UNIVERSITÉ PARIS DESCARTES

Inserm
Institut national de la santé et de la recherche médicale

THÈSE

pour l'obtention du grade de

Docteur de l'Université Paris V

Discipline: Génétique

École Doctorale Frontières du Vivant (ED474)

Présentée et soutenue publiquement le 6 novembre 2012

par

Alexandre Bolze

La découverte de l'origine génétique de l'asplénie congénitale isolée chez l'homme

The genetic dissection of isolated congenital asplenia in humans

Composition du jury

Président: Dr. Corinne Antignac

Rapporteur: Dr. Agnel Sfeir

Rapporteur: Dr. Jonathan R. Warner

Examinateur: Dr. Licia Selleri

Examinateur: Dr. Albert Libchaber

Directeur de thèse: Dr. Jean-Laurent Casanova

THE GENETIC DISSECTION OF ISOLATED CONGENITAL ASPLENIA IN HUMANS

ACKNOWLEDGEMENTS

I thank Jean-Laurent Casanova for giving me the opportunity to do this work. I am particularly grateful for the freedom I had during these four years. Jean-Laurent also taught me how to ask a question, how to plan an experiment, how to write, as well as some diplomacy. I also have to thank Jean-Laurent for his contagious passion about science and human genetics, and his support for the next steps in my scientific career.

I thank the PhD school and the University Paris Descartes for allowing me to do my PhD at Rockefeller and being so tolerant. I thank Francois Taddei for all the advice. I hope I will be able to think outside the box more, and combine more often mathematics with biology.

I thank the Rockefeller University for the great environment. I was very lucky to be here!

I thank all members of the laboratory in New York for all the discussions, the ideas, the critics and the protocols. I am particularly grateful to my room703-mates: Amandine Crequer, Bertrand Boisson, Sophie Cypowyj, Minji Byun and Giraldina Trevejo. Minji suggested to perform many of the experiments in this thesis, and was the harshest critic about this work. Minji also made millions of edits on all papers related to this thesis.

I thank all members of the laboratory in Paris, especially Anne Puel. Anne was my direct supervisor at the beginning and taught me cell culture, biochemistry, sequencing etc. Anne was in charge of this project before I joined the laboratory, and followed this project during four years, making sure I was not going completely off track.

I thank all the physicians that participated in the project. Thank you for collaborating with us, for going back and back and back in the case reports of the patients, and for coping with all the IRB paperwork. Nizar Mahlaoui was the physician who was in charge of the recruitment of all patients from France.

I thank the patients for participating in this project and being so collaborative and understanding. Without the huge help of patients and physicians, this project would not be possible.

I thank Licia Selleri and all her lab. Licia collaborated with us throughout this project. Firstly Licia suggested candidate genes involved in spleen development. Secondly, Licia with Matt Koss, a student in her lab, elucidated the role of Nkx2-5 in the development of the spleen in the mouse model, and helped to design the experiments to test the function of the NKX2-5 mutant we identified. Licia then taught me the basics of mouse genetics and allowed me to keep the Rpsa+/- mice in her lab. More importantly Licia gave many suggestions and critics that were very important for this project and I hope that we will keep collaborating on this project in the future.

I thank Nikolaus Trede and his lab. Nikolaus taught me the basics of genetics with zebrafish. He performed all the experiments using morpholinos to knockdown Rpsa in the zebrafish.

I thank Steven Ellis for his experience with RPSA and ribosomopathies. Steve made the Rpsa knockout mouse a couple of years ago to look for a DBA phenotype. He also performed the rRNA processing experiments. I hope we will continue to work together on this project.

I thank Daniel Meruelo and the members of his laboratory for the discussions on RPSA or LAMR. I haven't had time to test most of the protocols they sent me, but I hope to do so in the near future.

I thank my thesis committee members: Albert Libchaber and Licia Selleri. I enjoyed much the few discussions with Albert on science in general, biology and complex systems. They always triggered a lot of thinking afterwards. I will try to remember them in my future career in science and use some theory.

I thank Agnel Sfeir and Jonathan Warner for reading and reviewing this thesis. I look forward to your comments. I thank Corinne Antignac for being the president of my thesis jury. I look forward to the discussion on November 6 and after.

I thank Jean-Pierre Abastado and Alessandra Nardin, from SIgN (Singapore Immunology Network) for teaching me the basics of research in biology and making me want to do a PhD. I thank Philippe Kourilsky for advising me to do my PhD under the direction of Jean-Laurent Casanova.

I thank Brad Rosenberg for being a great example as a PhD student at Rockefeller and for giving me ideas on what to do after the PhD.

I thank Shaheen Kabir for pushing me to work more and for challenging me all the time.

Finally I would like to thank Anne Puel, Shaheen Kabir, Minji Byun, Sophie Cypowyj and Amandine Crequer for reading and editing this thesis.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 7

ABSTRACT OF THE DISSERTATION

The Genetic Dissection of Isolated Congenital Asplenia in Humans

By Alexandre Bolze

Université Paris Descartes

Thesis director: Jean-Laurent Casanova

Isolated congenital asplenia (ICA) is a rare primary immunodeficiency, first described in 1956, that typically manifests in childhood with sudden, life-threatening, invasive bacterial disease. Patients with ICA do not display any other overt developmental anomalies. The genetic etiology of ICA has remained elusive. I hypothesized that ICA results from single-gene inborn errors of spleen development. I aimed to decipher the molecular genetic basis of ICA by pursuing a genome-wide approach, based on the sequencing of the whole-exome and the detection of copy number variations in all patients of our cohort. I found that heterozygous mutations in *RPSA*, ribosomal protein SA, were present in more than half of ICA patients (19/33). I then showed that haploinsufficiency of *RPSA* led to ICA in one kindred at least. RPSA is a protein involved in pre-rRNA processing and is an integral part of the ribosome. The challenge is, now, to understand the pathogenesis of the disease. How does a mutation in a ubiquitous and highly expressed gene lead to a spleen specific phenotype? This discovery will set the basis for a broader understanding of the development of the spleen in humans and the function of a ribosomal protein. This discovery will also be beneficial to the families of patients with ICA, guiding genetic counseling. It will lead to prevention of infections in newborns with mutations in *RPSA*. Finally the method we used to analyze the exomes of the ICA cohort will be useful to discover the genetic etiology of other genetic diseases.

LIST OF PUBLICATIONS

1. **Bolze A**, Byun M, McDonald D, Morgan NV, Abhyankar A, Premkumar L, Puel A, Bacon CM, Rieux-Laucat F, Pang K, Britland A, Abel L, Cant A, Maher ER, Riedl SJ, Hambleton S, Casanova JL. 2010. "Whole-exome-sequencing-based discovery of human FADD deficiency", ***American Journal of Human Genetics*** 87(6):873-81.

2. Mahlaoui N, Minard-Colin V, Picard C, **Bolze A**, Ku CL, Tournilhac O, Gilbert-Dussardier B, Pautard B, Durand P, Devictor D, Lachassinne E, Guillois B, Morin M, Gouraud F, Valensi F, Fischer A, Puel A, Abel L, Bonnet D, Casanova JL. 2011. "Isoclated congenital asplenia: a French nationwide retrospective survey of 20 cases", ***Journal of Pediatrics*** 158(1):106-12.

3. Liu L, Okada S, Kong XF, Kreins AY, Cypowyj S, Abhyankar A, Toubiana J, Itan Y, Audry M, Nitschke P, Masson C, Toth B, Flatot J, Migaud M, Chrabieh M, Kochetkov T, **Bolze A**, Borghesi A, Toulon A, Hiller J, Eyerich S, Eyerich K, Gulacsy V, Chernyshova L, Chernyshov V, Bondarenko A, Grimaldo RM, Blancas-Galicia L, Beas IM, Roesler J, Magdorf K, Engelhard D, Thumerelle C, Burgel PR, Hoernes M, Drexel B, Seger R, Kusuma T, Jansson AF, Sawalle-Belohradsky J, Belohradsky B, Jouanguy E, Bustamante J, Bue M, Karin N, Widbaum G, Bodemer C, Lortholary O, Fischer A, Blanche S, Al-Muhsen S, Reichenbach J, Kobayashi M, Rosales FE, Lozano CT, Kilic SS, Oleastro M, Etzioni A, Traidl-Hoffmann C, Renner ED, Abel L, Picard C, Marodi L, Boisson-Dupuis S, Puel A, Casanova JL. 2011. "Gain-of-function human STAT1 mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis", ***Journal of Experimental Medicine*** 208(8):1635-48.

4. **Bolze A**[#], Abhyankar A, Grant AV, Patel B, Yadav R, Byun M, Caillez D, Emile JF, Pastor-Anglada M, Abel L, Puel A, Govindarajan R, de Pontual L, Casanova JL. 2012. "A mild form of SLC29A3 disorder: a frameshift deletion leads to the paradoxical translation of an otherwise noncoding mRNA splice variant", ***PLoS One*** 7(1):e29708

5. Koss M, **Bolze A***, Brendolan A*, Saggese M, Capellini TD, Bojilova E, Boisson B, Prall OWJ, Elliott D, Solloway M, Lenti E, Hidaka C, Chang CP, Mahlaoui N, Harvey RP, Casanova JL, Selleri L. 2012. "Congenital asplenia in mice and humans with mutations in a Pbx/Nkx2-5/p15 module", ***Developmental Cell*** 22(5):913-26.

*: equal contributions

[#]: corresponding author

LIST OF ABBREVIATIONS

By alphabetical order

Only the genes mentioned multiple times in the thesis appear on this list.

AD	autosomal dominant
ALPS	auto-immune lymphoproliferative syndrome
AR	autosomal recessive
AS	asplenia syndrome
BCE	before the common era
Bp	base pair
BWA	Burrows-Wheeler aligner
CBC	complete blood count
cDNA	complementary DNA
CE	common era
CHD	congenital heart disease
CHH	cartilage hair hypoplasia
CMC	chronic mucocutaneous candidiasis
CNV	copy number variant
Ctrl	control
Ct	threshold cycle
DBA	Diamond-Blackfan anemia
DD	death domain
DED	death effector domain
DISC	death-inducing signaling complex
DKC	X-linked dyskeratosis congenita
DNT	$CD4^-CD8^-TCR\alpha\beta^+$ T cells
DSC	differential scanning calorimetry
EBV	Epstein-Barr virus
EBV-B	EBV-transformed B cells
ES cell	embryonic stem cell

FADD	Fas-associated via death domain
GATK	genome analysis toolkit
gDNA	genomic DNA
HD	homeodomain
H&E	hematoxylin and eosin
HJB	howell-jolly bodies
HLA	human leukocyte antigen
HSE	herpes simplex encephalitis
ICA	**isolated congenital asplenia**
IFN	interferon
IPD	invasive pneumococcal disease
IRB	institutional review board
LDH	lactate dehydrogenase
LOF	loss-of-function
MIM	number in the OMIM database
MLPA	multiplex ligation-dependent probe amplification
MMR	measles mumps rubella
MOI	multiplicity of infection
mRNA	messenger RNA
MSKCC	Memorial Sloan-Kettering Cancer Center
MSMD	mendelian susceptibility to mycobacterial disease
NA	not assessed
NCBI	national center for biotechnology information
NKX2-5	NK2 homeobox 5
NMD	nonsense-mediated mRNA decay
NS	not stimulated
NT	not treated
OMIM	online mendelian inheritance in man
PBL	peripheral blood lymphocytes
PBMC	peripheral blood mononuclear cell
PCD	primary ciliary dyskinesia

PF	passing filtering
PHA blasts	phytohemagglutinin-stimulated peripheral blood lymphocytes
PID	primary Immunodeficiency
PM	post mortem
Q-PCR	quantitative RT PCR: quantitative real-time polymerase chain reaction
RP	ribosomal protein
RPSA	ribosomal protein SA
rRNA	ribosomal RNA
RU	Rockefeller University
SCID	severe combined immunodeficiency
SDS	Schwachman-Diamond syndrome
SEM	standard error of the mean
SNP	single nucleotide polymorphism
SVC	superior vena cava
TB	tuberculosis
TCS	Treacher Collins syndrome
TLX1	T-cell leukemia homeobox 1
SLC29A3	solute carrier family 29, member 3
uq	unique
US	ultrasound
UTR	untranslated region
VSV	vesicular stomatitis virus
VZV	varicella zoster virus
WES	whole-exome sequencing
WT	wild type
XLA	X-linked agammaglobulinaemia
YRD	tyrosine-rich domain

CHAPTER 1

INTRODUCTION

1.1 The Genetic Theory of Infectious Diseases

Albert Libchaber, a respected physicist at Rockefeller University, once told me that biology and medicine were not real sciences in his opinion. Biologists were telling independent stories rather than working on a theory. As an argument, he highlighted the fact that life expectancy barely increased in developed countries in the last thirty years despite all the resources invested in life sciences. He had a point, although I would argue that mathematics would not be a science either if the evolution of the Dow Jones was the only criteria considered. In any case, life expectancy has dramatically increased at one very specific point in History, at the beginning of the 20th century (**Figure 1**)[1]. This was primarily due to the wide acceptance of the germ theory of disease, first established by Louis Pasteur, and the consequent or parallel improvements in hygiene, as well as the development of vaccines and antibiotics. The theory suggested that microbes were the cause of many infections and diseases in general.

However, the germ theory of disease could not account for the fact that some people were infected without presenting any symptoms. Charles Nicolle was the first to notice this in 1911 in Tunis, while working on typhus. It intrigued him that some guinea pigs developed fever while others did not despite being inoculated the same dose of typhus. Strikingly when he inoculated the blood from a guinea pig that had contracted 'inapparent typhus' (i.e. with no signs of fever) to another guinea pig, the other guinea pig always presented signs of fever, if the dose of inoculum was sufficient[2]. This experiment proved that the first guinea pig had contracted typhus, in spite of the lack of fever. Later Nicolle also demonstrated that native children, particularly very young ones, contracted a benign form of typhus without developing any apparent symptoms, which could only be detected by experimental transmission of typhus by inoculating their blood in the monkey or the guinea pig[2]. This was the first example of inapparent infections. Another example is that some children develop a very severe infection after vaccination with Bacille Calmette-Guerin (BCG)[3], a phenomenon called BCG-itis. Since all the BCG vaccines are identical, what could explain that some children develop these severe infections whereas the vast majority does not develop any symptoms?

Almost 40 years after the discovery of Nicolle, the first explanations for inapparent infections started to arise. In 1954, Anthony Allison showed that genetic mutations could confer resistance to malaria. Allison was troubled by the very high percentage, up to 20% in some African tribes, of carriers of the sickle-cell allele at the heterozygous level, despite the fact that being a homozygous carrier led to an early death from anemia. He hypothesized that carrying the sickle-cell allele led to a selective advantage and showed that indeed, children from an East African tribe carrying the sickle-cell allele were less susceptible to infection by *Plasmodium Falciparum* compared to the children from the same tribe not carrying the sickle-cell allele[4]. However, this resistance could only explain 'inapparent malaria' for a small percentage of the population. This was the beginning of the 'population genetics' branch in the field of human genetics of infectious diseases. The dominant hypothesis in this branch was that common infections (at the population level) with virulent microbes in otherwise normally resistant (to infections) patients resulted from the polygenic (complex) inheritance of common susceptibility alleles. In parallel, Ogden Bruton described the first primary immunodeficiency (PID): X-linked agammaglobulinaemia (XLA, OMIM #300755) in one eight-year old boy with multiple infections[5]. Forty years passed between the clinical identification of the specific loss of gamma globulin as the cause for multiple infections in one child and the subsequent identification of the gene involved in XLA[6; 7]. This marked the rise of the 'Mendelian' branch in the field of human genetics of infectious diseases. The investigations tested successfully the hypothesis that rare infections with a broad range of weakly virulent microbes and/or multiple, recurrent infections in an individual were the result of a rare monogenic PID.

However none of these two mainstream hypotheses could account for the fact that half of the BCG-itis patients were not susceptible to any other infections[3]. Jean-Laurent Casanova and Laurent Abel put forward an alternative hypothesis in the 1990s. They hypothesized that life-threatening infectious diseases striking otherwise healthy children may also result from single-gene inborn errors of immunity[1; 8-10]. As the aim of this thesis is to test this hypothesis, I will highlight the words I thought were important. (i) Life-threatening: the focus is on extreme cases. This is one difference with the diseases usually looked at by population geneticists. (ii) Otherwise healthy: patients with PIDs

were thought to be susceptible to multiple pathogens such as the patients with Severe Combined Immunodeficiency (SCID, OMIM #601457) that were susceptible to fungal, viral and pyogenic pathogens. The hypothesis of the laboratory is different as it focuses on patients with susceptibility to only one type of weakly virulent pathogen, BCG vaccine for example. (iii) Children: The hypothesis is formulated in the case of children. Children have less time to acquire immunodeficiencies, and thus a defect is more likely to be congenital rather than acquired. (iv) Single-gene: one mutation in one gene is enough to cause the disease. At the population level however, a specific infection could be caused by a collection of highly diverse and possibly immunologically related single-gene lesions according to this hypothesis. In conclusion, this model, if true, could bridge the two previous fields and support a genetic theory of pediatric infectious diseases (**Figure 2**). This new model also led to a considerable expansion of the definition of a PID.

The lab had been testing this hypothesis on different types of infectious diseases including mycobacterial disease[11], invasive pneumococcal disease (IPD)[12], herpes simplex encephalitis (HSE)[13] and chronic mucocutaneous candidiasis (CMC)[14]. Interestingly a subgroup of patients suffering from recurrent IPD was also diagnosed with isolated congenital asplenia (ICA, OMIM #271400). Conversely, ICA patients only suffer from encapsulated bacterial infections, mostly IPD. In addition, ICA challenged many of the conventional thoughts on PIDs as it is not a hematopoietic defect and the prognosis spontaneously improves after early childhood[10; 15]. The genetic etiology of ICA was unknown. The choice was made; I would test the theory of the lab on this disease.

1.2 Isolated Congenital Asplenia

Asplenia can be functional (i.e. the spleen does not work) as in sickle cell disease[16]. Asplenia can also be acquired after trauma or surgery, or it can be congenital. Congenital asplenia is often associated with complex visceral defects, as part of the 'heterotaxy' syndromes[17]. The best known among these heterotaxy syndromes is the asplenia/polysplenia syndrome (AS; OMIM #208530), which was fully documented by Ivemark in 1955[17]. Patients suffering from this syndrome have a

complete absence of the spleen (right-sided isomerism), or in some cases two spleens (left-sided isomerism) (**Figure 3**). They also suffer from various anomalies of the heart and great vessels, which are usually the most severe aspects of the condition. The incidence is about 1/10 000 to 1/40 000 live births[15]. Mutations leading to AS have been identified in various genes controlling left-right laterality[18]. In contrast the patients with ICA lack other developmental anomalies, particularly with respect to their cardiovascular systems.

Since Myerson *et al.* reported the first case in 1956[19], only 53 patients with ICA have been reported in the medical literature in over 50 years[19-42] apart from our own recent series of French patients[15]. ICA is thus an extremely rare PID. However, ICA is likely under-diagnosed, for at least three reasons. First, it is a rare disease, and as with most rare diseases, awareness is insufficient. Second, this rare disease is not well known when compared with other PIDs or other rare diseases, in part, because it has received little attention from clinicians and investigators. PIDs were mostly thought to be hematopoietic defects and the spleen was/is not the most fashionable organ. Third, ICA strikes suddenly and swiftly, with unexpected lethal infections that do not provide medical professionals the opportunity to identify ICA. Among the 53 patients reported, 27 cases were sporadic and 26 were familial, belonging to 11 unrelated families (total of 38 families worldwide), suggesting there could be a genetic factor. ICA was diagnosed in 34 cases in children and in 15 cases in adults; for four cases, the age of the patient was not available. There were 24 males and 22 females; the gender was not specified in seven cases.

Sparse reports indicate that most known children with ICA die of fulminant bacterial disease – particularly IPD[43]. There were at least 49 episodes of severe bacterial infection in the patients reported with ICA. Twenty-seven patients suffered from IPD and five from *Haemophilus influenza* type b invasive disease. Overall, nine patients had several episodes of invasive bacterial disease: two episodes in five patients; four of whom died at the second episode, and three episodes in four patients; three of whom survived the third infection. Up to 22 of the 53 patients died of overwhelming bacterial sepsis. This is thought to result from the combination of two

phenomenons. Firstly, asplenic patients have low numbers of IgM memory B cells in their blood[44]. This low number of IgM memory B cells is associated with low levels of polysaccharide-specific antibodies. The B cells from the splenic marginal zone are important for the secretion of these antibodies[44]. Secondly, there is a lack of phagocytosis of circulating opsonized bacteria in the absence of splenic macrophages[43]. There is no curative treatment for ICA. Therefore patients with ICA are protected from infectious complications of asplenia by reinforced and repeated conjugated and unconjugated vaccinations against encapsulated bacteria (haemophilus, pneumococcus, meningococcus), by daily antibiotic prophylaxis (penicillin), and through education of the families and clinicians (in case of sudden fever). From these reports, we can only conclude that ICA is a life-threatening birth defect and that it remains poorly understood.

1.3 Aim

The aim of this thesis is to decipher the molecular genetic basis of ICA.

1.4 Hypothesis

I hypothesized that isolated congenital asplenia (ICA) results from single-gene inborn errors of spleen development.

Figure 1: Mortality curves at various periods of human history, from the Palaeolithic period (<10,000 BCE) to modern times (2000 CE).

Contemporary data for the United Kingdom and Mozambique are available from the WHO site (www.who.int/topics/global_burden_of_disease). Life tables for the Palaeolithic and Neolithic periods are based on skeleton examinations, assuming that 60% of newborn infants survived to the age of five, because few very young skeletons were found in the burial grounds. The gradual adjustment of the immune system by natural selection did not increase life expectancy until the end of the 19th century, due to the coevolution of microorganisms and the emergence of new infectious threats. Thus, the increase in life expectancy in the 20th century does not reflect the sudden and global natural selection of high-quality immune genes. The area between the four ancient curves and the curve for the UK in 2000 corresponds to ~65% of individuals currently alive. Most of these individuals have retained specific immunodeficiencies masked by medical progress.

Figure 1: Mortality curves at various periods of human history, from the Palaeolithic period (<10,000 BCE) to modern times (2000 CE).

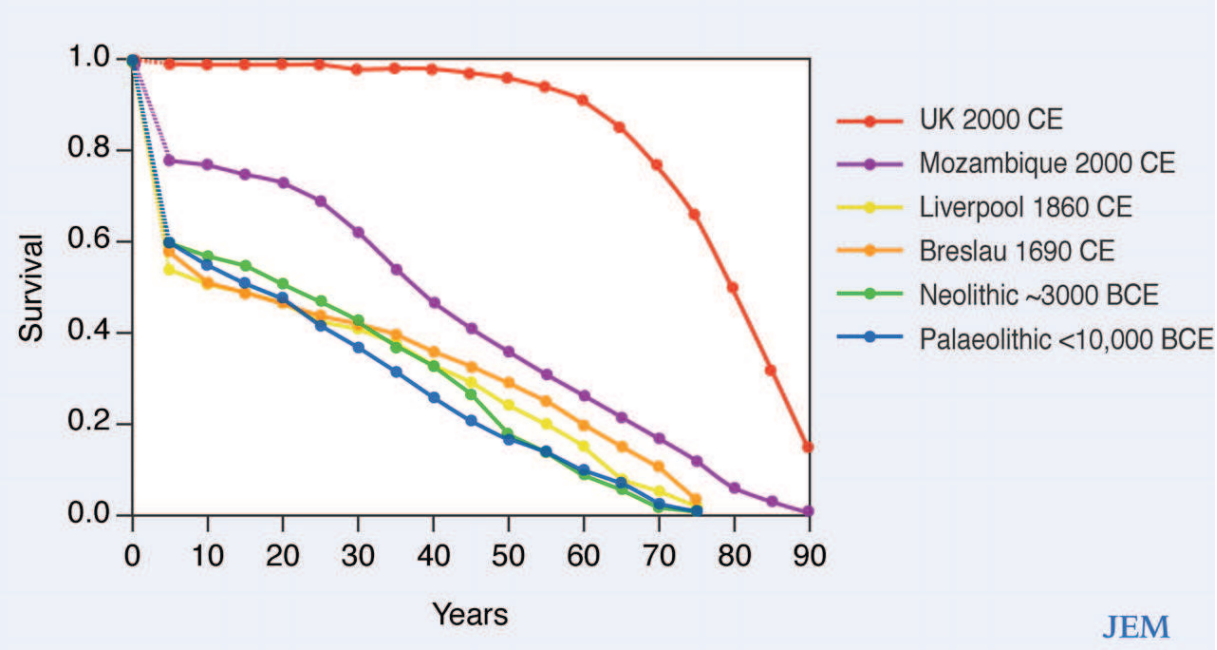


Figure 2: Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity?

Schematic representation of a hypothetical, age-dependent, human genetic architecture of infectious diseases. The hypothesis of the laboratory suggests that single-gene human variants make an important contribution to the determinism of life-threatening infectious diseases of childhood, in the course of primary infection. By contrast, predisposition to severe infectious diseases in adults, in the course of microbial reactivation from latency or secondary infection, is less influenced by germline human genetic variations, resulting in a more complex and, perhaps, polygenic contribution. Somatic and epigenetic processes are likely to play a greater role with advanced age.

Figure 2: Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity?

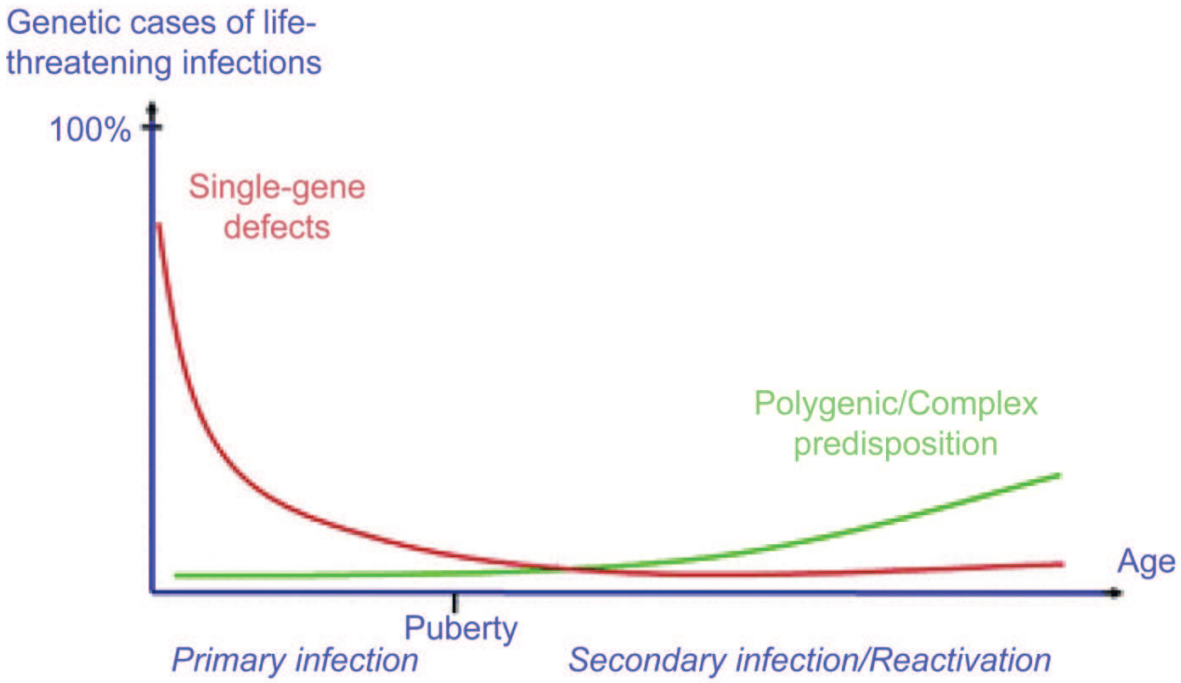
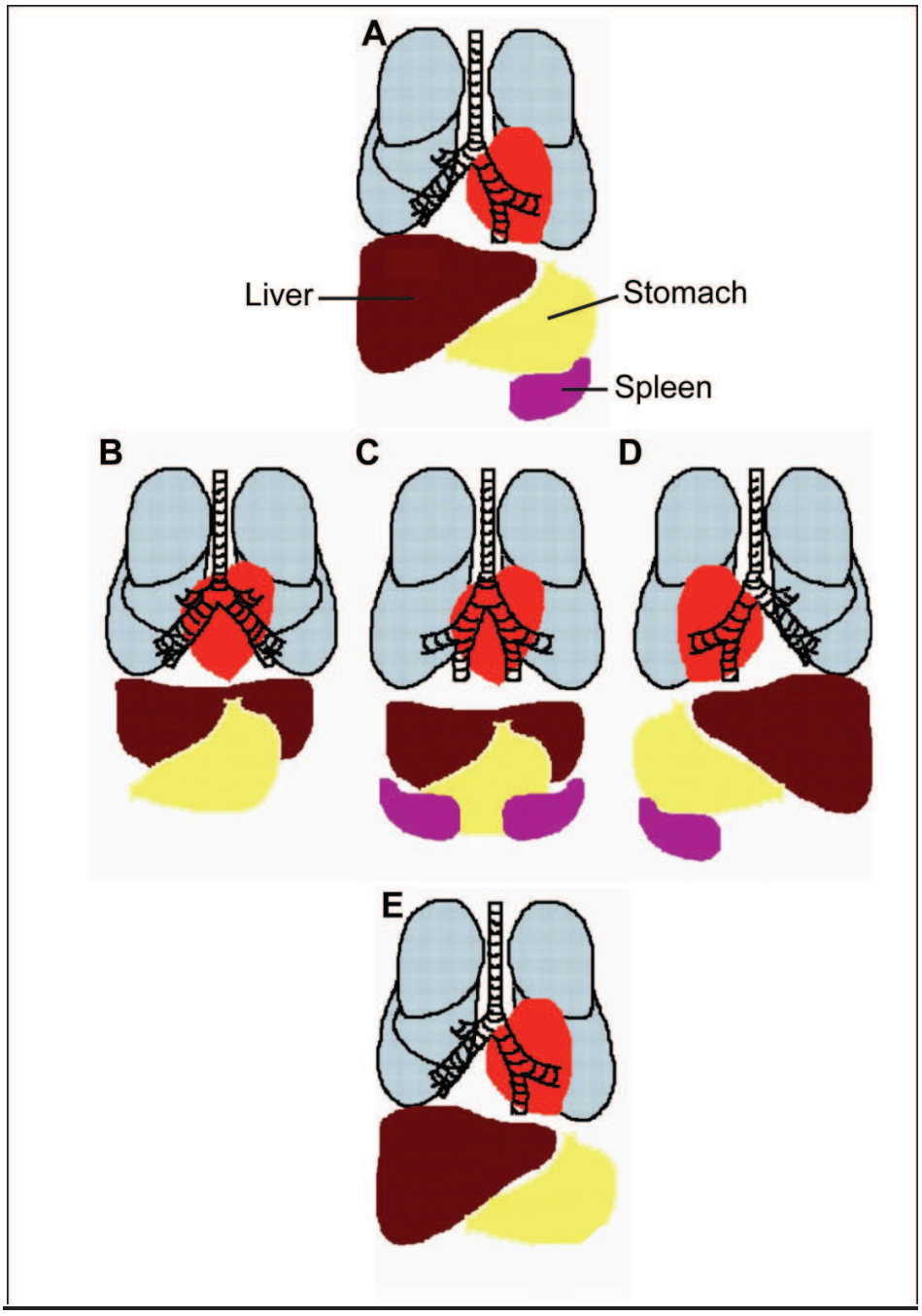


Figure 3: The anatomic spectrum of organ laterality.

Figure adapted from[45] (**A**) Situs solitus. Normal position of the organs in the human body. (**B**) Right atrial isomerism. The liver is midline, there are 2 eparterial bronchi, the position of the stomach and cardiac apex is indeterminate, and there is asplenia. (**C**) Left atrial isomerism. The liver is midline, there are 2 hyparterial bronchi, the position of the stomach and cardiac apex is indeterminate, and there are multiple spleens. (**D**) Situs inversus. (**E**) Isolated congenital asplenia (ICA).

Figure 3: The anatomic spectrum of organ laterality.



CHAPTER 2

COHORT OF PATIENTS WITH ICA

In order to test my hypothesis, I needed biological material, gDNA at least, from ICA patients. The more patients we would recruit, the more powerful the genetic study could be and the more likely I was going to be able to decipher the molecular genetic basis of ICA. This was the most important part of the project.

2.1 French retrospective survey

I was lucky because as I started in the lab, they conducted a French national retrospective survey, by mailing questionnaires to chiefs of pediatric services at public hospitals affiliated with immunological, infectious, or intensive care societies to learn more about ICA. The survey covered the years 1979-2008[15]. To be included in the survey, a patient had to be diagnosed with ICA based on the presence of Howell-Jolly bodies on peripheral blood smears (sign that the spleen is not functional), the lack of detectable spleen by abdominal ultrasound (US), and no detectable cardiovascular malformation by echocardiography. The data were collected from 294 of 443 pediatric units solicited. Twenty patients from 10 kindreds, neither related to each other nor consanguineous, were included[15]. Four cases were sporadic, 16 were familial, and the sex ratio was balanced (12:8). Out of the six multiplex families, four suggested autosomal dominant inheritance. Median age was 12 months at first infection (range: 2-516), 11 months at diagnosis of asplenia (range: 0-510), and 9.9 years at last follow up (range: 0.7-52). Fifteen patients suffered from 18 episodes of invasive bacterial infection, mainly caused by *Streptococcus pneumonia* (61%). The outcome was poor, as nine patients (45%) died of fulminant infection. This data clearly suggested that ICA is more common than initially estimated, with an incidence of about one per million live births (0.51 per million). This would mean that there is a minimum of 3,500 cases of ICA in the world.

2.2 International recruitment

This survey was a retrospective survey. Therefore we were able to gather clinical data, the key to understand the epidemiology of the disease, but not biological samples from the patients, which are the keys to undertake our genetic analysis. We, thus, tried

to recruit more patients from world-wide collaborations. The strategies used were the following:

(i) Contacting corresponding authors of papers reporting a patient with ICA. We were able to recruit only one family this way (kindred G in **Figure 4**) because there were only few recent papers, and the majority never responded to my emails.

(ii) Contacting physicians that already collaborate with us, in particular physicians that collaborate with us on the IPD project, or that already recruited ICA patients.

(iii) Publishing papers on ICA to alert physicians that one team is working on it. This was one of the goals of our paper in The Journal of Pediatrics reporting the results of the French national retrospective survey[15]. However few if none have contacted us after this publication. More success was encountered after the paper reporting the role of the Pbx/Nkx2-5/p15 pathway in the development of the spleen[46]. Two unrelated patients, one from the US and one from Australia, contacted us directly to participate in the study. They read the paper because it was advertised on many news websites thanks to Weill Cornell Medical College media office.

(iv) Presenting the project at conferences such as ESID (European Society of primary Immuno Deficiencies). Jean-Laurent is relentlessly traveling all around the World to recruit patients to test the hypothesis. However ICA being a 'newer' project in the laboratory, he did not talk frequently about asplenia.

In 3 years, we were able to recruit 24 kindreds with at least one ICA patient (**Figure 4**, **Table 1**). In total, there were 37 ICA patients and 7 possible ICA patients in these 24 kindreds. People were declared possible ICA patients when the diagnosis was unclear, or when the patients deceases due to severe bacterial infections and had a family history of ICA but their spleen was not analyzed. We were able to obtain gDNA from 35 of these patients (34 + 1 possible). The ratio male/female was 21/16. ICA segregates as an autosomal dominant (AD) trait in five kindreds (A-E), whereas it looks to segregate

as an autosomal recessive (AR) trait in one kindred: F. For the two other multiplex families, and the 16 sporadic cases, it is hard to conclude anything before the identification of the genetic etiology of ICA. Finally the clinical information remains unequal depending on where the patient was recruited. Therefore it would be inexact to conclude on the number of patients presenting severe infections etc…

2.3 Ideas

The recruitment of patients was the longest part of this thesis. For example, we were able to contact a physician in Toulouse following a family with 2 ICA patients. The patients agreed to participate but after a year of emails, phone calls and paperwork we are still not able to enroll them in the study because the Toulouse hospital has not yet been approved as a collaborative center for us. Other patients were recruited but it took more than a year to have the physicians send detailed case reports. Moreover I had to exclude some samples after discovering they were not ICA cases.

This seems inherent to any new project and there is nothing abnormal. I was actually very lucky to be able to recruit so many families. At the time of Twitter, Facebook and the Do It Yourself generation, there must be a faster approach. This approach would link directly the patient with the researcher. Patients interested in research would be able to contact the laboratory of their choice working on their disease. They would be able to fully participate in the study. If a patient wants to participate and discuss the literature on the topic with the researchers, why would he not have the right to? Why does the patient, frequently a tax payer, is not able to have access to the result of the research performed on him? In our lab some patients declined participating to studies, after being initially very interested, when told we would not be able to communicate any results with them. Without a change in the IRB policies and the way human genetics is being pursued in academia, my guess is that all future discoveries in the field will be made by biotech companies. For example, 23andMe has started research projects using the pool of customers wanting to have their genotyping done. They named it: 23andWe: https://www.23andme.com/research/ . They were able to recruit more than 5,000 patients with Parkinson disease and already published results in open access journals[47; 48]. 23andMe will soon have the genotype and DNA of

more than a million people, with a majority willing to participate in research and learn more about them. This is a great opportunity for them and for us as well. Patients will have their genome in hand. Giving them the opportunity to share it with the laboratory of their choice would be a great step forward.

Figure 4: Pedigrees of the ICA cohort.

Black symbols represent ICA patients. Gray symbols represent individuals not diagnosed with ICA, but likely to be ICA patients, due to their infectious disease history and familial history. Capital letters above pedigrees refer to a specific kindred (these letters will be used to refer to one family if needed later in the thesis). Numbers preceded with “JL” indicate patients sequenced by exome and their exome number. These will be used in some of the exome results’ tables.

Figure 4: Pedigrees of the ICA cohort.

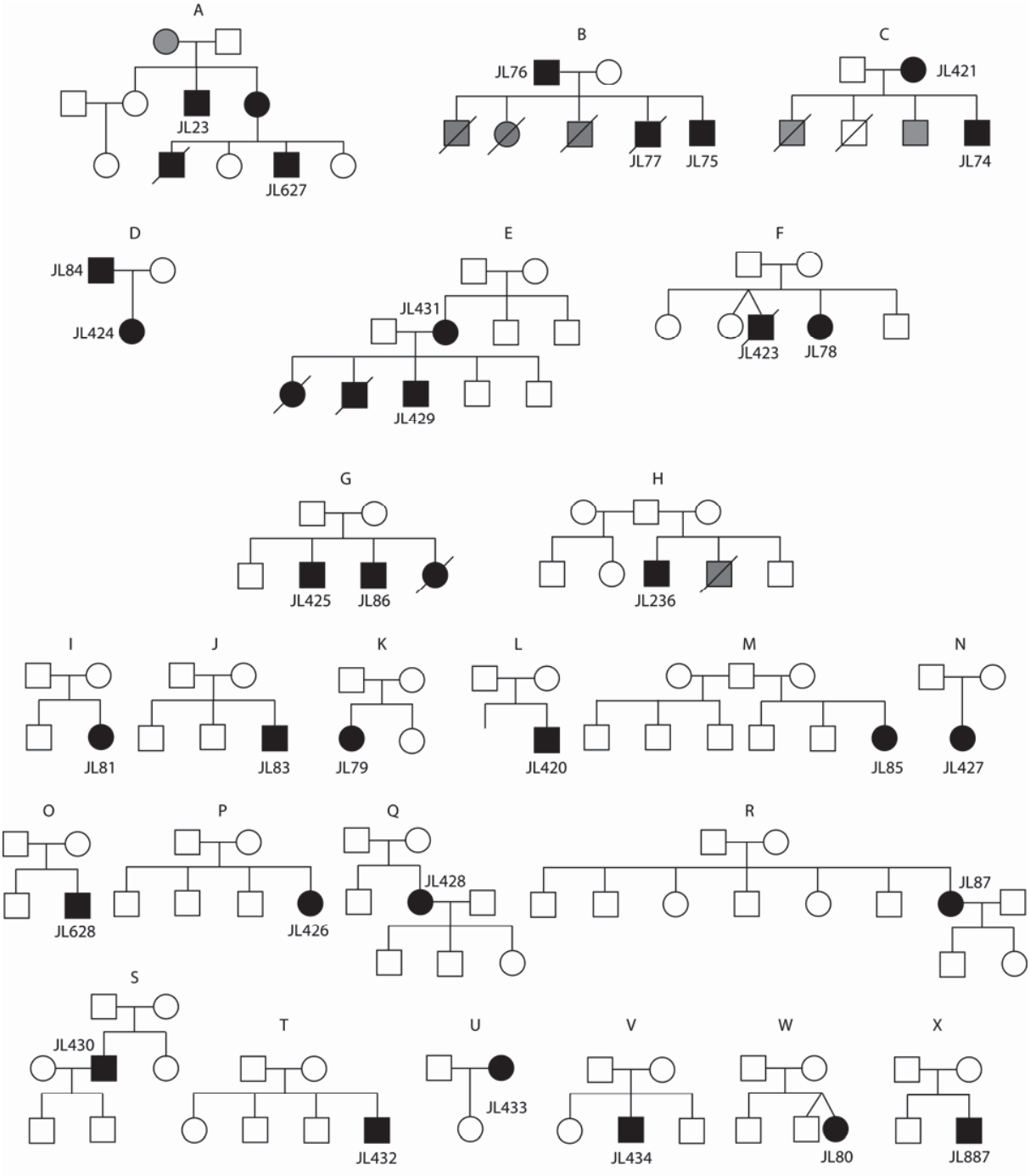


Table 1: Geography of the ICA cohort.

Kindred	Country[1]	Origin of the mother	Origin of the father	Mode of inheritance hypothesized
A	United Kingdom	Caucasian	Caucasian	AD
B	France	Subsaharan Africa (Congo)	Subsaharan Africa (Congo)	AD
C	Reunion island	Caucasian	Tamoul	AD
D	Italy	Caucasian	Caucasian	AD
E	US	Caucasian	Caucasian	AD
F	France	Caucasian	Caucasian	AR
G	United Kingdom	Caucasian	Caucasian	AR
H	France	Caucasian	Caucasian	AR
I	France	Caucasian	no information	?
J	Spain	Caucasian	Caucasian	?
K	France	Western Africa (Cameroun)	Caucasian	?
L	Colombia	Colombian	unknown	?
M	France	Caucasian	Caucasian	?
N	Israel	Israel (jewish)	Israel (jewish)	?
O	France	Caucasian	Caucasian	?
P	Israel	Israel (jewish)	Israel (jewish)	?
Q	United Kingdom	Caucasian	Caucasian	AR
R	France	Caucasian	Caucasian	?
S	Sweden	Caucasian	Caucasian	?
T	Sweden	Caucasian	Caucasian	?
U	Sweden	Caucasian	Caucasian	?
V	Sweden	Caucasian	Caucasian	?
W	France	Caucasian	Caucasian	AR[2]
X	France	Caucasian	Caucasian	?

1: country where they live and where they were diagnosed.

2: all grandparents from the same village.

AD: autosomal dominant

AR: autosomal recessive

CHAPTER 3

CANDIATE GENE APPROACH

To test my hypothesis I had to look for mutations that might be disease-causing in the 35 ICA patients of our cohort for whom I had gDNA. At the beginning of my PhD, the only way to do it was by Sanger sequencing (see **Material and Methods**). This method required to sequence each exon of each gene one by one. I had to select good candidate genes as it was impossible to sequence the 20,000 genes of the genome.

3.1 Genes involved in the development of the spleen

The genetic etiology of ICA was unknown in humans. The generation of genetically engineered mouse models has led to the discovery of a limited number of genes, mostly encoding transcription factors, which are required for the accurate temporal and spatial coordination of cell-fate specification, cell proliferation, and differentiation during embryonic spleen development. These cellular processes are disrupted in embryos deficient for *Nkx2-3*, *Nkx3-2*, *Pbx1*, *Sox11*, *Wt1*, and *Tcf21*[49-54], all of which are normally expressed within the spleen primordium, and also disrupted in embryos deficient for *Barx1* and *Elavl1*[55; 56]. The human orthologs of these genes whose mutations cause asplenia in mice are candidate genes in patients with ICA (**Table 2**). However these mutant mice all exhibit spleen agenesis or hypoplasia, accompanied by other organogenesis defects. There is a stronger candidate gene for human ICA, as deficiency in *Tlx1* (also known as *Hox11*) causes ICA in the mouse model[57].There is therefore at least one gene that seems to be specific for spleen genesis, at least in the mouse model (**Table 2**).

3.2 Genes mutated in patients with syndromic asplenia

A number of patients not included in our cohort present syndromic asplenia, i.e. asplenia in the context of other developmental defects. For example, about a third of patients with heterotaxy syndrome have no spleen[58]. The genetic etiologies of some of these traits have been partially identified[18]. We hypothesized that ICA might be allelic with these syndromes, i.e. that some ICA patients might carry mutations in these syndromic asplenia-causing genes listed in **Table 2**. Human *SLC29A3* deficiency is one example to back up this hypothesis. Homozygous loss of function mutations in

SLC29A3 have been shown to lead to a wide spectrum of syndromes such as the H-syndrome, type I diabetes, or familial Faisalabad histiocytosis[59]. Even more striking was the fact that siblings with the same disease-causing mutation could have different clinical phenotypes. For example one brother could present histiocytosis, hyperpigmentation and type I diabetes whereas the sister only presented one of the clinical phenotypes: histiocytosis[59]. It was therefore possible to hypothesize that some ICA patients were mutated in genes involved in heterotaxy syndrome or other syndromes with asplenia and presented a partial clinical phenotype.

These syndromes include the heterotaxy syndrome with or without primary ciliary dyskinesia (PCD). Since the first case report of a patient with ciliary abnormalities and asplenia in 2003[60], a study showed that 6.3% of the PCD patients had heterotaxy, some of these possibly having no spleen[61]. In humans, mutations in 10 genes had been shown to cause heterotaxy syndrome with PCD[62-70] (Table 2). Heterotaxy can also be caused by mutations in genes that are not involved in PCD[71-80] (**Table 2**). Most of these genetic defects are autosomal and seem to be dominant, making the corresponding genes good candidates for autosomal dominant or sporadic ICA. Others are X-linked recessive traits, and hence good candidates for sporadic ICA in young males. Other syndromes where asplenia was observed were: autoimmune polyglandular syndrome type I due to mutations in *AIRE*[81], hyper-IgE syndrome due to DOCK8 deficiency[82], ATR-X syndrome due to mutations in *ATR-X*[83] and finally HO-1 deficiency[84]. Human heme oxygenase-1 (HO-1) deficiency is particularly striking as the 3 cases reported have the same phenotype including asplenia, hemolysis and nephritis[84-86].

3.3 Failure of the approach

In total I identified 34 genes that could lead to ICA according to the knowledge available in 2009 (**Table 2**). However, only *TLX1* appeared, to me, to be a very good candidate. Some of the other genes felt unlikely. At least, it seemed unlikely that all ICA cases would only be phenocopies or incomplete phenotypes of a syndromic asplenia unless physicians misdiagnosed some of the patients.

I sequenced *TLX1*, *PBX1*, *SOX11*, *TCF21*, *NKX3-2*, *BAPX1*, *BARX1*, *OFD1*, and *CRELD1* in this order. The sequencing did not lead to any interesting mutation, or more specifically to any coding non-synonymous, and not common variation. The logic would have been to continue sequencing one by one the remaining genes. However the technique of whole-exome sequencing started to be available at the end of 2009. I was very lucky because the genomic facility of Rockefeller was a pioneer in the use of this technique and Jean-Laurent decided we would start using this technique instead of Sanger sequencing. We would keep the list of candidate genes, as genes to look for with special attention when analyzing the exome sequences.

Table 2: List of candidate genes to be sequenced in ICA patients.

Isolated Congenital Asplenia in the mouse model	Asplenia and developmental syndromes other than heterotaxy in the mouse model	Heterotaxy with Primary ciliary dyskinesia (PCD)	Heterotaxy without PCD	Other syndromes with asplenia
Tlx1	*Nkx2-3*	*DNAI1*	*ZIC3*	*AIRE*
	Nkx3-2	*DNAH5*	*ACVR2B*	*DOCK8*
	Pbx1	*DNAH11*	*LEFTYA*	*ATRX*
	Sox11	*DNAI2*	*CFC1*	*HO-1*
	Tcf21	*TXNDC3*	*CRELD1*	
	Wt1	*RPGR*	*NODAL*	
	Barx1	*KTU*	*GDF1*	
	Elavl1	*RSPH9*	*TDGF1*	
		RSPH4A	*INV*	
		OFD1	*NKX2-5*	
			BCOR	

CHAPTER 4

WHOLE-EXOME-SEQUENCING-BASED DISCOVERY OF HUMAN FADD DEFICIENCY

In our grant application for the March of Dimes in the fall of 2009 we mentioned we would use whole-exome sequencing (WES) to reach our aim. They rejected our application and our genome-wide strategy arguing we had no proof this approach could work. It was true at that time as the first paper identifying the genetic etiology of a mendelian disorder using exome sequencing was in 2010[87]. We thus decided to try WES on an easier case as a proof-of-principle that the technique could work in our hands. Mike Horn trained by crossing Greenland before embarking on a non-motorized tour of the Arctic Circle. He took nine days to cross Greenland on skis with the help of a kite.

4.1 Clinical features of patients from a large consanguineous family

An 'easy' case for us meant to test WES on a family where the hypothesis that the disease would be a mendelian trait was the strongest and the number of genes to look at the smallest. Therefore we investigated the molecular basis of a condition affecting at least four members of an extended consanguineous kindred of Pakistani origin (**Figure 5A**) with a complex infectious phenotype, with both viral and bacterial infections, congenital cardiovascular malformations, biological features of auto-immune lymphoproliferative syndrome (ALPS, MIM 601859) (high circulating $CD4^{-}CD8^{-}TCR\alpha\beta^{+}$ T-cell (DNT) counts, and elevated IL-10 and FasL serum levels) but no clinical features of ALPS (**Table 3**).

The affected children suffered from recurrent, stereotypical episodes of fever, encephalopathy and mild liver dysfunction (modestly elevated transaminases without cholestasis, metabolic derangement or synthetic defects), sometimes accompanied by generalized seizures that were difficult to control. Episodes lasted several days, sometimes requiring intensive care, and cranial imaging in three patients (P1, P3 and P4 [IV.2, IV.4 and IV.5 in **Figure 5**]) suggested atrophy, despite subsequent recovery in two of these patients (P1 and P3). For some of the episodes, it was possible to identify a viral trigger: varicella zoster virus (VZV), measles mumps rubella (MMR) attenuated vaccine, parainfluenza virus and Epstein-Barr virus (EBV). The index case, P3, died at the age of four years, during such an episode. In addition, fatal invasive pneumococcal disease occurred in two children (P1 and P2), and abnormally high counts of Howell-

Jolly bodies were detected in the two remaining patients (P3 and P4), despite the presence of a spleen. The identification of Howell-Jolly bodies on peripheral blood smears is the most reliable indicator of impaired splenic phagocyte function (including functional or anatomical asplenia)[15]. Neither P3 nor P4 suffered from IPD, probably because they were on antibacterial prophylaxis treatment from early infancy. These observations suggest that functional hyposplenism underlied IPD in patients P1 and P2. This clinical syndrome had never before been described (no MIM number). In the previous generation, another five family members had died in childhood, two with "epilepsy" and two from infection (pneumonia and measles, respectively) (data not shown), suggesting that there may have been up to nine patients with this disorder in these two generations of this kindred.

4.2 Whole-exome sequencing and Linkage analysis

We hypothesized that this syndrome would segregate as an autosomal recessive trait. We investigated its genetic etiology, by combining genome-wide linkage analysis by homozygosity mapping with whole-exome sequencing[87]. For the linkage analysis, we genotyped three affected patients (P2, P3, P4) their parents (III.1, III.2, III.3, III.4) and one healthy sibling (IV.1) with the Affymetrix 500K SNP array or the Affymetrix 6.0 array. No material was available for P1. Homozygous regions were detected with the HomozygosityMapper genome browser. Two regions (>1 Mb) were homozygous in the three affected patients and heterozygous in other members of the family: an 8 Mb region on chromosome 11 (between rs4930243 and rs7926553, bp position: g.68,499,874 – g.76,470,079, hg18, NCBI 36.1) and a 9 Mb region on chromosome 18 (between rs9959449 and rs12327522, bp position: g.60,703,632 – g.70,387,399, hg18, NCBI 36.1). The regions were further confirmed by typing microsatellite markers (**Figure 6**).

In parallel to the linkage analysis, we sequenced the exome of P3 with the SureSelect human All Exon kit (Agilent Technologies). The sequences obtained were aligned to the reference genome (hg18 build), with the BWA aligner. Three open-source packages were used for downstream processing and variant calling: the Genome analysis toolkit (GATK)[88], SAMtools[89] and Picard Tools. Substitution calls were made

with GATK UnifiedGenotyper. All calls with a read coverage ≤4x and a phred-scaled SNP quality of ≤30 were filtered out. All the variants were annotated with SeattleSeq SNP annotation. In total, we identified 23,146 variations. 67 variants were found in the chromosome 11 candidate region, and 14 variants were found in the chromosome 18 candidate region (**Table 4**). Comparisons with the NCBI dbSNP build 129, the 1000 Genomes project database and our in-house database (composed of 70 exomes at that time), identified only one nonsynonymous variant that had not been reported before and mapped to the chromosome 11 region. No previously unreported nonsynonymous variant mapping to the linked region on chromosome 18 was identified. The variant identified was a missense homozygous c.315T>G in exon 2 of *FADD* (NM_003824.3, MIM 602457) that changes Cysteine at amino acid position 105 to Tryptophan, p.C105W (C105W) (**Figure 5B** and **Figure 7**). Cysteine at amino acid position 105 is highly conserved throughout evolution (**Figure 5C**). We validated this variant by Sanger sequencing on genomic DNA from peripheral blood and on cDNA from EBV-transformed B cells (EBV-B). This variant segregated with the disease status in all family members examined (**Figure 5A**) and was not found in 282 Pakistani controls, suggesting that it is not an irrelevant polymorphism.

4.3 In vitro functional characterization the C105W FADD mutation

FADD was first described as an adaptor protein interacting with the apoptosis-inducing surface receptor Fas[90]. It has since been shown to interact with various partners and to participate in many other cellular processes, such as autophagy, inflammation, innate immunity, cell proliferation, and tumor development[91; 92]. Thus, we hypothesized that the FADD mutation was responsible for the complex clinical phenotype of the affected individuals. To test this hypothesis, we assessed the expression of *FADD* mRNA levels in EBV-B cells by quantitative RT PCR. We found that *FADD* mRNA levels in the EBV-B cells from P3 (C105W/C105W) were similar to those in the EBV-B cells from IV.1 (WT/WT) (data not shown). However, FADD protein levels were clearly lower in primary fibroblasts from P4 (~16%) than in primary fibroblasts from a healthy control (**Figure 5D, E**). Similarly, FADD protein levels were lower in the EBV-B cells from P3 (~21%) and III.1 (WT/C105W) (~62%) than in the

EBV-B cells from IV.1 (**Figure 5D, F**). Residue C105 is located in alpha-helix 1 of the FADD death domain (DD), at the interface of the Fas/FADD complex[93]. We tested the effect of the C105W mutation on FADD folding stability and Fas-FADD interaction with recombinant Fas DD and FADD DD proteins. The C105W mutant protein was folded, as shown by differential scanning calorimetry (DSC), but the folding stability of the mutant protein was lower than that of WT FADD, by ~10°C (**Figure 5G**). We then assessed the effect of the mutation on Fas-FADD complex stability using a gel co-purification assay in the presence of increasing concentrations of NaCl (**Figure 5H**). Binding levels for C105W FADD and Fas were clearly lower than those for WT FADD, suggesting that the primary Fas-FADD complex was less stable. Thus, the C105W mutation in FADD strongly decreases steady-state protein levels and impairs the interaction of the residual FADD protein with Fas.

FADD forms a critical bridge between the death receptor Fas and caspase-8 (MIM 601763), in a multimolecular array known as the death-inducing signaling complex (DISC)[94]. We therefore studied Fas-induced apoptosis in the patients' cells. Phytohemagglutinin-stimulated peripheral blood lymphocytes (PHA blasts) from P4 displayed impaired apoptosis (viability >50% higher than for controls) when challenged with cross-linked Fas ligand (FasL), similar to Fas-deficient cells (**Figure 8A**). In addition, EBV-B cells from P3 displayed impaired apoptosis when Fas was cross-linked with agonistic Fas-specific antibody (Apo1.3) (**Figure 8B**). In keeping with this finding, investigations of the surviving patient P4 revealed several features classically associated with ALPS due to Fas[95] or FasL[96] deficiency, and more rarely associated with deficiencies of caspase-8 or caspase-10 (MIM 601762)[97; 98]. P4 displayed an expansion of the population of mature $CD4^-CD8^-TCR\alpha\beta^+$ T cells (DNT) in peripheral blood as well as elevated serum levels of IL10 and soluble FasL (**Table 3** and **Figure 8C, D**). Our data confirm that the FADD C105W variant impairs apoptotic function both *in vitro* and *in vivo*, resulting in biological features of ALPS despite the lack of clinically evident autoimmunity or lymphoproliferation. T-cell proliferation in response to various stimuli was normal (**Figure 8E**), contrasting with findings for caspase-8 deficient individuals[97] and mice with FADD deficiency specific to T or B cells[99].

The laboratory findings for the FADD-deficient patients were typical of ALPS, but clinical presentation was dominated by features absent from classical ALPS, leading to early death. In addition to IPD, FADD-deficient patients suffered from recurrent episodes of encephalopathy and liver dysfunction with fever. These were sometimes triggered by viral infection (**Table 3**). Previous studies in mouse embryonic fibroblasts (MEFs) suggested that FADD plays a crucial role in type I interferon (IFN) -dependent antiviral immunity[91; 100]. We therefore tested the hypothesis that the unusually severe illnesses associated with viral infections in FADD-deficient patients might be due to an impaired type I IFN-dependent antiviral response, by assessing the antiviral effect of IFN-α in vesicular stomatitis virus (VSV)-infected P3 and IV.1 EBV-B cells. STAT1-deficient EBV-B cells, lacking a key type I IFN responsive pathway[100; 101], were included as controls. Prior treatment with IFN-α protected both P3 and IV.1 cells from VSV-induced cell death until 36 h post infection (**Figure 9A**). However, IFN-α-treated P3 cells eventually succumbed to VSV infection, whereas IFN-α-treated IV.1 cells remained protected until 60 h post infection (**Figure 9A, B**). These results suggest that the IFN-α-responsive pathway is intact, but that the positive feedback of IFN signaling is impaired in FADD-deficient cells. Interferon regulatory factor 7 (IRF7), induced by autocrine type I IFNs, is known to play a critical role in the positive feedback of IFN signaling, by inducing IFN-α gene family members. We found that the induction of IRF7 in response to VSV infection was impaired in P3 cells, despite levels of IFN-β induction similar to those in IV.1 cells (**Figure 9C**). These data provide a molecular basis for the increase in susceptibility to viral diseases in FADD-deficient patients.

4.4 Human FADD deficiency and functional asplenia

By contrast, we can only speculate about the mechanisms underlying the developmental abnormalities within this kindred. P3 displayed pulmonary atresia (MIM 178370) and P4 had a left-sided superior vena cava (SVC) draining into the left atrium, both of which are rare as sporadic defects and were probably caused by FADD insufficiency. Indeed, mice with a null mutation in *FADD* are not viable, owing to embryonic defects, including cardiac abnormalities, such as thinning of the ventricular myocardium. Moreover, *FADD* expression in mice, at E11.5, is detected not only in the

myocardium[102], but also in the brain, liver and developing vertebrae[102]. Spinal anatomy appeared normal in all FADD-deficient patients, but hepatic and cerebral dysfunctions were clearly key elements of the disease phenotype. The mild portal fibrosis observed on liver biopsy in two patients (**Figure 10 J-L**), should be considered as sequelae of FADD deficiency. Finally, FADD seems to be important for the development of a functional spleen in humans, as demonstrated by the presence of Howell-Jolly bodies – the hallmark of a lack of splenic function – in the blood smears of P3 and P4. In P2, a limited post mortem examination of lymph nodes, spleen and thymus revealed no histological features of ALPS or other lymphoid tissue abnormalities (**Figure 10**). However the spleen showed an expanded, congested red pulp containing increased neutrophils, consistent with sepsis. Moreover, the normal cord-sinus architecture of the red pulp was not readily apparent; reticulin deposition appeared disordered, normal sinusoidal "barrel-hoop" fibres were not identified and macrophages appeared haphazardly arranged (**Figure 10 C-I**). Examination of spleen histology in other affected individuals is unlikely to be possible, thus precluding firm conclusions upon relationship between human FADD deficiency and disordered red pulp microarchitecture. The prominence, but incomplete penetrance of cardiac, hepatic and splenic abnormalities in the clinical phenotype strongly suggests the existence of organ-specific effects of FADD deficiency worthy of further study in suitably designed mouse models.

4.5 The identification of human FADD deficiency is a proof of principle that WES can be applied to discover the genetic etiology of ICA

In conclusion, we described the clinical features of an autosomal recessive inherited disorder with biological features of ALPS, susceptibility to bacterial and viral infections, and developmental abnormalities. Furthermore, we have determined its genetic etiology, resulting in FADD deficiency in the affected children, by genome-wide linkage analysis and whole-exome sequencing. Finally, we have dissected the molecular mechanisms of disease for several aspects of this phenotype, including bacterial susceptibility (functional hyposplenism), viral susceptibility (impaired IFN immunity) and features of ALPS (impaired Fas-dependent apoptosis). Other phenotypes, particularly those of a developmental nature, remain unexplained but are

probably due to the disruption of Fas-independent pathways, as they were not documented in Fas-deficient patients. These unanticipated findings highlight the power of exome sequencing used in combination with genome-wide linkage analysis within consanguineous families. In parallel Minji Byun, my benchmate, had also used successfully WES without linkage analysis to discover that STIM1 deficiency leads to predisposition to Kaposi Sarcoma in humans[103]. With these two proofs-of-principles we were ready to use this approach on the ICA cohort.

Figure 5: Characterization of an inherited FADD mutation.

(**A**) Pedigree of the Pakistani family. Black symbols indicate patients. Gray symbols indicate siblings in the parents' generation who died early in childhood with clinical symptoms related to the disorder observed in P1 to P4. Haplotypes of FADD are indicated, as M stands for c.315T>G and WT for the wild type allele. (**B**) Schematic diagram of FADD protein showing the death effector domain (DED) and the death domain (DD). The location corresponding to the mutation is indicated by an arrow and the predicted amino-acid substitution is shown. (**C**) Evolutionary conservation of the FADD region containing amino acid residue C105 (indicated by the arrow). (**D**) FADD immunoblot in primary fibroblasts and EBV-B cells from patients and controls. Experiments were carried out with two different antibodies: a mouse monoclonal antibody against the C-terminus of FADD (#610399, BD Biosciences) and a rabbit polyclonal antibody against the residues surrounding Ser194 in human FADD (#2782, Cell Signaling). Similar results were obtained with both antibodies. GAPDH was used as a loading control. A representative blot with the monoclonal antibody is shown (n=6). (**E-F**) FADD protein levels in primary fibroblasts (**E**) and EBV-B cells (**F**) determined on the basis of the intensity of the signal on immunoblots and normalized with respect to GAPDH levels. A mean of six experiments is shown. Error bars indicate the SEM. (**G-H**) Effect of the C105W mutation on the stability of FADD DD folding and the Fas – FADD complex. (**G**) Differential scanning calorimetry (DSC) analysis of WT and C105W His_6-FADD DD proteins. (**H**) Fas/FADD complex stability assay. The retention of FADD DD (WT or C105W) with immobilized His_6-Fas DD was assessed with various concentrations of NaCl, as previously described[19] (E is imidazole elution of remaining complex after the final NaCl concentration).

Figure 5: Characterization of an inherited FADD mutation.

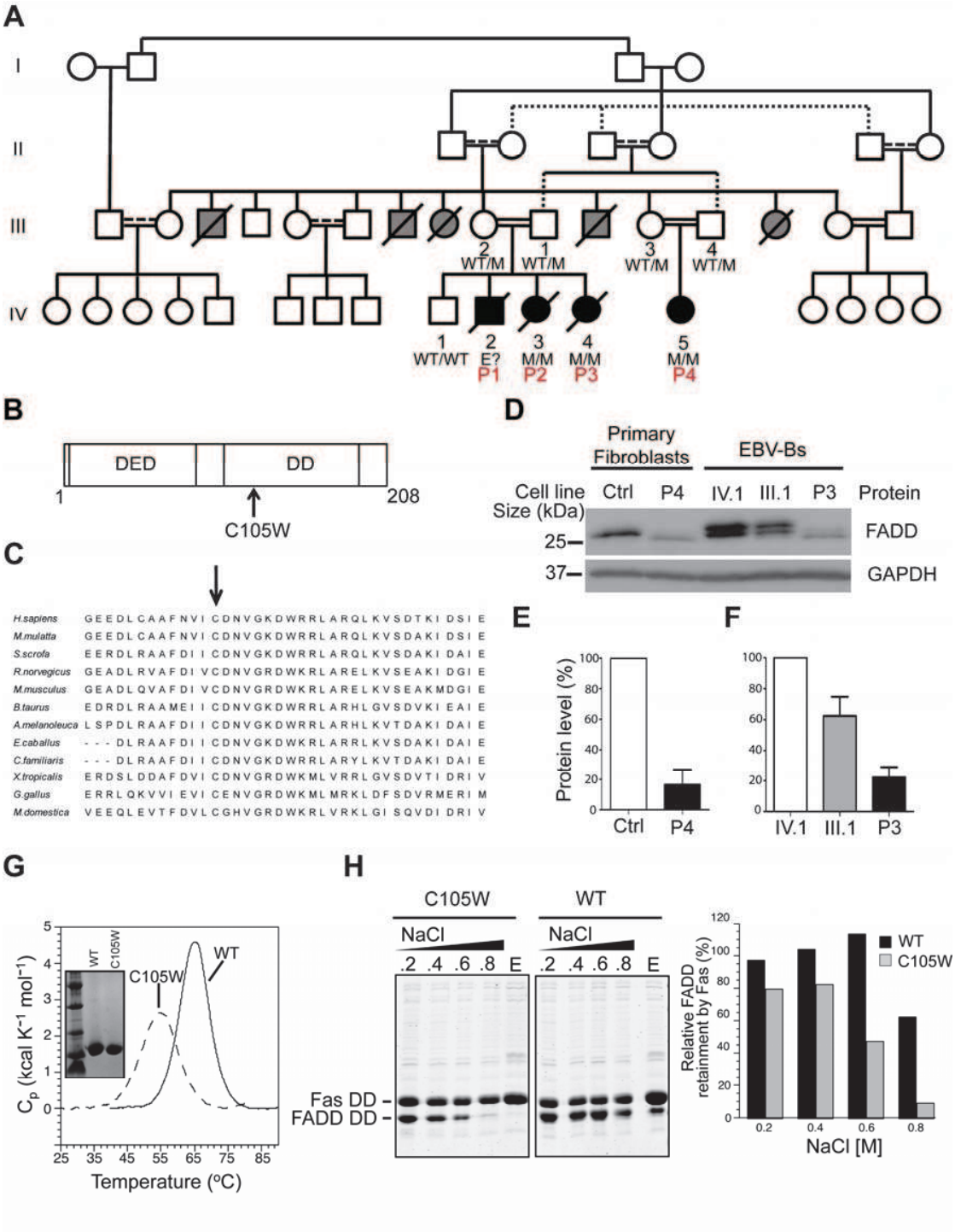
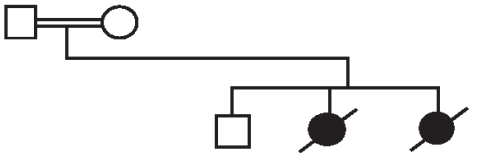


Figure 6: Haplotypes of microsatellites sequenced inside the two linked regions on chromosome 11 and chromosome 18.

18 microsatellites were sequenced in the chromosome 11 region and 2 in the chromosome 18 region. In bold, microsatellites with a homozygous haplotype. And inside frames are homozygous stretches found by linkage analysis. The black line is here to locate the position of FADD

Figure 6: Haplotypes of microsatellites sequenced inside the two linked regions on chromosome 11 and chromosome 18.



FADD mutation		C105 W/WT	C105W /WT	WT/WT	C105W /C105 W	C105W/ C105W	C105W /WT	C105W /WT	C105W /C105 W	
Individual		III : 1	III : 2	IV : 1	P2	P3	III : 4	III : 3	P4	
Marker	Physical position									
D11S4459	56514519	248-254	248-254	254-254	**248-248**	**248-248**	248-254	**248-248**	248-254	
D11S2006	59721888	317-321	317-321	317-317	**321-321**	**321-321**	317-321	317-321	317-321	
D11S4191	59999717					**100-100**		92-100	92-100	
D11S4076	61363135	148-154	**148-148**	148-154	**148-148**	**148-148**	**148-148**	**148-148**	**148-148**	
D11S1889	67313143	196-200	196-200	196-196	**200-200**	**200-200**	196-200	196-200	196-200	
D11S987	67893340	127-129	119-129	120-127	**129-129**	**129-129**	120-129	120-129	120-129	
D11S4087	68099158	**157-157**	**157-157**	**157-157**	**157-157**	**157-157**	**157-157**	**157-157**	**157-157**	
D11S4113	68765634	230-238	230-238	230-238	**230-230**	**230-230**	**230-230**	**230-230**	**230-230**	
D11S4136	69615711	187-189	178-189	178-187	**189-189**	**189-189**	178-189	178-189	**189-189**	*FADD*
D11S4196	70185224	200-228	200-228	228-228	**200-200**	**200-200**	200-228	200-228	**200-200**	
D11S4139	70504269	147-182	147-182	182-182	**147-147**	**147-147**	147-182	147-182	**147-147**	
D11S1314	72323249	**96-96**	**96-96**	**96-96**	**96-96**	**96-96**	**96-96**	**96-96**	**96-96**	
D11S2371	73505131		195-203	195-195	**203-203**	**203-203**	195-203	195-203	**203-203**	
D11S4081	75196447		185-187	187-187	**185-185**	**185-185**	185-187	179-185	**185-185**	
D11S4179	76396260	**235-235**	**235-235**	**235-235**	**235-235**	**235-235**	**235-235**	**235-235**	**235-235**	
D11S937	77854320	165-169	157-179	169-179	157-165	157-165	**157-157**	165-169	157-165	
D11S2002	79965403	235-243	231-235	235-235	231-243	231-243	**231-231**	235-243	231-243	
D11S901	81844785	**316-316**	320-324	216-320	316-324	316-324	**324-324**	316-324?	316-324	
D18S61	65587036	**215-215**	**215-215**	**215-215**	**215-215**	**215-215**	215-237	213-215	**215-215**	
ATA82B02	67064497	188-191	188-191	188-191	**191-191**	**191-191**	188-191	179-188	**188-188**	

Figure 7: Illumina sequencing reads displayed for patient P3.

Reads overlapping the mutation in exon 2 of FADD (bp position g.69,729,880 – g.69,729,949, hg18, NCBI 36.1) show the homozygous T>G base pair substitution.

Figure 7: Illumina sequencing reads displayed for patient P3.

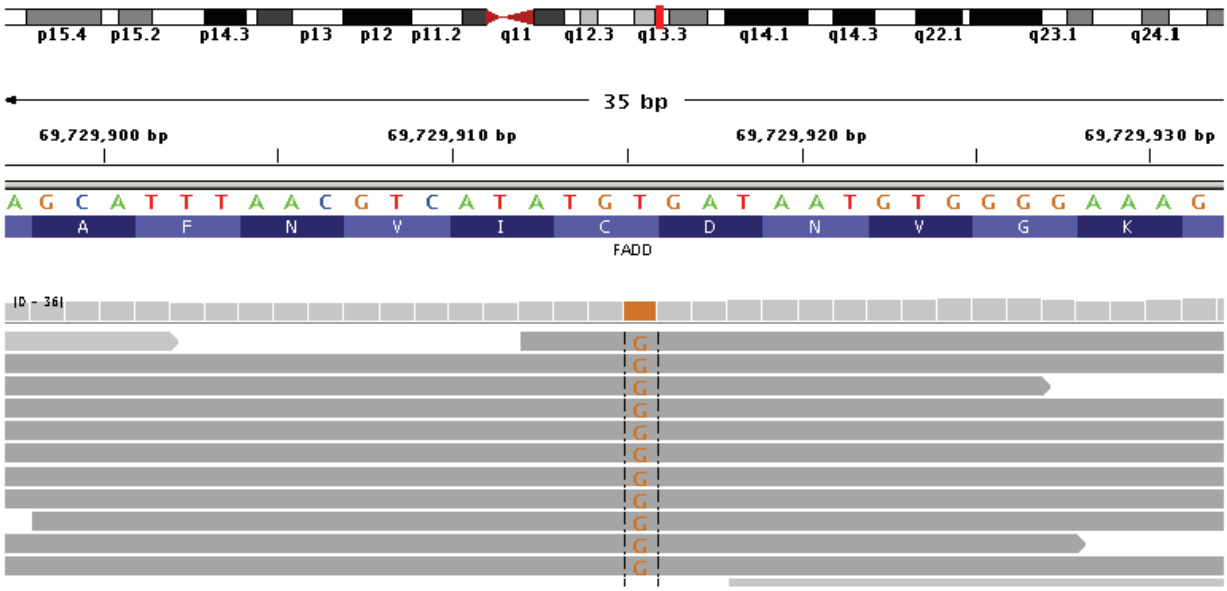


Figure 8: Characterization of the ALPS phenotype.

(**A**) Activated PBLs from normal volunteers (Ctrls), patient P4 and a heterozygous Fas deficient patient were stimulated with recombinant FasL (r-FasL) and cell viability was analyzed. The proportion of live cells in the presence of r-FasL is displayed as a percentage of live cells in the presence of enhancer alone, for each sample. (**B**) EBV-B cells from P3, IV.1 and a Fas-deficient patient were treated with an antibody against Fas (Apo-1.3), in the presence of rabbit anti-mouse Ig (Ramig), for 6 h and analyzed for early apoptosis. The percentage of cells that were apoptotic was calculated as the percentage of cells positive for AnnexinV and negative for propidium iodide staining. NS indicates that the cells were incubated with Ramig only. A mean of three independent experiments is shown. Error bars indicate the SEM. (**C**) FasL levels in the serum of P4 and eight healthy controls (Ctrls). (**D**) IL-10 levels in the serum of P4 and eight healthy controls. IL-10 values were normalized with respect to the mean of the control samples. (**E**) Proliferation assay. Proliferative responses to various stimuli (PHA, 5 µg/ml; IL-2, 500 U/ml; concanavalin A, 5 µg/ml; anti-CD3 (OKT3), 0.1 µg/ml; PMA, 50 ng/ml; ionomycin, 0.1 µg/ml). P3 and P4 responded similarly compared with the two contols: C1 and C2.

Figure 8: Characterization of the ALPS phenotype.

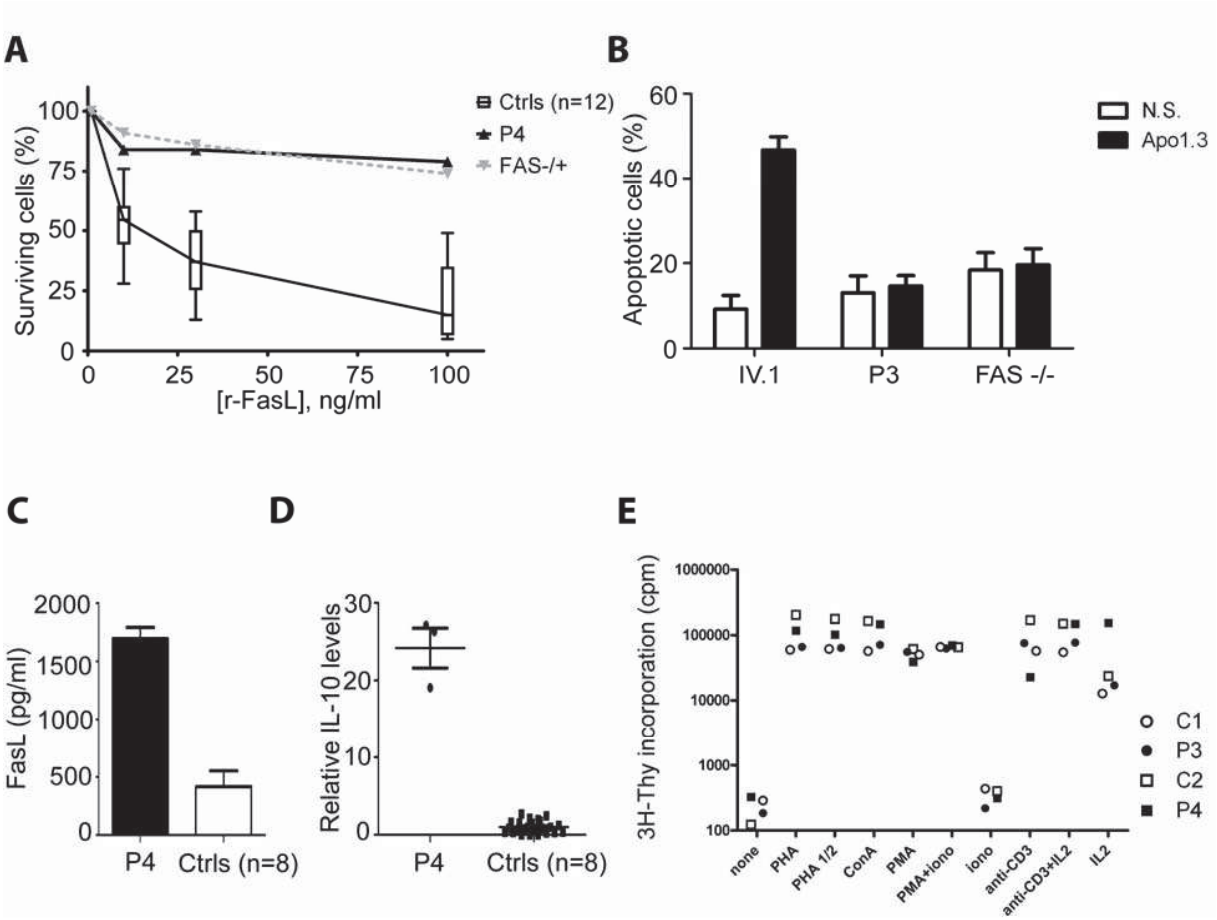


Figure 9: Impaired antiviral immunity.

(**A**) EBV-B cells from P3, IV.1, and a STAT1-deficient patient were infected with VSV at a multiplicity of infection (MOI) of 1, with or without prior treatment with 100 IU/ml of IFN-α for 24 h. Cell mortality was assessed by measuring the amount of LDH released into the medium. For STAT1-deficient cells, results are shown only until 36 h post infection, due to high levels of spontaneous LDH release at later time points. A mean of three independent experiments is shown. Error bars indicate the SEM. (**B**) EBV-B cells from P3, IV.1 and a STAT1-deficient patient were either left untreated (NT, bars in white) or treated with 100 IU/ml IFN-α for 24 h (IFN-α, bars in black), prior to infection with VSV at a MOI of 0.1, and were harvested 48 h post infection. VSV titers were determined on Vero cells. A mean of three independent experiments is shown. Error bars indicate the SEM. (**C**) EBV-B cells from P3 or IV.1 were infected with VSV at a MOI of 10. Cells were harvested at the indicated time points, and total RNA was extracted. *IFN-β* and *IRF7* mRNA levels were determined by quantitative RT-PCR. Threshold cycles normalized with respect to those of GUS (ΔC_T) are plotted as $2^{-\Delta CT}$. A mean of three independent experiments is shown. Error bars indicate the SEM.

Figure 9: Impaired antiviral immunity.

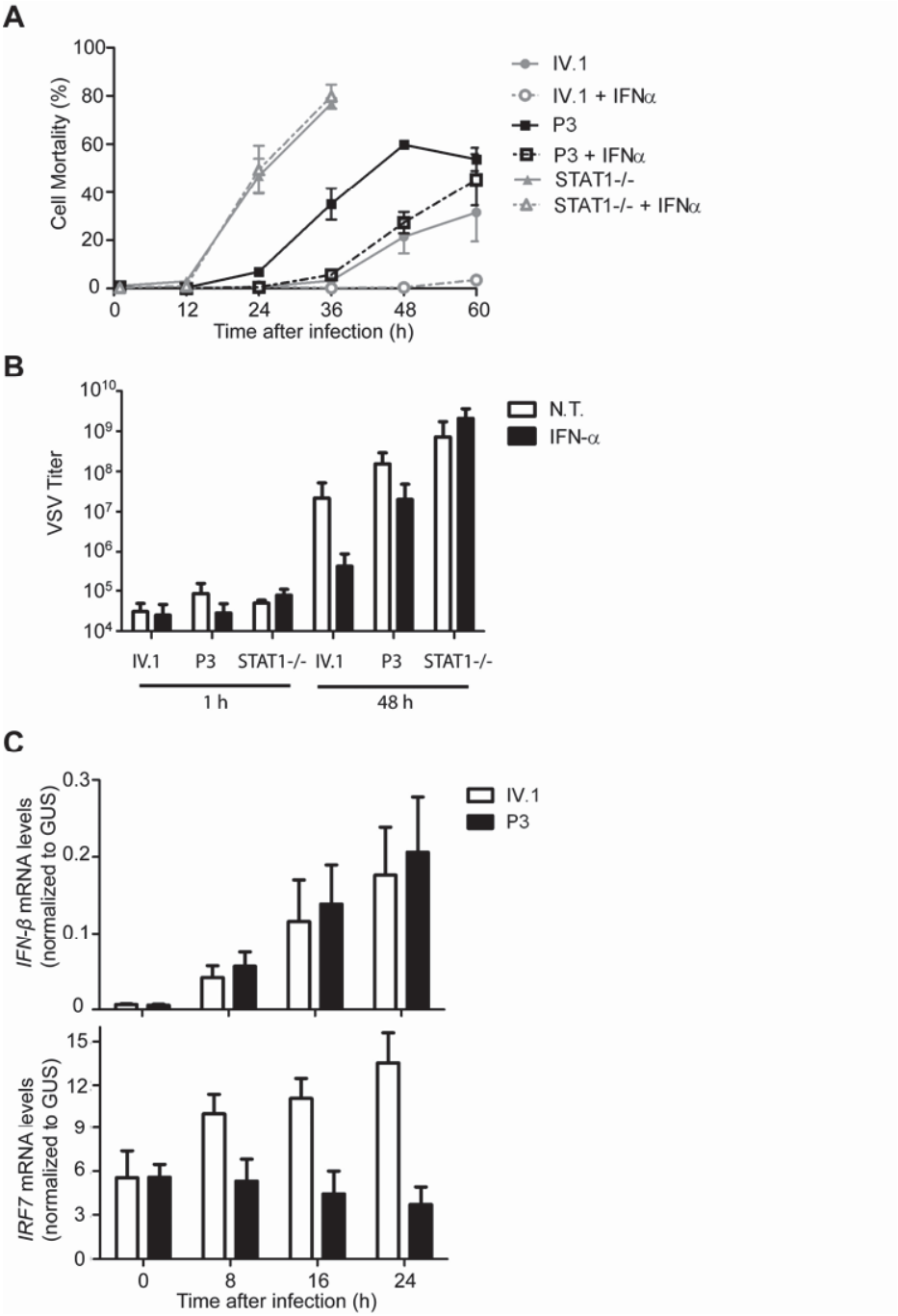


Figure 10: Histopathology of P2 (A-K, post-mortem, 14 months) and P4 (L, 22 months).

(**A**) Normal thymic architecture; increased cortical tingible body macrophages consistent with a stress response (P2, H&E stain, original magnification x100). (**B**) Normal-appearing reactive lymph node with follicular hyperplasia (P2, H&E stain, original magnification x40). (**C**) Low power view of the spleen showing expanded, congested red pulp with relatively reduced white pulp (P2, H&E stain, original magnification x20). (**D**) White pulp nodule containing a follicle with a reactive germinal centre, but no marginal zone hyperplasia, arising from a peri-arteriolar lymphoid sheath (P2, H&E stain, original magnification x200). (**E**) Immunohistochemistry for CD3 shows a normal distribution of CD3-positive T-cells in a white pulp nodule (P2, Ventana Benchmark XT immunostainer, monoclonal antibody LN10, original magnification x100). (**F**) Immunohistochemistry for CD20 shows a normal distribution of CD20-positive B cells in a white pulp nodule (P2, Ventana Benchmark XT immunostainer, monoclonal antibody L26, original magnification x100). (**G**) Congested, disorganized red pulp containing increased neutrophils consistent with sepsis (P2, H&E stain, original magnification x400). (**H**) Disorganized reticulin staining pattern of splenic red pulp. Inset: “barrel hoop” arrangement of sinusoidal reticulin fibres in a normal control spleen (P2, Gordon and Sweet’s reticulin stain, original magnification x400). (**I**) Immunohistochemistry for CD68 shows a haphazard arrangement of macrophages in the splenic red pulp. Inset: orderly distribution of cordal macrophages in a normal control spleen (P2, Ventana Benchmark XT immunostainer, monoclonal antibody PG-M1, original magnification x400). (**J**) Liver showing portal-portal linkage and expansion of portal tracts by a moderate chronic inflammatory cell infiltrate without interface activity (P2, H&E stain, original magnification x100). (**K**) Sirius red staining of liver shows collagen deposition (red) within portal-portal bridges. Elastin could also be demonstrated (not shown) consistent with established portal-portal fibrosis (P2, Sirius red / Fast green stain, original magnification x100). (**L**) Liver showing mild expansion of portal tracts by a chronic inflammatory cell infiltrate but no bridging fibrosis (P4, H&E stain, original magnification x200).

Figure 10: Histopathology of P2 (A-K, post-mortem, 14 months) and P4 (L, 22 months).

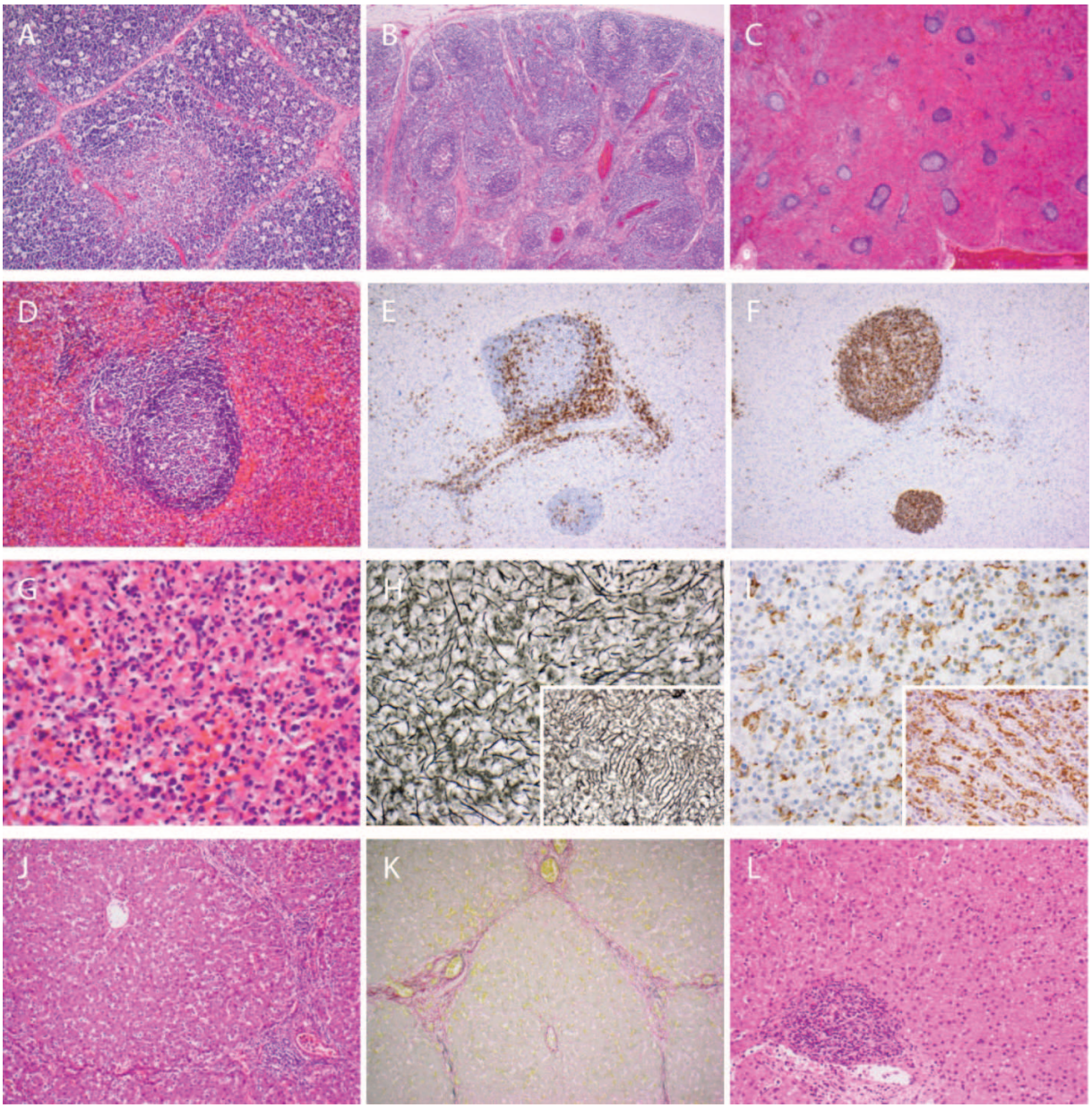


Table 3: Clinical features of four patients in the family.

Patient		P1	P2	P3	P4
gender		male	female	female	female
cardiovascular malformation		none	none	yes (pulmonary atresia) + ventricular septal defect	yes (L-sided SVC draining to left atrium)
functional hyposplenism	**spleen size**	NA	normal	normal	normal
	Howell-Jolly bodies	NA	NA	yes	yes
	invasive pneumococcal infection	yes (meningitis, 4 mo)	yes (septicemia, 14 mo)	no[a]	no[a]
features of febrile episodes	**documented viral trigger**	VZV		VZV HHV6	MMR vaccine Astrovirus Parainfluenza virus 2 EBV
	encephalopathy	yes	yes	yes	yes
	seizures	yes	yes	yes	yes
	liver dysfunction (maximal ALT, IU/L)	yes (NA)	yes (NA)	yes (1042)	yes (383)
ALPS phenotype	**autoimmune disease**	no	no	no	no
	presence of autoantibodies	NA	NA	no[b]	yes[b] anti-erythrocyte antibodies intermittently present
	lymphadenopathy or splenomegaly	no	no	no	no
	% of $CD4^-CD8^-TCR\alpha\beta^+$ T cells	NA	NA	NA	7.8
	lymphocyte apoptosis	NA	NA	NA	impaired
	serum FasL	NA	NA	NA	high
	serum IL10	NA	NA	NA	high
liver histology		NA	mild chronic portal inflammation with bridging fibrosis	NA	mild chronic portal inflammation with mild fibrosis
follow-up data		died at 4 mo	died at 14 mo	died at 4 yr and 4 mo.	alive and currently 2 yr and 9 mo.

NA, not assessed

[a] On antibacterial prophylaxis from early infancy

[b] Autoantibody screen including anti-nuclear, anti-smooth muscle, anti-mitochondrial, anti-gastric parietal cell antibodies, direct antiglobulin test.

Table 4: Whole-Exome sequencing results.

Total number of reads		24,989,074
Uniquely Mapped reads		17,816,126
Total variants called		23,146
Substitutions	Total (novel[a])	21,224 (1796)
	Synonymous	9,936
	Nonsense	71
	Missense	8,121
	Splice-site	37
	Non-coding RNAs	58
Indels		1,922
dbSNP rate[b]		91.54%
dbSNP concordance[b]		99.71%
chromosome 11 region	Total[c] (novel[a])	67 (8)
	Synonymous (novel[a])	26 (1)
	Nonsense (novel[a])	0 (0)
	Missense (novel[a])	15 (1)
	Splice-site (novel[a])	0 (0)
	Indels	0
Chromosome 18 region	Total[c] (novel[a])	14 (0)
	Synonymous (novel[a])	6 (0)
	Nonsense (novel[a])	0 (0)
	Missense (novel[a])	8 (0)
	Splice-site (novel[a])	0 (0)
	Indels	0

[a]Number of variants not found in dbSNP129, 1000 Genomes and our in-house database of 70 exomes.

[b]dbSNP rate and concordance are based on dbSNP129.

[c]Homozygous variants.

CHAPTER 5

MUTATION IN NKX2-5 IDENTIFIED IN ONE KINDRED WITH ICA

I was now ready to use WES on the ICA cohort. I thus started to investigate the most informative family of the cohort, or at least the one I have the most gDNA samples: family B.

5.1 Identification of a mutation in NKX2-5 in 3 patients from one kindred

Family B is an African kindred (Family B, **Figure 4**) with 3 ascertained cases (I.1, II.4 and II.5, **Figure 11A**) and 3 probable cases (II.1, II.2 and II.3, **Figure 11A**) of ICA. I hypothesized that ICA segregated as a fully penetrant, autosomal dominant (AD), Mendelian trait in this family. I analyzed the genomic DNA (gDNA) of the 3 ascertained cases by WES. The number of reads and the exome coverage metrics (**Table 5**) show inferior quality of the assembly for the exome of patient II.4 compared to the 2 other patients (42% of target bases covered at 10X compared to 72% and 71%), likely due to the suboptimal quality of II.4 gDNA, which was extracted from necropsy samples. Therefore I selected candidate variations from the annotated data, focusing on nucleotide substitutions, insertions or deletions (indels) either present in all three patients (I.1, II.4 and II.5), or present in I.1 and II.5 and not covered by WES in II.4. The breakdown of the variants identified by WES in each patient is in **Table 6**. After filtering out known polymorphisms reported in the National Center for Biotechnology Information (NCBI) dbSNP build 134, the 1,000 Genomes Project database (http://www.1000genomes.org), and our own database of 400 exomes at that time, I identified only 32 variants that could potentially underlie ICA in this family (**Table 7**).

Strikingly, a substitution in *NKX2-5* was one of the 32 candidate variants identified. The other 31 variants did not affect genes known to be involved in spleen development and/or heterotaxy, whether in humans or in the mouse model (**Tables 2 and 7**). The variant identified in *NKX2-5* is a missense heterozygous c.707C>A in protein-coding exon 2 of *NKX2-5* (NM _004387.3, MIM600584) that changes Proline at amino acid position 236 to Histidine, p.P236H (referred to as P236H hereafter) (**Figure 11 B, C**). I validated this variant by Sanger sequencing of gDNA from peripheral blood and of cDNA from EBV-transformed B cells (EBV-B) from ICA patients. This variant segregated with ICA in all family members examined, as all patients and none of the

healthy relatives were heterozygous for P236H (**Figure 11A**). The variant was not found in 1,052 additional healthy individuals from 52 ethnic groups from the Centre d'Etude du Polymorphisme Humain and Human Genome Diversity panels, including 127 samples from Sub-Saharan Africa. Moreover, there were 200 samples from Sub-Saharan Africa in the 1,197 samples sequenced by the 1,000 Genomes project that we used to filter out polymorphisms. Altogether, these results suggest that the P236H variant is a rare and possibly ICA-causing variant rather than an irrelevant polymorphism, since it was not found in a total of 2,261 control individuals of diverse ethnic backgrounds. Lastly, Proline at amino acid position 236 is well-conserved throughout evolution, although an Alanine is present at this position in *Mus musculus* and *Rattus norvegicus* (**Figure 11C**). Notably, there is no Histidine at this position in any species in which *Nkx2-5* has been sequenced. In parallel to WES, I examined copy number variants (CNV) in II.5 throughout the genome, using the Affymetrix 6.0 chip. I did not observe any CNV larger than 50 kb that was not present in the DGV database (http://projects.tcag.ca/variation/) or our own database of 150 samples. Overall, these genetic data suggested that P236H might underly ICA in this multiplex kindred

5.2 Nkx2-5 is involved in spleen development in the mouse model

We had already identified *NKX2-5* as a candidate gene because one patient with a heterozygous frameshift deletion at the N-terminus of *NKX2-5* presented both congenital heart disease and asplenia[104]. However it is the work of Matthew Koss, a PhD student in the laboratory of Licia Selleri at Cornell Medical School that made this result particularly interesting. In the mouse model, Nkx2-5, like Tlx1, is detected in E10.5 splenopancreatic mesenchyme and in spleen progenitors thereafter[46]. Nonetheless its role in spleen organogenesis was unclear due to early in utero lethality of null mutants[105]. Different Nkx2-5 partial loss-of-function (LOF) alleles including Nkx2-$5^{+/-}$ mice[105], Nkx2-$5^{Cre/GFP}$ hypomorphic embryos[106], and Nkx2-$5^{Y\text{-}A:IRESLacZ/+}$ wild-type chimeras, a dominant negative model of Nkx2-5 deficiency conferred by mutation of a conserved tyrosine-rich domain[107], showed hyposplenia (**Figure 12**). These mouse models demonstrated that Nkx2-5 dosage and mutations in different parts of the protein yield spleen hypoplasia. It was thus interesting to look whether the mutation identified in

NKX2-5 in patients from family B had an impact on the protein or its function and could be disease-causing.

5.3 Functional characterization of the NKX2-5 mutation

Because POLYphen II[108] predicted that the P236H mutation is benign and residue 236 is outside the homeodomain of NKX2-5, we hypothesized that the mutation may not impair the production of the protein or its binding to DNA. Indeed, Western blot analysis and EMSA assays confirmed that NKX2-5 P236H is normally produced and binds to DNA similarly to WT protein[46]. Instead, we hypothesized that the mutation may disrupt transactivation by NKX2-5 likely through interaction with essential spleen-specific cofactors. Notably, P236 lies immediately adjacent to the first of nine Tyrosines that define a conserved Tyrosine-rich domain (YRD; residues 237-275) within NKX2-5 (**Figure 11C**), which previous work of our collaborators established as a critical domain for its *in vivo* function, as well as transcriptional activity in a heterologous context[107]. Significantly, mouse chimeras composed partly of mutant cells in which critical Tyrosines in the Nkx2-5 YRD have been replaced by Alanines[107], develop hypoplastic spleens (**Figure 12F**). Thus, we posited that the P236H mutation could impact the function of the YRD, potentially affecting the interaction of NKX2-5 with splenic cofactors.

In order to test the effects of the P236H mutation on the transactivation potential of Nkx2-5, we employed a heterologous system that we previously used to study Nkx2-5 function in mouse heart development[107]. The C-terminal domain of human NKX2-5, containing either the WT or the mutant sequence, was fused to the Gal4 DNA-binding domain (DBD), and co-transfected into HEK293T cells with a Luc reporter (pGL4.31) downstream of multiple UAS elements. While the WT construct induced an increase in Luc expression of about 30-fold *versus* control, the construct bearing the P236H mutation induced significantly lower expression (about one-third of WT) (**Figure 13A**). Notably, mouse constructs bearing mutations of critical Tyrosines within the YRD domain showed low activity in this assay, consistent with our previous report[107]. Additionally, since *Nkx2-5* is subject to autoregulation in heart development[106], we tested whether the murine *Nkx2-5* spleen and stomach enhancer[46] could be

transactivated in spleen mesenchyme by full-length WT and/or mutant NKX2-5. We transfected SPCL2 cells, a cell line derived from embryonic mouse spleen, with expression vectors for WT or P236H *NKX2-5*, with or without *Tlx1*, which plays critical role in mouse spleen development as mentioned earlier[57]. *NKX2-5* WT co-transfected with *Tlx1* activated the *Nkx2-5* Luc reporter approximately 2.5-fold above control, whereas this transactivation was abolished when using the mutant *NKX2-5* P236H (**Figure 13B**). The same *Nkx2-5/Tlx1* co-transfection experiment performed with HEK293T cells did not lead to activation of the *Nkx2-5* Luc reporter, highlighting that this co-operative transactivation of *Nkx2-5*, as well as its requirement for P236, are spleen-specific. Altogether, these results demonstrated that the P236H mutation impairs transactivation by NKX2-5, both in a heterologous transactivation assay and in the context of spleen mesenchymal progenitors. Given the findings in our mouse models, in which reduced *Nkx2-5* gene dosage or YRD function yield hyposplenia, and the observed functional deficiency of the *NKX2-5* P236H mutation identified in ICA patients, we concluded that Nkx2-5 plays a central role in the development and growth of the spleen in mice and likely in humans.

The basic knowledge acquired in the mouse has helped define the etiology of ICA in one kindred. Indeed, Pbx target genes identified in the mouse spleen anlage in the study done in the laboratory of Licia Selleri guided the analysis of WES. This approach led to the identification of a novel missense mutation in *NKX2-5*, which is a central component within the Pbx-directed module in the mouse. This novel *NKX2-5* variant segregates with ICA in all family members examined, and it is not present as a polymorphism in large and diverse human populations, including Sub-Saharan ethnic groups. Notably, the P236H mutation significantly diminished transactivation by *Nkx2-5* in both a heterologous transcriptional assay and in cultured spleen cells. In the latter assay, transactivation occurs only in the presence of the spleen-specific cofactor Tlx1, suggesting that in this context spleen-specific interactions are disrupted by the P236H mutation. Interestingly, all known congenital heart disease (CHD) patients with previously-identified *NKX2-5* mutations do not display asplenia (Harvey et al., 2002), indicating that the P236H mutation would preferentially compromise spleen development. Moreover the subsequent Sanger sequencing of all exons of *NKX2-5* in

all other ICA patients of our cohort did not lead to the identification of another novel mutation in this gene. Together these results suggested that a mutation in *NKX2-5* might lead to ICA in humans but that other mutations in at least one other gene were likely to be the major genetic etiology of ICA.

Figure 11: Genetics of an inherited *NKX2-5* Mutation in ICA patients.

(**A**) Pedigree of the family. Black symbols indicate ICA cases. Gray symbols indicate siblings who died early in childhood with clinical symptoms similar to those of II.4. Haplotypes of *NKX2-5* are indicated: M stands for c.707C>A, WT for wild-type allele, E? for unknown haplotype. (**B**) Illumina sequencing reads displayed for patient I.1. Reads are reversed. Reads overlap the mutation in exon 2 of *NKX2-5* (bp position g.172,659,827 – g.172,659,852 , hg19). (**C**) Schematic diagram of the NKX2-5 protein (drawn to scale) showing conserved and/or functional domains (homeodomain: HD; tinman/Nkx2-5 domain: TN; NK2 specific domain: NK; tyrosine rich domain: YRD). Two other conserved domains, the NKX2-5 box and the GIRAW motif, are indicated by yellow and green boxes, respectively. Evolutionary conservation of the NKX2-5 region containing amino acid P236 is represented below.

Figure 11: Genetics of an inherited *NKX2-5* Mutation in ICA patients.

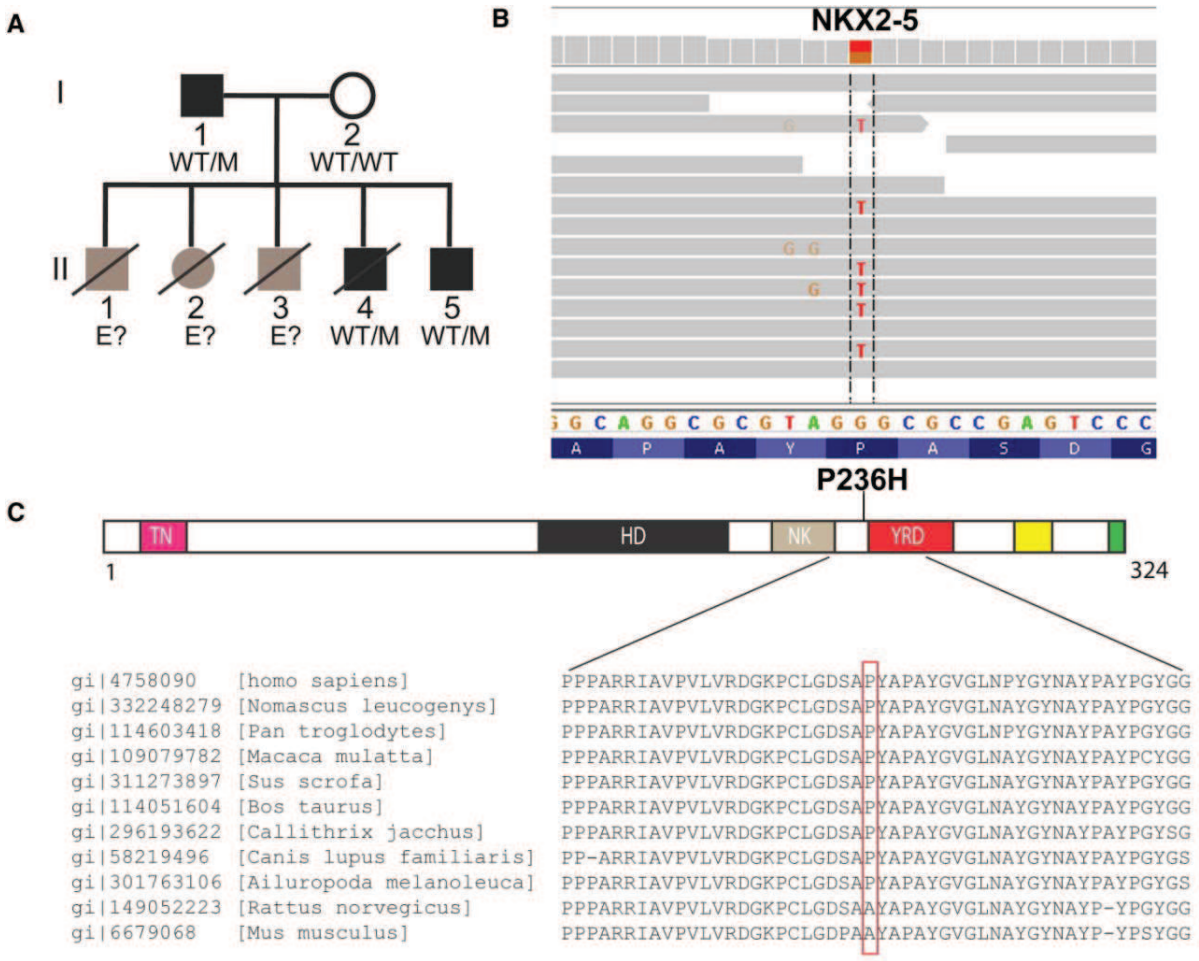


Figure 12: Reduced Nkx2-5 levels lead to hyposplenia in the mouse model.

(**A-D**) Reduced Nkx2-5 levels in Nkx2-5$^{+/-}$ (**B**), and Nkx2-5$^{Cre/GFP}$ (**D**) mice result in smaller spleens versus controls (**A** and **C**: P0 and E17.5). (**E-F**) Depicted are spleens taken from *Nkx2-5*$^{Y\text{-}A:IRESlacZ/+}$↔wildtype chimaeric mouse pups at P0[106]. Chimaeric spleens (right panel), with reduced Nkx2-5 function, are smaller and dismorphic, compared to WT (left panel).

Figure 12: Reduced Nkx2-5 levels lead to hyposplenia in the mouse model.

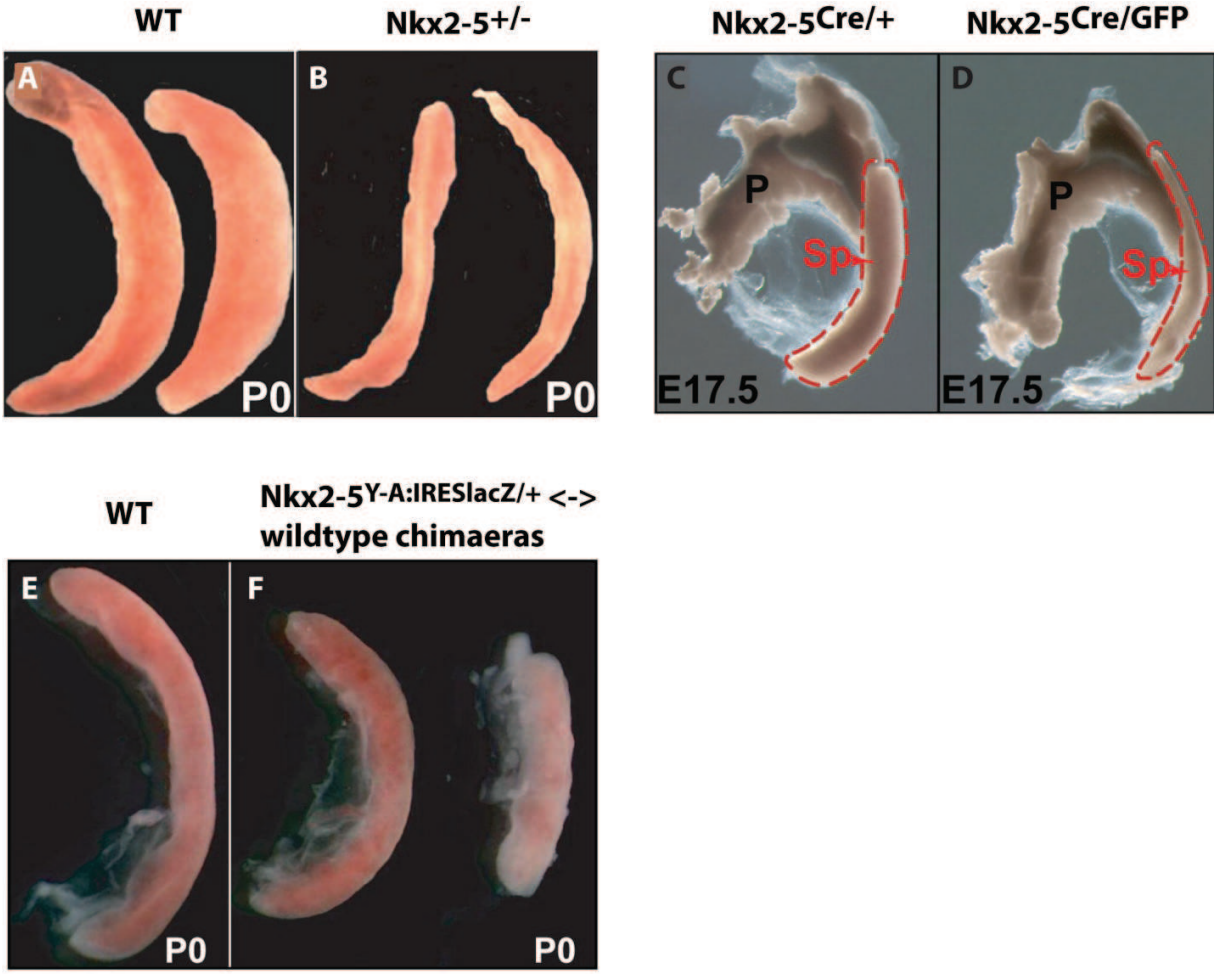


Figure 13: P236H mutation impairs NKX2-5 functional activity.

(**A**) HEK293T cells are transfected with a UAS-Luc reporter and different expression vectors consisting of the Gal4 DNA-binding domain fused to the C-terminus of human NKX2-5. NKX2-5 WT: WT NKX2-5 C-terminus. NKX2-5 P236H: the C-terminus of NKX2-5 with the P236H mutation. NKX2-5 Y->A and NKX2-5 Y->F: fusion constructs in which all Tyrosines in the YRD have been substituted with Alanines or Phenylalanines respectively[107], used as negative controls. Data represent average from two independent experiments. Error bars indicate S.E.M. (**B**) SPCL2 cells are transiently transfected and co-transfected with *Tlx1* expression vector, a Luc reporter containing the *Nkx2-5* spleen-stomach enhancer (pGL3-Nkx2-5-Luc) and different human NKX2-5 expression vectors. Data represent average from four independent experiments. Error bars ± S.E.M.

Figure 13: P236H mutation impairs NKX2-5 functional activity.

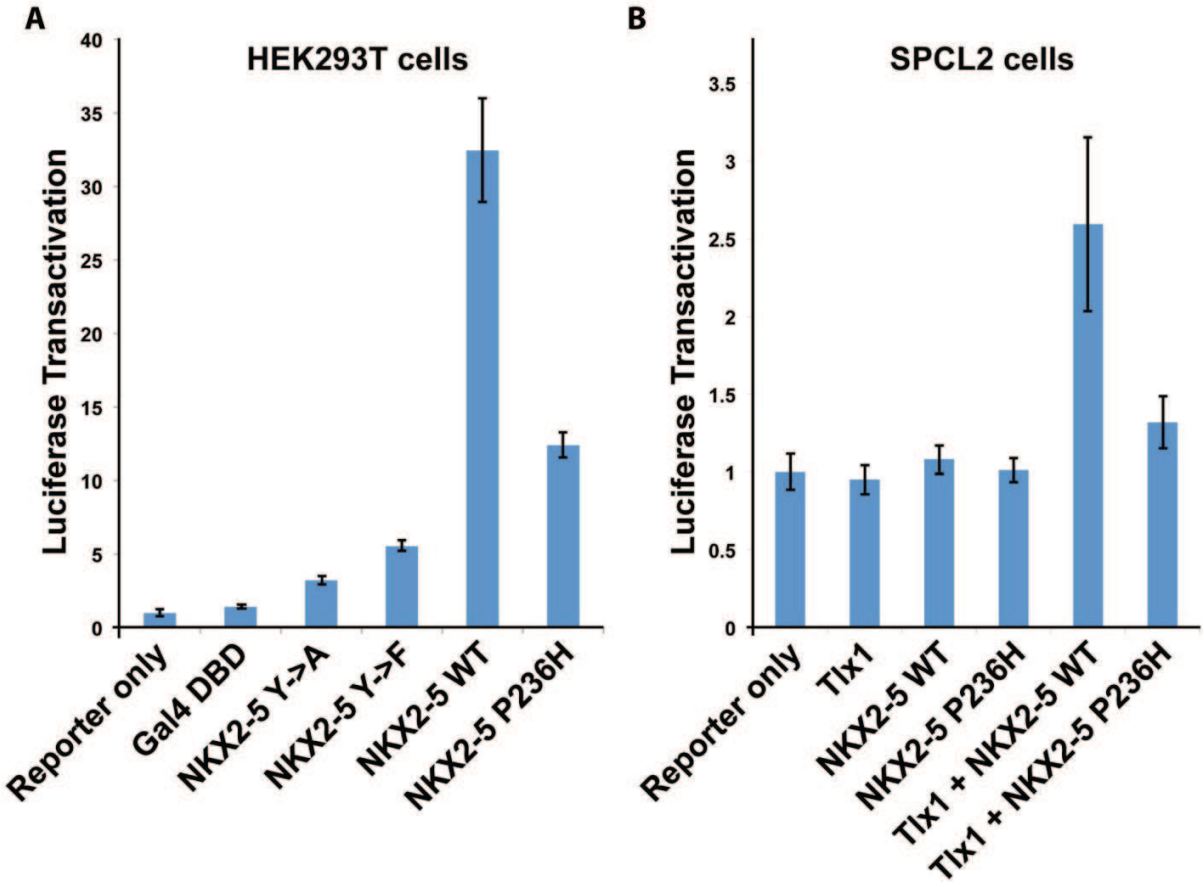


Table 5: Exome sequence alignment and coverage data for family B.

Metric	I.1	II.5	II.4	Legend
Genome size	3,095,677,412	3,095,677,412	3,095,677,412	The number of bases in the reference genome used for alignment
Bait territory	38,925,418	38,925,418	38,925,418	The number of bases which have one or more baits on top of them
Total reads	40,784,315	37,764,604	38,145,467	The total number of reads in the SAM or BAM file examine
PF unique reads	35,608,020	33,530,320	31,969,422	The number of reads that are not marked as duplicates and passed quality filters (PF)
On target bases	870,807,186	791,839,532	372,842,179	The number of PF aligned bases that mapped to a targetted region of the genome
Mean target coverage	23.13	20.95	9.98	The mean coverage of targets that recieved at least coverage depth = 2 at one base
Pct target bases 2X	92	93	86	The percentage of ALL target bases acheiving 2X or greater coverage
Pct target bases 10X	72	71	42	The percentage of ALL target bases acheiving 10X or greater coverage
Pct target bases 20X	48	43	12	The percentage of ALL target bases acheiving 20X or greater coverage
Pct target bases 30X	29	23	2	The percentage of ALL target bases acheiving 30X or greater coverage

Table 6: Breakdown of variants identified by WES of 3 patients from family B.

Variants	I.1	II.5	II.4	I.1 & II.5	I.1 & II.5 - dbSNP[5]	I.1 & II.5 - dbSNP - 1k genomes[6]	I.1 & II.5 Novel[7]	I.1 & II.4 & II.5 Novel[8]
TOTAL [1]	24,630	25,009	17,956	16,881	484	184	101	48
Substitutions	23,932	24,367	17,541	16,480	415	162	97	47
Synonymous	10,116	10,402	7,955	7,296	109	21	18	7
Nonsense[2]	104	103	74	71	9	3	1	0
Missense	8,849	8,975	6,923	6,261	217	54	44	24
Splice-site[3]	2,240	2,268	1,594	1,526	49	40	15	8
Non-coding RNAs	262	230	186	139	7	3	1	0
UTRs	3,086	3,122	1,379	1,690	54	41	18	9
Indels[4]	698	642	415	401	69	22	4	1

[1]: Heterozygous or homozygous variations are included. Variations that mapped in intergenic, or intronic regions that were not ranked as splice-site regions were excluded.

[2]: All variations leading a stop gain or a start/stop loss were included in this category.

[3]: All variations within 1-3 bp into exon or 1-8 bp into intron.

[4]: Frameshift and non-frameshift insertions or deletions are included.

[5]: Not found in the NCBI dbSNP build 134.

[6]: Not found in the dbSNP 134 or in the last release of the 1,000 genomes project (July 2011) including 1,197 whole genomes.

[7]: Not found in our in-house database of 400 exomes or in the dbSNP 134 or in the last release of the 1,000 genomes project.

[8]: Novel variants identified in I.1, II.5 and II.4 or I.1, II.5 and with 2 or less reads at this position for II.4.

Table 7: List of non-reported variations shared by I.1, II.4 and II.5

Chr.	Position (hg19 build)	Allele	Gene	Accession no.	Functional impact
1	24417388	C/T	*MYOM3*	NM_152372.3	Missense (p.R444Q)
1	226555934	T/C	*PARP1*	NM_001618.3	Missense (p.K748R)
2	196791265	T/C	*DNAH7*	NM_018897.2	Missense (p.Y1166C)
3	52823663	T/C	*ITIH1*	NM_0022152.2	Splicing
3	65365233	T/C	*AC121493.1*	NM_001033057.1	Splicing
3	119133310	A/T	*ARHGAP31*	NM_020754.2	Missense (p.D845V)
4	190947127	A/C	*FRG2*	NM_001005217.1	Missense (p.D142E)
5	9054321	G/C	*SEMA5A*	NM_003966.2	Missense (p.T856R)
5	112227446	G/A	*ZRSR1*	ENST00000391338	Missense (p.R37Q)
5	115336788	A/T	*AC034236.1*	NM_173800.4	Missense (p.I558F)
5	128302174	A/G	*SLC27A6*	NM_014031.3	Missense (p.E115G)
5	172659840	G/T	*NKX2-5*	NM_004387.2	Missense (p.P236H)
6	26468409	A/G	*BTN2A1*	NM_007049.2	Missense (p.S406G)
6	40400032	C/T	*LRFN2*	NM_020737.1	Missense (p.R274H)
6	42626002	A/T	*UBR2*	NM_015255.1	Missense (p.R1003W)
8	141675010	C/T	*PTK2*	NM_005607.3	Splicing
9	114825283	G/A	*SUSD1*	NM_022486.3	Missense (p.T595M)
10	95072806	C/A	*MYOF*	NM_133337.2	Missense (p.A1941S)
10	105815761	G/A	*COL17A1*	NM_000494.3	Missense (p.A489V) and Splicing
11	3111787	C/T	*OSBPL5*	NM_020696.3	Splicing
11	17191035	T/C	*PIK3C2A*	NM_002645.2	Missense (p.K85R)
11	123813978	C/T	*OR6T1*	NM_001005187.1	Missense (p.G190R)
12	82750746	T/C	*CCDC59*	NM_014167.3	Missense (p.T153A)
13	46425657	AC/TA	*SIAH3*	NM_198849.2	Missense (p.C36L)
14	21109553	T/C	*OR6S1*	NM_001001968.1	Missense (p.I100V)
14	65209852	G/A	*PLEKHG3*	NM_015549.1	Missense (p.G975S)
15	74554773	C/T	*CCDC33*	NM_025055.3	Splicing
16	1995622	G/T	*RPL3L*	NM_005061.2	Splicing
17	5085443	ACT/-	*ZNF594*	NM_032530.1	Deletion S703
17	27861129	G/C	*TAOK1*	NM_020791.1	Splicing
19	2111779	G/A	*AP3D1*	NM_003938.5	Missense (p.R884W)
22	50705585	C/T	*MAPK11*	NM_002751.5	Missense (p.V156M)

CHAPTER 6

HAPLOINSUFFICIENCY OF RPSA CAUSES ICA

6.1 Sequencing of all the exomes of the cohort

We thus sequenced 28 additional exomes for all the patients in our cohort for whom we had enough gDNA (**Figure 4**). In total we had the exome of 31 patients. The first 16 exomes (3 of them from family B) were sequenced with the Agilent Human All Exon 38Mb kit. The remaining exomes were sequenced using the newer kit: the 50Mb kit. Because all the patients shared the same clinical phenotype, we hypothesized that different mutations in one other gene than *NKX2-5* would be disease-causing in at least two kindreds. We thus compared and analyzed the sequencing data of all the exomes (n=31) together and looked for candidate disease-causing genes mutated in multiple patients. Firstly, we decided to analyze all the patients as if they were unrelated because the number of reads and coverage metrics for each exome highlighted a significant difference of quality between the two Agilent kits (**Table 8**), as well as a lower quality for the two patients for whom gDNA was obtained from necropsy samples (patient JL77 and JL423, **Figure 4**). Therefore the WES might pick up a variation in one member of one kindred and not the other one simply because the coverage was not identical in the 2 exomes. This is why we decided not to consider any relationship between samples at first. Secondly, due to the severity of the phenotype and the rarity of ICA, about 1/1,000,000 births[15], we hypothesized that the disease-causing variants would have an impact at the protein level and would not be common polymorphisms. Any polymorphism with a frequency >0.1% would not fit with the frequency of the disease, even with an AR mode of inheritance, if we hypothesize that the penetrance of the trait is high. Therefore we filtered out all variants that were present in the NCBI dbSNP 134 database, or in the database of the 1000 genomes project, or in our in-house database of 1000 exomes. We also filtered out the variants that were intronic, located in the untranslated regions or synonymous and were not in essential splice sites. Finally we crossed the results of the 31 exomes and looked at genes carrying novel variants in more than one patient. Fourteen genes were mutated in more than 6 patients (**Table 9**).

6.2 Identification of heterozygous mutations in RPSA

Some genes, in particular the very big genes, have statistically more chance to carry novel variations. Therefore we wanted to test if the high number of variations observed in *RPSA* (MIM #150370) or *TTN* (MIM #188840) were statistically significant. Ideally we would need to compare the number of novel variants in one gene in our cohort and the number of novel variants in the same gene in a control cohort. As a test experiment, I performed the exact same analysis I did for the ICA patients on exomes from 6 other projects in the laboratory. I selected projects of a similar sample size and projects where the patients do not present any invasive pneumococcal disease as some of these might present an undiagnosed ICA. Strikingly, the majority of the genes frequently mutated in ICA were also frequently mutated in the other projects confirming the bias for some genes (**Table 9**). The gene with the highest number of novel mutations in the ICA cohort, and that was not frequently mutated in the other cohorts is *RPSA*.

Even though this analysis validated our hypothesis, a more mathematical and systematic method to evaluate the significance of the results was needed. In particular for smaller projects, or projects where one gene does not clearly stand out, a method to calculate the probablility of one gene to be mutated in X samples among Y samples was needed. Iuliana Ionita-Laza from Columbia University developed an algorithm that would calculate such probablility[109]. The algorithm would rank the genes according to two main factors. The first factor is the number of novel variations in one gene in the 'cases' exomes. The second factor depends on the number of novel variations in one gene in the 'cases' exomes vs the number of novel variations in the same gene in the 'control' exomes. The more 'cases' and 'controls' you have, the more powerful this method is. I tried to use this joint-rank approach using 650 exomes from other projects in the laboratory as controls (I excluded again all patients suffering from IPD as these might have an undiagnosed ICA). The first results did not show anything significant due to the high number of false positive variants left in our analysis. I used the raw sequencing results to perform this analysis. Therefore variants in introns would also be taken into account. Removing these intronic variants in *RPSA* for our cohort and the

control cohort led to a very significant result for *RPSA* (**Figure 14A**). However a new analysis and new calculations using the algorithm will be done as soon as the raw data from the WES are realigned, annotated and the variants called again using a much more stringent filter. Given the fact that we now have a large number of samples it is probably better to have false negatives, i.e. mutations that are missed because not called, rather than false positives, i.e. mutations that are called although they are merely alignment errors.

RPSA being the highest ranked gene in our empirical statistical approach, we validated the mutations in *RPSA* by Sanger sequencing. We then resequenced the coding exons of *RPSA* in all patients. In total 16 patients from 8 kindreds, including the patients carrying the mutation in *NKX2-5*, were mutated in the coding region of *RPSA* (**Figure 14B**). Not all of the mutations were identified by WES because the coverage of *RPSA* in the 38Mb kit was poor due to the numerous pseudogenes. The advent of paired-end sequencing increased the coverage of this gene and enabled the identification of mutations in the recently sequenced ones. 7 different mutations, not irrelevant polymorphisms, were identified including one nonsense mutation (p.Q9X) and one frameshift insertion (p.P199Sfs) (**Table 10**). The five missense mutations affected residues that were conserved in mammals, plants, and prokaryotes (**Figure 14C**). We confirmed that the mutations followed the familial segregation in all kindreds (**Figure 14B**). Strikingly the mutations appeared *de novo* in the 2 sporadic cases, and the p.T54N mutation probably resulted from germline mosaicism from the father or the mother in kindred F. Microsatellite analysis, or HLA typing confirmed familial segregation in these families. Taken together these results strongly suggest that heterozygous mutations in *RPSA* lead to ICA in humans, with a complete penetrance, and in an autosomal dominant way.

6.3 The combination of non-coding mutations in *RPSA* leads to ICA in two multiplex families

We continued our genetic dissection of ICA by searching for non-coding mutations in *RPSA* in the patients not yet mutated in this gene. We hypothesized that novel mutations (i.e. not reported and not found in controls) in the 5'-UTR, 3'-UTR,

introns or promoter region of *RPSA* might lead to ICA. Sanger sequencing of these regions led to the identification of two novel non-coding mutations. One mutation is a substitution of a C to a T in the 5'-UTR (c.-48C>T, 3:39448242) in one family with one ICA patient and a possible ICA patient. The other novel non-coding mutation identified is a substitution in the intron 5 base pairs after exon 1, c.-35+5G>C (**Table 11**, **Figure 15A**). This position is at the donor site and might affect the splicing of exon 1.These two mutations were not found in any database. For example they were not reported in the 1000 genomes browser on 09/02/2012. None of the 2178 samples sequenced carried these mutations. These mutations were confirmed by Sanger sequencing and the other family members were sequenced. Familial segregation revealed that one healthy individual carried the splice mutation whereas two healthy individuals carried the substitution in the 5'-UTR (**Figure 15B**).

All patients carrying a heterozygous coding mutation in *RPSA* presented the disease (complete penetrance) (**Figure 14B**), indicating that environmental factors had a small impact on the onset of the disease, at least in the individuals mutated in *RPSA*. Therefore we hypothesized that a modifier variation or polymorphism in *RPSA* might be required in addition to the non-coding mutations to lead to ICA in patients from families G and H. In other words these non-coding mutations might reduce the function or expression of RPSA by 40%, which would not be disease-causing. However if there is another polymorphism that might reduce the expression of RPSA by 10%, the addition of these two variants would be disease-causing. We thus sequenced and listed all the polymorphic variants of *RPSA* in all members of family G and H. Strikingly the 20bp deletion (rs139686560 or rs199844419) upstream of exon 1 of RPSA segregates well with the disease when combined with the splice mutation or the 5'-UTR mutation (**Figure 15B**). Moreover the frequency of this 20bp deletion is 0.006 (13/2178) according to the 1000 Genomes browser (www.ncbi.nlm.nih.gov/variation/tools/1000genomes/). This frequency is too high for the deletion to be ICA-causing alone but is low enough that it could play a role in the disease.

Finally the mutations need to have an impact on RPSA either at the mRNA level or the protein level in order to be disease-causing. We first tested the impact of the splice mutant. We extracted mRNA from EBV-B cells from the members of family G, made cDNA and sequenced it. The sequences reveal the presence of a splice variant at the mRNA level (**Figure 15C**). This variant has an extra 43bp at the end of exon1. The intensity of the peaks suggests that about 40% of the isoforms of RPSA in the patients from family G have this extra 43bp in the 5'-UTR. However a precise quantitation needs to be done using the TA cloning technique. Deciphering the impact of these extra 43bp in the middle of the 5'-UTR, as well as the impact of the 20bp deletion upstream the 5'-UTR, and the impact of the mutation in the 5'-UTR in family H will be necessary. They might also give more insights in the pathogenesis of ICA. All together we identified novel mutations in *RPSA* in 19 ICA patients from 10 unrelated kindreds (**Figure 16**). This represents 57% of the ICA patients for whom we had gDNA and therefore we could sequence *RPSA*. Assuming we had gDNA for all the possible cases in gray and that the possible cases in families where the ascertained cases are mutated in *RPSA* would be mutated in *RPSA*, then we could have 30 out of 44 ICA patients mutated in *RPSA* (**Figure 16**). These results show that *RPSA* is the major disease-causing gene of ICA in humans.

6.4 Haploinsufficiency of RPSA leads to ICA

To go further in the genetic dissection of ICA, we asked whether the mutations in *RPSA* led to ICA by haploinsufficiency or dominance negative effect. We did not have access to cells from patients with the p.Q9X mutation, but we had blood samples from the patients carrying the frameshift mutation p.P199Sfs. Sequencing the cDNA from RNA extracted from PBMCs of these patients did not reveal any mutation (**Figure 17**). This highlights the fact that the mutated allele is degraded, most likely through nonsense-mediated mRNA decay, and only one copy of *RPSA* is present at the mRNA level in patients from family C. These results suggest that haploinsufficiency of *RPSA* leads to ICA in at least one family. The diversity of the mutations, as well as the fact that non-coding mutations affecting about half of the *RPSA* transcripts also lead to ICA suggest that haploinsufficiency of *RPSA* leads to ICA in the majority of the families.

Therefore we hypothesized that a large deletion, encompassing the entire gene, or some exons of *RPSA*, might lead to ICA. Since large deletions cannot be detected by WES, we used the Multiplex Ligation-dependent Probe Amplification (MLPA) method[110] and assessed the copy number of RPSA in 16 patients and in 15 controls. One patient from family G, as well as one patient from family H were included and used as controls. One probe upstream of each exon of *RPSA* was designed in addition to four control probes. No deletion or copy number was identified at the *RPSA* locus except for the loss of one copy of the probe upstream of exon 1 in the two patients from family G and H (**Figure 18**). This result was expected as the probe upstream of exon 1 overlapped with the 20bp deletion described above and identified in these two patients. We can conclude that no large deletion in the *RPSA* locus causes ICA in the patients of our cohort.

Figure 14: WES identifies mutations in RPSA in ICA patients.

(**A**) Manhattan plot representing the *p*-values (P) for 21,185 genes. *P*-values were calculated by a weighted sum approach, comparing novel variants identified by WES in 30 cases and 593 controls. (**B**) Pedigrees of all the kindreds carrying coding mutations in *RPSA*. Black symbols indicate ICA cases. Gray symbols indicate possible ICA patients. WT means wiltd-type allele. M represents the mutated allele. (**C**) Schematic diagram of the RPSA protein (drawn to scale) showing the different parts encoded by different exons. Mutations identified in ICA patients are marked in red. Evolutionary conservation of the RPSA regions containing amino acids mutated in ICA patients are represented below.

Figure 14: WES identifies mutations in RPSA in ICA patients.

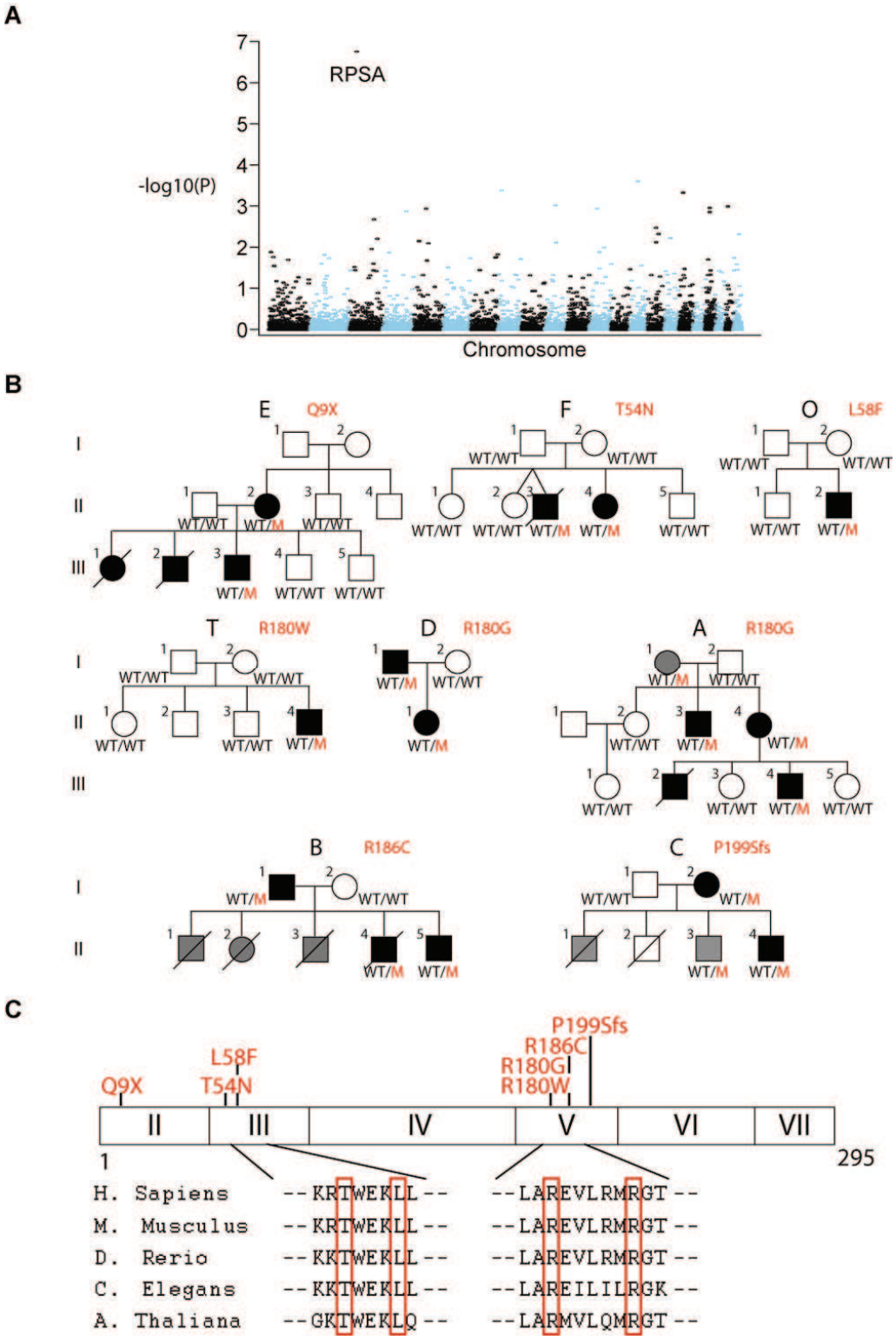


Figure 15: Non-coding mutations in *RPSA*.

(**A**) Schematic of the *RPSA* gene. Exons are represented by rectangles and introns by a line. Smaller rectangles indicate the untranslated regions. Size of introns and exons are at scale. Below the scheme is a zoom of the C-terminal part of exon 1. The blue rectangle represents the end of exon 1. Above the rectangle, letters indicate the base pairs. Below the two non-coding mutations are indicated. (**B**) Pedigrees of family G and H indicating the genotypes for the mutation in the 5'-UTR (family H) and the genotypes for the splice variant (family G). Genotypes for the 20bp deletion upstream of exon 1 (rs139686560) are also indicated. Black symbols represent ICA patients. Gray symbols indicate possible ICA patients. (**C**) Scheme of the WT transcript of RPSA and of the spliced variant transcript. The red rectangle represents the extra 43bp added in the 5'-UTR (corresponding to the first 43bp of the intron between exons 1 and 2.

Figure 15: Non-coding mutations in *RPSA*.

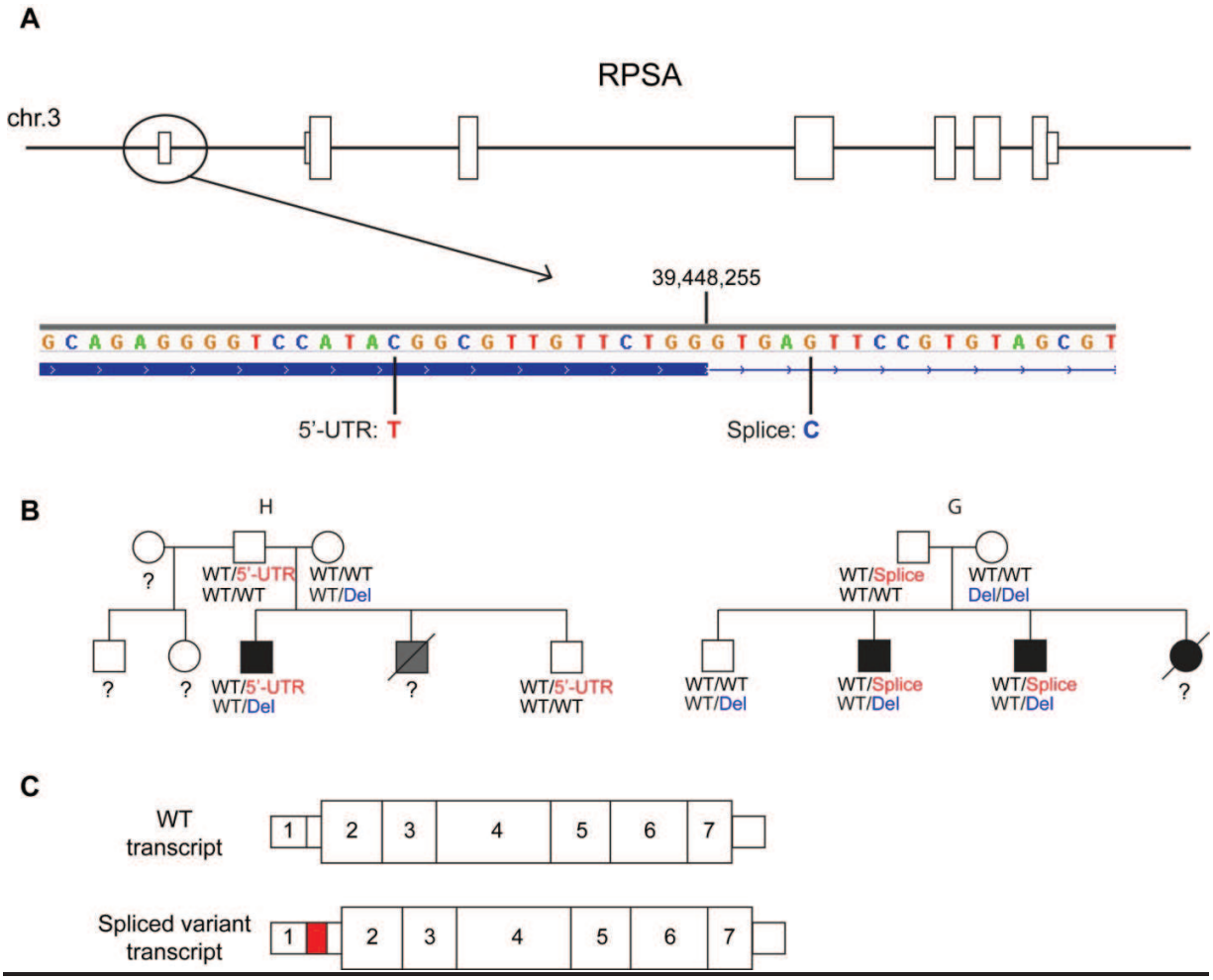


Figure 16: More than 50% of the ICA patients are mutated in *RPSA*.

Black symbols indicate ICA patients. Gray symbols indicate possible ICA patients. Pedigrees circled in red indicate families carrying mutations in *RPSA*.

Figure 16: More than 50% of the ICA patients are mutated in *RPSA*.

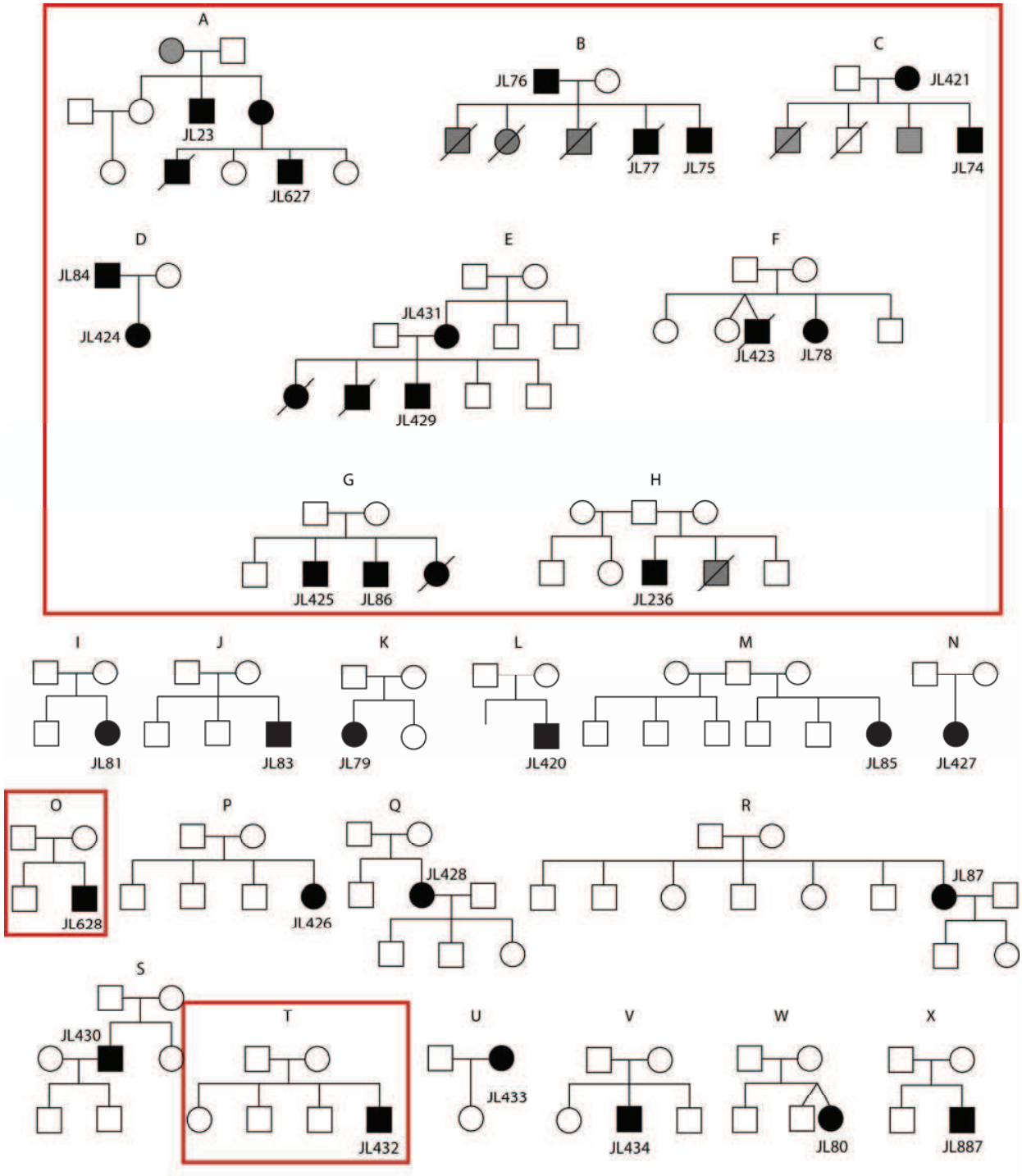


Figure 17: Haploinsufficiency of *RPSA* in one ICA patient.

Automated sequencing profile. The two top profiles correspond to the sequencing of the gDNA of patient I.2 in family C (mother) and shows heterozygous peaks due to the heterozygous insertion of 5 bp (TCATG). The two rows correspond to the sequencing using a forward or reverse primer. The two profiles on the bottom correspond to the sequencing of the cDNA of patient I.2 in family C. No heterozygous peaks are apparent, showing that the allele carrying the insertion is not present at the mRNA level.

These profiles are illustrative of three independent experiments where the gDNA and cDNA of the two patients and the healthy father were extracted from PBMCs or ficoll, and later amplifie and sequenced using different pairs of primers around the insertion. Primers for the gDNA bind in the introns around exon 5. Primers for the cDNA bind to the coding region of RPSA. DNAse treatment was performed before amplification and RT negative controls did not show any sequence.

Figure 17: Haploinsufficiency of *RPSA* in one ICA patient.

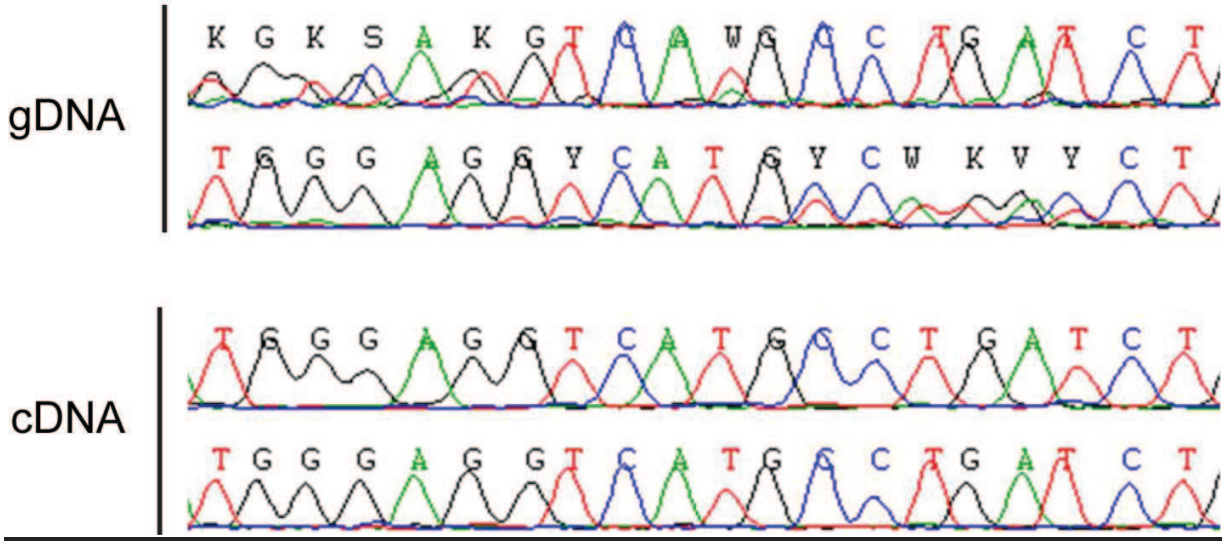


Figure 18: MLPA results showing deletion of one copy of probe R1.

Example of the fragment analysis results after MLPA method in one control, G-I.1 (father in family G), and one ICA patient, G-II.2 (JL0425 in family G). Horizontal axis represents the size of each probe. Vertical axis represents the intensity of the peaks. Each probe is represented by one peak. Control probes: X, 18, Y and IF hybridize in *BCAP31*, *SS18*, *SRY* and *IFNγ* respectively. Probes in RPSA are represented by R1, R2, R3, R4, R5, and R7. In red R1 is decreased compared to the other peaks in the patients' sample.

Figure 18: MLPA results showing deletion of one copy of probe R1.

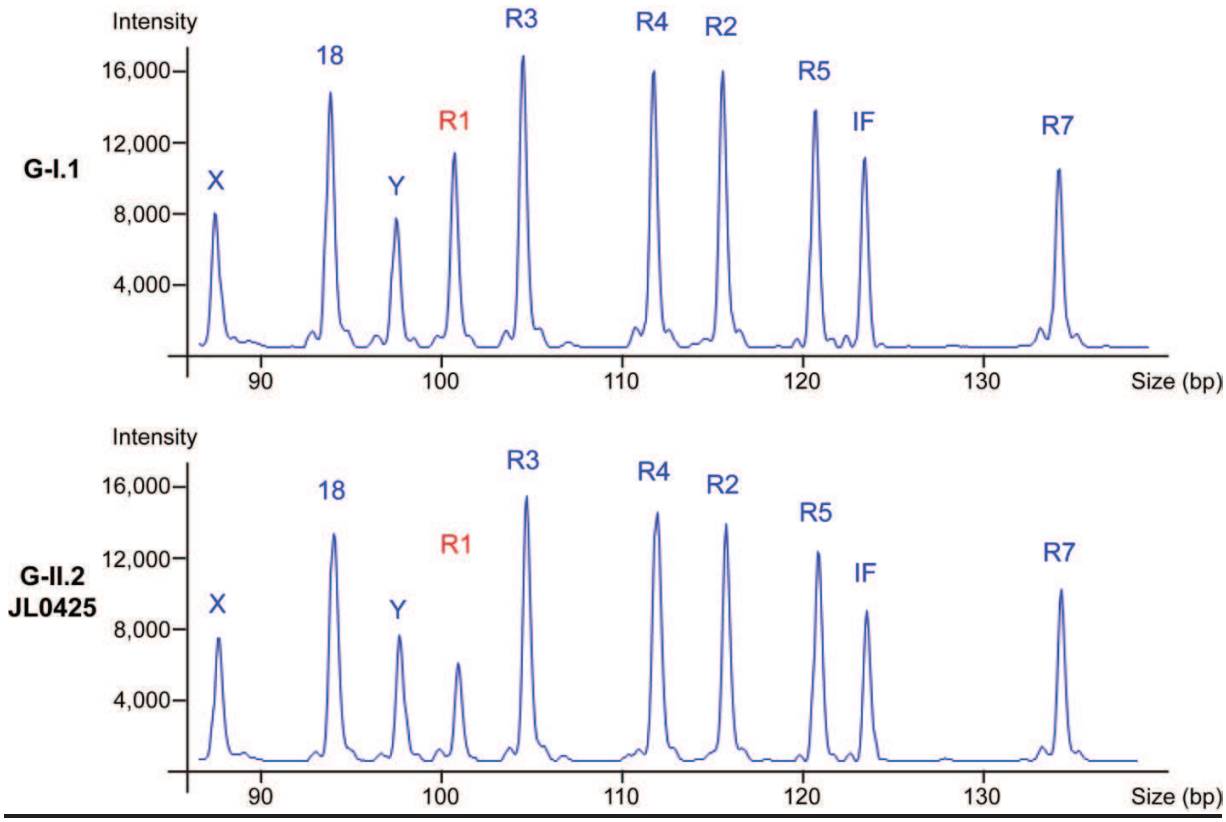


Table 8: Metrics and coverage for the 31 exomes of the ICA cohort.

Sample	**JL0023**	**JL0074**	**JL0075**	**JL0076**	**JL0077**
Single/Paired-end	single-end	single-end	single-end	single-end	single-end
Bait set	38mb	38mb	38mb	38mb	38mb
Total reads	17,236,982	46,309,265	40,784,315	37,764,604	38,145,467
PF uq reads aligned	10,107,021	21,996,482	20,089,344	17,806,372	7,410,366
PF uq bases aligned	768,010,436	1,759,448,314	1,606,927,209	1,424,306,381	592,624,602
Mean bait coverage	14.295475	24.739707	22.371171	20.34248	9.578373
Pct target bases 2X	0.909743	0.924441	0.918023	0.93181	0.860445
Pct target bases 10X	0.590847	0.750323	0.722219	0.712834	0.421068
Pct target bases 20X	0.268231	0.530087	0.484732	0.428057	0.115878
Pct target bases 30X	0.104084	0.342634	0.293734	0.234513	0.023321

Sample	**JL0078**	**JL0079**	**JL0080**	**JL0081**	**JL0082**
Single/Paired-end	paired-end	single-end	single-end	single-end	single-end
Bait set	50mb	50mb	50mb	50mb	50mb
Total reads	107,126,412	40,420,470	43,328,673	41,294,667	83,706,065
PF uq reads aligned	84,546,171	25,184,279	25,840,648	25,465,508	39,537,580
PF uq bases aligned	8,485,940,747	1,913,831,850	1,963,698,167	1,935,215,309	3,992,605,795
Mean bait coverage	67.418656	22.723002	23.597151	22.862263	45.813742
Pct target bases 2X	0.960521	0.909346	0.909479	0.908707	0.938292
Pct target bases 10X	0.902727	0.741285	0.745674	0.738741	0.855381
Pct target bases 20X	0.840115	0.507457	0.523226	0.507219	0.758379
Pct target bases 30X	0.7673	0.299268	0.320089	0.302696	0.647906

Sample	**JL0083**	**JL0084**	**JL0085**	**JL0086**	**JL0087**
Single/Paired-end	single-end	single-end	single-end	single-end	single-end
Bait set	50mb	50mb	50mb	50mb	50mb
Total reads	78,783,938	31,199,061	32,809,507	30,654,762	40,377,837
PF uq reads aligned	37,235,703	18,972,448	20,448,833	18,669,724	23,842,750
PF uq bases aligned	3,760,154,743	1,441,770,850	1,553,960,059	1,418,746,611	1,811,876,919
Mean bait coverage	43.565747	16.629181	18.20785	17.181194	20.917354
Pct target bases 2X	0.933597	0.892851	0.899239	0.902562	0.901283
Pct target bases 10X	0.84679	0.65101	0.679147	0.667414	0.710284
Pct target bases 20X	0.743303	0.342122	0.389697	0.353665	0.462034
Pct target bases 30X	0.625841	0.153176	0.190392	0.161803	0.259161

Sample	JL0236	JL0420	JL0421	JL0424	JL0425
Single/Paired-end	paired-end	paired-end	paired-end	paired-end	paired-end
Bait set	50mb	50mb	50mb	50mb	50mb
Total reads	125,029,032	117,243,000	140,492,902	103,098,554	85,285,190
PF uq reads aligned	110,617,562	98,100,474	121,287,351	91,843,102	75,544,358
PF uq bases aligned	11,081,506,521	9,816,627,457	12,141,343,988	9,226,836,010	7,591,196,405
Mean bait coverage	84.436359	72.119993	81.998927	63.497999	54.41728
Pct target bases 2X	0.969456	0.928401	0.965584	0.959868	0.957073
Pct target bases 10X	0.926532	0.837049	0.910576	0.88943	0.877033
Pct target bases 20X	0.875383	0.745354	0.850711	0.809445	0.780905
Pct target bases 30X	0.818594	0.657251	0.787337	0.717467	0.668404

Sample	JL0426	JL0427	JL0428	JL0429	JL0430
Single/Paired-end	paired-end	paired-end	paired-end	paired-end	paired-end
Bait set	50mb	50mb	50mb	50mb	50mb
Total reads	45,334,202	89,145,496	124,471,982	102,668,730	124,254,726
PF uq reads aligned	40,233,025	79,459,443	105,828,597	89,271,170	110,269,877
PF uq bases aligned	4,040,271,697	7,982,074,597	10,595,467,142	8,954,595,385	11,071,632,524
Mean bait coverage	30.19325	54.872578	77.966611	63.012581	71.986333
Pct target bases 2X	0.933296	0.953551	0.931559	0.959286	0.963892
Pct target bases 10X	0.794005	0.872634	0.84806	0.886683	0.900822
Pct target bases 20X	0.594626	0.778277	0.764386	0.805487	0.830957
Pct target bases 30X	0.404496	0.669789	0.682501	0.713138	0.753709

Sample	JL0431	JL0432	JL0433	JL0434	JL0627
Single/Paired-end	paired-end	paired-end	paired-end	paired-end	paired-end
Bait set	50mb	50mb	50mb	50mb	50mb
Total reads	109,856,484	79,120,912	98,829,284	72,651,514	21,841,098
PF uq reads aligned	95,581,558	69,749,207	87,207,926	63,845,042	20,931,942
PF uq bases aligned	9,539,384,937	6,998,934,247	8,744,717,784	6,410,188,375	2,101,848,864
Mean bait coverage	67.133081	49.687267	59.470436	45.48222	11.575381
Pct target bases 2X	0.942637	0.950422	0.955405	0.950631	0.842905
Pct target bases 10X	0.852382	0.855911	0.877213	0.853847	0.432141
Pct target bases 20X	0.760625	0.74382	0.789479	0.731924	0.173824
Pct target bases 30X	0.673637	0.620407	0.689303	0.59457	0.076029

Sample	JL0628	Legend
Single/Paired-end	paired-end	
Bait set	50mb	
Total reads	61,441,428	The total number of reads in the SAM or BAM file examine
PF uq reads aligned	58,012,862	The number of PF unique reads that are aligned with mapping score >0 to the reference genome
PF uq bases aligned	5,787,492,782	
Mean bait coverage	31.352339	The mean coverage of all baits in the experiments
Pct target bases 2X	0.918495	The percentage of ALL target bases achieving 2X or greater coverage.
Pct target bases 10X	0.755845	The percentage of ALL target bases achieving 10X or greater coverage.
Pct target bases 20X	0.543446	The percentage of ALL target bases achieving 20X or greater coverage.
Pct target bases 30X	0.379732	The percentage of ALL target bases achieving 30X or greater coverage.

Table 9: Genes mutated in six or more ICA patients after analysis of 31 exomes.

Gene	# patients mutated	# novel mutations	Significant[1]
TTN	15	18	No
MUC4	14	20	No
CUBN	9	8	No
NEB	8	8	No
MUC5B	7	11	No
FRAS1	7	6	No
RPSA	7	5	Yes
IGHV2-70	6	13	No
USH2A	6	8	No
KIAA0284	6	6	No
FBN3	6	6	No
MUC16	6	6	No
ANK2	6	4	No
OSBPL5	6	3	Yes

[1]: In this case, a gene is called significant if the gene was not mutated in more than 3 patients in one other project of a similar size (between 20 to 45 samples) from the laboratory using the exact same filters. Projects included were: HSE, MSMD, TB, CMC, Flu and Whipple. Projects where patients are susceptible to IPD were not included. The analysis was done in December 2011.

Table 10: Summary of all coding mutations in RPSA.

Family	Ethnic origin (country of residence)	Position (hg19 build)	Mutation	Protein impact	De novo	SIFT (score)[1]	Polyphen-2 HumDiv (score)[2]	Polyphen-2 HumVar (score)[2]
E	Caucasian (USA)	3 :39,449,169	c.25C>T	p.Gln9X Q9X	N/A	N/A	N/A	N/A
F	Caucasian (France)	3:39,450,124	c.161C>A	p.Thr54Asn T54N	Yes *	DAMAGING (0)	BENIGN (0.147)	BENIGN (0.433)
O	Caucasian (France)	3:39,450,135	c.172C>T	p.Leu58Phe L58F	Yes	DAMAGING (0.01)	BENIGN (0.167)	BENIGN (0.257)
T	Caucasian (Sweden)	3 :39,453,179	c.538C>T	p.Arg180Trp R180W	Yes	DAMAGING (0.02)	POSSIBLY DAMAGING (0.479)	BENIGN (0.081)
D	Caucasian (Italy)	3 :39,453,179	c.538C>G	p.Arg180Gly R180G	N/A	DAMAGING (0.02)	BENIGN (0.390)	POSSIBLY DAMAGING (0.579)
A	Caucasian (UK)	3 :39,453,179	c.538C>G	p.Arg180Gly R180G	N/A	DAMAGING (0.02)	BENIGN (0.390)	POSSIBLY DAMAGING (0.579)
B	Subsaharan Africa (Congo-Brazaville)	3 :39453197	c.555C>T	p.Arg186Cys R186C	N/A	DAMAGING (0.03)	BENIGN (0.003)	BENIGN (0.009)
C	Caucasian (France)	3 :39453230	c.589_593 dup	p.Pro199SerfsX26 P199Sfs	N/A	N/A	N/A	N/A

[1]: SIFT score were calculated on http://sift.jcvi.org/ on 09/02/2012.

[2]: Polyphen scores were calculated on http://genetis.bwh.harvard.edu/pph2 on 09/02/2012

CHAPTER 7

RPSA IN THE DEVELOPMENT OF THE SPLEEN

RPSA also called laminin receptor 1 (*LAMR*) encodes for ribosomal protein SA. In yeast the homolog of RPSA is RPS0, and it is RPS2 in prokaryotes. RPSA is ubiquitously expressed and has never been associated with the development of the spleen[46; 111]. In order to understand the mechanism of disease or to link *RPSA* with the spleen, we could tackle the problem from two sides: at the molecular scale or at the organism scale.

7.1 Clinical details of human *RPSA* haploinsufficiency

We started to look at the human level. We asked if there was any specific phenotype shared by the patients carrying RPSA mutations. Was there a phenotype we had not noticed, or was there a phenotype that was present in these samples and not in the other ICA patients of our cohort that were not carrying an identified RPSA mutation? Extensive case reports were sent by the physicians from these ten kindreds and are summarized in **table 11**. These data confirmed that the patients had no phenotype outside the spleen. They had no heart malformations as expected. Moreover complete blood counts from all patients and some of their siblings confirmed that the patients had no hematologic phenotype. None had anemia.

On the other hand, it was striking that the four post mortem analyses from 4 unrelated patients carrying 4 different mutations in RPSA and followed by four different physicians described the same spleen phenotype (**Table 11**). The spleen was minuscule and a large part of the spleen was a fibrotic nodule. At the periphery of the fibrotic nodule there was hemosiderin deposition (or presence of Gandy-Gamna bodies depending on the reports) whereas the remaining parenchyma was mostly red pulp with very few lymphoid follicles (**Figure 19**). The pathologists saw the sign of a hemorrhagic infarct of the spleen. A late infarction of the spleen might explain why some patients were only diagnosed at adult age and did not suffer any severe infections in childhood. The other possibility to explain these adult-diagnosed patients could merely be a bias from our recruitment. However, some patients were diagnosed as early as 3 days old (**Table 11**). This would mean that the infarct happened prenatally, which is unlikely. Therefore two hypotheses seem possible. The pathologists suggest that the spleen developed normally and suddenly infarcted. The non-pathologists suggest the spleen

never developed properly and this led to this small characteristic spleen of patients mutated in RPSA.

One way we might answer the question whether the spleen developed normally and then infarcted, or never developed properly is by looking at the development of the spleen in humans. In family A, the mother of JL0627 is pregnant again. The monitoring of the spleen of the baby during pregnancy will be interesting. The interest will also depend on the genotype of the baby. The answer will come in a couple of months.

7.2 *RPSA* haploinsufficiency in animal models

One other way to test the hypothesis that the spleen never developed completely and perhaps the architecture of the vascular system of the spleen was impaired and led to this fibrosis and iron deposits, we started to look at animal models: the mouse model and the zebrafish model taking advantage of the conservation of RPSA in different species. The amino acid sequence of Rpsa in mouse is 99.3% identical to its human homolog and the amino acid sequence of Rpsa in zebrafish is 82% identical to its human homolog.

Our collaborator Steven R. Ellis from the University of Louisville created a Rpsa+/- mouse model by inserting a neo-cassette replacing the first four exons of *Rpsa* (Material and Methods). The mice are currently being phenotyped at the Laboratory of Comparative Pathology at the Memorial Sloan-Kettering Cancer Center (MSKCC). Sixteen mice were sent for phenotyping at five weeks of age: 4 control WT mice, 8 mice Rpsa+/-, 2 mice Pbx1+/- and 2 mice Rpsa+/-,Pbx1+/-. The complete blood count results did not show any phenotype in the Rpsa+/- mice. The histology results on the spleens of these mice are not available yet. The results on the architecture and the vasculature of the spleen will be very important. In the meantime we dissected one Rpsa+/- female mouse at the age of five weeks, in the laboratory of Licia Selleri. The Rpsa+/- mouse presented a normal spleen in terms of size and weight (**Figure 20**). This is only one mouse and the results of the phenotyping at MSKCC will help to draw more conclusions. The complete knockout mouse dies at E3, well before the development of the spleen and thus does not allow any interpretable observation.

In addition to the mouse model, we took advantage of the zebrafish model organism. The zebrafish is an ideal system to study developmental defects as all the organs, the spleen in particular, form within 5 days and the fish remains transparent for that period of time[112]. In the laboratory of Nikolaus Trede at the Univeristy of Utah, we knocked down *Rpsa* using morpholinos (**Figure 21**). The first results were hard to interpret and the optimal levels of knocking down need to be optimized before we can assess the impact of Rpsa decrease on the zebrafish spleen using a Tlx1 probe at 5 days of age[113]. Our work to elucidate the pathogenesis of ICA by looking at the macroscopic level is thus ongoing.

7.3 Impact of RPSA haploinsufficiency at the molecular level

A complementary approach to understand the role of RPSA in the mechanism of ICA is to look at the molecular level. Its best characterized function is its ribosomal function which can be divided into two. Firstly, RPSA is important in pre-rRNA processing[114]. Secondly, it is also an integral part of the small subunit of the ribosome[115]. The best described function of RPSA is its role in the processing of 18S rRNA (**Figure 22A**). In vitro studies have shown that RPSA is important in the pre-rRNA processing, particularly in the late stages for the processing of the 21S to the 18S rRNA[114]. We thus tested the pre-rRNA processing in the patients' cells. We only had frozen peripheral blood mononuclear cells (PBMCs) available. In order to have enough cells to extract total RNA we activated these PBMCs with PHA for 3 days. In these conditions the T cells proliferate. The analysis by Northern blot of the RNA extracted after 72 hours did not reveal a strong phenotype. The intensity of the band corresponding to the 18SE form on the rRNA was as strong as the controls (**Figure 22B**). Actually the 18SE band might be more intense when compared to the other bands (the 45S for example) and this might indicate a phenotype. The experiment needs to be repeated. Looking at other time points such as 24 or 96 hours might reveal a phenotype. Activating the PBMCs with PHA and then IL-2 should lead to an increased proliferation of T cells, and should thus be better conditions to observe a rRNA processing phenotype. Experiments looking more downstream of the pre-rRNA processing will be performed next. In particular it would be interesting to analyze the

cytoplasmic ribosomes on a sucrose gradient and look at the quantity of the 40S ribosomal subunit for example. One more experiment would be to look at the nucleolar morphology and chromatin condensation by FISH in the patients' cells[114].

Finally we tried to understand how mutations in such a ubiquitous gene led to a spleen specific phenotype. We hypothesized that it was due to the level of expression of the protein in the spleen. Even though RPSA is ubiquitously expressed, it might be expressed at particularly low or high levels in the fetal spleen. For example if it were the ribosomal protein with the lowest expression in the fetal spleen, a supplementary decrease due to haploinsufficiency might lead to a defect in the spleen, whereas it would be not impair any function in cell types where RPSA is more expressed. We thus purchased mRNA extracted from human fetal spleen, human fetal heart, human fetal brain and human adult spleen and compared by Q-PCR the levels of *RPSA*. The first results did not show a major difference (**Figure 23**). However these results varied greatly depending on the gene used as a control (GAPDH or VCL for example). In order to decrease the impact of these control genes, we need to test more genes and probably more organs. Taken together our results need to be repeated and many experiments need to be performed to understand the impact of the mutations on the function of RPSA. Ultimately these results might help to understand the link with the spleen.

Figure 19: Spleen of a patient with a mutation in *RPSA*.

Spleen from patient JL0423 (family F). The spleen was 1cm large and contained a large fibrous nodule with many iron deposits and few lymphoid follicles (Perls staining)[116].The spleen is representative of the four spleens described from the post mortem analyses of ICA patients with mutations in *RPSA*.

Figure 19: Spleen of a patient with a mutation in *RPSA*.

Figure 20: Rpsa+/- mouse.

The spleen of a control WT mouse (top) and the spleen of a Rpsa+/- mouse (bottom). Both mice were 5 weeks old when the spleen was analyzed. The weight and size of both spleens are comparable.

Figure 20: Rpsa+/- mouse.

Figure 21: Knockdown of rpsa in zebrafish.

(**A**) Scheme of the rpsa gene in zebrafish. Rpsa is 308 amino acids in the zebrafish and has 8 exons. Exon 7 is different from human RPSA .The size of the exons and introns respect the scale. Shorter (in height) restangles represent the untranslated regions. Bigger (in height) rectangles represent the coding region. The localization of the morpholinos used to knockdown rpsa in the zebrafish is represented in red. From left to right: Rpsa-5UTR-MO-1, Rpsa-intr2ex3-MO-2, Rpsa-ex3intr2-MO-3. (**B**) Example of in situ hybridization results at day 5 with a Tlx1 probe[113]. The spleen of the zebrafish is indicated by the arrow.

Figure 21: Knockdown of rpsa in zebrafish.

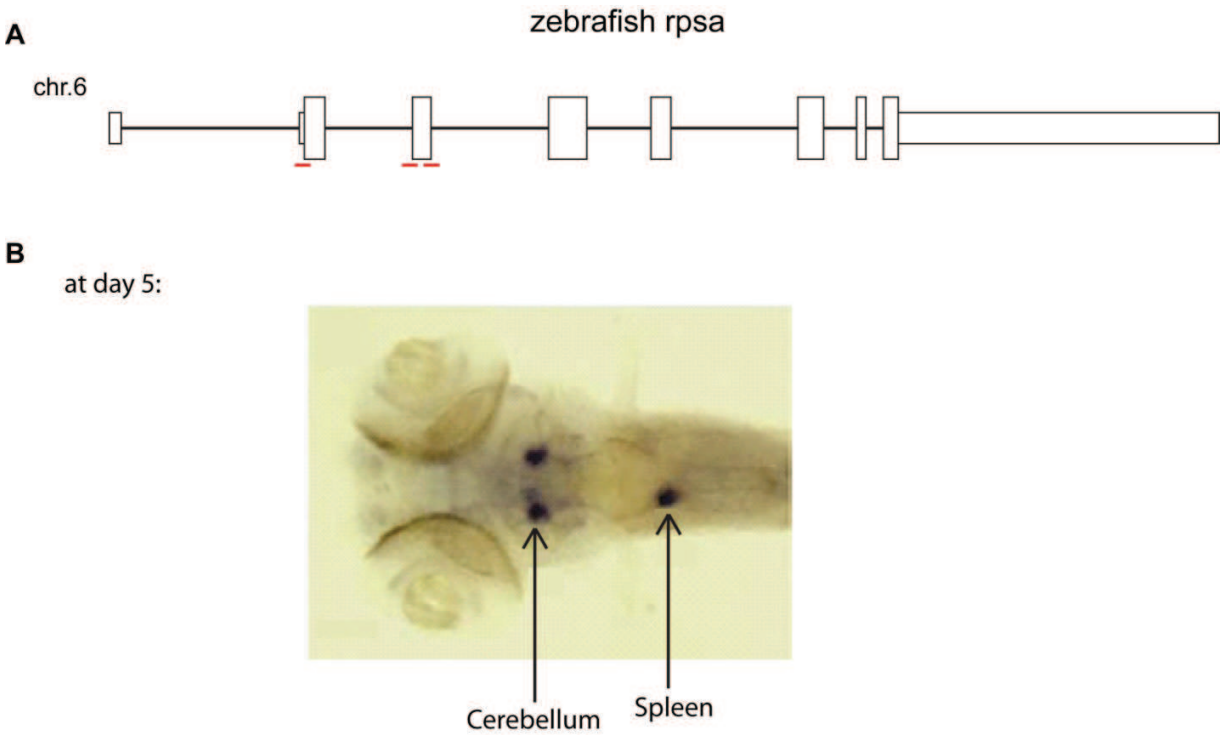


Figure 22: RPSA is involved in pre-rRNA processing.

(**A**) Pre-rRNA processing pathway in human cells. The main pathway is described[114]. The 18S-E pre-rRNA is converted to 18S rRNA in the cytoplasm. (**B**) Northern blot. Total RNA from PBMCs of controls and patients were extracted 72 hours after activation by PHA, resolved on a 1% agarose gel, and transferred to a nylon membrane, which was hybridized with 5'-ITS1 ^{32}P-labeled probe. Ratios of the intensity of the band 45S/18S is indicated underneath.

Figure 22: RPSA is involved in pre-rRNA processing.

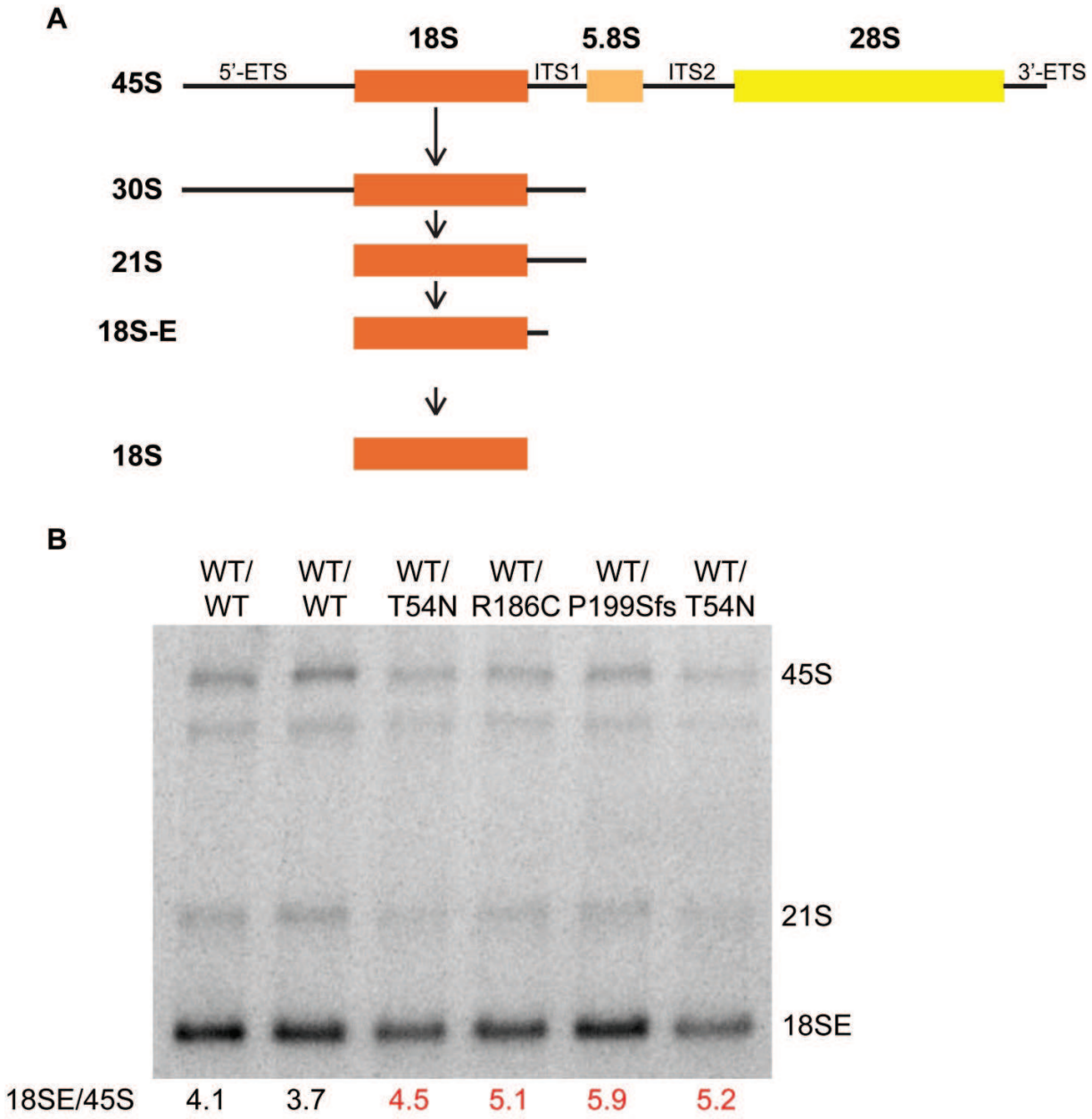


Figure 23: Expression of *RPSA* in different tissues.

Poly A+ RNA was purchased from Clontech and was used to generate cDNA. Human fetal spleen (cat #636154), human fetal heart (cat #636156), human fetal brain (cat #636106) and human adult spleen (cat #636121). (**A**) To control that the human fetal spleen library came from the spleen, and that the heart library was not a spleen library *TLX1* (specifically expressed in fetal spleen) mRNA levels were determined by quantitative RT-PCR using the Taqman assay. Threshold cycles normalized with respect to those of GUS (ΔC_T) are plotted as $2^{-\Delta CT}$. Results are then normalized with human fetal heart score. (**B**) *RPSA* mRNA levels were determined by quantitative RT-PCR by the TaqMan assay (cat #Hs03046712_g1, Applied Biosystems). Threshold cycles normalized with respect to those of GUS (ΔC_T) are plotted as $2^{-\Delta CT}$. (**C**) *RPSA* mRNA levels were determined by quantitative RT-PCR by the TaqMan assay (cat #Hs03046712_g1, Applied Biosystems). Threshold cycles normalized with respect to those of VCL (ΔC_T) are plotted as $2^{-\Delta CT}$.

Figure 23: Expression of *RPSA* in different tissues.

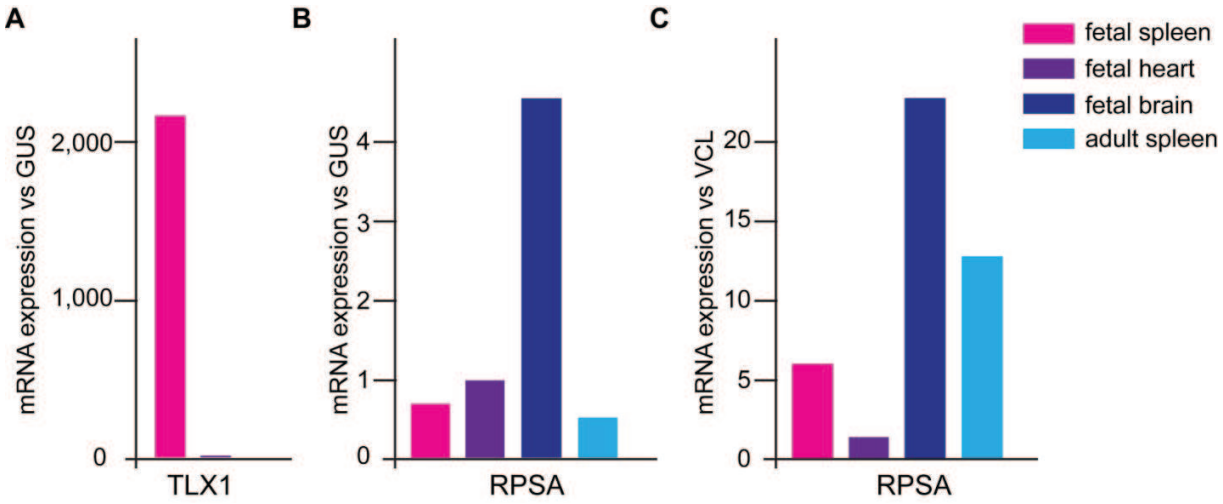


Table 11: Clinical phenotype of patients carrying a mutation in RPSA.

Family:Patient[1]	Age of Diagnosis	Spleen (US/PM)	HJB	Severe Infections (age of 1st infection)	Heart	CBC	Other
A : JL0023	20 years	no (US)	yes	no	normal	haemoglobin 4.4g/dl at age 20y	dilated veins in the jejunum and rectosigmoid. angiodysplasia
A : JL0627	early childhood	no (US)	yes	no[2]	normal	normal	/
B : JL0076	adult	no (US)	yes	no	normal	normal	/
B : JL0077	post mortem (23 mo)	1cm long (PM)	NA	yes (died at 23mo)	normal	NA	/
B : JL0075	at birth	no (US)	yes	no[2]	NA	normal	/
C : JL0421	?	no	?	?	normal	normal	/
C : JL0074	early childhood	1.86cm x 0.5cm (US)	yes	no[2]	normal	normal	/
D : JL0084	adult	no (US)	yes	no	normal	normal	/
D : JL0424	19 mo	no (US)	yes	no	normal	normal	born at 29 weeks
E : JL0431	adult	no (US)	yes	no	normal	normal	/
E : JL0429	7 mo	no (US)	yes	no[2]	normal	normal	/
F : JL0423	post mortem (22 mo)	1cm x 0.5cm	NA	yes (died at 22mo)	normal	NA	/
F: JL0078	4 days	no (US)	yes	no[2]	normal	normal	osteosarcoma at age of 12y (left tibia and left knee)
G : JL0425	childhood	no (US)	yes	no	normal	normal	/
G : JL0086	childhood	no (US)	yes	no	normal	normal	/
H : JL0236	20 mo	small (US)	?	yes (10mo)	normal	normal	/
O : JL0628	20 mo	no (US)	yes	yes (2mo)	normal	normal	/
T : JL0432	2 years	no (US)	yes	yes (8mo)	normal	normal	/

US: ultrasound. PM: post mortem analysis. HJB: Howell-Jolly bodies. CBC: complete blood count. NA: not assessed. y: year. mo: month.

[1]: the JL number in Figure 4 is used to identify the patients. Only certain ICA patients with a RPSA mutation are included.

[2]: under prophylaxis treatment since early childhood.

CHAPTER 8

DISCUSSION

8.1 ICA compared to PIDs, developmental defects and ribosomopathies

We showed that *RPSA* is the major genetic etiology of ICA. More than 50% of the ICA patients in our cohort are mutated in this gene, and most notably 100% of the familial cases. This result is comparable to the recent discovery that 56% (15/27) of the patients with pancreas agenesis (the lack of a pancreas, an organ physically close to the spleen) present *GATA6* haploinsufficiency[117]. This is however a very high percentage compared to the other PIDs studied in the laboratory. For example, mutations in *TBK1* explain less than 1% of HSE cases[118]. The highest percentages observed in the lab were: 21% for the MSMD patients mutated in *IL-12Rβ1*[119], and close to 50% for the CMC patients mutated in *STAT1*[120]. The higher percentage of ICA patients mutated in *RPSA* might be explained by a more homogeneous cohort compared to other PIDs. This homogeneity highlights the accuracy of diagnosis by the physicians we collaborated with. Besides the great work of the physicians it also reflects the fact that it is easier to diagnose a patient with invasive pneumococcal disease and miss the ICA phenotype rather than the contrary. Therefore one would expect more heterogeneity in the IPD cohort than the ICA cohort. The biggest risks for a phenocopy of ICA are syndromic asplenia, meaning asplenia associated with other developmental defects. In this study the case reports of the patients recruited showed in one case gut malformations and in another case a heart defect. We removed these two cases from the study.

Although we did not expect such a percentage of ICA cases mutated in the same gene, the biggest surprise came from the nature and function of RPSA. When we started the project we first hypothesized that mutations in *TLX1* would be disease-causing for the majority of the patients as the *Tlx1*-/- mice present ICA[57]. This was not the case. We then thought it could be allelic mutations in genes involved in asplenia syndrome or syndromic asplenia. This is why we investigated the mutation in *NKX2-5*[46] (chapter 5). The experiments performed in the mouse model showed that Nkx2-5 is an important transcription factor involved in the development of the spleen. NKX2-5 is probably also involved in the development of the spleen in humans. However its role might be redundant with another gene or transcription factor. At least it is unlikely that

the P236H heterozygous mutation in *NKX2-5* identified in family B is responsible for the ICA phenotype due to the identification of a mutation in *RPSA* in this family. Finally, when we were analyzing the exome results for all patients we first hypothesized that a transcription factor would be identified. TLX1 (ICA in the mouse) is a transcription factor. GATA6 (pancreas agenesis) is a transcription factor. NKX2-5 (congenital heart defects) is a transcription factor. From what we know RPSA is not a transcription factor and unlike most transcription factors, it is expressed ubiquitously. From the perspective of spleen development this result is very surprising.

RPSA is a ribosomal protein[121]. ICA could thus be classified as a new ribosomopathy. From the ribosome perspective, this result also stands out. So far, human disorders of ribosome dysfunction include Diamond-Blackfan anemia (DBA), Schwachman-Diamond syndrome (SDS), X-linked dyskeratosis congenita (DKC), cartilage hair hypoplasia (CHH), Treacher Collins syndrome (TCS) and the 5q- syndrome[122]. Most of these disorders involve bone marrow failure, anemia and/or craniofacial or other skeletal defects. Haploinsuffciency in ten different ribosomal proteins including *RPS19*, *RPS17*, *RPS24* and *RPL5* explain about 60% of DBA cases[122-125]. DBA is a congenital bone marrow failure syndrome characterized by hypoproliferative anemia and is associated with physical abnormalities. It was thus striking that our patients had no hematologic phenotype, in particular no anemia. Moreover some DBA patients also carry a broad range of developmental defects. For example the majority of patients with mutations in RPL5 have cleft palates or lip[126], but none of the patients reported with haploinsufficiency in a ribosomal protein was diagnosed with ICA or with a spleen abnormality. By comparison, ICA is the ribosomopathy with the narrowest clinical phenotype in humans. Another major difference concerning the developmental defects in other ribosomopathies and ICA is the clinical penetrance. The penetrance is 100% in ICA whereas the penetrance for the physical abnormalities is very low in DBA patients. For example the R94X mutation in *RPS19*, identified in two sisters and their mother, led to thumb malformations and duplicated ureter in one sister, to congenital glaucoma in the other sister and had no effect in the mother[123].

8.2 Hypotheses regarding the pathogenesis

Understanding the mechanism of pathogenesis and the link between RPSA and the spleen became the next challenge. The most relevant cells for these studies would be the spleen cells from human fetuses, not easily available for research purposes. Looking at the developing spleen in other organisms such as the mouse might be the most relevant experiment. We made the experiment. On one hand, haploinsufficiency of RPSA led to ICA in humans, a result that fits with the severity of the phenotype and the rare (n=1) non-synonymous polymorphisms in *RPSA*. On the other hand, haploinsufficiency of *Rpsa* did not lead to any striking abnormality in the spleen in the mouse, although more phenotyping needs to be completed. Interestingly, haploinsufficiency of different ribosomoal proteins (RP), such as RPS19, in the mouse does not lead to DBA[127]. Similarly, *Gata6* haploinsufficiency did not lead to pancreas agenesis in the mouse[117]. These observations suggest it might be required to knock down the genes by more than 50% to see a phenotype in the mouse. Therefore one alternative approach is to use transgenic models or conditional alleles allowing a complete deletion of Rpsa or a further reduction of Rpsa levels as compared to haploinsufficiency in an organ specific manner[127]. This would however be further away from imitating the human pathogenesis. Moreover the extreme phenotype (death at E3) of the *Rpsa*-/- mouse suggests that knocking out Rpsa completely in any organ would lead to the agenesis of this organ. We would thus need to conditionally knockout Rpsa in the spleen and at least one other organ as a control, making this experiment very costly. Moreover, an irreproducible phenotype in the mouse might still provide insight into the mechanism of pathogenesis. We can speculate that the difference between the two species is due to the promoter region as the protein sequences are almost identical. The human *RPSA* promoter might be weaker in the spleen.

While in vivo studies are in process, we zoomed in and looked at the cellular level. The amino acids mutated in the patients carrying missense mutations are conserved in bacteria, yeast, plants and mammals. The high conservation of these amino acids even in prokaryotes suggests that they are involved/required for the

ribosomal function of the protein. This is why it will be important that the next experiments test extensively the ribosomal function of RPSA in the patients' cells. Pre-rRNA processing defects in human cells are associated to a red blood cell or erythropoiesis phenotype[125]. If we see a phenotype in vitro, we could speculate that the red blood cells have a direct involvement in ICA as the spleen is a blood filter, and the graveyard of red blood cells. The pathologists analyzing the necropsy of the four ICA patients in our cohort described the spleens as 'infarcted', similar to what happens in sickle cell anemia patients. Finally the fact that patients with HO-1 deficiency (3 cases) present anemia, hemolysis and asplenia[84-86] might be one more link between anemia and asplenia. However asplenic patients did not present with anemia or a red blood cell phenotype making this hypothesis farfetched unless this phenotype is very transient, ahead of any infection and thus not detected by physicians.

On the other hand ribosomal proteins are known to have extraribosomal functions[128]. RPSA was first discovered as a laminin receptor at the cell surface[129]. It was shown to be a receptor for prions or viruses[130-132]. It was also associated with several diseases in humans, cancer particularly[133-135]. However none were very strong associations[135]. The most interesting phenotype observed might be that a functional retroposon of Rpsa caused right ventricular dysplasia in mice[136]. The authors showed that the retroposon was expressed in both ventriculars, but it led to a one-sided phenotype. Moreover the retroposon led to degradation of the myocardium even when injected in adult mice suggesting ICA might not come from a developmental defect but a later defect (the pathologist school). Finally they showed that the Rpsa retroposon could bind to HP1 and could lead to the change of expression of some specific genes[136]. This study invites us to consider the role of tissue specific isoforms and/or pseudogenes in the disease. There are more than 2,000 pseudogenes for all human RPs. Their role is unknown. It also suggests that ribosomal proteins might influence the expression of a specific subset of genes. Microarray on the patients' cells vs. their healthy siblings might bring novel insights in the role of RPSA. Moreover the lab of Maria Barna showed that *Rpl38* haploinsufficiency specifically impaired the translation of Homeobox mRNAs, leading to axial skeleton defects[137]. Ribosome profiling experiments such as the ones performed in the lab of Davide Ruggero[138] might answer this question, providing the

cells available are relevant. We believe using primary fribroblasts from some patients might be relevant, especially for the patient in family F who also developed osteosarcoma during childhood, knowing that osteoblasts and fibroblasts are very similar.

One alternative way to identify which molecular pathway is involved in the pathogenesis of ICA is to identify a second etiology of ICA. By exome-sequencing we did not identify one gene significantly more mutated in the remaining exomes when compared to controls. As ALL 'unexplained' cases are sporadic cases, we made three different hypotheses. The first one is that there is second AD etiology, which equals that some of the 'unexplained' cases might be due to *de novo* mutations. Therefore we sequenced the exome of 7 trios, meaning we sequenced the exomes of 7 couples of parents of sporadic ICA cases (on top of the 7 patients' exomes). Mutations present in the patient but not in the parents will be analyzed. The exomes are still being sequenced at the New York Genome Center for the time being. The second hypothesis is that some of these cases are due to AR mutations. Indeed if the second etiology was also AD, then some of the familial cases showing an AD mode of inheritance should have been mutated in the second hypothetical gene. It is not the case. Therefore it seems logical to look for recessive mutations. The analysis of the exomes with this AR perspective has not been successful so far. The third hypothesis is that some sporadic ICA cases are not due to a genetic lesion.

The recruitment of additional patients would also increase our statistical approach to the analysis of exomes and might lead to a second genetic etiology. We need to maintain our efforts into the recruitment of additional patients.

8.3 The consequences at the clinical level of this discovery

This work was mainly done at the Rockefeller University, where the goal is to do Science for the benefit of humanity. The discovery of the genetic etiology of ICA has important clinical consequences. ICA is a life-threatening disease unless diagnosed at birth. Patients diagnosed early will be treated with preventive prophylaxis, protecting the patient from bacterial infections. However abdominal US to look at the presence of the

spleen are not routine procedures and thus physicians only diagnose ICA after a severe bacterial infection. If we could diagnose ICA using the whole genome sequence of the new-born, we would save the lives of all ICA patients. As every new-born will have its genome sequenced in the near future, this should save the lives of many ICA patients. I hope *RPSA* will be incorporated into the genetic screens performed in new-borns and that red flags will be raised when a mutation is detected in *RPSA*. Otherwise it might be worth to start a company selling a *RPSA* genotyping. At least this work set the basis for improved diagnosis and treatment of ICA in the future. One side effect of this discovery will be the increase of the number of disease-causing *RPSA* variants in the population!!

What is true for ICA is true for all PIDs. If one knows that one is susceptible to one specific pathogen, one can take preventive treatment. This is not the case for most of the other genetic diseases where we still cannot change the outcome despite knowing the genetic etiology. This is also why the genetic theory of infectious diseases is so important clinically. The upper estimates suggest that six million people may be living with a PID worldwide[139], but the single-gene theory of pediatric infectious diseases remains not too popular and often appears provocative. This is due to two reasons. First, the germ theory of diseases holds such sway that there is still some reluctance to think of infectious diseases in terms of human genetics. Second, the popularity of the "common disease-common variant" hypothesis and the fact that infectious diseases are the most common pediatric illnesses have resulted in resistance to thinking about pediatric infections as monogenic diseases. Rare pediatric illnesses are thought to be Mendelian, whereas common illnesses are thought to be infectious. However, many individual infectious diseases in children are less common than some well-known genetic diseases. Severe infectious diseases are globally common because there are many of them and some are common. The discovery of the genetic etiology of ICA argues in favor of the hypothesis that severe pneumococcal infections are the result of a single-gene lesion.

CHAPTER 9:

MATERIAL AND METHODS

Recruitment of study participants

Our research protocol was approved by the Institutional Review Board of Rockefeller University. Written informed consent for participation in the study was obtained from all patients and family members studied.

DNA extraction

Genomic DNA was isolated from the peripheral blood mononuclear cells (PBMCs) or the EBV-B cells of the patients and the family members using standard procedures.

Whole-exome sequencing

Genomic DNA (3µg) extracted from EBV-B cells from the patient was sheared with a Covaris S2 Ultrasonicator (Covaris). An adapter-ligated library was prepared with the Paired-End Sample Prep kit V1 (Illumina). Exome capture was performed with the SureSelect Human All Exon kit (Agilent echnologies). Single-end sequencing or paired-end sequencing was performed on an Illumina Genome Analyzer IIx (Illumina), generating 72-base reads. Table 8 provides more details on the kit used as well as the coverage for each exome.

Sequence alignment, variant calling and annotation

We used BWA aligner[140] to align the sequences with the human genome reference sequence (hg19 build). Downstream processing was carried out with the Genome analysis toolkit (GATK)[88], SAMtools[89] and Picard Tools (http://picard.sourceforge.net). Substitution calls were made with a GATK UnifiedGenotyper, whereas indel calls were made with a GATK IndelGenotyperV2. All calls with a read coverage ≤2x and a Phred-scaled SNP quality of ≤20 were filtered out. All the variants were annotated with the GATK Genomic Annotator. For filtering purposes we used 1000 genomes (May 2011 release based on 1094 individuals), dbSNP134 and an in-house database of about 600 individuals.

Analysis of whole-exome sequencing

P-values for each gene were calculated based on the script of Ionita-Laza[109]. The data from 623 individuals was used. 592 were controls, 30 were cases and one was removed. A total of 1,859,802 variants were analyzed. All the genes with more than 2,000 variants were removed (n=20). The weighted sum with all rare variants (WS-R) was used to calculate the p-values for 21,185 genes.

Sanger Sequencing of the coding region of RPSA

The coding exons and intron/exon boundaries of RPSA were amplified by PCR and then sequenced using the Big Dye v3.1 kit on a 3730DNA Analyzer (Applied Biosystems).

Primers used for sequencing of *RPSA*.

Exon	Forward/Reverse	Sequence (5' to 3')
1	Forward	CCATACCATCTGTTCAATCTTG
1	Reverse	CCGAAGGGATAGCAACACAG
2	Forward	TGTGCCCTACTTAACTCAATGG
2	Reverse	CAGAACTAACCACGAGAATAGC
3	Forward	GCCTCTATGACTGGAGTTTGG
3	Reverse	TGTCTCACAATGTCTGTAAGCT
4	Forward	GGAGTAGTAGTGATTAAGTTGGC
4	Reverse	CCTACACTTGCGAGTTCATCG
5	Forward	GATGTAAGAGCCAGGAAGGTGC
5	Reverse	TCCTAGAAGCACAGTCCAAGTC
6	Forward	CTTGTGGTTACATAAGCAAATTGG
6	Reverse	GCAGTCTATCAGATGCTGTAGC
7	Forward	GCTACAGCATCTGATAGACTGC
7	Reverse	GAAGTACAACTCATGTCGAGAC

Multiplex Ligation-dependent Probe Amplification (MLPA)

For the MLPA experiment we used the EK1-FAM reagent kit from MRC-Holland. We followed their protocol. 6 probes in the RPSA locus were designed as well as 4 control probes: 1 in BCAP31 on chr. X, 1 in SRY on chr. Y, 1 in SS18 on chr. 18, and 1 in IFNγ on chr.12. Fragment analysis was done on Applied Biosytem 3730 Analyzer.

Gene (exon)	Left / Right	Sequence
BCAP31	L	GGGTTCCCTAAGGGTTGGAACAGGAGCCTGAAGGCTGACCT
BCAP31	R	GCAGAAGCTAAAGGACGAGCTGGACTTCTAGATTGGATCTTGCTGGCAC
SS18	L	GGGTTCCCTAAGGGTTGGACAGCAGCCACCTATGGGAATGATG
SS18	R	GGTCAAGTTAACCAAGGCAATCATATGATGTCTAGATTGGATCTTGCTGGCAC
SRY	L	GGGTTCCCTAAGGGTTGGACAGCGAAGTGCAACTGGACAACAGGTTGTA
SRY	R	CAGGGATGACTGTACGAAAGCCACACACTCTAGATTGGATCTTGCTGGCAC
RPSA (1)	L	GGGTTCCCTAAGGGTTGGAAAAAGCCCAGTCCCGGCAGCGTTCTT
RPSA (1)	R	CTCCGGCTCCGCCCTTCTTCCGCTCGACTTTCTTTTCTAGATTGGATCTTGCTGGCAC
RPSA (2)	L	GGGTTCCCTAAGGGTTGGAGACTGACTGACTGACTGACATGAAAAGAAATAGGTTGCTCTCTTATTTCA
RPSA (2)	R	GATTCCCGTCGTAACTTAAAGGGAAATCTAGATTGGATCTTGCTGGCAC
RPSA (3)	L	GGGTTCCCTAAGGGTTGGACAAGTGTTACAAATCCTTCTGCCCTCACTTAG
RPSA (3)	R	GCATCTATATCATAAATCTCAAGAGGACCTGGGTCTAGATTGGATCTTGCTGGCAC
RPSA (4)	L	GGGTTCCCTAAGGGTTGGAGAATATCGAGTACCACTAACTTTTAAATTCTTCAAAG
RPSA (4)	R	AGGGCTGTGCTGAAGTTTGCTGCTGCCACTGGAGCTCTAGATTGGATCTTGCTGGCAC
RPSA (5)	L	GGGTTCCCTAAGGGTTGGAGGTCATGCCTGATCTGTACTTCTACAGAGATCCTGAAGAG
RPSA (5)	R	GTAAGCTTCTCCAAAGGCTTGTGGTTACATAAGCAAATTGGTCTAGATTGGATCTTGCTGGCAC
RPSA (7)	L	GGGTTCCCTAAGGGTTGGATACATTGAGGACAGAGCTGATGGCTTTTTTTTGGTATTCTCTTAACAG
RPSA (7)	R	AAGACTGGAGCGCTCAGCCTGCCACGGAAGACTGGTCTGCAGCTCCCTCTAGATTGGATCTTGCTGGCAC
IFNγ	L	GGGTTCCCTAAGGGTTGGATTGACTTCTCAGACTCTAAGGTCAAGATTAGCATTAAAAAGGT
IFNγ	R	AATAGGAAATGTTTACAAATTAAAGTCAAAAAGGTCCTTAAATCTAGATTGGATCTTGCTGGCAC

Multiple species alignment

Conservation of RPSA [GenBank ID: NP_002286.2] was assessed by aligning all orthologs of RPSA obtained from release 65 of Homologene (http://www.ncbi.nlm.nih.gov/homologene, Homologene ID= 68249).

Quantitative RT-PCR

Poly A+ RNA was purchased from Clontech and was used to generate cDNA with the SuperScript III First-Strand Synthesis System (Invitrogen). Human fetal spleen (cat #636154), human fetal heart (cat #636156), human fetal brain (cat #636106) and human adult spleen (cat #636121) were purchased. The expression of RPSA was assessed in the TaqMan gene expression assay (cat. Hs03046712_g1, Applied Biosystems), with normalization of the results as a function of GUS (β-glucuronidase) or VCL gene expression.

Western Blot

Laminin-R (H-141) (sc-20979) was used for all blots to detect endogenous RPSA.

Rpsa +/- mouse

The strategy used to create a *RPSA* knockout mouse was designed and implemented by InGenious Targeting Laboratories. A 14kb DNA fragment flanked by Not1 restriction sites containing exons 1-7 of the mouse *RPSA* gene was isolated from a 129/SvEv lambda genomic library. A *Neo* cassette was used to replace sequences including exons 1-4 of *RPSA*. The sequence replaced begins with 5′-GCAATACCACGGCCCAGCAGCAGTTAG-3′ and ends with 5′-GAGTGGCTCCTGTGGCAGCAGCAA-3′. The targeting vector was linearized by Not1 and electroporated into iTL1 (129/SvEv) embryonic stem cells. After G418 selection surviving colonies were screened by Southern blot analysis using a 570 bp PCR fragment produced using the primer pair: PB1 (4.3kb upstream of exon 1) 5′-CTTTCTCCCTTGTGTCCATCACAG-3′ and PB2 (3.7kb upstream of exon1) 5′-TTTCCTTGCTGAGGCATCTCACTG-3′. For Southern blotting mouse genomic DNA was cleaved with HindIII and hybridized with the probe described above. The expected

size of the wild-type fragment hybridizing with this probe was 7.8kb and the knockout fragment 2.8kb.

ES cell lines containing the *RPSA* targeted construct were microinjected into C57BL/6J blastocysts and the resulting chimeric mice were set up for germline transmission of the disrupted *RPSA* allele. Mice were screened using the Southern blot strategy outlined above or using two PCR primer pairs. One primer pair was to the sequences within the *Neo* cassette: 5′-CTCTTCAGCAATATCACGGGTAGCCA-3′ and 5′-GCTCCTGCCGAGAAAGTATCCATCAT-3′ which gives a 371bp fragment in the knockouts and 5′-TTACACAATGTCCGGAGCCCTTGA-3′ and 5′-ACAAGCCAGAGTAGCACTCCAACA-3′ which gives a 240bp fragment from wild-type mice.

Mice were housed in Licia's Selleri mouse room at Weill Cornell Medical school and were phenotyped by the laboratory of Comparative Pathology at MSKCC. Phenotyping included complete blood count with blood smear, measure and weigh of the spleen, analysis of the spleen shape and histology of the spleen.

Rpsa +/- zebrafish

Morpholinos were designed to target either the 5'-UTR of Rpsa, before the ATG or the splicing of exon 3. The number of base pairs in exon 3 is not a multiple of 3 and thus impairing the splicing of this exon should lead to a premature STOP. The sequences are:

Rpsa-5UTR-MO-1: CCACTTTCCCTGCGTAACGAGTTTA
Rpsa-intr2ex3-MO-2: ACGCCTGCAGAAATACAATTAGAGC
Rpsa-ex3intr3-MO-3: ACGCCTGCAGAAATACAATTAGAGC

Morpholinos knocking down p53 were also used to prevent excessive toxicity.

Pre-rRNA processing

These experiments were performed by Steven R Ellis using the protocol described previously[141].

CONCLUDING REMARKS

The field of human genetics of infectious diseases is a small subgroup of the fast-growing field of human genetics. In four years, I became fascinated by human genetics and the impact that small changes in the DNA or RNA can have on the human body. I was also very fortunate to be part of this field during a major revolution: the advent of whole-exome sequencing. This technology changed the outcome of this project, and is changing the entire field. The pace of discovery of the genetic etiology of Mendelian diseases is increasing exponentially, and there is no doubt that exome sequencing will come to the clinic soon. A likely, most interesting change that I predict is on the horizon is the merge of the fields of the human population genetics and Mendelian genetics. In my opinion one of the main dividing factors between the two schools was purely technical. It was too expensive for population geneticists to sequence genes entirely in a large cohort. That is now possible. The Mendelian geneticists will have to adapt. Firstly, there is a need to find more efficient methods to recruit patients for Mendelian disorders. Secondly, there is a need to use more statistics to be on par with the population geneticists. The example of *NKX2-5* is the perfect example to illustrate this. Despite all the functional work that 'proves' the mutation to be disease-causing in family B, the genetic work on a larger number of patients showed that in fact the mutation was not disease-causing.

Clinically, the low cost of sequencing will also improve diagnosis and prognosis. Rare diseases are usually not well diagnosed. For example the lab started to recruit patients with rhinoscleroma, an infectious disease affecting the upper respiratory tract. After sequencing the exome of two patients from the same family, we found that these patients suffered from SLC29A3 deficiency[59], a disease that has nothing in common with rhinoscleroma. Actually we might not need physicians in the future for the diagnosis, at least for pediatric infections or pediatric diseases in general. Everyone (or a cloud-based program) will be able to proffer a diagnosis based on the patient's genome. The availability of more genomes will also enable research on many more traits, in particular non disease-causing traits. If a platform where individuals can share their genome, as well as their phenotype is equipped with open-access genome analysis programs, it might lead to many and very interesting discoveries.

REFERENCES

1. Casanova JL, Abel L (2005) Inborn errors of immunity to infection: the rule rather than the exception. J Exp Med 202:197-201
2. Nicolle C (1928) "Physiology or Medicine 1928 - Presentation Speech". http://wwwnobelprizeorg/nobel_prizes/medicine/laureates/1928/presshtml
3. Casanova JL, Jouanguy E, Lamhamedi S, Blanche S, Fischer A (1995) Immunological conditions of children with BCG disseminated infection. Lancet 346:581
4. Allison AC (1954) Protection afforded by sickle-cell trait against subtertian malareal infection. Br Med J 1:290-294
5. Bruton OC (1952) Agammaglobulinemia. Pediatrics 9:722-728
6. Tsukada S, Saffran DC, Rawlings DJ, Parolini O, Allen RC, Klisak I, Sparkes RS, Kubagawa H, Mohandas T, Quan S, et al. (1993) Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. Cell 72:279-290
7. Vetrie D, Vorechovsky I, Sideras P, Holland J, Davies A, Flinter F, Hammarstrom L, Kinnon C, Levinsky R, Bobrow M, et al. (1993) The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. Nature 361:226-233
8. Alcais A, Quintana-Murci L, Thaler DS, Schurr E, Abel L, Casanova JL (2010) Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity? Ann N Y Acad Sci 1214:18-33
9. Casanova JL, Abel L (2007) Human genetics of infectious diseases: a unified theory. EMBO J 26:915-922
10. Casanova JL, Abel L (2007) Primary immunodeficiencies: a field in its infancy. Science 317:617-619
11. Bogunovic D, Byun M, Durfee LA, Abhyankar A, Sanal O, Mansouri D, Salem S, et al. (2012) Mycobacterial Disease and Impaired IFN-gamma Immunity in Humans with Inherited ISG15 Deficiency. Science
12. von Bernuth H, Picard C, Jin Z, Pankla R, Xiao H, Ku CL, Chrabieh M, et al. (2008) Pyogenic bacterial infections in humans with MyD88 deficiency. Science 321:691-696
13. Zhang SY, Jouanguy E, Ugolini S, Smahi A, Elain G, Romero P, Segal D, et al. (2007) TLR3 deficiency in patients with herpes simplex encephalitis. Science 317:1522-1527
14. Puel A, Cypowyj S, Bustamante J, Wright JF, Liu L, Lim HK, Migaud M, Israel L, Chrabieh M, Audry M, Gumbleton M, Toulon A, Bodemer C, El-Baghdadi J, Whitters M, Paradis T, Brooks J, Collins M,

Wolfman NM, Al-Muhsen S, Galicchio M, Abel L, Picard C, Casanova JL (2011) Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. Science 332:65-68
15. Mahlaoui N, Minard-Colin V, Picard C, Bolze A, Ku CL, Tournilhac O, Gilbert-Dussardier B, Pautard B, Durand P, Devictor D, Lachassinne E, Guillois B, Morin M, Gouraud F, Valensi F, Fischer A, Puel A, Abel L, Bonnet D, Casanova JL (2011) Isolated congenital asplenia: a French nationwide retrospective survey of 20 cases. J Pediatr 158:142-148, 148 e141
16. Rogers ZR, Wang WC, Luo Z, Iyer RV, Shalaby-Rana E, Dertinger SD, Shulkin BL, Miller JH, Files B, Lane PA, Thompson BW, Miller ST, Ware RE (2011) Biomarkers of splenic function in infants with sickle cell anemia: baseline data from the BABY HUG Trial. Blood 117:2614-2617
17. Ivemark BI (1955) Implications of agenesis of the spleen on the pathogenesis of conotruncus anomalies in childhood; an analysis of the heart malformations in the splenic agenesis syndrome, with fourteen new cases. Acta Paediatr Suppl 44:7-110
18. Belmont JW, Mohapatra B, Towbin JA, Ware SM (2004) Molecular genetics of heterotaxy syndromes. Curr Opin Cardiol 19:216-220
19. Myerson RM, Koelle WA (1956) Congenital absence of the spleen in an adult; report of a case associated with recurrent Waterhouse-Friderichsen syndrome. N Engl J Med 254:1131-1132
20. Ahmed SA, Zengeya S, Kini U, Pollard AJ (2009) Familial isolated congenital asplenia: case report and literature review. Eur J Pediatr 169:315-318
21. Almoznino-Sarafian D, Dotan E, Sandbank J, Gorelik O, Chachashvily S, Shteinshnaider M, Cohen N (2007) Unusual manifestations of myelofibrosis in a patient with congenital asplenia. Acta Haematol 118:226-230
22. Araujo AR, Maciel I, Lima L, Chacim I, Barbot J (2008) Congenital asplenia and severe visceral toxocariasis. Pediatr Infect Dis J 27:478
23. Chanet V, Tournilhac O, Dieu-Bellamy V, Boiret N, Spitz P, Baud O, Darcha C, Travade P, Laurichesse H (2000) Isolated spleen agenesis: a rare cause of thrombocytosis mimicking essential thrombocythemia. Haematologica 85:1211-1213
24. Dyke MP, Martin RP, Berry PJ (1991) Septicaemia and adrenal haemorrhage in congenital asplenia. Arch Dis Child 66:636-637
25. Feder HM, Jr., Pearson HA (1999) Assessment of splenic function in familial asplenia. N Engl J Med 341:210-212
26. Germing U, Perings C, Steiner S, Peters AJ, Heintzen MP, Aul C (1999) Congenital asplenia detected in a 60 year old patient with septicemia. Eur J Med Res 4:283-285
27. Gilbert B, Menetrey C, Belin V, Brosset P, de Lumley L, Fisher A (2002) Familial isolated congenital asplenia: a rare, frequently hereditary dominant condition, often detected too late as a cause of overwhelming pneumococcal sepsis. Report of a new case and review of 31 others. Eur J Pediatr 161:368-372
28. Gillis J, Harvey J, Isaacs D, Freelander M, Wyeth B (1992) Familial asplenia. Arch Dis Child 67:665-666
29. Gopal V, Bisno AL (1977) Fulminant pneumococcal infections in 'normal' asplenic hosts. Arch Intern Med 137:1526-1530
30. Halbertsma FJ, Neeleman C, Weemaes CM, van Deuren M (2005) The absent and vanishing spleen: congenital asplenia and hyposplenism--two case reports. Acta Paediatr 94:369-371
31. Honigman R, Lanzkowsky P (1979) Isolated congenital asplenia: an occult case of overwhelming sepsis. Am J Dis Child 133:552-553
32. Kanthan R, Moyana T, Nyssen J (1999) Asplenia as a cause of sudden unexpected death in childhood. Am J Forensic Med Pathol 20:57-59
33. Katcher AL (1980) Familial asplenia, other malformations, and sudden death. Pediatrics 65:633-635
34. Kevy SV, Tefft M, Vawier GF, Rosen FS (1968) Hereditary splenic hypoplasia. Pediatrics 42:752-757

35. Lindor NM, Smithson WA, Ahumada CA, Michels VV, Opitz JM (1995) Asplenia in two father-son pairs. Am J Med Genet 56:10-11
36. Rose C, Quesnel B, Facon T, Fenaux P, Jouet JP, Bauters F (1993) [Congenital asplenia, a differential diagnosis of essential thrombocythemia]. Presse Med 22:1748
37. Schutze GE, Mason EO, Jr., Barson WJ, Kim KS, Wald ER, Givner LB, Tan TQ, Bradley JS, Yogev R, Kaplan SL (2002) Invasive pneumococcal infections in children with asplenia. Pediatr Infect Dis J 21:278-282
38. Takahashi F, Uchida K, Nagaoka T, Honma N, Cui R, Yoshioka M, Morio Y, Suzuki T, Tominaga S, Takahashi K, Fukuchi Y (2008) Isolated congenital spleen agenesis: a rare cause of chronic thromboembolic pulmonary hypertension in an adult. Respirology 13:913-915
39. Thiruppathy K, Privitera A, Jain K, Gupta S (2008) Congenital asplenia and group B streptococcus sepsis in the adult: case report and review of the literature. FEMS Immunol Med Microbiol 53:437-439
40. Uchida Y, Matsubara K, Wada T, Oishi K, Morio T, Takada H, Iwata A, Yura K, Kamimura K, Nigami H, Fukaya T (2011) Recurrent bacterial meningitis by three different pathogens in an isolated asplenic child. J Infect Chemother
41. Waldman JD, Rosenthal A, Smith AL, Shurin S, Nadas AS (1977) Sepsis and congenital asplenia. J Pediatr 90:555-559
42. Imashuku S, Kudo N, Kubo K, Takahashi N, Tohyama K (2012) Persistent thrombocytosis in elderly patients with rare hyposplenias that mimic essential thrombocythemia. Int J Hematol 95:702-705
43. Picard C, Puel A, Bustamante J, Ku CL, Casanova JL (2003) Primary immunodeficiencies associated with pneumococcal disease. Curr Opin Allergy Clin Immunol 3:451-459
44. Kruetzmann S, Rosado MM, Weber H, Germing U, Tournilhac O, Peter HH, Berner R, Peters A, Boehm T, Plebani A, Quinti I, Carsetti R (2003) Human immunoglobulin M memory B cells controlling Streptococcus pneumoniae infections are generated in the spleen. J Exp Med 197:939-945
45. Brueckner M (2007) Heterotaxia, congenital heart disease, and primary ciliary dyskinesia. Circulation 115:2793-2795
46. Koss M, Bolze A, Brendolan A, Saggese M, Capellini TD, Bojilova E, Boisson B, Prall OW, Elliott DA, Solloway M, Lenti E, Hidaka C, Chang CP, Mahlaoui N, Harvey RP, Casanova JL, Selleri L (2012) Congenital asplenia in mice and humans with mutations in a Pbx/Nkx2-5/p15 module. Dev Cell 22:913-926
47. Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, Mountain JL, Goldman SM, Tanner CM, Langston JW, Wojcicki A, Eriksson N (2011) Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. PLoS Genet 7:e1002141
48. Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide BM, Schjeide LM, et al. (2012) Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. PLoS Genet 8:e1002548
49. Brendolan A, Ferretti E, Salsi V, Moses K, Quaggin S, Blasi F, Cleary ML, Selleri L (2005) A Pbx1-dependent genetic and transcriptional network regulates spleen ontogeny. Development 132:3113-3126
50. Hecksher-Sorensen J, Watson RP, Lettice LA, Serup P, Eley L, De Angelis C, Ahlgren U, Hill RE (2004) The splanchnic mesodermal plate directs spleen and pancreatic laterality, and is regulated by Bapx1/Nkx3.2. Development 131:4665-4675
51. Herzer U, Crocoll A, Barton D, Howells N, Englert C (1999) The Wilms tumor suppressor gene wt1 is required for development of the spleen. Curr Biol 9:837-840

52. Lu J, Chang P, Richardson JA, Gan L, Weiler H, Olson EN (2000) The basic helix-loop-helix transcription factor capsulin controls spleen organogenesis. Proc Natl Acad Sci U S A 97:9525-9530
53. Pabst O, Zweigerdt R, Arnold HH (1999) Targeted disruption of the homeobox transcription factor Nkx2-3 in mice results in postnatal lethality and abnormal development of small intestine and spleen. Development 126:2215-2225
54. Sock E, Rettig SD, Enderich J, Bosl MR, Tamm ER, Wegner M (2004) Gene targeting reveals a widespread role for the high-mobility-group transcription factor Sox11 in tissue remodeling. Mol Cell Biol 24:6635-6644
55. Katsanou V, Milatos S, Yiakouvaki A, Sgantzis N, Kotsoni A, Alexiou M, Harokopos V, Aidinis V, Hemberger M, Kontoyiannis DL (2009) The RNA-binding protein Elavl1/HuR is essential for placental branching morphogenesis and embryonic development. Mol Cell Biol 29:2762-2776
56. Kim BM, Miletich I, Mao J, McMahon AP, Sharpe PA, Shivdasani RA (2007) Independent functions and mechanisms for homeobox gene Barx1 in patterning mouse stomach and spleen. Development 134:3603-3613
57. Roberts CW, Shutter JR, Korsmeyer SJ (1994) Hox11 controls the genesis of the spleen. Nature 368:747-749
58. Nagel BH, Williams H, Stewart L, Paul J, Stumper O (2005) Splenic state in surviving patients with visceral heterotaxy. Cardiol Young 15:469-473
59. Bolze A, Abhyankar A, Grant AV, Patel B, Yadav R, Byun M, Caillez D, Emile JF, Pastor-Anglada M, Abel L, Puel A, Govindarajan R, de Pontual L, Casanova JL (2012) A mild form of SLC29A3 disorder: a frameshift deletion leads to the paradoxical translation of an otherwise noncoding mRNA splice variant. PLoS One 7:e29708
60. Raman R, Al-Ali SY, Poole CA, Dawson BV, Carman JB, Calder L (2003) Isomerism of the right atrial appendages: clinical, anatomical, and microscopic study of a long-surviving case with asplenia and ciliary abnormalities. Clin Anat 16:269-276
61. Kennedy MP, Omran H, Leigh MW, Dell S, Morgan L, Molina PL, Robinson BV, Minnix SL, Olbrich H, Severin T, Ahrens P, Lange L, Morillas HN, Noone PG, Zariwala MA, Knowles MR (2007) Congenital heart disease and other heterotaxic defects in a large cohort of patients with primary ciliary dyskinesia. Circulation 115:2814-2821
62. Bartoloni L, Blouin JL, Pan Y, Gehrig C, Maiti AK, Scamuffa N, Rossier C, Jorissen M, Armengot M, Meeks M, Mitchison HM, Chung EM, Delozier-Blanchet CD, Craigen WJ, Antonarakis SE (2002) Mutations in the DNAH11 (axonemal heavy chain dynein type 11) gene cause one form of situs inversus totalis and most likely primary ciliary dyskinesia. Proc Natl Acad Sci U S A 99:10282-10286
63. Budny B, Chen W, Omran H, Fliegauf M, Tzschach A, Wisniewska M, Jensen LR, Raynaud M, Shoichet SA, Badura M, Lenzner S, Latos-Bielenska A, Ropers HH (2006) A novel X-linked recessive mental retardation syndrome comprising macrocephaly and ciliary dysfunction is allelic to oral-facial-digital type I syndrome. Hum Genet 120:171-178
64. Castleman VH, Romio L, Chodhari R, Hirst RA, de Castro SC, Parker KA, Ybot-Gonzalez P, Emes RD, Wilson SW, Wallis C, Johnson CA, Herrera RJ, Rutman A, Dixon M, Shoemark A, Bush A, Hogg C, Gardiner RM, Reish O, Greene ND, O'Callaghan C, Purton S, Chung EM, Mitchison HM (2009) Mutations in radial spoke head protein genes RSPH9 and RSPH4A cause primary ciliary dyskinesia with central-microtubular-pair abnormalities. Am J Hum Genet 84:197-209
65. Duriez B, Duquesnoy P, Escudier E, Bridoux AM, Escalier D, Rayet I, Marcos E, Vojtek AM, Bercher JF, Amselem S (2007) A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia. Proc Natl Acad Sci U S A 104:3336-3341

66. Loges NT, Olbrich H, Fenske L, Mussaffi H, Horvath J, Fliegauf M, Kuhl H, Baktai G, Peterffy E, Chodhari R, Chung EM, Rutman A, O'Callaghan C, Blau H, Tiszlavicz L, Voelkel K, Witt M, Zietkiewicz E, Neesen J, Reinhardt R, Mitchison HM, Omran H (2008) DNAI2 mutations cause primary ciliary dyskinesia with defects in the outer dynein arm. Am J Hum Genet 83:547-558
67. Moore A, Escudier E, Roger G, Tamalet A, Pelosse B, Marlin S, Clement A, Geremek M, Delaisi B, Bridoux AM, Coste A, Witt M, Duriez B, Amselem S (2006) RPGR is mutated in patients with a complex X linked phenotype combining primary ciliary dyskinesia and retinitis pigmentosa. J Med Genet 43:326-333
68. Olbrich H, Haffner K, Kispert A, Volkel A, Volz A, Sasmaz G, Reinhardt R, Hennig S, Lehrach H, Konietzko N, Zariwala M, Noone PG, Knowles M, Mitchison HM, Meeks M, Chung EM, Hildebrandt F, Sudbrak R, Omran H (2002) Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. Nat Genet 30:143-144
69. Omran H, Kobayashi D, Olbrich H, Tsukahara T, Loges NT, Hagiwara H, Zhang Q, Leblond G, O'Toole E, Hara C, Mizuno H, Kawano H, Fliegauf M, Yagi T, Koshida S, Miyawaki A, Zentgraf H, Seithe H, Reinhardt R, Watanabe Y, Kamiya R, Mitchell DR, Takeda H (2008) Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. Nature 456:611-616
70. Pennarun G, Escudier E, Chapelin C, Bridoux AM, Cacheux V, Roger G, Clement A, Goossens M, Amselem S, Duriez B (1999) Loss-of-function mutations in a human gene related to Chlamydomonas reinhardtii dynein IC78 result in primary ciliary dyskinesia. Am J Hum Genet 65:1508-1519
71. Bamford RN, Roessler E, Burdine RD, Saplakoglu U, dela Cruz J, Splitt M, Goodship JA, Towbin J, Bowers P, Ferrero GB, Marino B, Schier AF, Shen MM, Muenke M, Casey B (2000) Loss-of-function mutations in the EGF-CFC gene CFC1 are associated with human left-right laterality defects. Nat Genet 26:365-369
72. Gebbia M, Ferrero GB, Pilia G, Bassi MT, Aylsworth A, Penman-Splitt M, Bird LM, Bamforth JS, Burn J, Schlessinger D, Nelson DL, Casey B (1997) X-linked situs abnormalities result from mutations in ZIC3. Nat Genet 17:305-308
73. Hilton EN, Manson FD, Urquhart JE, Johnston JJ, Slavotinek AM, Hedera P, Stattin EL, Nordgren A, Biesecker LG, Black GC (2007) Left-sided embryonic expression of the BCL-6 corepressor, BCOR, is required for vertebrate laterality determination. Hum Mol Genet 16:1773-1782
74. Karkera JD, Lee JS, Roessler E, Banerjee-Basu S, Ouspenskaia MV, Mez J, Goldmuntz E, Bowers P, Towbin J, Belmont JW, Baxevanis AD, Schier AF, Muenke M (2007) Loss-of-function mutations in growth differentiation factor-1 (GDF1) are associated with congenital heart defects in humans. Am J Hum Genet 81:987-994
75. Kosaki K, Bassi MT, Kosaki R, Lewin M, Belmont J, Schauer G, Casey B (1999) Characterization and mutation analysis of human LEFTY A and LEFTY B, homologues of murine genes implicated in left-right axis development. Am J Hum Genet 64:712-721
76. Mohapatra B, Casey B, Li H, Ho-Dawson T, Smith L, Fernbach SD, Molinari L, Niesh SR, Jefferies JL, Craigen WJ, Towbin JA, Belmont JW, Ware SM (2009) Identification and functional characterization of NODAL rare variants in heterotaxy and isolated cardiovascular malformations. Hum Mol Genet 18:861-871
77. Peeters H, Devriendt K (2006) Human laterality disorders. Eur J Med Genet 49:349-362
78. Robinson SW, Morris CD, Goldmuntz E, Reller MD, Jones MA, Steiner RD, Maslen CL (2003) Missense mutations in CRELD1 are associated with cardiac atrioventricular septal defects. Am J Hum Genet 72:1047-1052
79. Roessler E, Ouspenskaia MV, Karkera JD, Velez JI, Kantipong A, Lacbawan F, Bowers P, Belmont JW, Towbin JA, Goldmuntz E, Feldman B, Muenke M (2008) Reduced NODAL signaling strength via

mutation of several pathway members including FOXH1 is linked to human heart defects and holoprosencephaly. Am J Hum Genet 83:18-29
80. Schon P, Tsuchiya K, Lenoir D, Mochizuki T, Guichard C, Takai S, Maiti AK, Nihei H, Weil J, Yokoyama T, Bouvagnet P (2002) Identification, genomic organization, chromosomal mapping and mutation analysis of the human INV gene, the ortholog of a murine gene implicated in left-right axis development and biliary atresia. Hum Genet 110:157-165
81. Pollak U, Bar-Sever Z, Hoffer V, Marcus N, Scheuerman O, Garty BZ (2009) Asplenia and functional hyposplenism in autoimmune polyglandular syndrome type 1. Eur J Pediatr 168:233-235
82. McDonald DR, Massaad MJ, Johnston A, Keles S, Chatila T, Geha RS, Pai SY (2010) Successful engraftment of donor marrow after allogeneic hematopoietic cell transplantation in autosomal-recessive hyper-IgE syndrome caused by dedicator of cytokinesis 8 deficiency. J Allergy Clin Immunol 126:1304-1305 e1303
83. Leahy RT, Philip RK, Gibbons RJ, Fisher C, Suri M, Reardon W (2005) Asplenia in ATR-X syndrome: a second report. Am J Med Genet A 139:37-39
84. Yachie A, Niida Y, Wada T, Igarashi N, Kaneda H, Toma T, Ohta K, Kasahara Y, Koizumi S (1999) Oxidative stress causes enhanced endothelial cell injury in human heme oxygenase-1 deficiency. J Clin Invest 103:129-135
85. Radhakrishnan N, Yadav SP, Sachdeva A, Pruthi PK, Sawhney S, Piplani T, Wada T, Yachie A (2011) Human heme oxygenase-1 deficiency presenting with hemolysis, nephritis, and asplenia. J Pediatr Hematol Oncol 33:74-78
86. Radhakrishnan N, Yadav SP, Sachdeva A, Wada T, Yachie A (2011) An interesting tetrad of asplenia, inflammation, hemolysis, and nephritis. Pediatr Hematol Oncol 28:723-726
87. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ (2010) Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42:30-35
88. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-1303
89. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079
90. Chinnaiyan AM, O'Rourke K, Tewari M, Dixit VM (1995) FADD, a novel death domain-containing protein, interacts with the death domain of Fas and initiates apoptosis. Cell 81:505-512
91. Balachandran S, Thomas E, Barber GN (2004) A FADD-dependent innate immune mechanism in mammalian cells. Nature 432:401-405
92. Tourneur L, Chiocchia G (2010) FADD: a regulator of life and death. Trends Immunol 31:260-269
93. Scott FL, Stec B, Pop C, Dobaczewska MK, Lee JJ, Monosov E, Robinson H, Salvesen GS, Schwarzenbacher R, Riedl SJ (2009) The Fas-FADD death domain complex structure unravels signalling by receptor clustering. Nature 457:1019-1022
94. Peter ME, Krammer PH (2003) The CD95(APO-1/Fas) DISC and beyond. Cell Death Differ 10:26-35
95. Rieux-Laucat F, Le Deist F, Hivroz C, Roberts IA, Debatin KM, Fischer A, de Villartay JP (1995) Mutations in Fas associated with human lymphoproliferative syndrome and autoimmunity. Science 268:1347-1349
96. Del-Rey M, Ruiz-Contreras J, Bosque A, Calleja S, Gomez-Rial J, Roldan E, Morales P, Serrano A, Anel A, Paz-Artal E, Allende LM (2006) A homozygous Fas ligand gene mutation in a patient causes a new type of autoimmune lymphoproliferative syndrome. Blood 108:1306-1312
97. Chun HJ, Zheng L, Ahmad M, Wang J, Speirs CK, Siegel RM, Dale JK, Puck J, Davis J, Hall CG, Skoda-Smith S, Atkinson TP, Straus SE, Lenardo MJ (2002) Pleiotropic defects in lymphocyte activation caused by caspase-8 mutations lead to human immunodeficiency. Nature 419:395-399

98. Wang J, Zheng L, Lobito A, Chan FK, Dale J, Sneller M, Yao X, Puck JM, Straus SE, Lenardo MJ (1999) Inherited human Caspase 10 mutations underlie defective lymphocyte and dendritic cell apoptosis in autoimmune lymphoproliferative syndrome type II. Cell 98:47-58
99. Wilson NS, Dixit V, Ashkenazi A (2009) Death receptor signal transducers: nodes of coordination in immune signaling networks. Nat Immunol 10:348-355
100. Balachandran S, Venkataraman T, Fisher PB, Barber GN (2007) Fas-associated death domain-containing protein-mediated antiviral innate immune signaling involves the regulation of Irf7. J Immunol 178:2429-2439
101. Chapgier A, Wynn RF, Jouanguy E, Filipe-Santos O, Zhang S, Feinberg J, Hawkins K, Casanova JL, Arkwright PD (2006) Human complete Stat-1 deficiency is associated with defective type I and II IFN responses in vitro but immunity to some low virulence viruses in vivo. J Immunol 176:5078-5083
102. Yeh WC, Pompa JL, McCurrach ME, Shu HB, Elia AJ, Shahinian A, Ng M, Wakeham A, Khoo W, Mitchell K, El-Deiry WS, Lowe SW, Goeddel DV, Mak TW (1998) FADD: essential for embryo development and signaling from some, but not all, inducers of apoptosis. Science 279:1954-1958
103. Byun M, Abhyankar A, Lelarge V, Plancoulaine S, Palanduz A, Telhan L, Boisson B, Picard C, Dewell S, Zhao C, Jouanguy E, Feske S, Abel L, Casanova JL (2010) Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. J Exp Med 207:2307-2312
104. Watanabe Y, Benson DW, Yano S, Akagi T, Yoshino M, Murray JC (2002) Two novel frameshift mutations in NKX2.5 result in novel features including visceral inversus and sinus venosus type ASD. J Med Genet 39:807-811
105. Lyons I, Parsons LM, Hartley L, Li R, Andrews JE, Robb L, Harvey RP (1995) Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes Dev 9:1654-1666
106. Prall OW, Menon MK, Solloway MJ, Watanabe Y, Zaffran S, Bajolle F, Biben C, McBride JJ, Robertson BR, Chaulet H, Stennard FA, Wise N, Schaft D, Wolstein O, Furtado MB, Shiratori H, Chien KR, Hamada H, Black BL, Saga Y, Robertson EJ, Buckingham ME, Harvey RP (2007) An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. Cell 128:947-959
107. Elliott DA, Solloway MJ, Wise N, Biben C, Costa MW, Furtado MB, Lange M, Dunwoodie S, Harvey RP (2006) A tyrosine-rich domain within homeodomain transcription factor Nkx2-5 is an essential element in the early cardiac transcriptional regulatory machinery. Development 133:1311-1322
108. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248-249
109. Ionita-Laza I, Makarov V, Yoon S, Raby B, Buxbaum J, Nicolae DL, Lin X (2011) Finding disease variants in Mendelian disorders by using sequence data: methods and applications. Am J Hum Genet 89:701-712
110. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30:e57
111. Brendolan A, Rosado MM, Carsetti R, Selleri L, Dear TN (2007) Development and function of the mammalian spleen. Bioessays 29:166-177
112. Renshaw SA, Trede NS (2012) A model 450 million years in the making: zebrafish and vertebrate immunity. Dis Model Mech 5:38-47

113. Langenau DM, Palomero T, Kanki JP, Ferrando AA, Zhou Y, Zon LI, Look AT (2002) Molecular cloning and developmental expression of Tlx (Hox11) genes in zebrafish (Danio rerio). Mech Dev 117:243-248
114. O'Donohue MF, Choesmel V, Faubladier M, Fichant G, Gleizes PE (2010) Functional dichotomy of ribosomal proteins during the synthesis of mammalian 40S ribosomal subunits. J Cell Biol 190:853-866
115. Armache JP, Jarasch A, Anger AM, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, Marquez V, Mielke T, Thomm M, Berninghausen O, Beatrix B, Soding J, Westhof E, Wilson DN, Beckmann R (2010) Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-A resolution. Proc Natl Acad Sci U S A 107:19748-19753
116. Ferlicot S, Emile JF, Le Bris JL, Cheron G, Brousse N (1997) [Congenital asplenia. A childhood immune deficit often detected too late]. Ann Pathol 17:44-46
117. Lango Allen H, Flanagan SE, Shaw-Smith C, De Franco E, Akerman I, Caswell R, Ferrer J, Hattersley AT, Ellard S (2011) GATA6 haploinsufficiency causes pancreatic agenesis in humans. Nat Genet 44:20-22
118. Herman M, Ciancanelli M, Ou YH, Lorenzo L, Klaudel-Dreszler M, Pauwels E, Sancho-Shimizu V, Perez de Diego R, Abhyankar A, Israelsson E, Guo Y, Cardon A, Rozenberg F, Lebon P, Tardieu M, Heropolitanska-Pliszka E, Chaussabel D, White MA, Abel L, Zhang SY, Casanova JL (2012) Heterozygous TBK1 mutations impair TLR3 immunity and underlie herpes simplex encephalitis of childhood. J Exp Med 209:1567-1582
119. de Beaucoudrey L, Samarina A, Bustamante J, Cobat A, Boisson-Dupuis S, Feinberg J, Al-Muhsen S, et al. (2010) Revisiting human IL-12Rbeta1 deficiency: a survey of 141 patients from 30 countries. Medicine (Baltimore) 89:381-402
120. Liu L, Okada S, Kong XF, Kreins AY, Cypowyj S, Abhyankar A, Toubiana J, et al. (2011) Gain-of-function human STAT1 mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis. J Exp Med 208:1635-1648
121. Auth D, Brawerman G (1992) A 33-kDa polypeptide with homology to the laminin receptor: component of translation machinery. Proc Natl Acad Sci U S A 89:4368-4372
122. Narla A, Ebert BL (2010) Ribosomopathies: human disorders of ribosome dysfunction. Blood 115:3196-3205
123. Draptchinskaia N, Gustavsson P, Andersson B, Pettersson M, Willig TN, Dianzani I, Ball S, Tchernia G, Klar J, Matsson H, Tentler D, Mohandas N, Carlsson B, Dahl N (1999) The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. Nat Genet 21:169-175
124. Farrar JE, Vlachos A, Atsidaftos E, Carlson-Donohoe H, Markello TC, Arceci RJ, Ellis SR, Lipton JM, Bodine DM (2011) Ribosomal protein gene deletions in Diamond-Blackfan anemia. Blood 118:6943-6951
125. Lipton JM, Ellis SR (2010) Diamond Blackfan anemia 2008-2009: broadening the scope of ribosome biogenesis disorders. Curr Opin Pediatr 22:12-19
126. Gazda HT, Sheen MR, Vlachos A, Choesmel V, O'Donohue MF, Schneider H, Darras N, Hasman C, Sieff CA, Newburger PE, Ball SE, Niewiadomska E, Matysiak M, Zaucha JM, Glader B, Niemeyer C, Meerpohl JJ, Atsidaftos E, Lipton JM, Gleizes PE, Beggs AH (2008) Ribosomal protein L5 and L11 mutations are associated with cleft palate and abnormal thumbs in Diamond-Blackfan anemia patients. Am J Hum Genet 83:769-780
127. Jaako P, Flygare J, Olsson K, Quere R, Ehinger M, Henson A, Ellis S, Schambach A, Baum C, Richter J, Larsson J, Bryder D, Karlsson S (2011) Mice with ribosomal protein S19 deficiency develop bone marrow failure and symptoms like patients with Diamond-Blackfan anemia. Blood 118:6087-6096

128. Warner JR, McIntosh KB (2009) How common are extraribosomal functions of ribosomal proteins? Mol Cell 34:3-11
129. Mecham RP, Hinek A, Griffin GL, Senior RM, Liotta LA (1989) The elastin receptor shows structural and functional similarities to the 67-kDa tumor cell laminin receptor. J Biol Chem 264:16652-16657
130. Jamieson KV, Hubbard SR, Meruelo D (2011) Structure-guided identification of a laminin binding site on the laminin receptor precursor. J Mol Biol 405:24-32
131. Scheiman J, Jamieson KV, Ziello J, Tseng JC, Meruelo D (2010) Extraribosomal functions associated with the C terminus of the 37/67 kDa laminin receptor are required for maintaining cell viability. Cell Death Dis 1:e42
132. Vana K, Weiss S (2006) A trans-dominant negative 37kDa/67kDa laminin receptor mutant impairs PrP(Sc) propagation in scrapie-infected neuronal cells. J Mol Biol 358:57-66
133. Ardini E, Sporchia B, Pollegioni L, Modugno M, Ghirelli C, Castiglioni F, Tagliabue E, Menard S (2002) Identification of a novel function for 67-kDa laminin receptor: increase in laminin degradation rate and release of motility fragments. Cancer Res 62:1321-1325
134. Givant-Horwitz V, Davidson B, Reich R (2004) Laminin-induced signaling in tumor cells: the role of the M(r) 67,000 laminin receptor. Cancer Res 64:3572-3579
135. Yun J, Jin HT, Lee YJ, Choi EK, Carp RI, Jeong BH, Kim YS (2011) The first report of RPSA polymorphisms, also called 37/67 kDa LRP/LR gene, in sporadic Creutzfeldt-Jakob disease (CJD). BMC Med Genet 12:108
136. Asano Y, Takashima S, Asakura M, Shintani Y, Liao Y, Minamino T, Asanuma H, Sanada S, Kim J, Ogai A, Fukushima T, Oikawa Y, Okazaki Y, Kaneda Y, Sato M, Miyazaki J, Kitamura S, Tomoike H, Kitakaze M, Hori M (2004) Lamr1 functional retroposon causes right ventricular dysplasia in mice. Nat Genet 36:123-130
137. Kondrashov N, Pusic A, Stumpf CR, Shimizu K, Hsieh AC, Xue S, Ishijima J, Shiroishi T, Barna M (2011) Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. Cell 145:383-397
138. Hsieh AC, Liu Y, Edlind MP, Ingolia NT, Janes MR, Sher A, Shi EY, Stumpf CR, Christensen C, Bonham MJ, Wang S, Ren P, Martin M, Jessen K, Feldman ME, Weissman JS, Shokat KM, Rommel C, Ruggero D (2012) The translational landscape of mTOR signalling steers cancer initiation and metastasis. Nature 485:55-61
139. Bousfiha AA, Jeddane L, Ailal F, Benhsaien I, Mahlaoui N, Casanova JL, Abel L (2012) Primary Immunodeficiency Diseases Worldwide: More Common than Generally Thought. J Clin Immunol
140. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760
141. Ebert BL, Pretz J, Bosco J, Chang CY, Tamayo P, Galili N, Raza A, Root DE, Attar E, Ellis SR, Golub TR (2008) Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. Nature 451:335-339