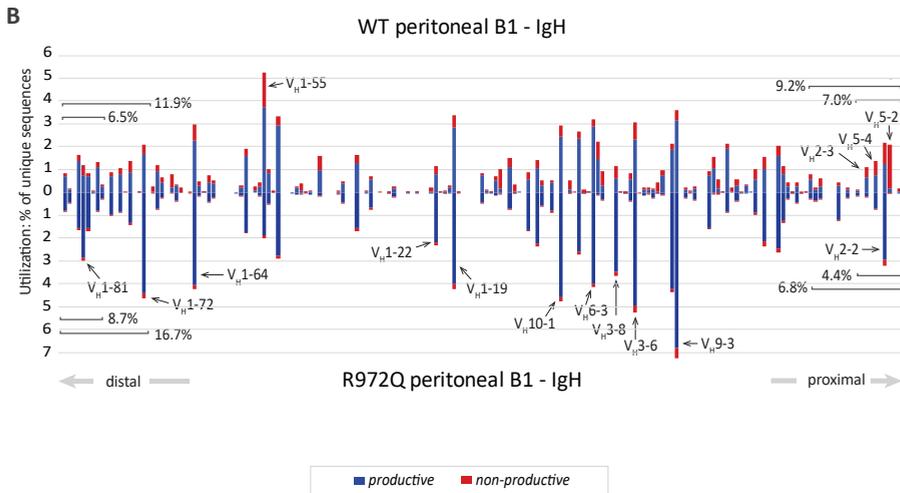
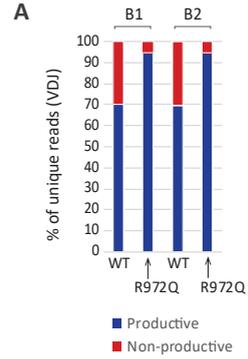
Figure S4: *Igh* repertoire of bone marrow and splenic B-cells
[A] V gene usage in pre-B-cells from WT (top, n=3) and R972Q (bottom, n=3) mice. **[B]** V gene usage in pre-B-cells from WT (top, n=3) and F971L (bottom, n=3) mice. **[C]** Distribution of P and N nucleotide addition in D-J and V-D junctions in bone marrow pro-B-cells and splenic B-cells. X-axis: number of P/N nucleotide additions, Y-axis: percentage of unique sequences with given number of P/N nucleotide additions. **[D]** D50, corresponding to the percentage of unique CDR3 sequences that account for 50% of the total number of sequences recovered of mature splenic B-cells from R972Q vs WT mice.

Figure S5: Igh repertoire of peritoneal B1 and B2 cells

[A] Bar graph showing proportion of productive (blue) and non-productive (red) reads among total unique Igh VDJ rearrangements in FACS-sorted peritoneal B1 (B220⁺ CD19⁺ IgM⁺ CD11b⁺) and B2 (B220⁺ CD19⁺ IgM⁺ CD11b⁻) cells.

[B] Frequency of V gene utilization in unique IgH VDJ rearrangements from peritoneal B1 (B220⁺ CD19⁺ IgM⁺ CD11b⁺) cells from WT (top) and R972Q (bottom) mice.

[C] Frequency of V gene utilization in unique IgH VDJ rearrangements from peritoneal B2 (B220⁺ CD19⁺ IgM⁺ CD11b⁻) cells of WT (top) and R972Q (bottom) mice.



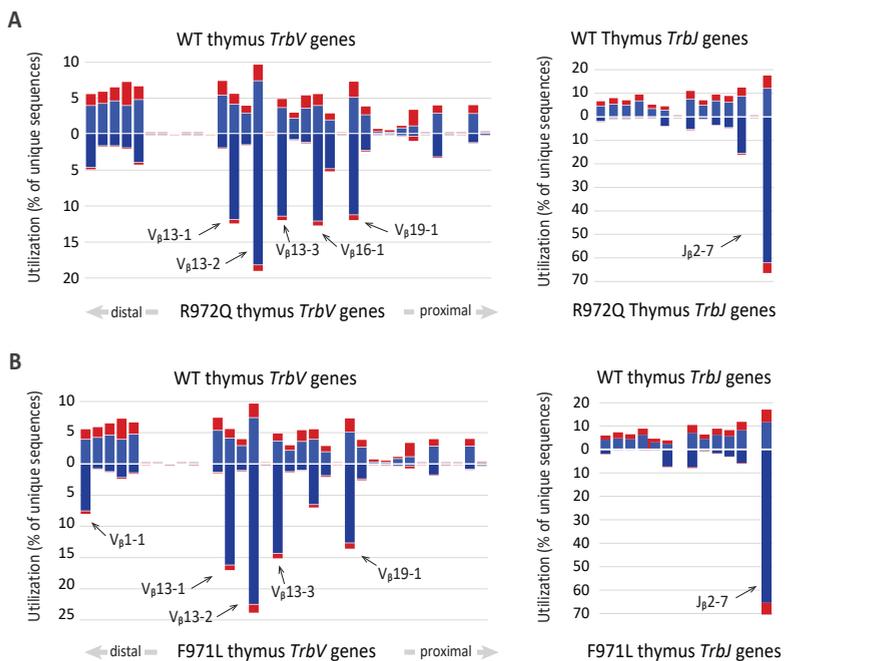


Figure S6: Thymocyte Trb repertoire

[A] Utilization of V (left) and J (right) genes in Trb VDJ rearrangements in unsorted thymocytes from WT (top, n=3), and R972Q (bottom, n=3) mice.

[B] Utilization of V (left) and J (right) genes in Trb VDJ rearrangements in unsorted thymocytes from WT (top, n=3), and F971L (bottom, n=3) mice.

[C] Shannon Entropy as a measure of diversity of the Trb repertoire of Treg and Tconv cells.

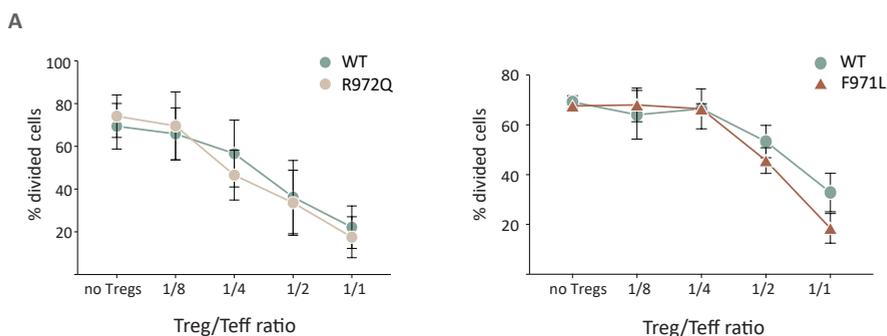


Figure S7: Treg suppressive function

A) Treg suppression assay comparing Treg cell function in WT, R972Q and F971L mice. Data points represent means and SEM of 2 independent experiments.

Table S1. Genotypes of generated F0 mice (related to figure S1)

	Allele 1	Allele 2
1	c. 2911T>C	6bp indel
2*	c. 2911T>C	c. 2911T>C
3 ¹	c. 2911T>C	c. 2911T>C
4 ²	c. 2911T>C	c. 2911T>C
5 ³	c. 2914C>T	c. 2911T>C
6	c. 2911T>C	c. 2911T>C
7	c. 2911T>C	2bp indel
8*	c. 2915G>A	c. 2915G>A
9 ⁴	c. 2915G>A	c. 2914C>T
10	c. 2915G>A	c. 2914C>T
11	c. 2911T>C	6bp indel
12	1bp deletion	c. 2914C>T
13	c. 2914C>T	156bp indel
14	c. 2911T>C	c. 2911T>C
15	c. 2911T>C	c. 2911T>C
16*	c. 2914C>T	c. 2914C>T
17*	c. 2915G>A	c. 2908C>A (silent mut)
18	c. 2915G>A	c. 2914C>T
19*	c. 2911T>C	c. 2914C>T
20 ⁵	c. 2911T>C	c. 2915G>A
21	c. 2914C>T	9bp deletion
22	c. 2914C>T	c. 2914C>T
23	c. 2911T>C	c. 2911T>C
24	c. 2911T>C	c. 2911T>C
25	WT	4bp deletion
26	c. 2911T>C	c. 2915G>A

¹ 17bp indel on one allele

² 1bp indel on one allele

³ Additional third sequence with c.2915G>A due to somatic chimerism

⁴ 9 bp indel on one allele

⁵ 1bp indel on one allele

* F0 mice that were selected to continue breeding with, to generate 2 lines per mutation

Table S2. Repertoire sequencing counts (related to figure 6 and 7 and figure S4 and S5)

Sample	VDJ		
	Total	Unique	Unique productive
pro-B WT	4439	3149	1729
pro-B R972Q (2)	1543	610	431
Pro-B F971L (2)	482	112	67
Spleen WT	129721	15617	10674
Spleen R972Q (2)	87132	10967	10170
pre-B WT (3)	317993	171193	124139
pre-B R972Q (3)	187524	56416	52571
Pre-B F971L (3)	20208	11030	10309
Thymus WT (3)	1505542	932332	768007
Thymus R972Q (3)	559557	94890	89866
Thymus F971L (3)	53583	4870	4152
DN4 WT	84084	56679	35104
DN4 R972Q	199	128	95
DP WT	1459836	1144368	795466
DP R972Q	515080	74453	70049
DP F971L (2)	74457	17501	16450
WT T conv	80705	65569	45410
WT Treg	44015	33666	22600
R972Q Tconv	63692	17664	16797
R972Q Treg	21781	7954	7502
F971L Tconv	37247	8820	8294
F971L Treg	10363	2486	2276

Sample	HTGTS total VDJ sequences ¹	HTGTS DJ sequences ²
pro-B WT	8178	9448
pro-B R972Q (2)	1031+1852	1884+4048
Pro-B F971L (2)	286+747	1280+4043

¹ used for VDJ/DJ ratio

² used for D graph

Gene	Cell	WT vs.	p-value
V	pro-B-cells	R972Q	1.922E-12
V	pro-B-cells	F971L	4.270E-22
V	pre-B-cells	R972Q	4.400E-02
V	pre-B-cells	F971L	8.360E-07
V	spleen B-cells	R972Q	1.980E-03
V	peritoneal B1 cells	R972Q	3.300E-02
V	peritoneal B2 cells	R972Q	2.400E-02
V	DN4 thymocytes	R972Q	3.340E-08
J	DN4 thymocytes	R972Q	1.870E-05
V	DP thymocytes	R972Q	1.584E-05
V	DP thymocytes	F971L	8.800E-07
J	DP thymocytes	R972Q	1.189E-04
J	DP thymocytes	F971L	1.870E-05
V	unsorted thymocytes	R972Q	2.270E-05
V	unsorted thymocytes	F971L	2.270E-05
J	unsorted thymocytes	R972Q	1.189E-04
J	unsorted thymocytes	F971L	1.870E-05
V	CD4+ Tconv	R972Q	9.680E-03
V	CD4+ Tconv	F971L	1.655E-03
J	CD4+ Tconv	R972Q	6.434E-04
J	CD4+ Tconv	F971L	1.189E-04
V	CD4+ Treg	R972Q	9.680E-03
V	CD4+ Treg	F971L	7.290E-05
J	CD4+ Treg	R972Q	6.434E+04
J	CD4+ Treg	F971L	1.189E-04

Table S3. Comparison of gene distributions of WT vs. mutant, using Kolmogorov-Smirnov

Table S4. IgH V genes in pro-B and splenic B-cells (related to figure 6)
p-values for wild type vs. mutation calculated with Chi square test:

*	$p \leq 0.05$
**	$p \leq 0.01$
** *	$p \leq 0.001$
** **	$p \leq 0.0001$

Table S4

V genes	pro-B unique			spleen unique	
	WT	R972Q	F791L	WT	R972Q
<i>Ighv1-86</i>	0.03	0.16	0.00	0.01	0.00
<i>Ighv1-85</i>	0.32	0.66	0.89	0.79	1.38 **
<i>Ighv1-84</i>	0.25	0.16	0.00	0.34	0.42
<i>Ighv1-83</i>	0.16	0.00	0.00	0.08	0.01 *
<i>Ighv1-82</i>	1.87	1.97	2.68	3.28	2.80 *
<i>Ighv1-81</i>	2.41	3.11	2.68	2.86	4.40 **
<i>Ighv1-80</i>	0.89	1.64	4.46	2.15	2.22 *
<i>Ighv1-79</i>	0.16	0.00	0.00	0.11	0.00 **
<i>Ighv1-78</i>	1.02	0.66	0.89	0.90	0.78

Table S4

V genes	pro-B unique			spleen unique	
	WT	R972Q	F791L	WT	R972Q
<i>Ighv1-77</i>	0.29	0.00	0.00	0.39	0.28
<i>Ighv8-16</i>	0.00	0.00	0.00	0.00	0.00
<i>Ighv1-76</i>	1.30	1.15	0.00	1.28	1.31
<i>Ighv8-15</i>	0.00	0.00	0.00	0.00	0.00
<i>Ighv1-75</i>	1.62	1.15	0.89	1.25	1.43
<i>Ighv8-14</i>	0.00	0.00	0.00	0.00	0.00
<i>Ighv1-74</i>	0.95	1.15	0.89	1.17	1.73 **
<i>Ighv1-73</i>	0.03	0.00	0.00	0.06	0.00
<i>Ighv8-13</i>	0.00	0.00	0.00	0.00	0.00

Continues on next page ...

Table S4 continued ...

Table S4	pro-B unique			spleen unique		
	V genes	WT	R972Q	F791L	WT	R972Q
<i>Ighv1-72</i>	2.06	4.26	*	1.79	2.44	3.72 **
<i>Ighv1-71</i>	0.00	0.00		0.00	0.00	**
<i>Ighv1-70</i>	0.51	0.16		1.79	0.21	0.00 **
<i>Ighv8-12</i>	1.17	0.49		0.89	1.09	0.80 *
<i>Ighv1-69</i>	0.83	0.82		0.89	1.96	0.89 **
<i>Ighv1-68</i>	0.00	0.00		0.00	0.00	**
<i>Ighv1-67</i>	0.98	0.82		0.89	0.63	0.01 **
<i>Ighv1-66</i>	0.41	0.16		0.00	0.54	0.31 *
<i>Ighv8-11</i>	0.25	0.33		0.89	0.14	0.01 **
<i>Ighv1-65</i>	0.00	0.00		0.00	0.00	**
<i>Ighv8-10</i>	0.00	0.00		0.00	0.00	**
<i>Ighv1-64</i>	2.54	1.80		2.68	3.49	4.17 *
<i>Ighv1-63</i>	0.57	0.33		0.00	0.47	0.14 **
<i>Ighv8-9</i>	0.03	0.00		0.00	0.04	0.00 **
<i>Ighv1-62-3</i>	0.57	0.33		0.00	0.60	0.28 **
<i>Ighv1-62-2</i>	0.00	0.00		0.00	0.00	**
<i>Ighv1-62-1</i>	0.00	0.00		0.00	0.00	**
<i>Ighv1-62</i>	0.06	0.00		0.00	0.07	0.01 *
<i>Ighv1-61</i>	0.19	0.16		0.00	0.65	0.19 **
<i>Ighv1-60</i>	0.00	0.00		0.00	0.00	**
<i>Ighv1-59</i>	0.60	0.82		0.00	0.79	0.41 **
<i>Ighv1-58</i>	0.41	0.16		0.00	0.49	0.16 **
<i>Ighv8-8</i>	1.43	2.46		1.79	1.76	1.87 **
<i>Ighv1-57</i>	0.00	0.00		0.00	0.00	**
<i>Ighv8-7</i>	0.51	0.16		0.89	0.22	0.01 **
<i>Ighv1-56</i>	0.10	0.00		0.00	0.10	0.05 **
<i>Ighv1-55</i>	2.92	1.97		0.89	2.88	2.39 *
<i>Ighv1-54</i>	0.60	0.82		0.89	0.71	0.44 *
<i>Ighv8-6</i>	0.00	0.00		0.00	0.06	0.00 *
<i>Ighv1-53</i>	1.68	1.15		0.00	2.43	2.18 **
<i>Ighv1-52</i>	0.35	0.49		0.00	1.04	0.56 **
<i>Ighv1-51</i>	0.00	0.16		0.00	0.00	0.00 **
<i>Ighv1-50</i>	0.64	0.66		0.89	1.14	1.09 **
<i>Ighv8-5</i>	0.22	0.16		0.00	0.17	0.03 **
<i>Ighv1-49</i>	0.51	0.33		0.00	0.33	0.04 **
<i>Ighv1-48</i>	0.00	0.00		0.00	0.01	0.00 **
<i>Ighv8-4</i>	0.10	0.00		0.00	0.06	0.00 *
<i>Ighv8-3</i>	0.00	0.00		0.00	0.00	0.00 **

Table S4 continued ...

Table S4	pro-B unique			spleen unique		
	V genes	WT	R972Q	F791L	WT	R972Q
<i>Ighv1-47</i>	0.92	0.33		0.00	0.56	0.16 **
<i>Ighv1-46</i>	0.03	0.00		0.00	0.00	0.00 **
<i>Ighv1-45</i>	0.00	0.00		0.00	0.00	0.00 **
<i>Ighv1-44</i>	0.00	0.00		0.00	0.00	0.00 **
<i>Ighv1-43</i>	0.00	0.00		0.00	0.04	0.01 **
<i>Ighv1-42</i>	0.64	0.33		0.00	0.81	0.62 **
<i>Ighv1-41</i>	0.00	0.00		0.00	0.00	0.00 **
<i>Ighv1-40</i>	0.06	0.00		0.00	0.02	0.00 **
<i>Ighv1-39</i>	1.17	1.64		1.79	1.82	2.09 **
<i>Ighv1-38</i>	0.32	0.16		0.00	0.13	0.00 **
<i>Ighv1-37</i>	0.06	0.00		0.00	0.09	0.00 *
<i>Ighv1-36</i>	0.44	0.66		1.79	0.47	0.46 **
<i>Ighv1-35</i>	0.00	0.00		0.00	0.00	0.00 **
<i>Ighv1-34</i>	0.13	0.33		0.00	0.21	0.47 **
<i>Ighv1-33</i>	0.13	0.16		0.89	0.38	0.16 *
<i>Ighv1-32</i>	0.06	0.00		0.00	0.05	0.00 **
<i>Ighv1-31</i>	0.19	0.16		0.00	0.18	0.20 **
<i>Ighv1-30</i>	0.00	0.00		0.00	0.00	0.00 **
<i>Ighv1-29</i>	0.00	0.00		0.00	0.00	0.00 **
<i>Ighv1-28</i>	0.03	0.00		0.00	0.04	0.00 **
<i>Ighv1-27</i>	0.00	0.00		0.00	0.00	0.00 **
<i>Ighv1-26</i>	3.49	4.75		0.89	5.48	6.23 *
<i>Ighv1-25</i>	0.10	0.33		0.00	0.06	0.00 *
<i>Ighv1-24</i>	0.00	0.00		0.00	0.00	0.00 **
<i>Ighv1-23</i>	0.41	0.00		0.00	0.32	0.00 **
<i>Ighv1-22</i>	0.95	0.82		2.68	1.68	2.59 **
<i>Ighv1-21</i>	0.00	0.16		0.00	0.02	0.00 **
<i>Ighv1-21-1</i>	0.10	0.00		0.00	0.05	0.00 **
<i>Ighv1-20</i>	0.19	0.00		0.00	0.15	0.03 *
<i>Ighv1-19</i>	2.03	4.26	*	3.57	1.52	3.15 **
<i>Ighv1-19-1</i>	0.06	0.00		0.00	0.02	0.00 **
<i>Ighv1-18</i>	2.41	2.46		6.25	2.71	3.33 *
<i>Ighv1-17</i>	0.00	0.00		0.00	0.00	0.00 **
<i>Ighv1-17-1</i>	0.00	0.00		0.00	0.00	0.00 **
<i>Ighv1-16</i>	0.00	0.00		0.00	0.00	0.00 **
<i>Ighv1-15</i>	0.64	0.66		0.89	1.18	0.81 *
<i>Ighv1-14</i>	0.03	0.16		0.00	0.08	0.00 *
<i>Ighv1-13</i>	0.03	0.16		0.00	0.04	0.00 **

Table S4 continued ...

Table S4 V genes	pro-B unique			spleen unique		
	WT	R972Q	F791L	WT	R972Q	
<i>Ighv1-12</i>	0.51	0.00	0.00	0.81	0.26	** **
<i>Ighv1-11</i>	1.59	0.66	0.89	0.76	0.09	** **
<i>Ighv1-10</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv1-9</i>	1.68	0.98	1.79	2.52	1.29	** **
<i>Ighv15-2</i>	0.76	0.33	0.00	0.60	0.16	** **
<i>Ighv1-8</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv10-4</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv1-7</i>	1.02	1.48	0.00	1.17	1.69	** *
<i>Ighv1-6</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv10-3</i>	0.95	1.15	3.57	1.12	1.42	*
<i>Ighv1-5</i>	1.30	1.48	2.68	1.02	1.18	
<i>Ighv10-2</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv1-4</i>	0.48	0.98	0.00	0.58	0.61	
<i>Ighv1-3</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv10-1</i>	1.65	3.28	* 1.79	1.08	1.62	** *
<i>Ighv1-2</i>	1.40	1.48	0.89	0.76	0.16	** **
<i>Ighv8-2</i>	1.05	1.31	1.79	0.44	0.01	** **
<i>Ighv6-7</i>	0.03	0.00	0.00	0.03	0.00	
<i>Ighv6-6</i>	0.86	1.31	1.79	0.93	1.81	** **
<i>Ighv6-5</i>	0.06	0.33	0.00	0.07	0.00	*
<i>Ighv6-4</i>	0.03	0.00	0.00	0.01	0.00	
<i>Ighv6-3</i>	1.27	1.97	2.68	1.20	3.10	** **
<i>Ighv12-3</i>	0.89	0.33	0.89	0.53	0.02	** **
<i>Ighv13-2</i>	1.05	0.49	1.79	0.58	0.25	** **
<i>Ighv8-1</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv3-8</i>	0.83	2.13	* 2.68	0.68	1.68	** **
<i>Ighv5-21</i>	0.00	0.00	0.00	0.02	0.00	
<i>Ighv3-7</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv9-4</i>	0.79	0.82	0.00	0.35	0.36	
<i>Ighv3-6</i>	2.38	4.10	* 4.46	2.73	5.27	** **
<i>Ighv13-1</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv3-5</i>	0.25	0.33	0.00	0.17	0.17	
<i>Ighv3-4</i>	0.25	0.16	0.00	0.15	0.05	*
<i>Ighv7-4</i>	0.10	0.16	0.00	0.10	0.07	
<i>Ighv3-3</i>	0.51	0.00	0.00	0.27	0.00	** **
<i>Ighv14-4</i>	0.83	0.16	0.89	1.22	0.23	** **
<i>Ighv15-1</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv7-3</i>	0.70	1.15	0.00	0.70	1.53	** **

Table S4 continued ...

Table S4 V genes	pro-B unique			spleen unique		
	WT	R972Q	F791L	WT	R972Q	
<i>Ighv9-3</i>	2.29	3.93	* 5.36	2.64	5.54	** **
<i>Ighv12-2</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv9-2</i>	0.51	0.82	0.00	0.26	0.45	*
<i>Ighv12-1</i>	0.10	0.00	0.00	0.06	0.01	*
<i>Ighv9-1</i>	0.38	0.00	0.89	0.53	0.81	*
<i>Ighv6-2</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv16-1</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv14-3</i>	0.92	1.31	1.79	0.95	1.34	*
<i>Ighv11-2</i>	0.44	0.16	0.00	0.35	0.03	** **
<i>Ighv3-2</i>	0.54	1.15	0.89	0.57	0.03	** **
<i>Ighv4-2</i>	0.60	0.66	0.00	0.36	0.02	** **
<i>Ighv14-2</i>	0.60	0.33	0.00	1.20	0.67	** **
<i>Ighv11-1</i>	0.38	0.33	0.00	0.13	0.00	** *
<i>Ighv3-1</i>	0.32	0.82	0.00	0.42	0.49	
<i>Ighv4-1</i>	1.33	0.98	0.00	0.62	0.48	
<i>Ighv14-1</i>	0.10	0.00	0.00	0.42	0.13	** **
<i>Ighv7-2</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv7-1</i>	0.98	1.48	1.79	0.60	0.61	
<i>Ighv5-19</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv2-9</i>	0.79	1.31	0.00	0.67	0.33	** *
<i>Ighv2-8</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv5-18</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv5-17</i>	0.76	0.98	1.79	1.56	1.50	
<i>Ighv5-16</i>	0.83	0.98	3.57	0.92	1.40	** *
<i>Ighv5-15</i>	0.32	0.66	0.00	0.24	0.17	
<i>Ighv2-7</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv2-6-8</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv5-13</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv2-9-1</i>	0.54	1.31	* 0.00	0.64	1.11	** **
<i>Ighv5-12-4</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv5-9-1</i>	0.51	0.16	0.00	0.45	0.28	*
<i>Ighv2-6</i>	0.76	0.82	0.00	0.44	0.41	
<i>Ighv5-12</i>	0.95	1.48	0.89	0.48	0.22	** *
<i>Ighv5-11</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv2-5</i>	0.86	0.49	0.89	0.56	0.57	
<i>Ighv5-10</i>	0.00	0.00	0.00	0.00	0.00	
<i>Ighv5-9</i>	0.35	0.33	1.79	0.24	0.56	** **

Trb V	Spleen Treg			Spleen Tconv			Thymus DN4		Thymus DP		
	WT	R972Q	F971L	WT	R972Q	F971L	WT	R972Q	WT	R972Q	F971L
01-1	5.81	6.52 *	14.00 **	6.92	6.51	11.20 **	4.81	4.69	4.90	4.08 **	7.91 **
02-1	6.88	1.93 **	0.46 **	6.90	2.08 **	0.66 **	5.29	2.34	5.46	1.46 **	0.45 **
03-1	8.55	3.66 **	2.19 **	7.65	3.22 **	1.75 **	5.66	0.78	5.99	1.22	0.87
04-1	4.83	1.32 **	2.11 **	6.01	1.42 **	1.68 **	6.69	2.34	7.19	1.98	2.56
05-1	7.83	5.78 **	2.03 **	7.29	5.45 **	1.48 **	6.29	9.38	6.64	4.15	1.40
06-1	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.78	0.00	0.00	0.01
07-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00 **	0.00 **
08-1	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
09-1	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00 *	0.00	0.00 **	0.00 **
10-1	0.06	0.03	0.00	0.08	0.00	0.00	0.09	0.00	0.06	0.00 **	0.00 **
11-1	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.78	0.00	0.00 **	0.01 **
12-1	4.69	3.08 **	2.40 **	2.67	0.82 **	0.53 **	8.93	1.56	7.92	2.05 **	1.29 **
13-1	5.08	8.01 **	9.98 **	5.85	9.13 **	13.71 **	5.37	9.38	5.42	11.68	16.60
12-2	3.07	2.23 **	1.16 **	1.72	0.98 **	0.52 **	4.88	1.56 **	4.22	1.51 **	0.89 **
13-2	11.22	17.03 **	23.15 **	11.98	18.33 **	24.51 **	8.90	13.28	8.85	19.44 **	26.25 **
12-3	0.01	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.00 **	0.00 **
13-3	5.61	11.34 **	12.71 **	6.13	12.73 **	15.55 **	4.64	10.94 **	4.38	12.30 **	13.85 **
14-1	1.57	0.20 **	0.37 **	1.91	0.40 **	0.42 **	2.74	3.13	3.01	0.72 **	1.00 **
15-1	4.66	1.39 **	1.28 **	3.94	1.30 **	0.94 **	5.38	7.03	5.40	1.24	1.00
16-1	5.43	13.61 **	9.23 **	5.07	12.46 **	6.67 **	5.10	11.72	5.15	12.74 **	6.03 **
17-1	1.54	1.92 *	0.33 **	1.61	1.81 **	0.53 **	2.63	5.47 *	2.63	5.25 **	1.86 **
18-1	0.00	0.01	0.04	0.00	0.00	0.01	0.00	0.78	0.00	0.00 **	0.01 **
19-1	7.70	14.95 **	14.04 **	8.62	16.42 **	14.79 **	6.60	7.03	6.85	12.37 **	12.62 **
20-1	4.14	2.57 **	2.36 **	4.26	2.60 **	2.70 **	3.61	0.00	3.62	2.53 **	2.57 **
21-1	0.37	0.01 **	0.00 *	0.42	0.00 *	0.02 *	0.60	0.00	0.53	0.00 **	0.00 **
22-1	0.34	0.00 **	0.00 *	0.24	0.00 *	0.00 *	0.29	0.00	0.26	0.00	0.00
23-1	0.93	0.28 **	0.00 **	1.01	0.22 **	0.11 **	0.94	0.00	0.93	0.18 **	0.06 **
24-1	3.12	0.08 **	0.00 **	2.85	0.11 **	0.02 **	3.71	1.56	3.03	0.20	0.08
25-1	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.78	0.00	0.00	0.01
26-1	3.95	3.42 *	1.74 **	3.74	3.38 **	1.72 **	3.66	0.78	3.77	3.42 **	1.73 **
27-1	0.01	0.08	0.21	0.00	0.02	0.04	0.00	1.56	0.00	0.01 **	0.10 **
28-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
29-1	2.23	0.42 **	0.04 **	2.98	0.45 **	0.26 **	3.00	2.34	3.53	1.33	0.74
30-1	0.32	0.13 *	0.04 *	0.15	0.16 *	0.14 *	0.19	0.00	0.24	0.13	0.10

Table S5: Trb V gene usage (related to figure 7)
p-values for wild type vs. mutation calculated with Chi square test.

*	$p \leq 0.05$	**	$p \leq 0.001$
**	$p \leq 0.01$	**	$p \leq 0.0001$

Trb J	DN4			DP			Spleen Treg			Spleen TConv								
	WT	R972Q	*	WT	R972Q	F971L	WT	R972Q	F971L	WT	R972Q	F971L						
1-01	10.55	3.91	*	10.36	1.56	**	2.03	**	8.46	0.68	**	1.11	**	4.27	0.67	**	0.89	**
1-02	6.80	0.78	*	7.01	0.17	**	0.31	**	5.59	0.05	**	0.05	**	7.12	0.09	**	0.09	**
1-03	6.08	1.56	*	6.22	0.30	**	0.23	**	6.01	0.16	**	0.05	**	6.91	0.20	**	0.12	**
1-04	8.22	0.00	*	8.56	0.10	**	0.18	**	7.12	0.11	**	0.18	**	8.49	0.07	**	0.14	**
1-05	3.82	0.00	*	4.07	0.12	**	0.22	**	3.14	0.05	**	0.09	**	4.00	0.07	**	0.10	**
1-06	4.17	4.69		4.29	3.28	**	6.62	**	2.97	5.42	**	11.33	**	2.87	4.30	**	8.56	
1-07	0.04	0.00		0.04	0.01	**	0.01	*	0.02	0.00		0.00		0.02	0.00		0.01	
2-01	9.71	2.34	*	9.95	4.75	**	7.29	**	9.96	4.01	**	5.73	**	10.01	4.34	**	6.05	**
2-02	5.86	4.69		6.14	0.37	**	0.68	**	6.40	0.48	**	0.82	**	6.77	0.49	*	0.93	**
2-03	8.88	3.13	*	9.14	2.97	**	1.48	**	8.07	1.85	**	1.63	**	8.98	2.33	**	1.55	**
2-04	7.97	4.69		8.10	3.86	**	2.70	**	9.35	4.88	**	3.77	**	8.15	4.36	**	3.12	**
2-05	10.67	15.63		11.35	14.54	**	6.52	**	12.48	15.80	**	7.12	**	12.50	15.53	**	7.34	**
2-06	0.03	0.00		0.03	0.01	**	0.02		0.02	0.00		0.00		0.02	0.00		0.02	
2-07	17.07	58.59	**	14.66	67.58	**	71.70	**	20.40	66.49	**	68.13	**	19.90	67.55	**	71.08	**

Table S6. Trb J gene usage (related to figure 7)

p-values for wild type vs. mutation calculated with Chi square test.

*	$p \leq 0.05$
**	$p \leq 0.01$
**	$p \leq 0.001$
**	$p \leq 0.0001$

R972Q >WT	Coding end	F971L >WT	Coding end
Ighv1-19	ga	Ighv1-19	ga
Ighv1-72	ga	Ighv1-18	ga
Ighv3-6	ga	Ighv3-6	ga
Ighv9-3	ga	Ighv9-3	ga
Ighv10-1	ca	Ighv5-16	ga
Ighv3-8	ta	Ighv3-8	ta
Ighv1-26	ga	Ighv10-3	ga
Ighv8-8	ag	Ighv1-22	ga
Ighv2-9-1	aa	Ighv5-9	ca
Ighv1-80	ga	Ighv1-80	ga

WT > R972Q	Coding end	WT > F971L	Coding end
Ighv14-4	ca	Ighv1-7	ga
Ighv8-12	ag	Ighv2-3	cc
Ighv1-9	ga	Ighv1-76	ga
Ighv1-64	ga	Ighv4-1	ca
Ighv1-11	gg	Ighv1-53	ga
Ighv1-55	ga	Ighv5-6	ca
Ighv2-3	cc	Ighv1-55	ga
Ighv5-4	ga	Ighv5-4	ga
Ighv2-2	aa	Ighv1-26	ga
Ighv5-2	ca	Ighv5-2	ca

WT = R972Q	Coding end	WT = F971L	Coding end
Ighv1-18	ga	Ighv1-69	ga
Ighv1-82	ga	Ighv14-4	ca
Ighv5-6	ca	Ighv2-5	aa
Ighv1-2	ga	Ighv12-3	ga
Ighv1-76	ga	Ighv1-74	ta
Ighv1-22	ga	Ighv5-12	ca
Ighv1-69	ga	Ighv1-67	ga
Ighv9-4	ga	Ighv1-78	ga
Ighv2-6	ga	Ighv1-72	ga
Ighv1-50	ga	Ighv8-12	ag

Table S7: Igh V gene Coding end comparison in pro-B-cells (related to discussion)

R972Q >WT	Coding end	F971L >WT	Coding end
TCRBV13-02	tg	TCRBV13-02	tg
TCRBV13-03	tg	TCRBV13-01	tg
TCRBV16-01	ga	TCRBV13-03	tg
TCRBV13-01	tg	TCRBV19-01	ag
TCRBV19-01	ag	TCRBV01-01	ga
TCRBV17-01	ga	TCRBV16-01	ga

WT > R972Q	Coding end	WT > F971L	Coding end
TCRBV12-02	tc	TCRBV14-01	tc
TCRBV24-01	ta	TCRBV26-01	tc
TCRBV05-01	ga	TCRBV29-01	tc
TCRBV02-01	ga	TCRBV24-01	ta
TCRBV15-01	gc	TCRBV12-02	tc
TCRBV03-01	gc	TCRBV15-01	gc
TCRBV04-01	ga	TCRBV04-01	ga
TCRBV12-01	tc	TCRBV02-01	ga

WT = R972Q	Coding end	WT = F971L	Coding end
TCRBV30-01	ga	TCRBV30-01	ga
TCRBV22-01	tc	TCRBV22-01	tc
TCRBV26-01	tc	TCRBV21-01	tc
TCRBV21-01	tc	TCRBV17-01	ga
TCRBV23-01	tc	TCRBV23-01	tc
TCRBV01-01	ga	TCRBV20-01	tg

Table S8: Trb V gene coding end comparison in DP cells (related to discussion)

Supplementary methods chapter 5

Mice

A guide RNA (gRNA) was designed to target nucleotides 2919 and 2920 of the mouse *Rag1* gene (CCDS16463.1), in a region within the C-terminus domain (CTD).

To introduce the desired mutations into the mouse *Rag1* locus, three single-stranded 160 bp DNA oligonucleotide (ssODN) donor templates (ordered from Integrated DNA Technologies, Coralville, IA) were used, that were highly homologous to the area flanking the cutting site, and containing the following protospacer regions, each containing the desired point mutation (in **red**):
ACAAGCTGTTTAGAAAGGTTTC**AG** (R972Q, or c.2915G>A),
ACAAGCTGTTTAGAAAGG**CTTC**GG (F971L, or c. 2911T>C),
ACAAGCTGTTTAGAAAGGTTTT**T**GG (R972W, or c.2914C>T).

In addition to the desired mutation, all ssODNs contained a silent mutation at c. 2908 C>A (in **blue**) to prevent additional cutting after the sequence of the ssODN has been incorporated into the targeted mouse genome.

100ng/ul of Cas9 mRNA (purchased from System Biosciences, Palo Alto, CA), 50ng/ul of gRNA (generated from PCR product using Megashortscript T7 kit, Life Technologies, Washington D.C.), and 150 ng/μl of a mixture with equal amounts of the three ssODNs were injected in 200 zygotes of C57BL/6 mice. Embryos were placed back into six recipient C57BL/6 mice, as reported previously¹.

F0 genotypes were analyzed using Sanger sequencing. F0 mice were bred with wt C57BL/6 mice to get heterozygous F1 mice, which were bred to obtain homozygous F2 mice. All mouse experiments were done with F2 or later generations. *Rag1*^{-/-} control mice and the *Foxp3*EGFPCre mice² were purchased from the Jackson Laboratory (Bar Harbor, ME). *Foxp3*EGFPCre mice were crossed with homozygous F2 mutants. Wild-type control mice were derived from wild-type F2 littermates. All mice described in this manuscript were on a C57BL/6 background. Offspring of different founders were compared to rule out any effects of offsite targeting. Animal work was conducted adhering to the institution's guidelines for animal use, and followed the guidelines and basic principles in the United States Public Health Service Policy on Humane Care and Use of Laboratory Animals, and the Guide for the Care and Use of Laboratory Animals by certified staff in an Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) International accredited facility.

Cell isolation

Thymic DP (CD4⁺ CD8⁺) and DN4 (CD4⁻ CD8⁻ CD44⁻ CD25⁻) cells, splenic Treg cells (CD4⁺ EGFP⁺) and Tconv cells (CD4⁺ EGFP⁻), bone marrow pro-B (B220^{low} IgM⁻ B220^{lo} IgM⁺ CD43⁺) and pre-B (B220^{low} IgM⁻ B220^{lo} IgM⁺ CD43⁻) and peritoneal B1 (CD19⁺ CD11b⁺), B2 (CD19⁺ CD11b⁻) and B1a (CD19⁺ CD11b⁺ CD5⁺) and B1b (CD19⁺ CD11b⁺ CD5⁻) cells were isolated on FACS Aria II (BD). Mature splenic IgM⁺ B-cells were purified using Miltenyi MACS bead purification.

Flow cytometry antibodies

Lymphocyte subsets were analyzed by flow cytometry using the BD LSRFortessa™ (BD Biosciences, San Jose, CA) with DIVA software, and Flowjo™. The following antibodies were used: CD3-PerCP-Cy 5.5 Clone 145-2c11 (BD Biosciences), CD4-PB clone RM 4-5 (eBioscience, San Diego, CA), CD8-APC-Cy7 Clone 53-6.7 (Biolegend, San Diego, CA), CD44-PE-Cy7 clone IM7 (Biolegend), CD25-Percp-cy5.5 and PE Clone PC61 (BD), CD44-FITC clone IM7 (BD Biosciences), CD62L-PE clone MEL-14 (Biolegend), B220-PB Clone RA3-6B2 (eBioscience), CD117-PE-Cy7 clone 2B8 (Biolegend), CD19-PB clone eBio1D3 (eBioscience), IgM-Percp-Cy5.5 and BV605 (eBioscience), CD43-PE and FITC clone eBioR2/60 (eBioscience), CD93-PE-Cy7 clone AA4.1 (eBioscience), CD21-FITC clone 7G6 (BD), CD23-PE clone B3B5 (eBioscience), lambda-PE clone RML-42 (Biolegend), γδ TCR-FITC clone GL3 (Biolegend), β TCR-PE-Cy7 clone H57-597 (eBioscience). For thymus staining, a lineage negative stain was used to exclude cells that were not thymocytes: B220-APC clone RA3-6B2 (Biolegend), CD11b-APC clone M1/70 (Biolegend), Ter119-APC clone TER119 (Biolegend). Annexin V eBioscience kit was used with Annexin V-APC and PI. For staining of peritoneal B-cells, the following antibodies were used: B220 APC-Cy7 clone C363-16A (Biolegend), CD5-APC clone 53-7.3 (Biolegend), CD3e Alexa Fluor 700 clone 500A2 (BD Pharmingen), CD11b PE clone M1/70 (BD Pharmingen), CD19 PerCP-Cy5.5 clone eBio1D3 (eBioscience). For staining of age-associated B-cells (ABCs), the following antibodies were used: CD19 FITC (clone eBio1D3, eBioscience), CD11c (clone HL3, BD PharMingen), 7-AAD as a vitality dye, IgG1κ Isotype control eFluor 660 (clone: P3.6.2.8.1, eBioscience) and T-bet eFluor 660 (clone 4B10, eBioscience). Protocol B (One-step protocol: intracellular (nuclear) proteins) of Thermofischer scientific was followed.

Immunoglobulin levels in naïve mice

Total IgG, IgM, and IgE immunoglobulin levels in serum of naïve mice were measured using ELISA as previously described³. Briefly, plates were coated overnight using the following antibodies: anti-mouse IgM (Southern Biotech, Birmingham, AL), anti-mouse IgG (Southern Biotech, Birmingham, AL), anti-mouse IgE (BD). Plates were blocked with 1-3% BSA for 2h and serum

dilutions were incubated overnight. For IgG detection, an IgG specific antibody conjugated to HRP was used (BD Biosciences). Rag2-deficient serum was used as a negative blank control. For IgM, an alkaline phosphatase-conjugated antibody was used (BD). For IgE determination, a biotinylated anti-IgE (BD Biosciences) primary antibody was used, followed by streptavidin-HRP. OD was measured at 405 nm and 450 nm for AP- and HRP-mediated reactions, respectively, using an ELISA reader (BioTek Instruments, Winooski, VT). A standard curve was included and exact concentrations could be determined.

Immunizations with TNP-Ficoll, TNP-KLH and pneumococcal WCA

To elicit a T-independent antibody response, 8-12 week old mutant mice and wild type mice were injected with 50 µg TNP(23)Ficoll (Biosearch Technologies, Petaluma, CA). TNP-specific IgM and IgG3 serum levels were determined by ELISA prior to immunization (d0) and 14 days post immunization (d14). IgM and IgG3 levels were measured by ELISA using goat anti-mouse antibody conjugated to AP specific for IgM and IgG3 followed by reaction with *p*-Nitrophenyl Phosphate, Disodium Salt (PNPP) and stopped by 3 N NaOH, and reading of OD at 405 nm. To elicit a T-dependent antibody response, 8-12 week old mice received a primary immunization with 100 µg TNP(18)KLH (Biosearch Technologies) mixed with aluminum adjuvant (volume ratio 2:1) and a booster immunization at d14 with 50 µg TNP(18)KLH with aluminum adjuvant. TNP-specific high affinity IgG was determined prior to immunization, at d14 (prior to booster) and at day 21 (d21) by coating the ELISA plate with TNP(5)BSA (Biosearch Technologies). To measure IgG levels, HRP-conjugated goat anti-mouse IgG (Sigma Aldrich) was used, followed by reaction with TMB that was stopped by sulfuric acid, and OD was read at 450 nm. For both TNP-Ficoll and TNP-KLH ELISAs five-fold serum dilutions were used: 1:100, 1:500, 1:2500, 1: 12500, 1:62500, 1:312000, 1:1562500.

To elicit a Th17 response, 8-12 week old mutant and wild type mice were immunized with 100 µg Pneumococcal whole cell antigen (WCA, a gift of Dr. Richard Malley, Boston Children's Hospital) in the presence of aluminum adjuvant at d0, followed by a booster immunization at d14 with 50 µg WCA without aluminum adjuvant. Antigen specific IgG was determined by coating the 96-well plate with WCA and determining specific IgG at d0, d14 and d21 with HRP-conjugated goat anti-mouse IgG (Sigma Aldrich, Saint Louis, MO), followed by reaction with TMB that was stopped by sulfuric acid, and reading of OD at 450 nm. Serum dilutions at 1:150, 1:450 and 1:1350 were used⁴.

Autoantibody microarray

Protein arrays were used to screen for a broad panel of IgM autoantibodies (University of Texas Southwestern Medical Center, Genomic and Microarray

Core Facility) as previously described⁵. Serum from naïve 8-12 week old mice was used. Values were normalized using positive IgM controls for each sample, generating a normalization factor. Each signal was multiplied by this factor. Values from negative control samples (wild-type mice) for each antigen were averaged and ratios were calculated between each sample and the average of negative controls plus 2SD. A heat map of these ratios was generated using Multi experiment viewer software (MeV, DFCI Boston, MA). Significant differences were assessed using the significance analysis of microarray (SAM, Stanford University Labs) with a false discovery rate <1.

Anti-cytokine autoantibody microarray

RayBio® mouse protein array G2 (Cat # PAM-G2, RayBiotech) was used to screen for cytokine autoantibodies (IgG) in mouse sera diluted 1:200. Protein array screening was performed in the recommended buffers according to the manufacturer's protocol. Arrays were scanned using a micro-array scanner (LuxScan HT24, BioCapital). GenePix pro microarray software (v5.1) was used for alignment and data acquisition. Results were quantile normalized to account for interplate variability⁶.

Next generation sequencing of *Trb*

T-cell receptor β (*Trb*) high throughput sequencing was performed by Adaptive Biotechnologies (Seattle, WA). The DNA was extracted from FACS-sorted lymphocyte populations. Multiplex Polymerase chain reaction (PCR), containing a mixture of all *V* and *J* gene primers, was used to amplify the rearranged CDR3 β ⁷. PCR products were sequenced using the Illumina HiSeq platform (Illumina, San Diego, CA). The sequences were aligned to a reference genome, and *Trb* *VDJ* gene definitions were based on the international ImMunoGeneTics (IMGT) system⁸. A complete, synthetic repertoire of TCRs was previously used to establish an amplification baseline and to correct for any amplification bias. Data deposited at: <https://doi.org/10.21417/B7MP7Q>.

Next generation sequencing of *Igh* rearrangements

The *Igh* locus of pre-B-cells (B220^{lo}IgM⁺CD43⁻) and peritoneal B1 cells (CD19⁺CD11b⁺), peritoneal B2 (CD19⁺CD11b⁻) cells, peritoneal B1a (CD19⁺CD11b⁺CD5⁺) and peritoneal B1b (CD19⁺CD11b⁺CD5⁻) cells was sequenced by Adaptive Biotechnologies using a similar technique as for the *Trb* sequencing.

However, with the technology of Adaptive Biotechnologies partial rearrangements (*DJ*) would not be picked up. Therefore, we sought a different in depth technology that would be able to show these partial products. This is particularly important at the pro-B-cell stage where *D*-to-*J* and *V*-to-

DJ rearrangements take place. The *Igh* locus of pro-B-cells was therefore sequenced using high-throughput genome-wide translocation sequencing-adapted repertoire sequencing (HTGTS-Rep-seq) as described^{9,10}. With this translocation-based sequencing technique, different *J* bait primers are used, so any sequence containing a *J* sequence can be analyzed, including *DJ* rearrangements that do not contain a *V* gene sequence. Adaptive data are deposited at: <https://doi.org/10.21417/B7MP7Q>. HTGTS-Rep-seq data are deposited at: <http://www.ncbi.nlm.nih.gov/bioproject/421212>.

PCR amplification of *Vκ-Jκ* rearrangements

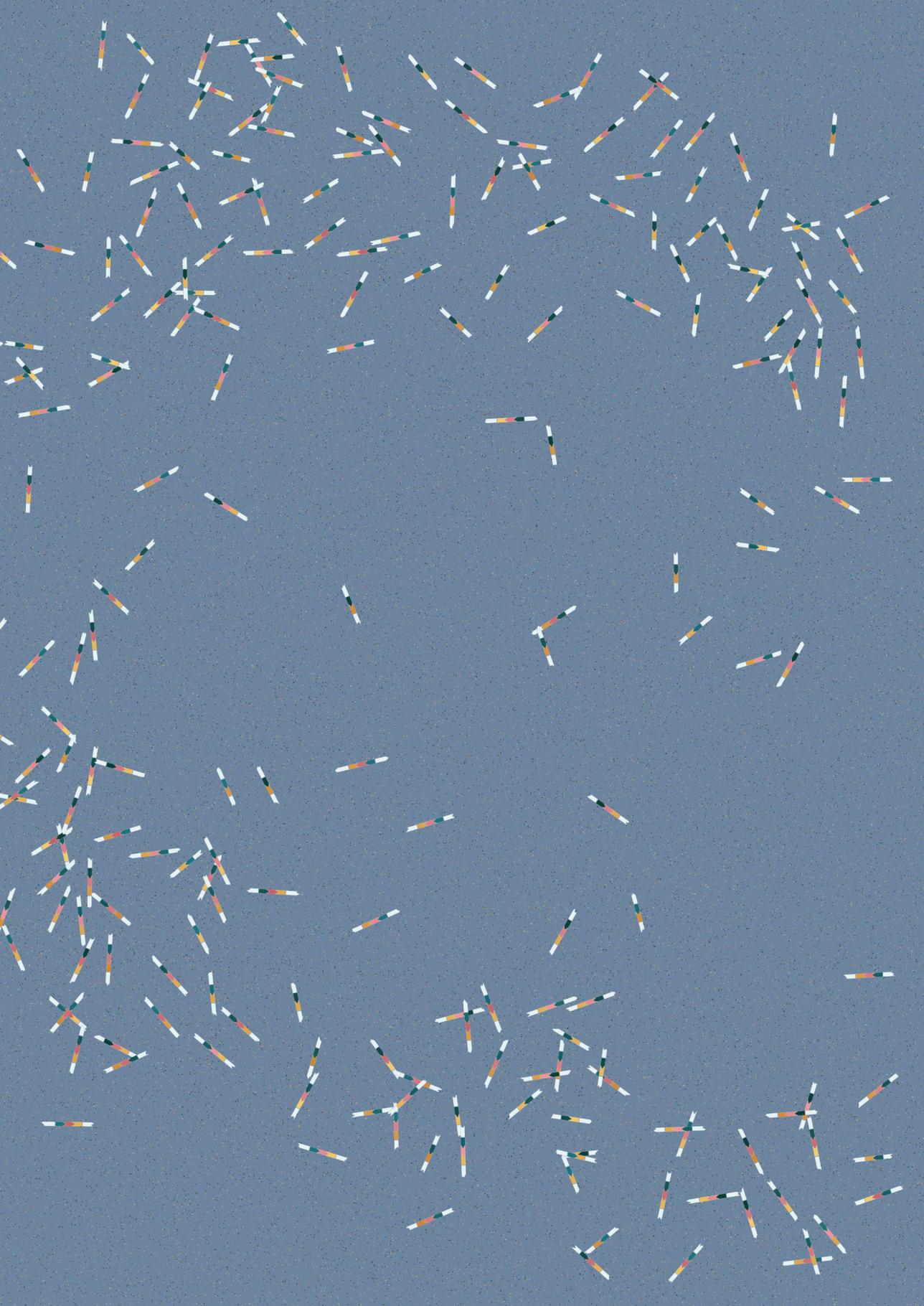
Vκ-to-Jκ rearrangement PCRs were performed as previously described¹¹ in pre-B-cells (B220^{lo}IgM⁻ CD43⁻) of wild-type and R972Q mutant mice. Primers flanking exon 6 of the *Dlg5* gene were used as a loading control under the same conditions. PCR products were gel electrophoresed and transferred to determine *VJ* recombination by Southern blotting.

Treg suppression assay

CD4⁺ T-cells were isolated with a CD4 negative isolation kit (Miltenyi). The FACSARIA II was used to separate Treg cells (CD4⁺EGFP⁺) and use these as suppressor cells. The remaining CD4⁺EGFP⁻ cells were labeled with CellTrace Violet Cell Proliferation dye according to the manufacturer's instructions (Life Technologies) and used as responder cells. *Rag1*^{-/-} spleen cells were used as feeder cells as previously described¹² with a 1:4 ratio of responder to feeder cells. Different ratios of suppressor to responder cells were used in serial dilutions (1:1, 1:2, 1:4, 1:8 and no suppressors). 15,000-45,000 responder cells/well were used. Cells were stimulated for 4-5 d with 2 μg/ml of soluble anti-CD3 and cultured in 96-wells, in round-bottomed plates in triplicate. The percentage of divided cells was determined by flow cytometry and the average of the three values was plotted.

References

1. Ott de Bruin L, Yang W, Capuder K, et al. Rapid generation of novel models of RAG1 deficiency by CRISPR/Cas9-induced mutagenesis in murine zygotes. *Oncotarget*. 2016;7(11):12962-12974.
2. Zhou X, Jeker LT, Fife BT, et al. Selective miRNA disruption in T reg cells leads to uncontrolled autoimmunity. *J Exp Med*. 2008;205(9):1983-1991.
3. Walter JE, Rucci F, Patrizi L, et al. Expansion of immunoglobulin-secreting cells and defects in B cell tolerance in Rag-dependent immunodeficiency. *J Exp Med*. 2010;207(7):1541-1554.
4. Moffitt KL, Gierahn TM, Lu YJ, et al. T(H)17-based vaccine design for prevention of *Streptococcus pneumoniae* colonization. *Cell Host Microbe*. 2011;9(2):158-165.
5. Volpi S, Santori E, Abernethy K, et al. N-WASP is required for B-cell-mediated autoimmunity in Wiskott-Aldrich syndrome. *Blood*. 2016;127(2):216-220.
6. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185-193.
7. Robins HS, Campregher PV, Srivastava SK, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T-cells. *Blood*. 2009;114(19):4099-4107.
8. Alamyar E, Duroux P, Lefranc MP, Giudicelli V. IMGT((R)) 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. *Methods Mol Biol*. 2012;882:569-604.
9. Lin SG, Ba Z, Du Z, Zhang Y, Hu J, Alt FW. Highly sensitive and unbiased approach for elucidating antibody repertoires. *Proc Natl Acad Sci U S A*. 2016;113(28):7846-7851.
10. Hu J, Meyers RM, Dong J, Panchakshari RA, Alt FW, Frock RL. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat Protoc*. 2016;11(5):853-871.
11. Guo C, Yoon HS, Franklin A, et al. CTCF-binding elements mediate control of V(D)J recombination. *Nature*. 2011;477(7365):424-430.
12. Charbonnier LM, Wang S, Georgiev P, Sefik E, Chatila TA. Control of peripheral tolerance by regulatory T cell-intrinsic Notch signaling. *Nat Immunol*. 2015;16(11):1162-1173.



6 General discussion

Mutations in *RAG1* and *RAG2* cause distinct clinical and immunological phenotypes, including T- B- severe combined immune deficiency (SCID), Omenn syndrome (OS), atypical/leaky SCID (LS), $\gamma\delta$ SCID and combined immunodeficiency with granuloma and/or autoimmunity (CID-G/AI). Previously, *in vitro* assays^{1,2} to test the recombination activity of mutated RAG proteins and structural models³ of the RAG complex have given insights into why some mutations cause a more severe phenotype than others. In particular, leaky mutations allow for more residual RAG activity and result in phenotypes such as leaky SCID or CID-G/AI¹. However, phenotypic variability remains even between patients with mutations that affect the same region of *RAG1*, showing the same residual RAG1 activity using *in vitro* models. How these mutations cause these different phenotypes remains an open question.

The aim of this thesis was to use genome-editing to generate novel models of RAG1 deficiency that make it possible to study this broad spectrum of clinical and immunological phenotypes, with a focus on the CID-G/AI phenotype. In addition, optimizing these genome-editing tools would be helpful for future treatment options.

In vitro modeling using induced pluripotent stem cells (iPSCs)

For **chapter 3** I generated and characterized induced pluripotent stem cells (iPSCs) derived from the fibroblasts of a RAG1 patient carrying a homozygous c.1428delC mutation which results in a frameshift and premature termination (p.N476Kfs*16) (**Chapter 3, Fig. 3**, p. 46). This patient showed a SCID phenotype with severe lymphopenia and very low immunoglobulins, resulting in recurrent lower respiratory tract infections since the first month of life. *In vitro* T-cell differentiation by the Zúñiga-Pflücker lab showed that patient cells were able to briefly develop into CD4⁺CD8⁺ (DP) T-cells but this was not sustained and TCR repertoire was severely restricted. The same results were found for another SCID patient and Omenn patient (**Chapter 3, Fig. 6**, p. 51), confirming impaired T-cell development.

In vitro T-cell differentiation provides a useful assay to study different disease phenotypes. In addition, as explained in **Chapter 2**, it can be used as a read-out to study the effect of introducing novel mutations in iPSCs of healthy donors or to test for successful correction of mutations in patient iPSCs using genome-editing (**Chapter 2, Fig. 3**, p. 27). This way, gene correction strategies can be validated in a pre-clinical setting. I have attempted to correct RAG1 patient iPSCs, and was able to target and disrupt the *RAG1* locus efficiently, but homologous recombination was inefficient and correction did not occur. Others have shown that *in vitro* T-cell differentiation was restored after using CRISPR/Cas9 to correct a SCID causing JAK3 mutation⁴. With the rapid advances in genome-editing, more efficient approaches will continue to develop.

In vivo modeling using mouse models

In patients or iPSCs, environmental factors cannot be controlled for. In mouse models these factors can be manipulated while studying the mechanisms underlying the broad phenotypic spectrum of mutations in *RAG1*. In addition, *in vitro* T-cell differentiation takes place in an artificial setting that does not reflect the full complexity of the immune system.

Therefore, we sought to generate novel *Rag1* mouse models. As homologous recombination is much more efficient in murine ESCs than in human iPSCs, I was able to introduce a *Rag1* missense mutation in murine ESCs using CRISPR/Cas9. Unfortunately, these ESCs showed an abnormal karyotype and could not be injected into murine blastocysts to develop a mouse model (see Chapter 2 for explanation on generation of mouse models).

Shortly afterwards, facilities became available to directly inject CRISPR/Cas9 into murine zygotes. In **Chapter 4** I show how CRISPR/Cas9 can be directly injected into the murine zygote to generate *Rag1* mouse models. Since I was able to disrupt the area around residue 838, a region previously not well studied, I could study the function of this area in more detail. It turned out that all frameshift mutations leading to an early stop, but also in-frame deletions of just one amino acid, resulted in a complete *Rag1*^{-/-} phenotype.

In addition, I showed that the H836Q missense mutation did not affect T- and B-cell development. This was consistent with structural modeling studies that suggested a neutral change. We found that somatic mosaicism can occur when, after pro-nucleus injection, CRISPR/Cas9-mediated cleavage happens multiple times at various cell number stages in the morula. We also noted the importance of introducing a silent mutation in the single strand DNA oligonucleotide (ssODN) to make sure CRISPR/Cas9 would stop cutting after the right sequence was successfully introduced. All of this was used to

ultimately generate mouse models with mutations found in RAG1 patients with a CID-G/AI phenotype.

Several mechanisms have been postulated to play a role in the autoimmune phenotype of RAG1 patients with a CID-G/AI phenotype⁵. Peripheral blood T- and B-cells of CID-G/AI patients and mice have a skewed TCR and BCR repertoire⁶⁻⁸. One hypothesis was that their mutated RAG1 proteins were more likely to select specific coding genes during V(D)J recombination, thereby skewing the repertoire in a way that would cause autoreactivity^{7,9}. Other possibilities were a defect in central or peripheral checkpoints of tolerance, directly or indirectly caused by the mutated RAG protein⁵. However, since most patients develop signs of autoimmunity after being exposed to viral infections or vaccinations, environmental factors seem to play an important role as well. These make it difficult to study the underlying mechanisms in patients.

In order to study these different mechanisms, I selected 3 mutations (F971L, R972Q, and R972W) corresponding to human mutations (F974L, R975Q, and R975W) previously described in patients with CID-G/AI. I generated and characterized these three models. I showed that they indeed recapitulate the phenotype in patients with partial T and B lymphocyte development, persistence of naïve cells, impaired antibody responses and presence of autoantibodies¹⁰⁻¹³. In the next paragraphs, I discuss some of our findings and give a few examples of how these mouse models are being used to explore the underlying mechanisms. The factors thought to contribute to the final phenotype are summarized in **Figure 1**, p. 132.

Abnormal T- and B-cell repertoire

We used our models to study lymphocyte development in the thymus and bone marrow, and in particular the TCR and BCR repertoire before antigenic stimulation, something that is difficult in patients. By using high-throughput sequencing, we identified marked skewing of *IghV* and *TrbV* gene usage in early progenitors, with a bias for productive *Igh* and *Trb* rearrangements after selection and with increased apoptosis of B-cell progenitors. Rearrangement at the Igk locus was impaired, and polyreactive immunoglobulin M antibodies were detected. This study provides novel insights into how hypomorphic Rag1 mutations alter the primary repertoire of T- and B-cells, setting the stage for the immune dysregulation frequently seen in patients.

Previously, these three mutations had been postulated to favor targeting of certain coding elements^{7,9}. We found no differences in the composition of the 2 base pair coding ends when comparing *IghV* and *TrbV* genes that were used differentially in mutant mice versus WT mice. The Manis lab is currently investigating the underlying mechanisms that contribute to the skewed

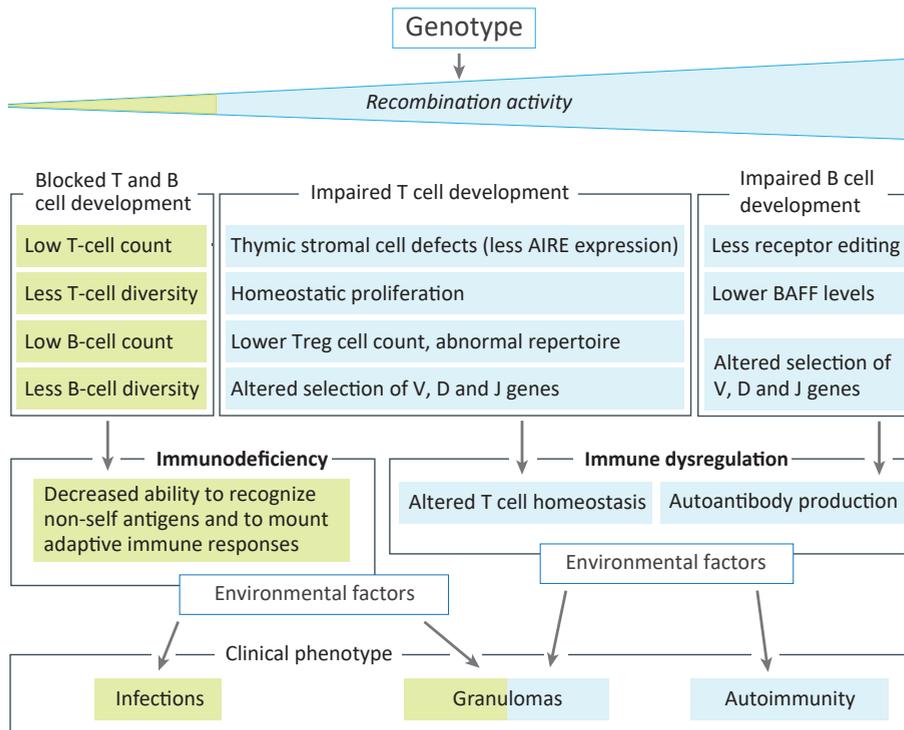


Figure 1: The recombination-activating gene (*RAG*) genotype determines the levels of recombination activity of the mutant *RAG* proteins. mutations with higher residual recombination activity are more likely to result in immune dysregulation. Environmental triggers affect the immune deficiency and immune dysregulation status of the patient, thereby determining the clinical phenotype. AIRE = autoimmune regulator; BAFF= B cell-activating factor; *D* = diversity; *J* = joining; Treg cell = regulatory T-cell; *V* = variable. Adapted from Notarangelo et al⁴.

primary repertoire reported in **Chapter 5**, and in particular, why the mutated *RAG1* seems to affect proximal *IghV* gene usage more than distal *IghV* gene usage (**Chapter 5, Fig. 6**, p. 102). To better understand this skewing, they will focus on the relation between *RAG1* induced cutting and cell cycle, and the role played by locus contraction and accessibility.

Abnormal Treg repertoire

Tregs of CID-G/AI patients show abnormal TCR repertoire⁸. The mouse models generated for this thesis show a comparable abnormal repertoire of Tregs. This abnormal repertoire could contribute to autoimmunity. Tregs of CID-G/AI patients also show less suppressive activity *in vitro*. In my models we did not detect a similar decrease in suppressive activity (**Chapter 5, Fig. S7**, p. 113).

This could mean several things. The Tregs of the RAG1 patients may have been activated T-cells instead of bona fide Tregs. Alternatively, the patient Tregs may have behaved differently because these patients had suffered from infections and had exhausted their Tregs, whereas the mice were kept in a semi-sterile environment, with limited exposure to antigens.

Autoantibody production

I showed that naïve 8 week old R972Q mice and to a lesser extent F971L mice both had IgM autoantibodies (**Chapter 5, Fig. 5C**, p. 96). I found that at 8 weeks the R972Q model already had significant levels of IgG autoantibodies (**Fig. 2**). At 6 months all three models showed IgG autoantibodies to a range of self-antigens (**Fig. 2**, p. 134). We and others¹⁴ showed that B-cell activating factor (BAFF) levels are increased in hypomorphic *Rag1* models (**Chapter 5, Fig. 5E**, p. 96), which is consistent with previous findings in other B-cell lymphopenic hosts¹⁵. Increased BAFF levels can potentially promote survival of immature B-cells producing autoreactive antibodies⁵. Another mechanism possibly contributing to autoimmunity in hypomorphic *Rag* models is reduced receptor editing, as indicated by the indirect measure of expressed lambda light chain (**Chapter 5, Fig. 5D**, p. 96). Receptor editing is important to rearrange autoreactive BCRs of immature B-cells, prior to exiting the bone marrow. If this is unsuccessful, the immature B-cell undergoes apoptosis. More in depth sequencing is needed to evaluate receptor editing more directly. In general, more study is needed to understand the autoantibody production in these models. IgM is generated in the bone marrow independent of antigen stimulation (**Chapter 1, Fig. 1**, p. 19), but several antigen driven processes such as somatic hypermutation and class switching are involved in the production of IgG.

Thymic stromal cell defects

As shown in **chapter 5**, the number of total thymocytes was drastically reduced in all three models. In the R972W model, the most severe model, very few double positive (DP) cells developed in the thymus. This model almost behaved like *Rag1*^{-/-} mice. In contrast, the R972Q and F971L models were able to develop double positive (DP), single positive CD4⁺ and CD8⁺ cells. Currently, Marita Bosticardo of the Notarangelo lab is performing an extensive evaluation of the thymic phenotype of the three models. The cross talk between T-cells and thymic epithelial cells (TEC) in the thymus is fundamental for the development and maturation of both types of cells. In the absence of mature T-cells, TECs cannot complete their maturation, and the medullary (mTEC) subset is virtually absent. These results were observed in the R972W mouse model and *Rag1*^{-/-} mice. In R972Q and F971L mice TECs were present, but markedly reduced and consisted mostly of cortical TECs (cTECs). Importantly,

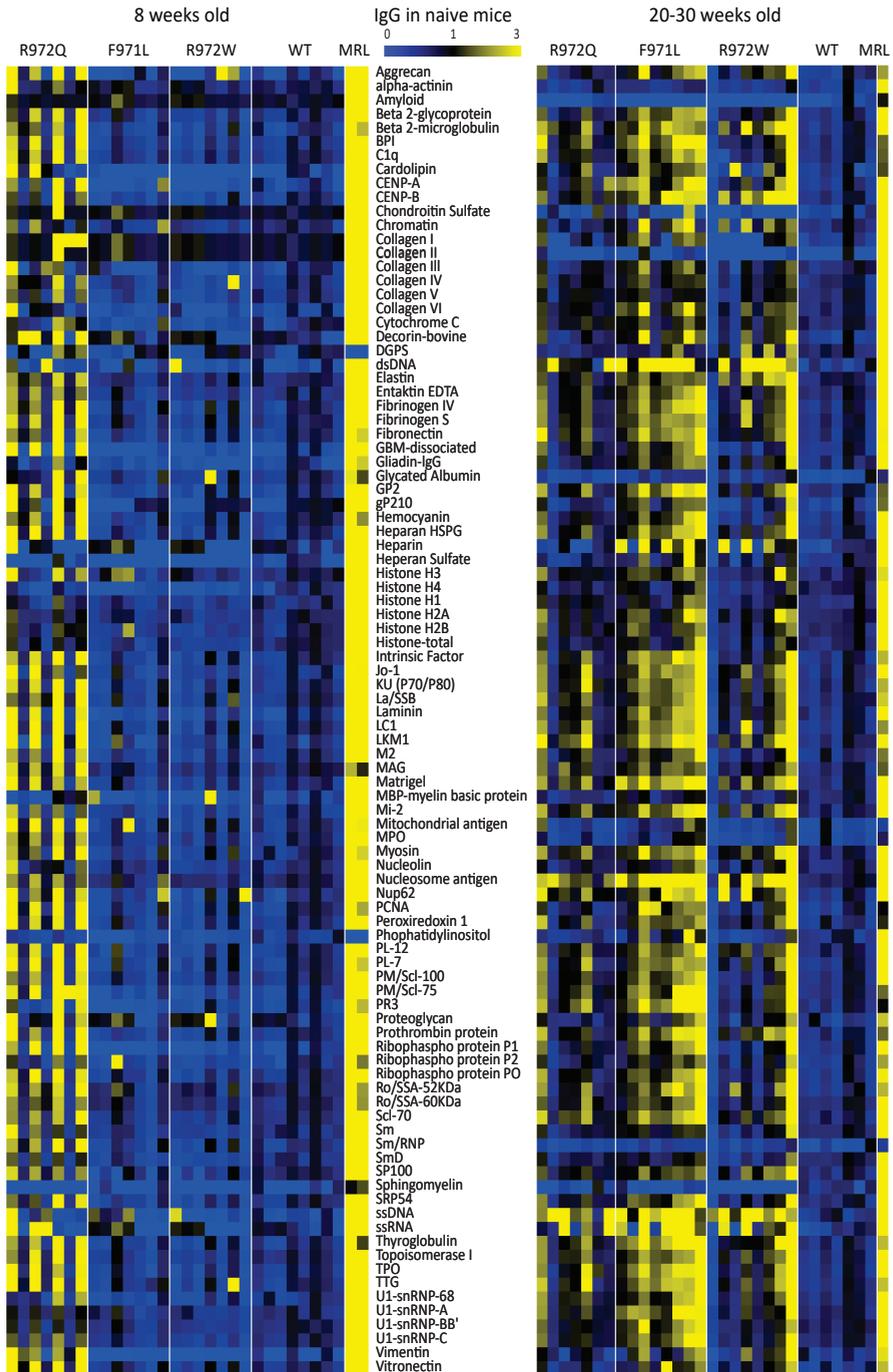


Figure 2: IgG autoantibodies in serum of naive mice at 8 and at 20-30 weeks, normalized to WT mean plus two standard deviations. See supplemental materials of **chapter 5** for the methods used.

mTECs are crucial for autoimmune regulator (AIRE) expression. AIRE enables expression of specific self-antigens (tissue-restricted antigens – TRAs). This is needed to test newly generated T-cells for autoreactivity. The functional consequences of impaired cross-talk between thymocytes and TECs in Rag1 mutant mice are currently being studied further by the Notarangelo lab.

Novel treatment options

The three mouse models I generated offer new tools to study treatment options for patients with *RAG1* mutations showing a CID-G/AI phenotype. Hematopoietic stem cell transplantation (HSCT) has been used for the treatment of hematologic malignancies and primary immunodeficiencies (PID) for several decades now, and with increasing efficacy. However, in some cases conditioning may be needed to make space in the bone marrow prior to a transplantation, to ensure engraftment of donor cells. This is particularly the case for patients with Omenn, Leaky SCID or CID-G/AI, where there is residual production of dysfunctional T and/or B-cells. In these treatments, toxicity related to conditioning regimens based on the use of chemotherapy and/or irradiation remains a significant problem and novel conditioning regimens or other treatment strategies are needed.

Recently, an alternative, potentially low-toxicity, conditioning approach has been proposed¹⁶, which makes use of immunotoxin targeting CD45-expressing cells. These include HSCs and mature leukocytes. Using one of the mouse models I generated, the F971L model, colleagues Marita Bosticardo and Enrica Calzoni in the Notarangelo lab have tested the efficacy of conditioning with an anti-CD45 immunotoxin (anti-CD45 Saporin or α CD45-SAP) alone or in combination with low irradiation (200rads; α CD45-SAP/200), and compared these regimens to a myeloablative dose of irradiation (800rads). Following conditioning, the F971L mice were transplanted with wild-type (WT) bone marrow (BM) lineage-negative cells and followed over time to evaluate immune reconstitution. Conditioning with α CD45-SAP alone or with α CD45-SAP/200 led to a consistent engraftment of donor cells. Successful donor engraftment was seen in bone marrow, thymus and peripheral T, B, NK and myeloid cells.

Finally, the three generated mouse models of CID-G/AI can be used to optimize gene therapy strategies, both in the form of genome editing using endonucleases and using viral gene therapy, as discussed in **chapter 2**. In 2016 the European Commission granted market approval to GlaxoSmithKline (GSK) for *ex vivo* hematopoietic stem cell gene therapy for the treatment of adenosine deaminase (ADA) SCID¹⁷. This is the second *ex vivo* stem cell gene therapy to receive regulatory approval in Europe. The first one was Glybera (alipogene tiparvovec) which delivers the lipoprotein lipase gene in an adeno-

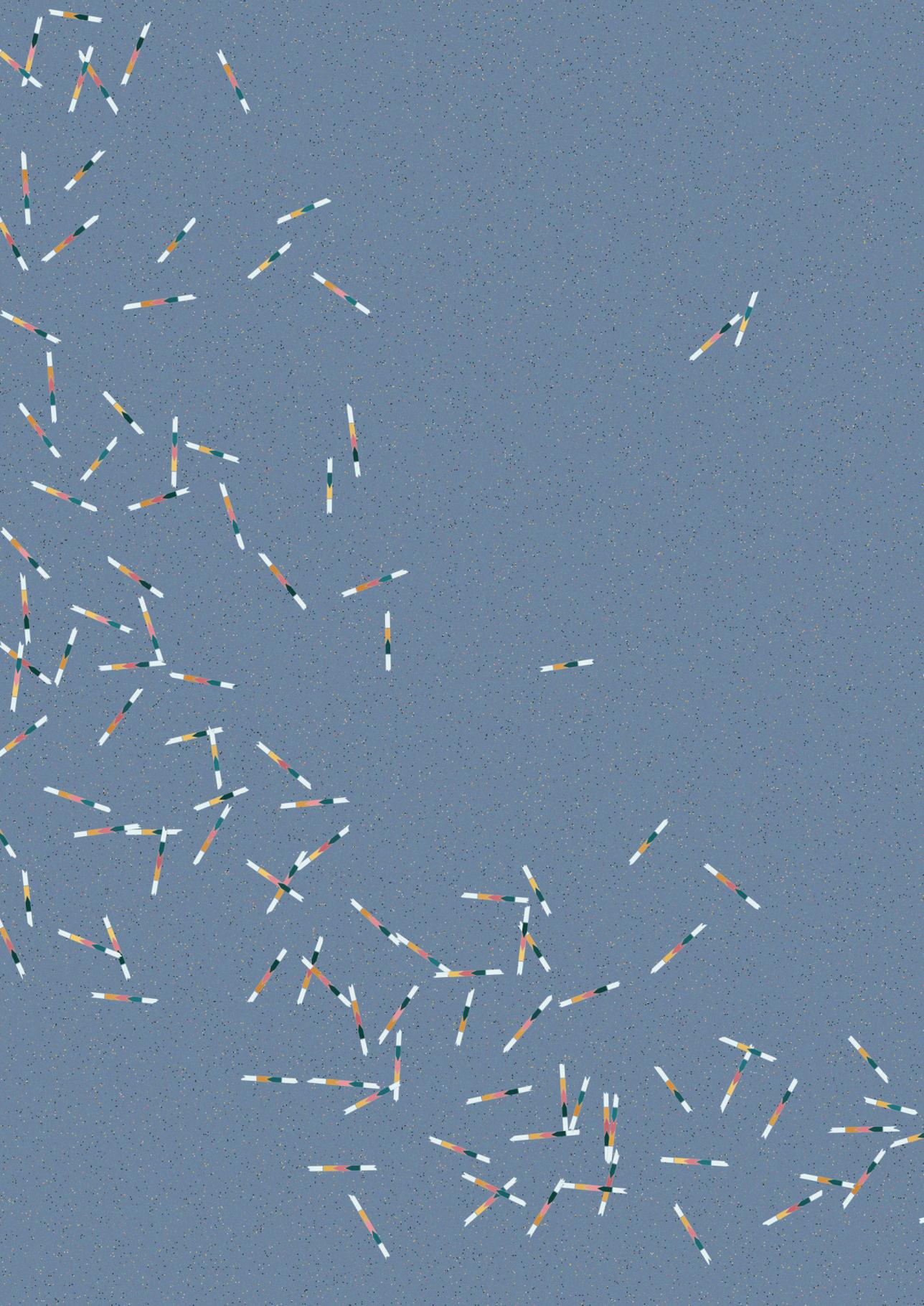
associated virus vector, but which has already been taken off the market again due to high costs¹⁸.

Concluding remarks

We have generated and used *in vitro* and *in vivo* models to study RAG1 deficiency. As *in vitro* T-cell differentiation assays improve, iPSC models can more easily be used to study primary immunodeficiencies and perform pre-clinical trials of gene correction strategies. However, mouse models will still be needed to really understand the different mechanisms leading to the distinct phenotypes of primary immunodeficiencies and distinguish between genetic and environmental factors.

References

1. Lee YN, Frugoni F, Dobbs K, et al. A systematic analysis of recombination activity and genotype-phenotype correlation in human recombination-activating gene 1 deficiency. *J Allergy Clin Immunol*. 2013.
2. Tirosh I, Yamazaki Y, Frugoni F, et al. Recombination activity of human recombination-activating gene 2 (2) mutations and correlation with clinical phenotype. *J Allergy Clin Immunol*. 2018.
3. Kim MS, Lapkouski M, Yang W, Gellert M. Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature*. 2015;518(7540):507-511.
4. Chang CW, Lai YS, Westin E, et al. Modeling Human Severe Combined Immunodeficiency and Correction by CRISPR/Cas9-Enhanced Gene Targeting. *Cell Rep*. 2015;12(10):1668-1677.
5. Notarangelo LD, Kim MS, Walter JE, Lee YN. Human mutations: biochemistry and clinical implications. *Nat Rev Immunol*. 2016;16(4):234-246.
6. Lee YN, Frugoni F, Dobbs K, et al. Characterization of T- and B-cell repertoire diversity in patients with RAG deficiency. *Sci Immunol*. 2016;1(6).
7. Khiong K, Murakami M, Kitabayashi C, et al. Homeostatically proliferating CD4 T-cells are involved in the pathogenesis of an Omenn syndrome murine model. *J Clin Invest*. 2007;117(5):1270-1281.
8. Rowe JH, Stadinski BD, Henderson LA, et al. Abnormalities of T-cell receptor repertoire in CD4+ regulatory and conventional T-cells in patients with mutations: Implications for autoimmunity. *J Allergy Clin Immunol*. 2017.
9. Wong SY, Lu CP, Roth DB. A *RAG1* mutation found in Omenn syndrome causes coding flank hypersensitivity: a novel mechanism for antigen receptor repertoire restriction. *J Immunol*. 2008;181(6):4124-4130.
10. Schuetz C, Huck K, Gudowius S, et al. An immunodeficiency disease with mutations and granulomas. *N Engl J Med*. 2008;358(19):2030-2038.
11. Henderson LA, Frugoni F, Hopkins G, et al. Expanding the spectrum of recombination-activating gene 1 deficiency: a family with early-onset autoimmunity. *J Allergy Clin Immunol*. 2013;132(4):969-971 e961-962.
12. Villa A, Sobacchi C, Notarangelo LD, et al. V(D)J recombination defects in lymphocytes due to mutations: severe immunodeficiency with a spectrum of clinical presentations. *Blood*. 2001;97(1):81-88.
13. Avila EM, Uzel G, Hsu A, et al. Highly variable clinical phenotypes of hypomorphic *RAG1* mutations. *Pediatrics*. 2010;126(5):e1248-1252.
14. Walter JE, Rucci F, Patrizi L, et al. Expansion of immunoglobulin-secreting cells and defects in B-cell tolerance in Rag-dependent immunodeficiency. *J Exp Med*. 2010;207(7):1541-1554.
15. Lavie F, Miceli-Richard C, Ittah M, Sellam J, Gottenberg JE, Mariette X. Increase of B-cell-activating factor of the TNF family (BAFF) after rituximab treatment: insights into a new regulating system of BAFF production. *Ann Rheum Dis*. 2007;66(5):700-703.
16. Palchaudhuri R, Saez B, Hoggatt J, et al. Non-genotoxic conditioning for hematopoietic stem cell transplantation using a hematopoietic-cell-specific internalizing immunotoxin. *Nat Biotechnol*. 2016;34(7):738-745.
17. Cicalese MP, Ferrua F, Castagnaro L, et al. Update on the safety and efficacy of retroviral gene therapy for immunodeficiency due to adenosine deaminase deficiency. *Blood*. 2016;128(1):45-54.
18. Senior M. After Glybera's withdrawal, what's next for gene therapy? *Nat Biotechnol*. 2017;35(6):491-492.



7 Summary

Nederlandse samenvatting

Acknowledgments

Curriculum vitae

List of publications

Introduction

In order to recognize and eliminate the many different pathogens that the immune system can encounter, it generates an enormous repertoire of different unique T-cell receptors and B-cell immunoglobulins. The main contributor to this diversity is the process of V(D)J recombination, in which first different *V*, (*D*) and *J* genes are combined, then more diversity is created at the junction sites between the *V*, (*D*) and *J* genes. Recombination-activating gene 1 (*RAG1*) and 2 (*RAG2*) are crucial both for the recognition and cleavage of the RSS flanking the *V*, *D* or *J* genes and to form the hairpins that allow the broken DNA ends to be ligated by the non-homologous end joining pathway (NHEJ).

Mutations in *RAG1* and *RAG2* are associated with a broad spectrum of distinct clinical and immunological phenotypes, including T- B- severe combined immune deficiency (SCID), Omenn syndrome (OS), atypical/leaky SCID (LS), $\gamma\delta$ SCID and combined immunodeficiency with granuloma and/or autoimmunity (CID-G/AI). Previous to this work, *in vitro* assays and structural modeling of the RAG complex have given some insights into why some mutations cause a more severe phenotype than others. In particular, mutations that allow for more residual RAG activity result in phenotypes such as leaky SCID or CID-G/AI. However, phenotypic variability remains even between patients with mutations that affect the same region of *RAG1*, showing the same residual

RAG1 activity using *in vitro* models. How these mutations cause these different phenotypes remains an open question.

The main aim of this thesis was to generate novel *in vitro* and *in vivo* models of RAG1 deficiency to study the mechanisms underlying the different RAG phenotypes. These models will also be useful to test novel treatment options for patients with mutations in *RAG1*. A promising novel treatment is gene correction, so improving the genome-editing strategy used to generate models of RAG1 deficiency should also contribute to the development of a gene therapy that corrects patient stem cells .

Genome editing using CRISPR/Cas9

In **chapter 2** we review the most important genome-editing strategies for primary immunodeficiencies. Currently, the mainstay of treatment for SCID or other severe primary immunodeficiencies is allogeneic hematopoietic stem cell transplantation, and although outcomes have improved over the years, this remains a high-risk procedure when HLA-matched donors are lacking. For this reason, correcting the patient's own hematopoietic stem cells with gene therapy would offer an attractive alternative.

Gene therapies currently being used in clinical settings insert a functional copy of the entire gene by means of a viral vector. Even though safer vectors continue to be developed, this strategy risks integration within oncogenes. In the case of RAG deficiency, gene addition is even more challenging, because RAG expression is so tightly regulated, and continuous expression can be harmful. A promising alternative is the use of endonucleases such as ZFNs, TALENs and CRISPR/Cas9 to introduce a double-stranded break in the DNA and thus induce homology-directed repair. With these genome-editing tools a correct copy can be inserted in a precisely targeted "safe harbor." They can also be used to correct pathogenic mutations *in situ* and to develop cellular or animal models needed to study the pathogenic effects of specific genetic defects found in immunodeficient patients.

The chapter focuses on CRISPR/Cas9 because of its efficacy and versatility.

Models created from iPSCs

Chapter 3 describes how we generated and characterized induced pluripotent stem cells (iPSCs) of one Omenn patient and two SCID patients. The Zúñiga-Pflücker lab used their *in vitro* assay to differentiate these iPSCs and the iPSCs of healthy donors into T-cells. The SCID and Omenn iPSCs showed a similar block in T-cell development. Subsequently, we compared the TCR repertoire of the precursor T-cells and found that cells derived from SCID and Omenn patients showed a restricted repertoire with reduced number of

rearrangements and skewed usage of *V* and *J* genes. In addition, T-cell receptor excision circle (TREC) levels, a measure of recombination activity also used for newborn screening, were reduced.

The *in vitro* differentiation of iPSCs from RAG1 patients into T-cells described in **chapter 3** will be useful to test gene-correction with CRISPR/Cas9. I was able to target and introduce a double stranded DNA break in the *RAG1* locus in iPSCs, but did not yet succeed in true gene correction due to the challenges of first getting enough CRISPR/Cas9 and repair DNA template (single stranded oligonucleotide) into the cells, and subsequently getting enough homologous recombination to correct the mutation. As yet no lab has succeeded in CRISPR/Cas9 mediated gene correction for RAG1 deficiency, though such corrections have been reported for some of the other SCID causing mutations.

Creating a *Rag1* mouse model using CRISPR/Cas9

Although they are useful to test true gene correction and the intrinsic effects of a mutation on the T-cell differentiation potential, *in vitro* models provide an artificial setting and cannot simulate the complexity of the full immune system with all its different components. In addition, the effect of environmental factors cannot be tested. To control for environmental factors in a realistic setting we need *in vivo* models.

For this reason, I generated several *Rag1* mouse models. For the study reported in **chapter 4** I showed that CRISPR/Cas9 can be used to generate *Rag1* mouse models in a single step, by directly injecting the Cas9 and *Rag1* specific gRNA into the zygote. Residue 838 was targeted with high efficiency, as none of the mice showed a wild type sequence. This way, we were able to evaluate the effects of in-frame deletions in residues 832-877 of *Rag1*, a region of unclear function that does not contain any catalytic residues nor is involved in zinc binding. It was found that even 1 amino acid deletions in this region resulted in a complete knock-out phenotype with a complete lymphocyte developmental block in thymus at DN3 and bone marrow at the pro-B-cell stage. In addition, a H836Q missense mutation had no effect at all on RAG function. This provided insights into the RAG1 structure and function. In addition, this study offered proof of principle that a similar strategy can be applied to develop novel knock-in mouse models with specific patient mutations.

Mouse models of *Rag1* mutations found in patients with CID-G/AI

The underlying mechanisms causing the autoimmunity in patients with CID-G/AI are unclear. Previously, it was found that peripheral blood T- and B-cells of RAG patients with CID-G/AI have an abnormal T-cell receptor and B-cell immunoglobulin repertoire. However, it was unknown if these changes

were already present in the primary repertoire in thymus and bone marrow, or may simply be due to antigenic stimulation in the periphery. Therefore, for **chapter 5** I used CRISPR/Cas9 to generate three mouse models with missense mutations equivalent to those found in patients with CID-G/AI.

Immunological characterization showed partial development of T and B lymphocytes, preserved serum immunoglobulin, partial immunity and presence of autoantibodies, thereby recapitulating the phenotype seen in patients with CID-G/AI. By using high-throughput sequencing, we identified marked skewing of *IghV* and *TrbV* gene usage in early progenitors, with a bias for productive *Igh* and *Trb* rearrangements after selection occurred and increased apoptosis of B-cell progenitors. Rearrangement of *light chain* was impaired, and polyreactive immunoglobulin M antibodies were detected. This study provides novel insights into how hypomorphic Rag1 mutations alter the primary repertoire of T- and B-cells, setting the stage for the immune dysregulation frequently seen in patients.

Conclusion

This thesis provides novel insights into the phenotypic spectrum of RAG1 deficiency, with a focus on the CID-G/AI phenotype. The mouse models I generated using CRISPR/Cas9 mediated genome-editing can be used for further investigations of autoimmunity in RAG1, as evidenced by several ongoing studies described in chapter 6. Mouse models such as these can also be used to develop novel treatment strategies and to optimize existing ones. For instance, hematopoietic stem cell transplantations for RAG patients with CID-G/AI are complicated by the requirement for conditioning regimens to get rid of the patient's own (dysfunctional) T- and B-cells. Toxicity related to conditioning regimens based on the use of chemotherapy and/or irradiation remains a significant problem and novel conditioning regimens or other treatment strategies are needed. Possible solutions can be safely tried on mouse models. The *in vitro* T-cell differentiation from patient iPSCs discussed in **chapter 3** can be used to further develop genome-editing correction strategies and offers a pre-clinical platform to prove true gene correction in human stem cells.

Nederlandse samenvatting

Introductie

Ons immuunsysteem moet pathogenen kunnen herkennen en doden. Dit is een uitdaging, omdat de populatie van pathogenen enorm varieert en er voortdurend nieuwe bijkomen. Om zich snel aan te kunnen passen aan deze veranderlijke bedreigingen, moet het immuunsysteem een zo groot mogelijk repertoire aan verschillende unieke T-cel receptoren en B-cel immunoglobulines kunnen genereren. De moleculaire structuur van die receptoren bepaalt op welke pathogenen ze “passen” en het is dus zaak zoveel mogelijk verschillende structuren te laten ontstaan en zodra er een blijkt te passen, deze snel te vermenigvuldigen (klonale expansie).

De structuur van een T- of B-cel wordt in hoge mate bepaald door de combinatie van een specifiek V, J en D gen. Op mensen- (en muizen-) DNA zijn tientallen varianten van deze genen te vinden. Deze genen komen tot expressie zodra er een enkel gen voor ieder van die typen over is en de rest verwijderd is. Er moeten dus genen worden weggeknipt uit het DNA, waarna de overblijvende stukken weer aan elkaar moeten worden geplakt: “cut and paste”. De resulterende diversiteit in DNA structuur wordt overigens niet uitsluitend bepaald door de keuze van V, D en J gen, tijdens het weer aan elkaar plakken van de delen kunnen ook kleine veranderingen ontstaan op de overgang tussen de gecombineerde genen: “junctional diversity”.

Recombination-activating Gene 1 (RAG1) en 2 (RAG2) zijn eiwitten die de verbindingen (RSS) tussen de V, D en J genen herkennen en knippen, en daarna de uiteinden zodanig vormgeven (als “hairpins”) dat deze middels het non-homologous end joining (NHEJ) systeem weer aan elkaar geplakt kunnen worden.

Mutaties in de genen voor deze eiwitten (de *RAG1* en *RAG2* genen) worden geassocieerd met een breed spectrum aan verschillende uitingen van aangeboren (“primaire”) immuundeficiënties, zoals T- B- Severe combined immunodeficiency (SCID), het Omenn syndroom (OS), atypische (“leaky”) SCID (LS), $\gamma\delta$ SCID en gecombineerde immuundeficiëntie met granulomen en/of auto-immuniteit (CID-G/AI). Eerdere *in vitro* testen en structurele modellen van het RAG complex hebben eerste inzichten gegeven in waarom sommige mutaties een ernstiger beeld veroorzaken dan anderen. Zo werd aangetoond dat mutaties die minder desastreus voor de werking van RAG zijn en een hogere resterende RAG activiteit vertonen, tot meer T- en B-cel ontwikkeling leiden en geassocieerd zijn met fenotypes als *leaky* SCID of CID-G/AI. Echter, het blijkt ook dat mutaties in eenzelfde *RAG1* domein die met *in vitro* testen

ook een zelfde activiteit vertoonden, bij patiënten toch voor uiteenlopende uitingen van ziekte kunnen zorgen. Hoe deze mutaties tot verschillende klinische beelden kunnen leiden bleef onduidelijk.

Het doel van dit promotie onderzoek was om verschillende *in vitro* en *in vivo* modellen te creëren om daarmee de mechanismen te kunnen onderzoeken die ten grondslag liggen aan het brede spectrum van klinische uitingen van RAG deficiëntie, met een focus op het CID-G/AI ziektebeeld. Daarnaast kunnen deze modellen gebruikt worden om nieuwe behandelingen te testen en optimaliseren. Het optimaliseren van de genome-editing technieken die gebruikt werden om deze modellen te maken, kan ook bijdragen aan het ontwikkelen van een nieuwe vorm van gentherapie waarbij de stamcellen van de patiënt direct gecorrigeerd kunnen worden.

Genome editing met CRISPR/Cas9

In **hoofdstuk 2** bespreken we de belangrijkste genome-editing strategieën voor primaire immuundeficiënties. Op dit moment worden SCID en andere ernstige primaire immuundeficiënties vooral behandeld met allogene hematopoietische stamcel transplantatie. De resultaten hiervan zijn erg verbeterd de afgelopen jaren, maar dit blijft een riskante procedure wanneer er geen donor met een HLA match gevonden kan worden. Daarom is het aantrekkelijk om de stamcellen van de patiënt zelf te kunnen corrigeren d.m.v. gentherapie.

Bij de gentherapie die momenteel gebruikt wordt in de kliniek wordt met behulp van een virale vector een functionele kopie van het gehele gen geïntroduceerd ergens in het genoom. Ook al worden er steeds veiligere vectoren ontwikkeld, men kan nooit volledig controleren waar de vector zich in het genoom invoegt. Als dit binnen of in de buurt van een oncogen gebeurt, bestaat er een reëel risico op maligniteit. In het geval van primaire immuundeficiënties die zijn veroorzaakt door mutaties in *RAG1* blijkt het introduceren van een extra *RAG1* kopie door een virale vector nog risicovoller, waarschijnlijk omdat *RAG1* expressie tijdens T- en B-cel ontwikkeling zo strak gereguleerd is en continue expressie schadelijk kan zijn.

Een veelbelovende nieuwe manier van gentherapie is het gebruik van endonucleases zoals ZFN's, TALENs en CRISPR/Cas9. Hiermee wordt eerst een dubbelstrengs DNA breuk geïntroduceerd, precies op de gewenste locatie, waarna door homologe recombinatie het DNA veranderd kan worden. Op deze manier kan men precies bepalen waar het gen geïntroduceerd wordt, en zorgen dat dit gebeurt op veilige plekken ("safe harbors") waar er geen risico is op het ontstaan van maligniteiten. Nog belangrijker is dat met deze technieken zelfs specifieke mutaties gecorrigeerd worden in cellen van patiënten.

Tenslotte kunnen zo ook dier en cel modellen met specifieke mutaties gemaakt worden, zodat deze in detail bestudeerd kunnen worden. De meeste aandacht wordt in hoofdstuk 2 gegeven aan CRISPR/Cas9, omdat deze variant zo efficiënt en veelzijdig is.

iPSC modellen

Hoofdstuk 3 beschrijft hoe we geïnduceerde pluripotente stamcellen (induced pluripotent stem cells of iPSCs) hebben gemaakt van cellen van een Omenn patient en twee SCID patiënten. Het Zúñiga- Pflücker lab heeft deze iPSCs en de iPSCs van gezonde mensen laten differentiëren tot T cellen in hun *in vitro* test opstelling. De SCID en Omenn iPSCs toonden een vergelijkbare blokkade in de T-cel ontwikkeling. Vervolgens hebben we het T-cel receptor repertoire (TCR) vergeleken met dat van gezonde mensen. De cellen van SCID en Omenn patiënten bleken een beperkt repertoire te hebben met minder succesvolle herschikkingen (rearrangements) en een afwijkend gebruik van V en J genen. Ook werden er minder T-cell receptor excision circles (TRECs) gevonden. TRECs zijn een bijproduct van recombinatie (de “weggesneden” stukken DNA) en geven dus een goede indicatie van de hoeveelheid recombinatie. TREC levels worden daarom steeds vaker ingezet als middel om immuundeficiënties tijdig bij pasgeborenen op te sporen.

De *in vitro* differentiatie van iPSC's van RAG1 patiënten tot T-cellen zoals beschreven in hoofdstuk 3 kan van nut zijn bij het testen van gen-correcties met CRISPR/Cas9. Ik ben erin geslaagd om in iPSC's een dubbelstrengs DNA breuk te introduceren op de plek van het *RAG1* gen, maar gen-correctie bleef nog buiten bereik door complicaties bij het introduceren in de cellen van voldoende CRISPR/Cas9 en repair DNA template (single stranded oligonucleotide), en het opwekken van voldoende homologe recombinatie om de mutatie ook daadwerkelijk te kunnen corrigeren. Ook op andere labs is gen correctie van RAG1 deficiënties met CRISPR/Cas9 tot nu nog niet gelukt, Er zijn wel geslaagde mutatie correcties gerapporteerd voor sommige andere genen die tot SCID leiden en de verwachting is dat dit spoedig ook voor RAG mogelijk zal zijn.

Een *Rag1* muis model gemaakt met CRISPR/Cas9

In vitro modellen zijn weliswaar heel bruikbaar bij het testen van gen correcties en het effect van een mutatie op de (variatie in) T-cel differentiaties, maar bieden door hun kunstmatige aard geen goede basis voor een realistische toetsing die recht doet aan de complexiteit van het immuunsysteem, waarin heel veel factoren samenkomen. Het effect van de vele relevante omgevingsfactoren kan alleen meegewogen worden in *in vivo* modellen zoals *Rag1* muismodellen.

Hoofdstuk 4 beschrijft de ontwikkeling van dergelijke muismodellen met CRISPR/Cas9. Ik kon deze modellen in een enkele stap genereren door zowel de Cas9 als de *Rag1*-specifieke gRNA rechtstreeks in de zygote (de bevruchte eicel) te injecteren. Dit was heel erg efficiënt en specifiek: in alle muizen was op de juiste plek (aminozuur 838) een dubbelstrengs DNA breuk geïntroduceerd. Zo konden we de effecten evalueren van in-frame deleties in residues 832-877 van *Rag1*, een deel van RAG1 met een onbekende functie. Het was onduidelijk of dit deel belangrijk is aangezien het geen aminozuren bevat die van belang zijn voor de katalitische activiteit van RAG1. Echter, zelfs een enkele aminozuur deletie in dit gebied resulteerde al in een volledig “knock-out” fenotype, zonder enige ontwikkeling van lymfocyten in de thymus voorbij het DN3 stadium of in het beenmerg voorbij het pro-B stadium. Een H836Q missense mutatie had geen enkel effect op de RAG functie. Naast nieuw inzicht in de structuur en functie van RAG, toonde dit onderzoek aan dat het in principe mogelijk was om nieuwe “knock-in” muismodellen te ontwikkelen met specifieke patiënt mutaties.

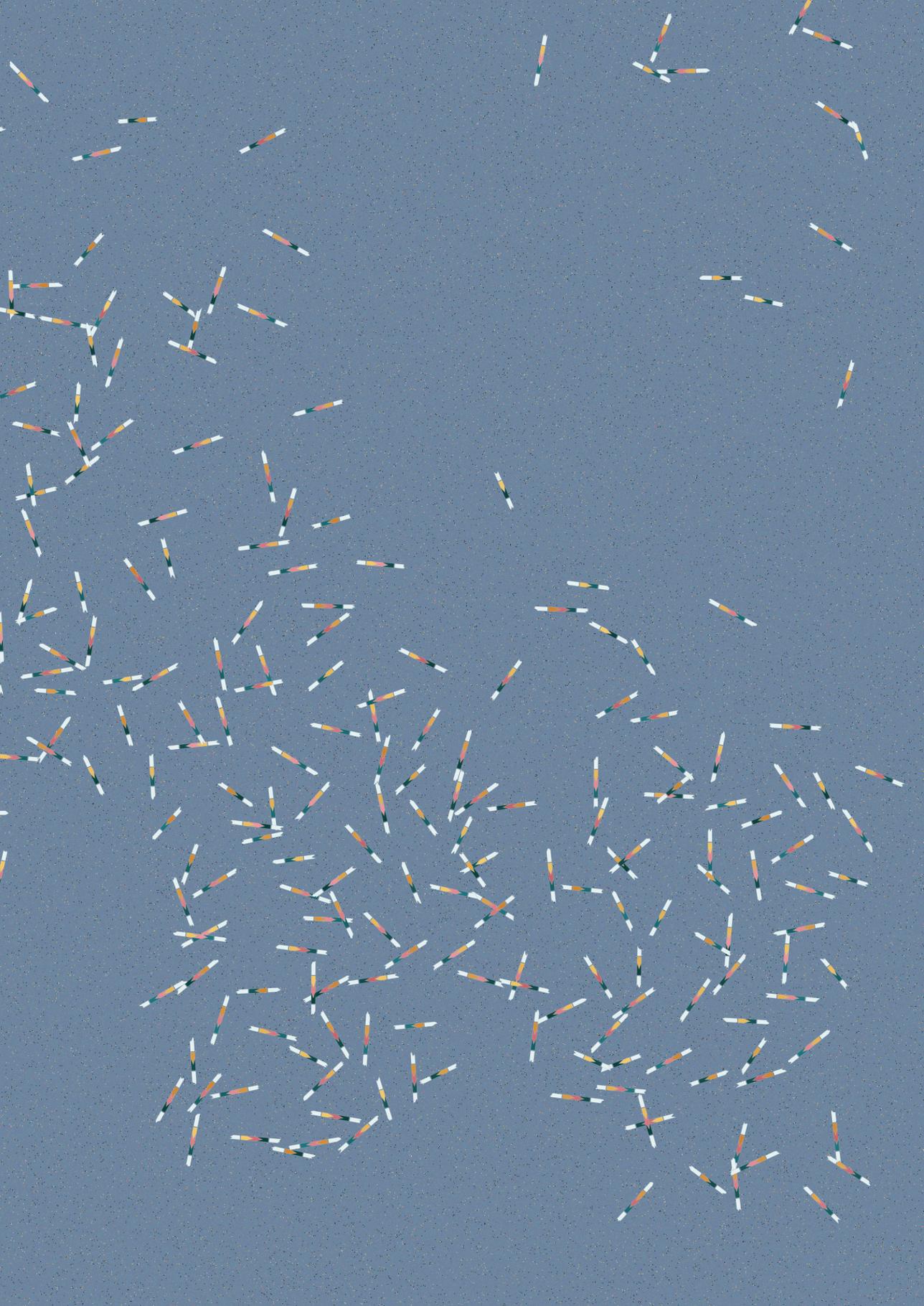
Rag1 muismodellen voor drie mutaties bij CID-G/AI patiënten

Het is niet duidelijk hoe de auto-immuniteit te verklaren is die optreedt bij patiënten met CID-G/AI. In het bloed van CID-G/AI patiënten is een abnormaal T-cel receptor en B-cel immunoglobuline repertoire aangetroffen, maar niet bekend was of deze afwijkingen al aanwezig zijn in het primaire repertoire in thymus en beenmerg, of slechts een gevolg zijn van antigene stimulatie in de periferie. Om dit te onderzoeken heb ik, als beschreven in **hoofdstuk 5**, CRISPR/Cas9 gebruikt om drie muismodellen te genereren met missense mutaties die zijn aangetroffen bij CID-G/AI patiënten.

Immuunfenotypering van deze modellen toonde een gedeeltelijke ontwikkeling van T en B lymfocyten, een normaal serum immunoglobuline, een gedeeltelijke immuniteit en de aanwezigheid van autoantistoffen. Dit is een replicatie van het fenotype van CID-G/AI patiënten. Met behulp van *high-throughput sequencing* konden we meerdere verschillen in het T-cel receptor en B-cel immunoglobuline repertoire van de *Rag1* mutanten ten opzichte van wild-type muizen aantonen: een duidelijk afwijkend gebruik van *IghV* en *TrbV* genen, een tendens richting meer productieve *Igh* en *Trb* herschikkingen na de selectie, een hogere apoptose van de B-cel voorlopers, minder herschikkingen van de *light chain*, en de aanwezigheid van polyreactieve immunoglobuline M antistoffen. Dit onderzoek biedt daarom nieuwe inzichten in de manier waarop hypomorfe *Rag1* mutaties tot een ander primair repertoire van T- en B-cellen leiden. Mogelijk draagt dit afwijkende primaire repertoire bij aan de ontregeling van het immuun systeem zoals gevonden bij patiënten met CID-G/AI.

Conclusie

Dit proefschrift biedt nieuwe inzichten in het fenotypische spectrum van RAG1 deficiëntie, met een nadruk op het CID-G/AI fenotype. De muismodellen die ik met CRISPR/Cas9 heb ontwikkeld kunnen worden ingezet voor verder onderzoek van auto-immuniteit bij RAG1 deficiëntie. Dit gebeurt al in diverse lopende onderzoeken die ik beschrijf in **hoofdstuk 6**. De muismodellen kunnen ook gebruikt worden bij de ontwikkeling van nieuwe behandelmethoden en optimalisering van huidige behandelingen voor RAG patiënten met CID-G/AI. Omdat er bij die beelden nog steeds behoorlijk wat eigen T- en B-cellen kunnen voorkomen, moeten deze bij hematopoietische stamcel transplantaties eerst uitgeschakeld worden door pre-transplantatie conditionering, wat tot extra complicaties kan leiden. Mogelijke oplossingen hiervoor kunnen veilig getest worden op muismodellen. De *in vitro* T-cel differentiatie van de iPSC's van patiënten zoals besproken in **hoofdstuk 3** kan worden toegepast bij de verdere ontwikkeling van genterapieën en biedt een preklinisch platform voor het testen van gen correcties in humane stamcellen.



Acknowledgments

Prof. Wulffraat, dear Nico, you were the first to show me what it was like to collaborate closely with international colleagues, learn from their expertise and enjoy international conferences to the fullest (including lots of dancing). The EULAR task force meetings in Zurich and the PReS conference in Valencia were a great experience especially as a medical student. This is how it all started and how I ended up in Boston.

Prof. Notarangelo, dear Gigi, I still remember how I didn't understand a word the first time we were meeting. Lab research was completely new to me, but still you gave me the opportunity to work in your lab. The Notarangelo lab was very warm and welcoming to me from the start. Your passion for research is contagious and I am proud to have worked for someone so brilliant. I really hope we will continue to work together.

Dr. Manis, dear John, without you I would have never been able to do any of my mouse experiments. Quite literally actually as the hypomorphic *Rag* pups generated for chapter 5 would have died because of PCP if you hadn't suggested to give them bactrimel. I started doing mouse experiments two years into my PhD. At that point, there was no one else in the Notarangelo lab doing mouse experiments. You helped me organize the colony, plan experiments and you taught me a lot about mouse immunology.

Dr. Boes, dear Marianne, thank you for bridging the Boston-Utrecht gap as you knew both worlds and what was required for a Dutch PhD. Your insight in the Harvard medical School community was super helpful to me. You really understood the lab struggles.

Colleagues from the Notarangelo lab, thank you for all your help, motivational talks, friendships, fun conferences, nights at the Squealing Pig, and for making the hard work more fun. The Notarangelo lab uses fundamental research to answer clinical questions. It was great to work with people that are both clinician and researcher. Stefano, your jokes and optimism helped me through the dark times. Thank you for becoming such a good friend. Lauren, your focus has been a true inspiration to me. Even though Stefano was determined to cause girl drama between us, you became a mentor and a friend. Kerstin, you are so sweet and from the day I first met you I was impressed by how much you know and how hard you work. Francesca, you surprised me with a present the first Christmas away from home and introduced me to the amazing 34 William street apartment where I would stay for my entire PhD. Katja, I enjoyed our talks sitting next to each other in the lab. You have an extremely busy life

combining pediatric hematology/oncology with lab work, but still found time to help me. Jolan, you are so inclusive and energetic, I really don't know how you do it all. Kerry, thank you for always keeping the lab running. Thanks also to the Pai lab for our discussions during joint lab meetings and to the colleagues from other labs for their help: Louis with injections, Treg staining and suppression assay; Erin with troubleshooting failed experiments; Ron, by always making time to sort my cells no matter how crazy your schedule was.

Kim Ching and Alessandro Barbieri from the Manis lab, you were a great help with the mouse experiments and are continuing the work since I left. Marita Bosticardo from the current Notarangelo lab at the NIH, thank you for helping with some key experiments for chapter 5 and for doing such great follow up work together with Enrica Calzoni. Thank you to all co-authors and other collaborators, in particular Patrick Brauer and Dr. Zúñiga-Pflücker for chapter 3 of this thesis.

Thank you, Dr. Derrick Rossi and Dr. Kiran Musunuru, for being on my Harvard thesis committee. Kiran, you taught me how to design and generate gRNA, and made it possible for me to work with CRISPR/Cas9 so early on. Thank you for always being available for questions and for letting me write the genome-editing review (chapter 2) with you. Dear Prof. Geha, thank you for taking the time to evaluate my work during our T32 progress meetings and to discuss my future. Prof. Terhorst, it was great to get advice from a Dutch professor based at Harvard. Dr. Thomas Fleisher, Dr. Sergio Rosenzweig, Dr. Steven Holland and all the other faculty and fellows of the 2016 CIS summer school in Miami, I am so grateful that I could take this course during my PhD to learn about PID and get to know all of you.

Dear Utrecht PhD advisory committee: Dr. Slaper-Cortenbach, Dr. Boelens, Prof. Frenkel. Thank you for taking the time to discuss my progress and thesis outline.

During my time in Boston I not only worked with great people, I also got support from many people outside work, who made my time there such a pleasure.

My US family: Marieke and José, I was happy to move close to you and see Lucas and Isabel grow up. At least now I got to experience Lucas in the phase where he would run towards me to hug me and not just say "Stuff" when asked "What's up?".

Alice, you gave me a home when I arrived in Boston the first time. Deborah, I enjoyed being your roommate in Jamaica Plain and have great memories of our walks with Lucille in the arboretum, yoga retreats and dinner nights.

Dear Roomies of 34 William street. Jonas, luckily, I passed the test and wasn't taller than you so you guys let me move in. Thank you for all the running, stressing, eating, laughing and Harlem shakes together. Many days we have been stuck in that apartment during a lock down, snow storm or hurricane but with you guys those days were great. Camilla, you were the one making it a real cozy family home for the rest of us, with Saturday brunches with homemade scones or Swedish pancakes, Christmas trees and Swedish Midsummer parties. You immediately became a partner in crime, already doing your PhD and immediately raising the bar for me with your incredible dedication. You were a true inspiration to me and I really enjoyed our discussions and practice talks. When Jonas moved after 2,5 great years of us living together, Josie made our home even more of a PhD apartment. I will never forget how one night we all came back at 3 am from mouse experiments, and how we had some serious PhD brainstorm sessions in the living room. Dear Josie, Josephine, we joked about it many times when we met new people: we both have a Dutch and American nationality, are medical doctors and were doing a lab PhD that involved working with CRISPR/Cas9, so we just had to become close friends, and we sure did. You were a huge example and support to me in Boston, but also when we were both back in the Netherlands. I'm so happy we share this adventure and can continue to motivate each other for future ones.

Kiki, Ariane, Stuart, Cesar you were all great people I met at the start. Cesar, your excitement about your past years in Boston as a postdoc made me really look forward to the years ahead. Raj, you introduced us to the Harvard Business School community and I loved being concert groupies with you. Aislinn, your sense of humor is impeccable. Vera, you were my UCU intro group mom and it was great to reconnect in Boston. Jesper Hjortnaes, your ambition to combine clinical work and basic research has been exemplary to me and your mentorship and friendship helped me on my way. Nico and Claire, thank you for your friendship and Nico in particular for holding my smelly shoes at the end of the Reach The Beach relay race.

All the different Dutchies who were living in Boston at some point. We celebrated holidays together and went on weekend trips to explore the area. Thomas, Klaas, Noor, Sjors, Louise and Sarah thank you and I'm looking forward to continue in the Netherlands. Pieter, a special thanks to you, together with Dominique we made a great team ("I haven't done a single useful thing today, but at least we really lived"). Tiago, you are almost a Dutchy too now. I admire you so much for all you have accomplished. The Netherlands should feel honored to have you now.

My book club, *Bookness casual*, in many ways represents what makes the Boston/Cambridge area so special. Thank you for letting me be part of this

diverse, articulate and motivating group of women. I learned a lot from our book discussions and felt supported and inspired by your friendship.

Noémie, we immediately connected. Thank you for starting the book club and for having become such a close friend. You will kick ass during your own PhD now. Erica, I'm glad you went to follow your dream to move from finance to food. Jess Goldfin, thank you for being so excited and warm, always introducing me to new people. Stevie, you are more articulate than anyone I know and I'm so glad you are in law school now. Pia and Josh, thank you for visiting me and Alidus in Paris and for sharing such a special surprise with us.

Jess Tollette, I am so thankful Katya introduced us many years ago. You have become a great friend and I am so proud of all you do in Spain now.

Julia von Metzsch, our friendship started when your parents Ernst and Gail, practically adopted me, giving me a home for Thanksgiving every year. George and Kerry Anne, thank you too for all those great memories of jokes about dogs and microwaves. It felt great to be part of the giggling kids at the table. Simaõ, you too have become family, thank you for all the times you lent me your car.

Ksenia and Travis, we met through different friend groups so were destined to become friends I think. Ksenia, I got to know you initially through your smart comments at our book club but soon you and Travis both became good friends. Thank you for always being up for everything. Your energy is amazing.

Katya Nick, we met in 1987 as babies but it turned out we also like each other as grown-ups. When I moved back to Boston in 2011 you immediately invited me to NYC even though we didn't know each other that well. After that, many NYC trips followed and you introduced me to a lot of great people.

Dominique, Do, I have known you since we were 14 years old and we have always been good friends, but our time in Boston made you really become a sister. You pushed me when needed, sat next to me during my first paper submission or danced the night away with me. You introduced me to polo which added a whole new world of happiness and excitement to my last 2 years in Boston.

Azza, Az, it feels like nothing I write does our friendship full justice. You were a part of it all. You know all the people of my different Boston phases, and in fact a lot of the great people I described above I even met through you. We met while dancing at the Middlesex and recognized each other from work. We decided that must mean we both liked going out and soon we became friends. We had lunch at work every single day during my PhD. We went out together, went on weekend trips, spent Christmas together in a group or just with the two of us, we laughed, cried and really talked about everything. When I think of

Boston I think of you.

Another important aspect of my Boston time was polo. Thank you to the Myopia polo community, BillieBo and all the horses. Phil, I can't thank you enough for letting me use your horses. At one point I was really living from polo practice to polo practice (including the burger afterwards) and it made the hard work in lab much more bearable. Leslie, you are a role model, thank you for all your generosity and for welcoming me into the BillieBo family. I really enjoyed our arena polo, beach rides, brunches and the polo tournament you hosted for my goodbye.

Dog-sitting was another important part of my Boston happiness. In particular Igby made me feel so special when at one point he didn't want to go outside with my roommates anymore. Thank you Brian!

Several people in the Netherlands have had a mentor role before going to Boston and stayed in touch or even visited me. Annet van Royen, you introduced me to Prof. Wulffraat and I am happy you are now the director of SUMMA. This program really helped with the start of my Boston adventure. It was great to see you and your family in Boston. Janneke Tekstra, you were my SUMMA mentor and helped me decide where to go for my PhD and even came to check it out for yourself. Marloes Heijstek, I am so thankful I got to work with you and learn from you during the EULAR task force. You became a role model and a friend. Joost Swart and Bas Vastert, thank you for making me excited about rheumatology and for making my first conference so fun. Joris van Montfrans, I appreciate our discussions about my future career plans and what steps to take next.

My aunt Marian van Opstal, you know the challenges of lab research, but also the excitement of good results. You even came to check out my poster presentation at the ESGCT meeting in The Hague.

January 2017, I moved back to the Netherlands to start as a pediatric resident in Rotterdam. Starting clinical work in a new hospital was made easier by my new colleagues. Prof. Matthijs de Hoog and Rogier de Jong gave me some time off to finish my PhD work. The neonatologists with whom I started appreciated my background. Prof. Irwin Reiss showed the Boston mindset and passion for research. Prof. Annemarie van Rossum and Dr. Pieter Fraaij were enthusiastic about my PhD work. Most importantly, my fellow Rotterdam residents were understanding and welcoming. I have been amazed how quickly we went from colleagues to good friends.

Another important person during this last phase was Margreet van Rixtel, my coach through the Erasmus mc Challenge and support program. You helped me

at a time when I was stuck with my final paper.

In January 2018 I continued my residency with ETZ Tilburg. Dear Charlie Obihara and the other pediatricians and residents in Tilburg, thank you for showing me how to become a good clinician and for making it possible for me to go to the CIS conference in Toronto where I could present my final paper.

After getting addicted to polo in Boston, I could now join Dominique in The Netherlands and get to know the rest of the Dutch polo community, in particular Polo Club Midden-Nederland. Playing with you helped me a lot to blow off steam and I really enjoyed our weekends together.

Dear Utrecht thesis committee: Prof. Edward Nieuwenhuis, Prof. Monique den Boer, Prof. Edmond Rings, Prof. Arjan Lankester, Dr. Mirjam van der Burg, thank you for evaluating my thesis. I enjoyed the opportunities I had to meet you and hope to work together in the future.

In addition to the support I received from my Boston friends, many friends from before kept in touch, visited me and supported me from a distance. I am happy to be closer to you all now and I am looking forward to have more time to see each other.

All my UCU friends. Chicas Frederique, Loes, Lotte and Katie, thank you for your unconditional friendship and support even though all of you have such busy lives. Loes, glad we got to watch that Red Sox game together. Vanya, Henrieke and Sanne, the three of you coming to visit me in Boston was the best. Vanya, thank you for the additional trips to DC, Baltimore, Philly and New Orleans and all your visits to Boston when you were in DC for work. I will never forget we were in DC for Obama's inauguration. Max and Moritz, we met on those couches during UCU intro week and I can't wait to have time to hang like that again.

SUMMA '08 friends thank you for being such unique medical doctors and for always making it so easy to join the group again as if I never left. I am so happy I once decided to take this route. Emma, thank you for always making an effort to see each other, even if it's just for a quick hug.

Utrecht Singelstraat roomies, my cozy home during SUMMA. You teach me about completely different fields and visited me in Boston. JC Toffee, I am happy so many of you flew across the pond for yoga retreats and Boston brunches. The Balisto gang, thank you for giving me such a warm welcome back to the Netherlands.

Selma, our first vacation without our parents was to the US when we were 15 and 16 years old. We were surprised about the giant fridges and coke

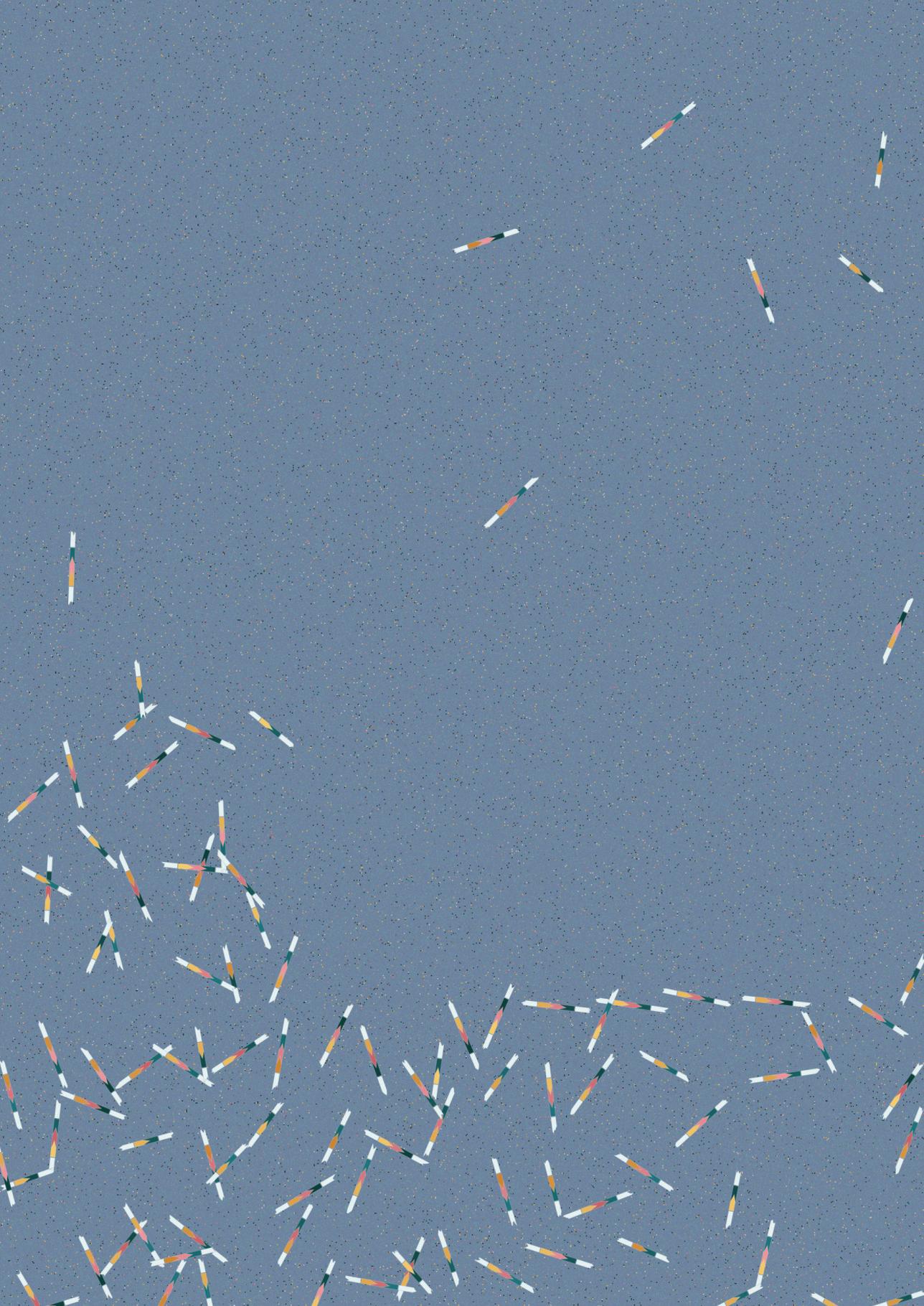
glasses. Over a decade later I was living in Boston and you came to run the NYC marathon. No matter how much time goes by without seeing each other, things never change between us. We just need to talk faster to cover everything. Ariane, my sweet and funny friend from next door. I have known you almost my entire life. Thanks for making the trip as well.

Finally, I want to thank Alidus and my family.

Alidus, thank you for letting me pursue my dream to go back to Boston for my PhD and giving me the extra push at other career decisions. You were always understanding when I had to cancel a trip because of my PhD work. Your optimism, relaxed attitude and sense of humor has helped me a lot. I love you.

Edgar, "little" brother, during my PhD in Boston you moved even further away, from Amsterdam to Singapore, and later to Hong Kong. But as you started an expat life, we in fact grew closer than before. I'm so proud of you. Thank you for your love and support.

My parents. You played a crucial part at different points: You motivated me to go to Boston the first time and when I moved back to the Netherlands you helped me finish my PhD thesis. After having lived abroad for a long time it has been great to stay with you on the weekends, eat your home cooked meals and discuss my PhD. Mom, as a kid I saw your PhD thesis on the book shelf and from that moment onwards I was determined to get a PhD (not really knowing what it entails). You helped me focus when I was getting distracted and took care of me when needed. Dad, you did a terrific job doing the entire layout for this book, made the most unbelievable figures, but most importantly you discussed my work. You made me excited again when I was completely done with it, you studied immunology books in order to understand my work and wrote algorithms to help with the bioinformatics. Thank you for everything!





Lisa Ott de Bruin - *Curriculum vitae*

Work

July 2017 – now

Pediatric resident (in training) - Sophia Children's Hospital – Erasmus MC, Rotterdam, The Netherlands (Jan 2018 - June 2019, ETZ Tilburg).

Jan 2017 – June 2017

Pediatric resident (not in training) - Sophia Children's Hospital – Erasmus MC, Rotterdam, The Netherlands.

Oct 2012 – Dec 2016

Postdoctoral research fellow, Notarangelo lab, Division of Immunology, Boston Children's Hospital, Harvard Medical School.

PhD

Oct 2012 – now

Infection and immunity (I&I) PhD program, UMC Utrecht. PhD supervisors Utrecht: Prof. Nico Wulffraat, Dr. Marianne Boes. PhD supervisors Boston: Prof. Luigi Notarangelo, Dr. John Manis. All research was performed in the Notarangelo lab. Title: Cut and paste: Using CRISPR/Cas9 to model RAG1 deficiency.

Oct 2016

Clinical Immunology Society (CIS) summer school on primary immunodeficiencies.

Jan 2014

Advanced immunology course, Utrecht University.

Medical school

2008 - 2012

Selective Utrecht Medical Master (SUMMA), Utrecht University, *GPA 4.0*.

- Electives: Pediatrics (hematology/rheumatology/immunology) and adult hematology.
- Aug. 2011 - Apr. 2012 **Research internship**, Boston Children's Hospital, Harvard Medical School.
Supervisor: Prof. Notarangelo. Primary immunodeficiencies (gene therapy for RAG deficiency).
- Mar. 2012 **United States Medical Licensing Examination (USMLE) Step 2 Clinical Skills (CS)**
- Dec. 2009 - Sept. 2010 **Research internship**, EULAR guidelines for vaccinating pediatric patients with rheumatic diseases. Supervisor: Dr. Nico Wulfraat, Wilhelmina Children's hospital. See publications.
- Summer 2009 Clinical rotation general surgery, Hospital Beneficiencia Portuguesa (UNAERP) in Ribeirão Preto, São Paulo, Brazil.

Bachelor

- 2005 - 2008 **University College Utrecht (UCU)**, Major: Science (pre-medical track), *Magna Cum Laude*.
- Fall semester 2007 **UC Berkeley**, CA. Endocrinology, Human Anatomy, Aging Human Brain.

High school

- 1999 - 2005 **VWO** (Life Sciences, Economics), Het Nieuwe Lyceum, Bilthoven. *Cum Laude*.

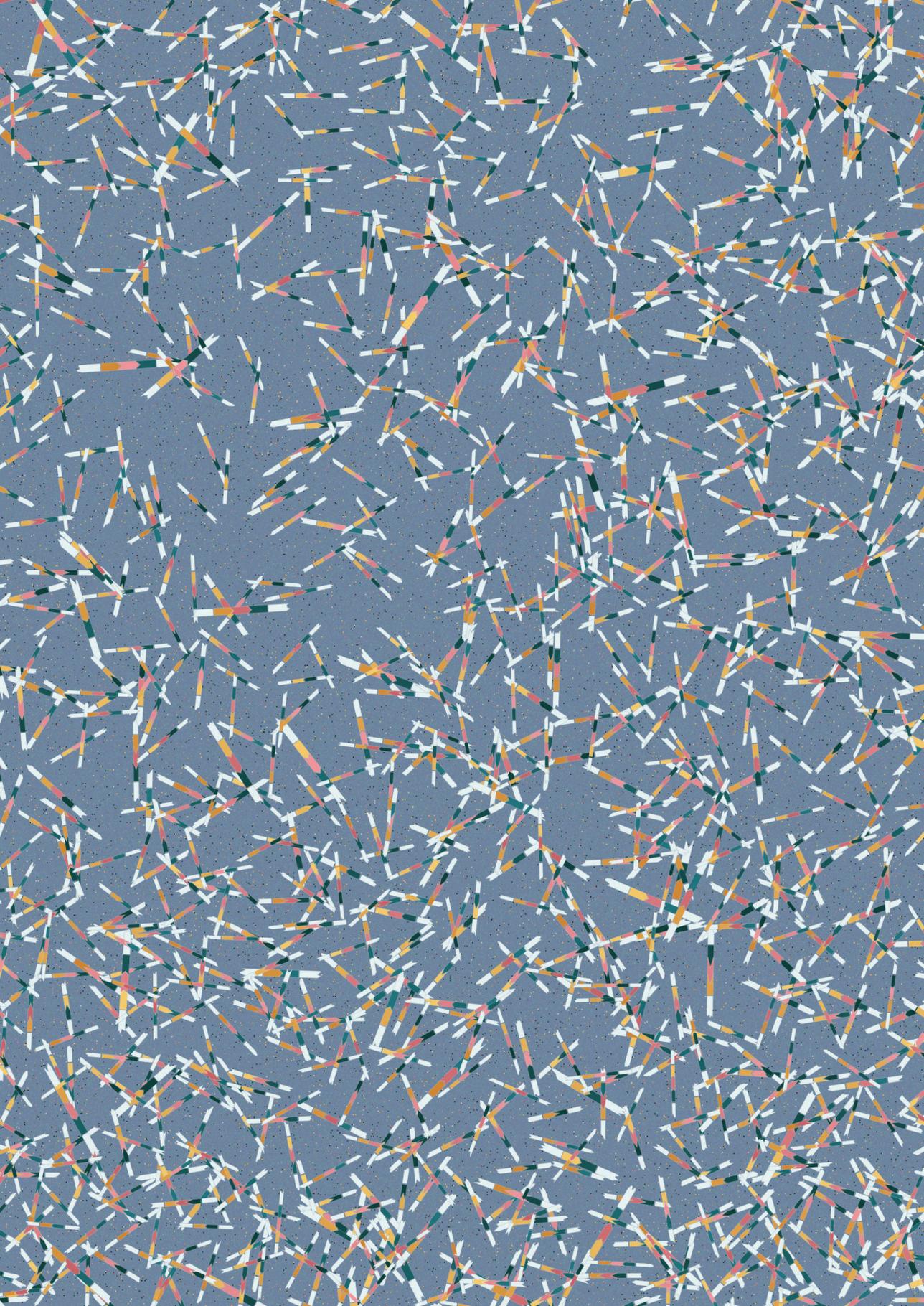
Presentations

- April 2018 **CIS:** Hypomorphic Rag1 mutations alter the pre-immune repertoire at early stages of lymphoid development
- Oct 2017 **Grand round Sophia Children's Hospital:** Genome editing to understand and correct primary immunodeficiencies.
- Aug. 2016 **AML1:** Novel Genome Editing Tools to Model Primary Immunodeficiencies (invited speaker)

- Oct. 2014 **ESID:** Correcting RAG1 deficiency using CRISPR/Cas9
- Oct. 2014 **ESGCT:** Generation of *Rag1* mouse models using CRISPR/Cas9 (poster)
- June 2014 **ISSCR:** Gene-editing of *RAG1* using CRISPR/Cas9 (poster)
- 2010 **PRes:** Vaccinations in pediatric patients with autoimmune inflammatory or rheumatic diseases (AIIRD) receiving immunosuppressive therapy (poster)

Grants/awards

- 2016 Contributed to R01: Immune Repertoire and Function in Typical and Atypical SCID
- 2013 - 2016 T32 NRSA Training Grant (NIH)
- 2014 Best Poster Award ESID meeting



List of publications

This thesis

Ott de Bruin, LM, S Volpi, K Musunuru. Novel Genome-Editing Tools Model and Correct Primary Immunodeficiencies. *Frontiers in Immunology*. 2015 May 21;6:250.

Brauer PM, Pessach IM, Clarke E, Rowe JH, **Ott de Bruin L**, Lee YN, Dominguez-Brauer C, Comeau AM, Awong G, Felgentreff K, Zhang YH, Bredemeyer A, A-Herz W, Du L, Ververs F, Kennedy M, Giliani S, Keller G, Sleckman BP, Schatz DG, Bushman FD, Notarangelo LD, Zúñiga-Pflücker JC. Modeling altered T-cell development with human induced pluripotent stem cells from patients with *RAG 1* mutations and distinct immunological phenotypes. *Blood*. 2016 Aug 11;128(6):783-93.

Ott de Bruin, Lisa, Wei Yang, Kelly Capuder, Yu Nee Lee, Maddalena Antolini, Robin Meyers, Martin Gellert, Kiran Musunuru, John Manis, Luigi Notarangelo. Rapid generation of novel models of RAG1 deficiency by CRISPR/Cas9-induced mutagenesis in murine zygotes. *Oncotarget*. 2016; 7(11):12962-12974.

Ott de Bruin LM, Bosticardo M, Barbieri A, Lin SG, Rowe JH, Poliani PL, Ching K, Eriksson D, Landegren N, Kämpe O, Manis JP, Notarangelo LD. Hypomorphic *Rag1* mutations alter the pre-immune repertoire at early stages of lymphoid development. *Blood*. 2018 May 9.

Other publications during PhD

Dobbs, K [and 51 others, including **L.M. Ott de Bruin**], L.D. Notarangelo. Natural Killer Cells from Patients with Recombinase-Activating Gene and Non-Homologous End Joining Gene Defects Comprise a Higher Frequency of CD56^{bright} NKG2A⁺⁺⁺ Cells, and Yet Display Increased Degranulation and Higher Perforin Content. *Frontiers in immunology*. 2017 Jul 17;8:798.

NK cells of patients with mutations in RAG1 have a different phenotype, which has implications for bone marrow transplantations to treat RAG deficient patients. Previously, it was thought that NK cells of RAG deficient patients were unaffected because they are not dependent on V(D)J recombination.

Rowe, J.H, B. Stadinski, L. A. Henderson, **L.M. Ott de Bruin**, O. Delmonte, Y.N. Lee, M. T. de la Moren, R. K. Goyal, A. Hayward, H. Chiung-Hui, M. Kanariou, A. King, T. Kuijpers, J. Yi Soh, B. Neven, J.E. Walter, E. S. Huseby, L. D. Notarangelo. Abnor-

malities of T-cell receptor repertoire in CD4⁺ regulatory and conventional T-cells in patients with mutations: implications for autoimmunity. *Journal of Allergy and Clinical Immunology*. 2017 Dec;140(6):1739-1743.

CD4⁺ regulatory and conventional T-cells from peripheral blood of patients with RAG mutations show restricted TCR repertoire and patient CD4⁺ regulatory T-cells showed impaired in vitro suppressive activity

Volpi S, Yamazaki Y, Brauer PM, van Rooijen E, Hayashida A, Slavotinek A, Sun Kuehn H, Di Rocco M, Rivolta C, Bortolomai I, Du L, Felgentreff K, **Ott de Bruin L**, Hayashida K, Freedman G, Marcovecchio GE, Capuder K, Rath P, Luche N, Hagedorn EJ, Buoncompagni A, Royer-Bertrand B, Giliani S, Poliani PL, Imberti L, Dobbs K, Poulain FE, Martini A, Manis J, Linhardt RJ, Bosticardo M, Rosenzweig SD, Lee H, Puck JM, Zúñiga-Pflücker JC, Zon L, Park PW, Superti-Furga A, Notarangelo LD. *EXTL3* mutations cause skeletal dysplasia, immune deficiency, and developmental delay. *J Exp Med*. 2017 Mar 6;214(3):623-637

A novel gene defect, EXTL3, causes skeletal dysplasia, immune deficiency, and developmental delay. The patients had an Omenn-like phenotype. EXTL3 conditional knock out zebrafish showed reduced thymic tissue and patient iPSCs showed impaired T-cell differentiation. I used CRISPR/Cas9 to successfully generate a mouse model with the same human mutation. Unfortunately, these mice did not show an immune phenotype, nor any evident skeletal dysplasia or developmental delay.

Yu Nee Lee, Francesco Frugoni, Kerry Dobbs, Irit Tirosh, Likun Du, Francesca A. Ververs, Heng Ru, **Lisa Ott de Bruin**, Mehdi Adeli, Jacob H. Bleesing, David Buchbinder, Manish J. Butte, Caterina Cancrini, Karin Chen, Sharon Choo, Reem A. Elfeky, Andrea Finocchi, Ramsay L. Fuleihan, Andrew R. Gennery, Dalia H. El-Ghoneimy, Lauren A. Henderson, Waleed Al-Herz, Elham Hossny, Robert P. Nelson, Sung-Yun Pai, Niraj C. Patel, Shereen M. Reda, Pere Soler-Palacin, Raz Somech, Paolo Palma, Hao Wu, Silvia Giliani, Jolan E. Walter, Luigi D. Notarangelo. Characterization of T- and B-cell repertoire diversity in patients with RAG deficiency. *Science Immunology*. 2016 Dec 16;1(6).

This study characterizes the repertoire of peripheral blood T- and B-cells of patients with RAG deficiency. It shows that patients with a CID-G/Al phenotype showed an abnormal repertoire, but not as abnormal as Omenn syndrome or Leaky SCID patients. As we wanted to understand the primary, pre-immune, repertoire, we generated 3 mouse models with mutations found in patients with CID-G/Al where we could focus on this repertoire (chapter 5).

Publications before PhD

M.W Heijstek, **L.M Ott de Bruin**, R. Borrow, F. van der Klis, I. Koné-Paut, A. Fasth, K. Minden, A. Ravelli, M. Abinun, G.Pileggi, M. Borte, M. Bijl, N.M Wulffraat. EULAR recommendations for vaccination in paediatric patients with rheumatic diseases. *Annals of the Rheumatic Diseases (ARD)*, 2011;70(10):1704-12.

M.W Heijstek, **L.M Ott de Bruin**, R. Borrow, F. van der Klis, I. Koné-Paut, A. Fasth, K. Minden, A. Ravelli, M. Abinun, G.Pileggi, M. Borte, M. Bijl, N.M Wulffraat. Vaccination in paediatric patients with auto immune rheumatic diseases: a systematic literature review for the European League Against Rheumatism evidence-based recommendations. *Autoimmunity Reviews*, 2011; 11(2):112-22.

These two publications lay at the basis of my Boston adventure. During medical school I had to do a literature study and Prof. Nico Wulffraat suggested I could help Marloes Heijstek, PhD student at the time, with the literature search and organization of the EULAR task force. This was a group of pediatric immunologists and vaccine experts from Europe and Brazil tasked with critically appraising the currently available literature and coming up with evidence based recommendations for vaccinating pediatric patients with rheumatic diseases. Here my interests in immunology grew as I wanted to understand more of the underlying mechanisms of what we were discussing. One of the experts was Prof. Fasth who later introduced me to Prof. Notarangelo.



