Paired Box 5 (PAX5) in B cell precursor acute lymphoblastic leukemia

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Franziska Auer

aus Nürnberg

Düsseldorf, Oktober 2016

aus der Klinik für Kinder-Onkologie, -Hämatologie und Klinische Immunologie

des Universitätsklinikums Düsseldorf

Referent: Prof. Dr. Arndt Borkhardt

Korreferent: Prof. Dr. Hermann Aberle

Gedruckt mit der Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Tag der mündlichen Prüfung: 28.11.2016

TABLE OF CONTENTS

Summary

Although significant improvements in the treatment of acute lymphoblastic leukemia (ALL) have been made over the last decade, with survival rates up to 90%, its treatment still requires intensive therapy and the prognosis for certain subtypes or relapse patients remains poor. In order to develop new individualized therapy approaches, which minimize toxicity and treat each patient according to their tumor-specific lesion rather than by a standard protocol, the genetic basis of the disease has to be elucidated. In this regard, inherited predisposing variants are of special interest, since the identification and study of these susceptibility loci enables potential therapeutic intervention even before a phenotypic manifestation of the disease.

In this study, we characterized three individuals with childhood B cell precursor ALL (BCP-ALL) from one family. We were able to identify a rare heterozygous germline variant *PAX5* c.547G>A with incomplete penetrance in the family using a whole exome sequencing (WES) approach. The mutation confers reduced transcriptional activity *in vitro*, by affecting the octapeptide domain of the paired box 5 (*PAX5*) gene, one of the most important master regulators in B cell development. Single nucleotide polymorphism (SNP) array analysis further revealed somatic deletions of large regions on chromosome 9p in all three children, which encompassed the tumor suppressor locus *CDKN2A*. Moreover, these chromosomal aberrations lead to tumor specific loss of the *PAX5* wildtype (WT) allele, retraining the mutant *PAX5* c.547G>A allele. The mutation is extremely rare and confined to distinct family backgrounds, since its presence was detected neither in a representative ALL cohort, nor in a healthy cohort of 1100 individuals.

We were able to further connect inherited *Pax5* susceptibility to leukemia development after exposure to infection, since we showed that BCP-ALL was initiated in *Pax5*[+/-] mice only when they were exposed to common pathogens. Strikingly, these murine BCP-ALLs closely resembled the human disease. We additionally identified somatic inactivating alterations affecting the remaining WT *Pax5* allele, as well as activating mutations of the tyrosine kinase Jak3. Therefore, we suggest that inactivating aberrations of *Pax5* promote leukemogenesis by creating an aberrant progenitor compartment that is susceptible to malignant transformation through constitutive active Jak3 kinase signaling.

Taken together, we were able to identify *PAX5* c.547G>A as a new inherited susceptibility variant for BCP-ALL. This scenario could be extended to a murine *Pax5*[+/-] model, which displayed BCP-ALL development after infection exposure, with additional

somatic aberrations affecting the remaining *Pax5* WT allele, as well as activating mutations in Jak3, which finally drive BCP-ALL in *Pax5* heterozygous mice.

Zusammenfassung

Im letzten Jahrzehnt konnten erhebliche Fortschritte in der Behandlung von akuter lymphatischer Leukämie (ALL) erzielt werden, welche sich in Überlebensraten von bis zu 90% widerspiegeln. Aufgrund intensiver Therapien, sowie schlechter Prognosen für Hochrisikogruppen und Patienten mit einem Rückfall, wird die Entwicklung neuer Behandlungen forciert, welche individuell auf den Patienten zugeschnitten sind. Hierfür ist die Aufklärung der genetischen Krankheitsgrundlage entscheidend. Vererbte, prädisponierende Mutationen sind dabei von besonderem Interesse, da die Identifizierung von Suszeptibilitätsregionen im Genom die Basis für präventive Therapien schafft.

Diese Arbeit beschäftigt sich mit der Charakterisierung von drei Kindern aus einer Familie, welche alle an einer B Zell Vorläufer ALL erkrankt sind. Mithilfe von Exom-Sequenzierung konnten wir eine seltene heterozygote Keimbahnmutation mit unvollständiger Penetranz in *PAX5* (c.547G>A) identifizieren. Das betroffene *PAX5* Gen stellt einen Hauptregulator der B Zell Entwicklung und Differenzierung dar. Wir konnten zeigen, dass es durch die in der Oktapeptiddomäne gelegene Mutation zu einer Verringerung der PAX5 Aktivität kommt. Des Weiteren wiesen die Tumore aller drei Kinder Deletionen auf Chromosom 9p auf, welche nicht nur den Tumorsuppressor Lokus CDKN2A, sondern auch das verbleibende Wild Typ (WT) *PAX5* Allel mit einschlossen. Darüber hinaus konnte die gefundene *PAX5* c.547G>A Mutation weder in einer Leukämiekohorte mit 9p Deletionen, noch in einer gesunden Kohorte mit 1100 Individuen gefunden werden, was ihre Seltenheit unterstreicht.

Außerdem konnten wir zeigen, dass Infektionsexposition zu Leukämieentwicklung mit unvollständiger Penetranz in *Pax5* heterozygoten (*Pax5*[+/-]) Mäusen führt. Die murinen Leukämien wiesen eine hohe Ähnlichkeit zu humanen B Zell Vorläufer ALL auf und zeigten ähnlich der analysierten Kinder eine reduzierte Pax5 Aktivität in den Tumoren. Darüber hinaus wurden in den murinen Leukämien aktivierende Mutationen in der Tyrosinkinase Jak3 akquiriert, welche zur malignen Transformation beitrugen.

Zusammenfassend konnten wir die erbliche *PAX5* Mutation c.547G>A als neue Suszeptibilitätsmutation für die B Zell Vorläufer ALL Entwicklung identifizieren. Das humane Szenario konnte des Weiteren auf ein *Pax5*[+/-] Mausmodell ausgedehnt werden in dem nur Infektion exponierte Mäuse eine B Zell Vorläufer ALL entwickelten. Diese murinen Leukämien wiesen zusätzliche somatische Mutationen in dem verbleibenden Pax5 WT Allel, sowie aktivierende Jak3 Mutationen auf, welche nachweislich an der malignen Zellproliferation beteiligt waren.

1 Introduction

1.1 Pediatric Cancers: Role of leukemia in childhood

Childhood cancer represents the leading cause for disease-based mortality among children under the age of 15, with over 12,000 patients newly diagnosed annually in the United States [1]. Since the 1960s, survival rates for children with cancer have been drastically improved from around 30% to over 80% [2]. However, over the last decade this positive trend has started to plateau, which emphasizes the need for further advances in cancer research and biology.

Leukemia is the most common pediatric cancer and is the cause of 29% of all cancer-related deaths in children between 0 and 14 years (**Figure 1**) [2]. Leukemia originates from the Greek words "leukos" and "emia" which refers to the term "white blood" and describes its characteristics as a cancer of the blood that manifests with high numbers of malignant white blood cells (leukocytes). These cells are no longer capable of executing their physiological role, but rather constitute rapidly proliferating leukemic blasts, which produce a disease phenotype by invading healthy tissues and organs.
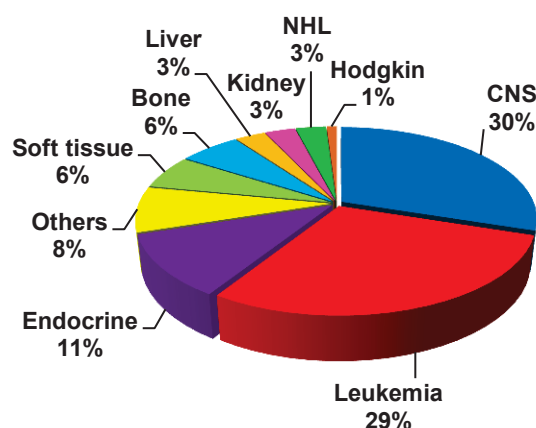
Figure 1: Cancer related deaths among children between 0 and 14 years. 29% of all cancer related deaths are caused by leukemia. CNS, central nervous system; NHL, non-Hodgkin lymphoma. Adapted from [2].

Leukemias can be divided in two major profiles – acute and chronic. While acute leukemias usually display a rapid accumulation of immature blast cells which requires

immediate treatment, chronic leukemias progress more slowly, mostly affect blood cells in mature developmental stages and are more frequently diagnosed in elderly patients. Further sub-classification is made according to the affected cell lineage. The malignant transformation can arise in cells either derived from common lymphoblastic or common myeloid progenitors (CLP or CMP), and are therefore differentiated in lymphoblastic or myeloid leukemias, respectively (**Figure 2**).
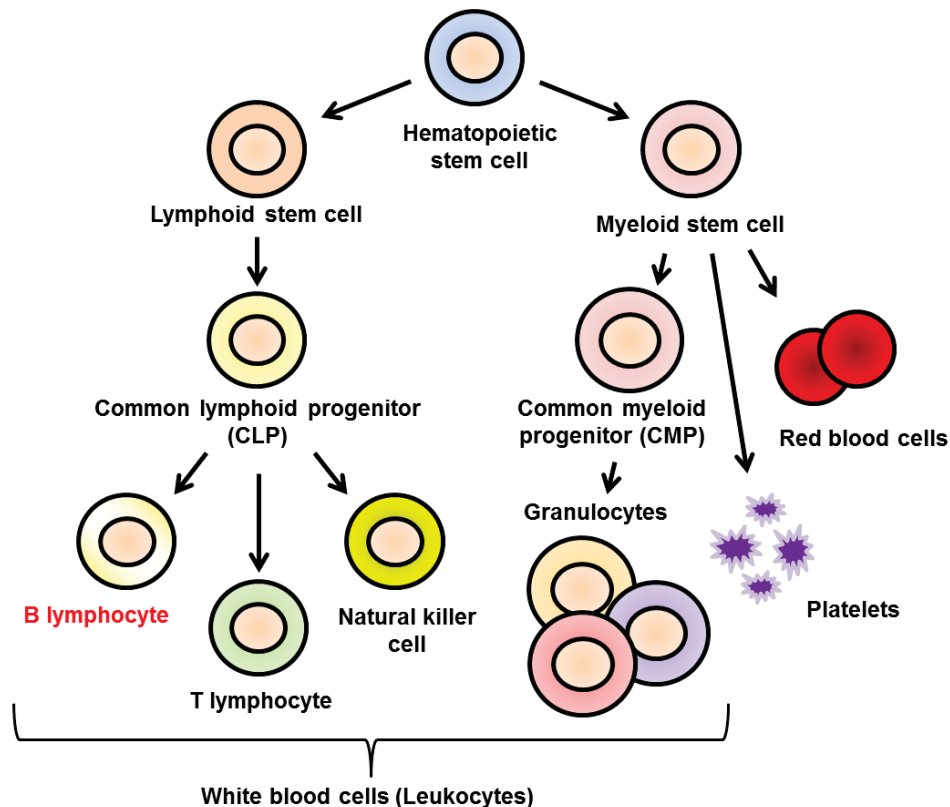
Figure 2: Hematopoiesis overview. The different hematopoietic lineages differentiate from a common hematopoietic stem cell. In this work, the focus is on malignancies arising from B lymphocytes (red font). Adapted from [3].

The combination of both characteristics allows the classification of leukemias in four distinct categories:

- Acute lymhphoblastic leukemia (ALL)

- Acute myeloid leukemia (AML)

- Chronic lymphoblastic leukemia (CLL)

- Chronic myeloid leukemia (CML)

In this doctoral thesis, the main research focus is on pediatric ALLs, with particular interest in the malignant transformation of B cell precursors (BCP).

1.2   Acute lymphoblastic leukemia (ALL) in children

1.2.1   Epidemiology

ALL is the most common pediatric cancer, being identified in 1/4 of all cancer diagnoses among children under the age of 15, while its incidence peaks in two to five year olds [4, 5]. There are significant differences in ALL incidence between the different ethnic groups, with Hispanics being most affected (40.9 cases per million), followed by Caucasians (35.6 cases per million) and people from African origin (14.8 cases per million) [6]. Moreover, pediatric ALL occurs more frequently in boys than girls (female to male ratio: 45% to 55%) [5]. In the United States, 30 to 40 ALL cases per million are recorded annually in children and adolescents younger than 20. This represents approximately 2,900 newly diagnosed ALLs each year, with incidences that have been gradually increasing over the past 25 years [4, 5].

1.2.2   Symptoms and diagnosis

The clinical presentation of ALL typically includes fatigue, bone and joint pain, pallor, bleedings, and recurrent infections [7]. These symptoms reflect the infiltration of the bone marrow with leukemic blast cells, which impacts normal hematopoiesis and leads to anemia, thrombocytopenia and neutropenia. Further clinical findings frequently encompass lymphadenopathy, hepatomegaly and splenomegaly, which are signs of blast accumulations in lymph nodes, liver and spleen, respectively, when blasts are spread from the bone marrow to the peripheral organs [7].

At disease presentation, approximately 50% of patients display elevated leukocyte counts (> 10,000 leukocytes per mm^3). In general, a bone marrow aspirate is necessary to establish an unambiguous diagnosis. To distinguish between ALL and the closely related lymphoblastic lymphoma, the percentage of blast cells in the bone

marrow is conventionally used, with more than 25% establishing a leukemia diagnosis [7]. Historically, ALL diagnosis relied solely on morphologic classification [8]. Over the years, this system has been proven to be error prone, since morphologic blast evaluation is subjective and difficult to reproduce or standardize. Although classification according to morphologic criteria can still be useful in urgent cases, e.g. to distinguish between ALL and AML, it has been replaced routinely by immunophenotyping which is now the standard method for ALL diagnosis [7]. Immunophenotyping utilizes fluorescence-activated cell sorting (FACS) technology to identify clonal blast populations on the basis of their distinct co-expression of surface antigens [9]. Different diagnostic panels are available for the characterization of the different hematopoietic subsets, including cells of T, B and myeloid origin (**Table 1**). This allows the precise determination of changes in the respective cell population, displayed by abnormal cell accumulations. Hence, immunophenotyping data output provides a classification of the lineage and developmental stage of the leukemic blast cells, which is essential for treatment stratification [7].

Lineage	Subtype	Marker
B lymphocyte	pro-B-ALL	CD19+, CD22+, CD79a+
	common ALL	CD10+
	pre-B-ALL	cytoplasmic heavy mu chain (IgM+)
	mature B-ALL	surface immunoglobulin light chains
T lymphocyte	Early T precursor ALL	CD5+, CD34+, CD117+, CD11b+, no Cd1a/CD8
	pro-T-ALL	CD3+, CD7+
	pre-T-ALL	CD3+, CD7+, CD5+, CD2+
	cortical T-ALL	cCD3+, Cd1a+
	mature T-ALL	cCD3+, sCD3+, no CD1a

Table 1: Immunophenotype of B and T-lineage ALL. Adapted from [10]. Subtype classification is based on the European Group for the Immunological characterization of Leukemia (EGIL) guidelines [11].

1.2.3 Prognostic Factors

To identify the correct treatment strategy for a child who is positively diagnosed with ALL, additional clinical and biological features related to the patient and the disease phenotype have to be taken into account. These so-called prognostic factors exhibit

prognostic value and allow a patient's stratification to specific risk groups. Besides the immunophenotype, important predictors of ALL prognosis include the patient's initial white blood cell count, age, karyotype, cytogenetic and molecular genetic characteristics [7].

1.2.4 Genetic basis of pediatric ALL (subtypes)

In children, ALL is in 85% of B lineage origin, with the major entity affecting B cell precursors, and 15% of T cell lineage origin and is further divided in several subgroups with regard to their distinct tumor specific fingerprint (**Figure 3**) (reviewed in [12, 13]). Somatic alterations of the subgroups can correspond to changes in the chromosome number (aneuploidy), to structural abnormalities of the chromosome, including chromosomal rearrangements, copy number variations (CNVs), small DNA insertions, and/or deletions (indels) or to DNA sequence mutations (reviewed in [14, 15]). Most of these aberrations lead to deregulation of the lymphoid development, affecting pathways involved in differentiation, proliferation, survival and tumor-suppression [16].
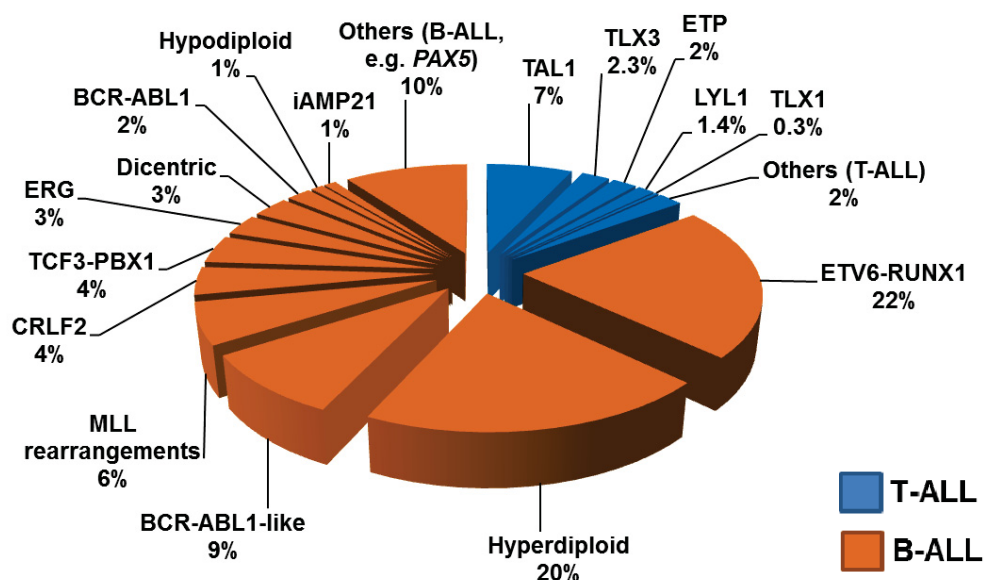
Figure 3: Different cytogenetic subtypes of pediatric ALL according to their frequency. Displayed are all major subtypes for B cell (orange) and T cell (blue) lineage ALL. Adapted from [12].

Aneuploidy frequently occurs in B cell precursor ALL (BCP-ALL), where it is sub-grouped in hyperdiploidy (chromosome number >50) and hypodiploidy (chromosome number <45) [17-20]. Another large cytogenetic subgroup is defined by structural abnormalities of chromosomes, including chromosomal translocations and intrachromosomal rearrangements (reviewed in [21]). These potentially leukemia-initiating events occur early in life, with some being already detected in neonatal blood samples [22], where they remain silent and clinically hidden until their presence is validated years later in leukemic blast cells of initial and relapse ALLs [23]. There are two possible mechanisms by which a translocation can be involved in leukemogenesis (reviewed in [13]). Structural alterations can relocate an oncogene into the regulatory region of an actively transcribed gene (e.g. translocations of CRLF2 in juxtaposition with IGH transcriptional enhancers (reviewed in [24])) or result in tumor-specific gene products, which produce a fusion gene by connecting genes that are normally located on two different chromosomes. An important fusion gene is the ETV6-RUNX1 fusion t(12;21)(p13;q22), which is the most common fusion, detected in 22% of all BCP-ALLs [25]. Moreover, it includes BCR-ABL1, a very well-characterized fusion with formation of the so-called Philadelphia (Ph) chromosome t(9;22)(q34;q11), which was the first report of a fusion protein more than 40 years ago [26]. In this scenario, the oncogenic BCR-ABL1 kinase is constitutively active, which promotes proliferation and survival in normal B cell precursors [27]. Other major examples for translocations in BCP-ALL include rearrangements involving the mixed lineage leukemia (MLL) gene [28], TCF3-PBX1 [29], as well as the intrachromosomal amplification of chromosome 21 (iAMP21) (reviewed in [30]). There are several other defined ALL subgroups, which show no single distinct chromosomal abnormality, but rather cluster together by specific gene expression profiles. Ph-like ALL affected children, for example, display gene expression profiles, that are similar to those of patients with Ph chromosome positive ALL, although no BCR-ABL1 fusion protein is detected. Instead of constitutive expression of the ABL1 kinase, Ph-like ALLs show a broad range of mutations that

activate tyrosine kinase signaling, involving fusions with ABL-class kinases and alterations of the JAK-STAT signaling [31].

1.2.5   Treatment of pediatric ALL

For more than 40 years, a variety of chemotherapy regimens have been successfully implemented for the treatment of ALL. Most contemporary therapies are based on an intensive multidrug combination induction protocol developed by Riehm *et al.* [32], which is the cornerstone for the efficacious Berlin-Frankfurt-Münster (BFM) chemotherapy regimen [33]. Since its introduction, this regimen has led to remarkable improvements, achieving up to 90% overall survival rates in recent clinical trials [34].

1.2.5.1   Contemporary Therapy

Although the various ALL therapies are adapted to the respective subtype and risk group, they share a common ground and can be divided in three distinct phases: (1) Induction therapy to achieve remission, which refers to the clearing of all clinical evidence of leukemia and the restoration of normal hematopoiesis; (2) Intensification and consolidation therapy to eradicate residual leukemic blast cells; (3) Maintenance therapy to retain the remission state and to reduce the risk of relapse (reviewed in [13]) (**Figure 4**).

ALL patients who are diagnosed as BCR-ABL1 positive receive a tyrosine kinase inhibitor (TKI) additional to chemotherapy, which has greatly increased the poor prognosis of this subtype in the past few years [35]. Another intensive but effective treatment option is allogeneic hematopoietic stem cell transplantation (alloHSCT). Although alloHSCT significantly cures about 50% of relapsed or initially therapy refractory patients, further improvements in controlling side effects such as toxicity, graft vs. host disease and immune reconstitution after transplantation are necessary.
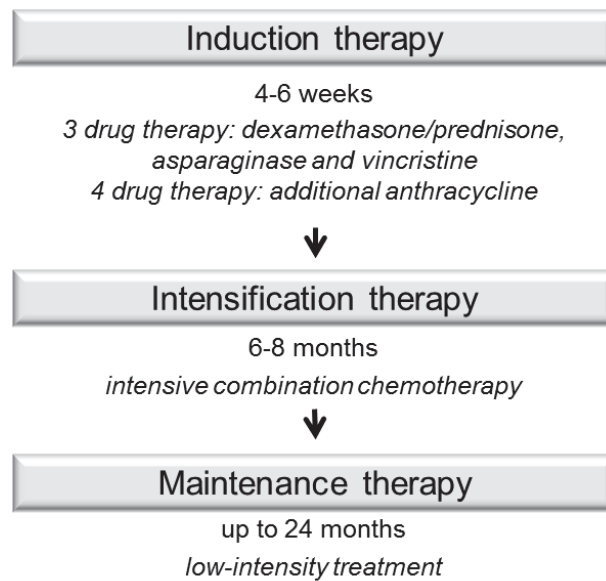
Figure 4: General treatment protocol for contemporary therapy of pediatric ALL. The treatment is divided in three distinct phases, which are accompanied by the administration of different drugs. Based on [13].

1.2.5.2 Targeted therapy and personalized medicine

Improvements in ALL treatment that were made solely by adapting and refining current chemotherapy protocols and agents have been optimized over the last 50 years. Nevertheless, cure rates for high-risk groups like BCR-ABL, MLL-AF4 fusions and iAmp21 remain modest [36-38], which strengthens the need for efficacious therapies with minimal toxicity (**Figure 5**).
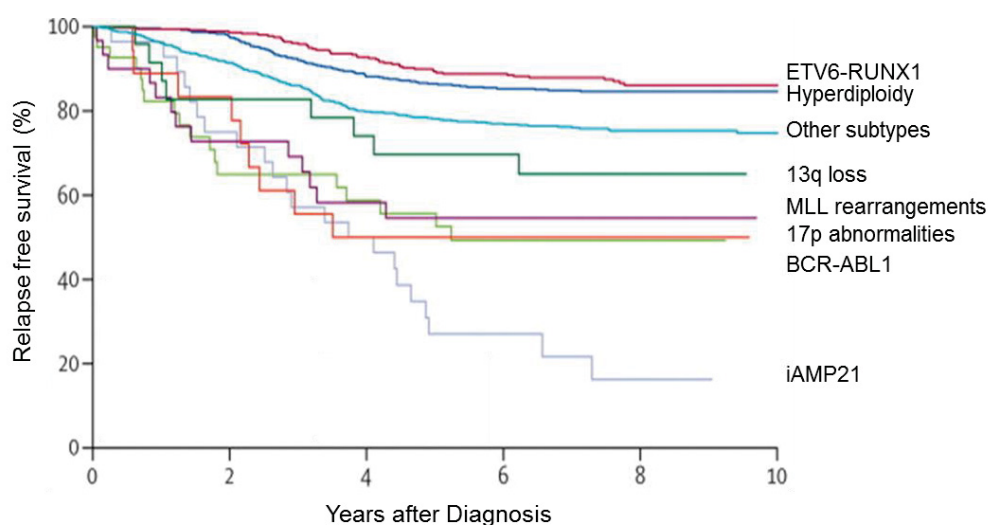
Figure 5: Relapse-free survival corresponding to different childhood leukemia subtypes. Good prognosis is achieved for ETV6-RUNX1 and hyperdiploidy affected patients, while the cure rates for high risk groups including iAMP21, BCR-ABL1 and MLL rearrangements, remain poor. Adapted from [39].

Therefore, new therapy approaches are needed that further define ALL subtypes and allow individualized treatment (reviewed in [13]). Exploration of the genetic basis of ALL continually identifies new potential targets, which either predispose to or drive leukemogenesis. Hence, personalized or precision medicine aims to design targeted therapies that are directed to patients' individual genetic background (reviewed in [40, 41]). The introduction of TKIs like imatinib, for example, has changed the therapy for BCR-ABL1-driven chronic myeloid leukemia from intensive treatments to an oral therapeutic regiment with tolerable side effects [42]. In combination with chemotherapy, TKIs are successfully applied to Ph-Chromosome positive BCP-ALLs, which greatly increased survival rates and moreover reduced the demand for alloHSCTs in the first remission [43]. New preclinical studies indicate a successful application of TKIs also for the Ph-like BCP-ALL subtype (reviewed in [44]).

Nevertheless, in order to design future clinical trials, we need to expand our knowledge and understanding of the molecular basis behind each specific ALL subtype, as well as about intrinsic and extrinsic factors that contribute to leukemia development.

1.3 Susceptibility to pediatric ALL

The majority of cancers can be associated with acquired somatic mutations in key regulators of cell growth, survival, differentiation and genome integrity [45]. Sometimes, however, inherited mutations can pose a high cancer risk either by themselves (Mendelian inheritance) or in combination with other genetic and environmental factors. Genome-wide association studies (GWASs) have revolutionized the field of ALL, extending the molecular understanding of the disease through genomic profiling (reviewed in [46]). The GWAS approach does not rely on prior knowledge to focus on a defined gene set, but utilizes next generation sequencing techniques to detect single nucleotide polymorphism (SNP) susceptibility loci by comparing the genome sequences of ALL cohorts to healthy control populations (reviewed in [47, 48]). Since the prevalence and identity of predisposing mutations among ALL affected children are

largely unknown, increasing knowledge of such potentially inherited or *de novo* mutations can improve the understanding of tumorigenesis and pave the way towards personalized medicine.

1.3.1 Inherited germline variants conferring ALL susceptibility

A potential genetic basis of ALL susceptibility was recently assessed in a study of about 4000 patients, revealing that children with affected siblings are three times more likely to develop ALL, while this risk is accelerated up to 163 times for monozygotic twins [49]. This extremely high frequency of twins does not necessarily arise from ALL predispositions, but rather comes from the fact that early leukemia events can occur *in utero*, where malignant cells are potentially passed to the monozygotic twin via the shared placenta circulation [50]. Inherited ALL susceptibility variants are germline mutations, which are genomic alterations that occur in a mature sexual reproductive cell (gamete), and are therefore passed to the entire organism of its offspring. A study analyzing pediatric cancers in 1120 children and adolescents evaluated a frequency of 8.5% germline mutations in cancer-predisposing genes. Non-CNS solid tumors displayed the highest prevalence of germline mutations (16.7%), followed by CNS tumors (8.6%) and leukemia (4.4%) [51]. Since such studies are based on the detection of known cancer genes, the real prevalence of susceptibility genes remains unknown. Although, historically, there has been little evidence of a major role for germline predispositions implicated in leukemogenesis, GWASs and the use of new research techniques started to reveal an increasing number of inherited susceptibility variants that contribute to ALL development (reviewed in [52]) (**Table 2**). Some of these predisposing variants were found in genes that are also often somatically mutated in ALL, including *ARID5B*, *CDKN2A*, *CEBPE* and *IKZF1* [53-55]. ARID5B is involved in embryonic development and the regulation of cell growth [56], the CCAAT/enhancer binding proteins (CEBPs) are transcription factors relevant for hematopoietic cell development [57], while IKZF1 is a transcriptional regulator of lymphoid development

[58]. The CDKN2A/B locus encodes the so-called ARF or INK4 proteins, which is a family of tumor suppressors and cell cycle regulators [59, 60].

Gene	Rs ID	Study Type	Reference
ARID5B	rs7089424	GWAS	[54]
	rs10821936	GWAS	[53]
	rs10994982	GWAS	[53]
IKZF1	rs4132601	GWAS	[54]
	rs11978267	GWAS	[53]
CEBPE	rs2239633	GWAS	[54]
	rs4982731	GWAS	[61]
CDKN2A	rs17756311	GWAS	[61]
	rs3731217	GWAS	[55]
PIP4K2A	rs7088318	GWAS	[61]
	rs10828317	GWAS	
GATA3	rs3824662	GWAS	[62]
		GWAS	[63]
TP63	rs17505102	GWAS	[64]
TP53	rs121913344	Familial ALL	[65]
ETV6	rs786205155	Familial ALL	[66]

Table 2: Summary of known germline genetic variants that are associated with ALL susceptibility. Also indicated are the respective rs IDs as well as the method by which the variants were identified. Adapted from [52].

Sometimes the relevance of identified susceptibility loci for leukemogenesis is not yet determined. For example, the mechanism of a predisposing variant in the BMI1-PIP4K2A locus and its role in lymphoid cell differentiation remains unknown [61]. Down syndrome (DS) patients display up to 20% higher leukemia incidence, while the etiology of this increased risk remains largely unresolved (reviewed in [67]). Some germline susceptibilities can even be linked to a specific ALL subtype or risk group, e.g. *GATA3* mutations in Ph-like ALL [63] and germline genetic variations in *TP53* and *PTPRJ*, which are frequently associated with ETV6-RUNX1 positive ALL [64]. Other variants are predominantly found in ALLs of specific ethnic backgrounds, like the

greater frequency of the high-risk *ARID5B* SNP rs10821936 genotypes in Caucasians compared to people from African origin [68]. Moreover, a genetic basis for leukemia development could be validated in families with high leukemia frequencies in both close and distantly related family members. Therefore, besides GWASs, such families with more than one case of leukemia (familial ALL) can also be used to screen for common inherited ALL related variants. It's a challenging method, since familial ALL cases are rare, but the analysis of available kindreds is highly informative and poses an opportunity to identify so far unknown ALL susceptibility genes. Using this approach, a variety of new susceptibility genes could be unmasked, including genes encoding for the tumor suppressor TP53 [65] and the transcription factor ETV6 [66]. The compilation and subsequent sequencing of further familial ALL kindreds will help to reveal the full spectrum of inherited susceptibility genes, which contribute to ALL development.

1.3.2 Environmental risk factors

The identification of new genetic susceptibility loci that contribute to ALL development helps to shed light on the underlying molecular basis of the disease. Nevertheless, some predisposing mutations are "weak" and their presence alone does not automatically result in a disease phenotype. One example for this are inherited susceptibility loci, which frequently display incomplete penetrance, e.g. out of two siblings carrying the same predisposing mutation, only one will eventually develop leukemia. The etiology of ALL in children is therefore believed to be multifactorial, including an initial event in combination with secondary extrinsic and intrinsic factors (reviewed in [69, 70]). Nevertheless, the precise role of environmental leukemogenesis in childhood leukemia remains undefined.

1.3.2.1 Radiation

One of the few established causal environmental factors for ALL development in children is ionizing radiation exposure. Leukemia incidence and survival after the atomic bomb explosions in Japan during World War II could be linked to the degree of

contamination, with the greatest risk appearing in highly irradiated individuals [71, 72]. Nevertheless, an epidemiological assessment of naturally occurring low-dose radiation is very difficult, and discussion of this topic is still causing controversy [73, 74]. Whether a substantial fraction of ALL cases are owing to ionization exposure should therefore be assessed with caution.

1.3.2.2 Chemicals

While chronic chemical exposure to benzene, which is the main toxin in the hydrocarbon fraction of motor vehicle exhaust, has been linked to AML incidence in adults, its association with childhood ALL is still contested (reviewed in [75]). Moreover, an increase of leukemia risk has been reported for early childhood exposures to pesticides [76]. Geographically, the industrialized West exhibits a high ALL incidence when compared to North Africa or the Middle East [77]. This phenomenon could be attributed to increased exposure to environmental leukemogens, and is backed up historically, since major pre-school peaks of ALL occurred simultaneously with periods of socioeconomic changes [78].

1.3.2.3 Parental influences

That parental smoking has implications during pregnancy was shown through the identification of tobacco-derived components in maternal placentas and milk as well as in the blood and urine of newborns. A causative role in childhood leukemia was attributed to paternal smoking due to a higher mutational risk in germ cells but, various study limitations challenge its direct contribution [79].

1.3.2.4 Infection

A possible relationship between childhood ALL and infection exposure was first postulated over a century ago and is gradually gaining more and more support. Various hypotheses regarding its influence have been proposed, which puts exposure to infection as one of the most promising foci in acute leukemia research at present (**Figure 6**).
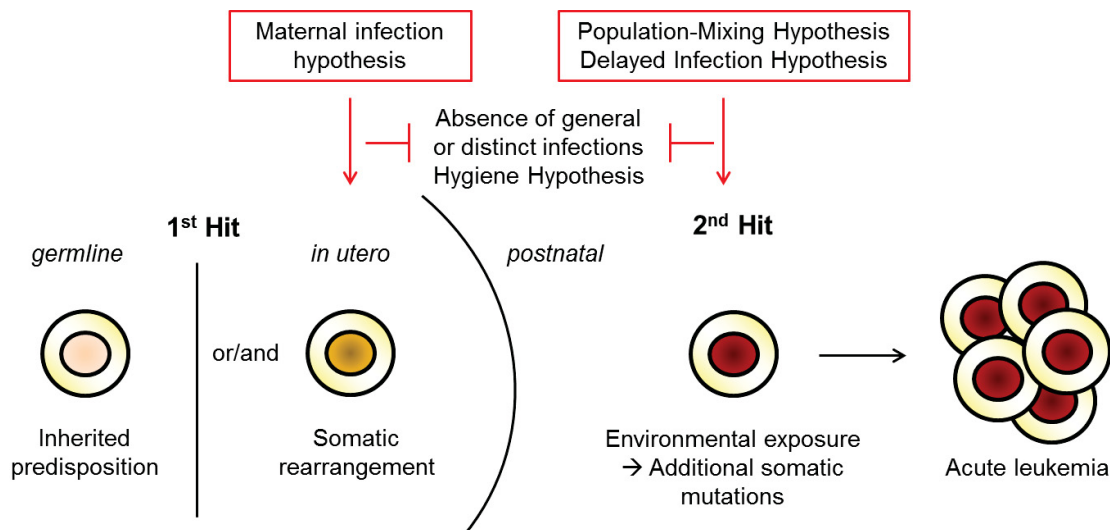
Figure 6: Overview of the main hypotheses that have been proposed for the influence of infections in the development of childhood ALL. It was postulated that the first hit can constitute an inherited germline predisposition or/and *in utero* occurring somatic alterations. Infection as environmental exposures is subsequently believed to represent the second hit, either through maternal infections during pregnancy or the absence of infectious agents in early childhood due to increased hygiene or population based isolation. Based on [70, 80-82].

Two main hypotheses have been postulated that link childhood infections and ALL susceptibility [80, 81]. Alexander Kinlen's "population mixing theory" is based on increased lymphoproliferative stress that arises through the immigration of previously isolated populations into new geographical regions, thereby introducing so far unencountered infection exposures to both sides [80]. In addition, Mel Greaves proposed the "delayed infections" or "two-hit hypothesis". The first hit describes an initial exposure *in utero* that is initiated through chromosomal alterations, while the second hit refers to a postnatal "exposure window", which allows the acquisition of additional genetic changes through internal and external factors, e.g. infection. He further postulates that, in the case of limiting viral or bacterial exposures during a child's first year of life, improper development of the immune system takes place, which increases the acquisition risk of subsequent leukemia-driving mutations when the immune system is triggered by infection [70, 81]. Although both hypotheses are compatible in the sense that childhood leukemia is a rare response to common infections in genetically susceptible individuals, the identification of specific infectious agents that lead to ALL development has so far been unsuccessful.

In the past, the prime rationale regarding how infection influences pediatric leukemia development was given by the age distribution of the disease, which peaks at the same time as when children encounter common childhood infections [70]. Hypotheses that point towards a disordered immune response at the time of infection have been backed up by the emergence of increased hygiene levels in western countries. For example, there are multiple reports that show an inverse correlation between hepatitis A infections and ALL incidence [78, 83]. Another immunological argument is postulated in the context of childhood allergies ("hygiene hypothesis"), which highlights the need for the induction of a distinct anti-inflammatory regulatory network through persistent challenges of the immune system during early life (reviewed in [84]). In agreement, attendance of day-care facilities in infancy protects against the development of allergies in later life [85]. This finding could be expanded to ALL, since most reports revealed a decrease in childhood leukemia incidence through an increased risk of exposure to infections in day-care and playgroups [86, 87]. An explanation for this reduced risk of childhood ALL is the "adrenal hypothesis", which postulates that infections in early life induce qualitative and quantitative changes in the hypothalamus–pituitary–adrenal axis, which in turn increases plasma cortisol levels and eliminates possible preleukemic cells [88]. Here, the timing seems important, since there are also suggestions of increased leukemia risk through maternal infections transmitted to the fetus during pregnancy [82]. Nevertheless, the analysis of other epidemiological factors that are linked to infection, has produced conflicting results, e.g. a potential protection against ALL through protracted breastfeeding or early vaccination (reviewed in [89]). Up to now, no specific infectious trigger could be pinpointed, which is mostly due to the fact that a variety of determinants that influence leukemia onset must be taken into account. These include age at exposure, immune status of the mother, family size, social and community factors, and whether a distinct infectious species at a single time point or a variety of multiple infectious agents over a longer period of time or at repeated intervals are relevant.

Case-control and cohort studies that use the classic epidemiological approach to depict the influence of causal exposures for childhood leukemia are valuable, but limited in their power due to the rarity of the disease and insufficient patient information. In order to effectively validate and implement the infection theory, specific cellular and biochemical mechanisms that facilitate leukemogenesis through infection exposure have yet to be discovered. High throughput screening of distinct signaling pathways that can be linked to infection (e.g. cytokines and immune response) could be a valuable tool, but further genetic evidence is required to fully grasp how childhood leukemia and infection exposure are intertwined.

1.4 Molecular mechanisms involved in BCP-ALL leukemogenesis

It is crucial to reveal the influence of environmental factors for ALL development, but most exposures are believed to act as additional or cooperating events on already existing preleukemic clones. Therefore, in order to fully understand the interplay between an individual's genetic repertoire and the environment, the molecular basis of the "first hit", which predisposes to disease development, has to be known and elucidated. While multiple factors are needed to give rise to an acute leukemia phenotype, the "first hit" occurs before birth in many cases. In this regard, the aberrant expression of main transcription factors, which can lead to abnormal proliferation or differentiation arrest of specific lymphoid cell subsets, was shown to be a hallmark of ALL pathobiology [90]. For example, the B-lymphoid transcription factors PAX5 and IKZF1 are essential for early B cell development [91]. Therefore, it is not surprising that they are genetically lesioned in over 80% of BCP-ALLs [45, 92]. However, despite their high frequency in BCP-ALL, the role and significance of these lesions is poorly understood. Hence, the following section will focus on PAX5, the main transcription factor in B cells, and its potential role for leukemogenesis.

1.4.1 PAX5 function during B cell development

The paired box 5 (*PAX5*) gene is located on chromosome 9p13 and belongs to the paired box (*PAX*) gene family. So far, 9 mammalian PAX transcription factors could be identified that play essential roles in the control of tissue-specific transcription, which is important for lineage determination in a variety of cell types (reviewed in [93]). PAX5 is the only one expressed in the hematopoietic system, while it is also detected in the midbrain-hindbrain boundary in the CNS [94, 95]. PAX5 acts as a master regulator for B cell development, commitment and identity, exerting functions as a transcriptional activator of B cell lineage genes (e.g. *CD19*), while simultaneously repressing lineage-inappropriate genes (e.g. *Notch1*, *Flt3*) (reviewed in [96-99]). Nevertheless, the mechanisms by which both activator and repressor functions are achieved concurrently in the same cell remain largely unresolved. Pax5 knockout mice (*Pax5$^{-/-}$*) display a complete block in B cell differentiation at a precursor stage [97]. In addition, precursor B cells of *Pax5$^{-/-}$* mice carry hematopoietic stem cell like features, including self-renewing capacity and multipotency [100]. Mouse analyses based on a reporter gene inserted into the *Pax5* locus validated the B cell confinement of Pax5, since its expression was absent in multipotent and the vast majority of common lymphoid progenitors [101]. In hematopoiesis, PAX5 expression is initiated at the pre-pro-B cell stage through concordant expression of E2A and EBF1 [102]. PAX5 levels steadily increase until the cells are uniformly PAX5 positive and enter the pro-B cell stage. This constant PAX5 expression is then maintained stably throughout the development and life of a B cell before it is downregulated during plasma cell differentiation (reviewed in [103]) (**Figure 7**). Remarkably, up to this point, PAX5 expression is essential for the maintenance of B cell identity, since in mice a conditional inactivation of Pax5 in already differentiated B cells results in the de-differentiation of a proportion of these cells, which in turn can be re-differentiated into the T cell lineage [104].
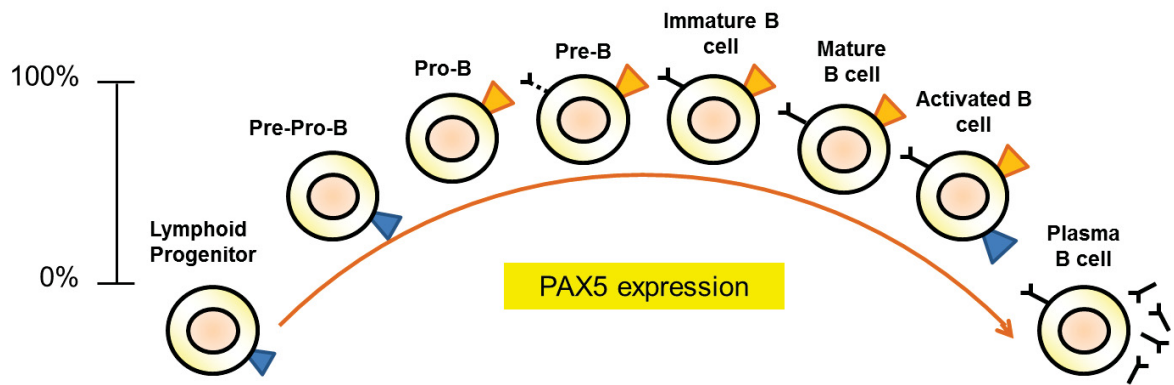
Figure 7: Simplified model depicting PAX5 expression during B cell development. PAX5 expression is first detectable at the pre-pro-B cell stage and maintained stably throughout B cell development, until it is again downregulated in plasma B cells. Its expression represses lineage-inappropriate genes (blue marks), while activating B cell lineage-specific genes (orange marks). Adapted from [103].

Combined approaches utilizing microarray data of various sources, including mice and humans with PAX5 alterations, identified more than 200 potential PAX5-activated genes. These PAX5 targets could be assigned to specific classes of transcription factors or cell cycle regulators, as well as groups of proteins involved in cell adhesion and migration [105, 106]. Hence, it is unsurprising that aberrant or uncontrolled PAX5 expression has major implications in normal hematopoiesis and poses a high risk for leukemia development.

1.4.2 PAX5 in childhood leukemia

Genetic alterations affecting the *PAX5* gene are one of the most common features present in BCP-ALL [45]. *PAX5* deletions are identified in around 30% of all BCP-ALL cases, while mutations and translocations occur in a smaller percentage of about 7% and 2-3%, respectively. This alteration spectrum is observed in both adult and childhood ALLs and is frequently associated with other genetic lesions, e.g. in combination with chromosomal rearrangements like ETV6-RUNX1 [45], BCR-ABL1 [107] or TCF3-PBX1 [108, 109]. Most PAX5 aberrations are somatically acquired and hemizygous, which means that they affect only one allele. Nevertheless, this haploinsufficiency already exerts functional consequences in reducing total PAX5 expression levels, while in the majority of cases the transcription of PAX5 target genes,

like CD19, is still detectable [45]. Moreover, through the inactivation of one allele, the remaining wildtype PAX5 allele is prone to acquiring additional mutations, further disrupting PAX5 protein functions. Deletions affecting PAX5 can correspond to either wide-range deletions or focal deletions, which results in a lack of the complete PAX5 protein or the production of truncated forms [45]. Point mutations can impair the DNA binding of PAX5, alter its transcriptional regulation or cause truncated proteins through frame shifts or interference with splice sites, as predicted by using crystal structure modelling of the PAX5 protein [110]. Like deletions, point mutations affecting the *PAX5* gene impair PAX5 expression and therefore generate hypomorphic PAX5 alleles that could lead to haploinsufficiency [45]. In addition to deletions and point mutations, PAX5 is involved in a growing number of chromosomal translocations with a variety of different partner genes, which results in the generation of newly formed chimeric proteins that antagonize the transcriptional activity of PAX5 [111]. In this regard, 26% of all 9p abnormalities involving PAX5 are dicentric chromosome events, mostly affecting partners on chromosomes 7, 12 and 20 [112]. While the fusion partners can constitute a variety of proteins, including proteins related to transcriptional regulation, structural proteins or kinases, the fusion with the transcription factor ETV6 (PAX5-ETV6), which originates from the translocation t(9;12) or dic(9;12), is the most common [45, 113].

1.5 Additional mutations driving tumorigenesis

1.5.1 Tumor suppressor locus CDKN2A

PAX5 is altered in 81% of BCP-ALL cases displaying 9p abnormality [112]. The majority of these 9p aberrations result in a wide-spread or even complete deletion of the chromosome's short arm. Besides PAX5, this region also encompasses the CDKN2A and CDKN2B tumor-suppressor loci, which are frequently inactivated in a variety of hematological malignancies. CDKN2A is known to be an important tumor

suppressor locus, since it encodes both p16 (INK4A), a negative regulator of cyclin-dependent kinases, and p14 (ARF1), an activator of the tumor suppressor p53. Deletions affecting the CDKN2A chromosomal region represent one of the most common genetic events in childhood BCP-ALL, with a frequency of around 40% [45, 114]. CDKN2A deletions (9p21.3) as secondary event are proposed to be the causal basis for tumorigenesis, since, in addition to leukemia, they are associated with a variety of cancers (melanoma, basal cell carcinoma and glioma) and other diseases (type 2 diabetes and coronary heart disease) (reviewed in [55]).

1.5.2 Tyrosine Kinases

Compared with many other tumor types, childhood BCP-ALL displays a relatively low mutation rate, with less than 10-20 non-silent mutations per case. Despite these few mutations, they recurrently involve key pathways, including transcriptional regulation of lymphoid development (e.g. PAX5), as well as cell cycle regulation and tumor suppression (CDKN2A) (reviewed in [115]). Another important group of frequently somatically affected genes in ALL are tyrosine kinases. In particular, somatic mutations in the Janus kinases JAK1, JAK2 and JAK3, which result in their constitutive activity, are commonly found in leukemia (reviewed in [116, 117]). These tyrosine kinases exhibit essential roles as signal transducers downstream of cytokine receptors, which themselves lack intrinsic protein kinase domains. Upon receptor activation, receptor-associated JAKs are trans-phosphorylated and in turn phosphorylate tyrosines within the cytoplasmic regions of the receptor. This generates docking sites for SH2-domain-containing adaptor and effector proteins, including the signal transducers and activators of transcription (STAT) proteins. Subsequently, depending on the receptor type, one or more STAT proteins (STAT1, STAT2, STAT3, STAT4, STAT5 or STAT6) are recruited. After being phosphorylated by JAK, STAT proteins translocate into the nucleus, where they activate a variety of transcriptional programs corresponding to cell proliferation, differentiation and survival. In addition to STAT activation, JAK signaling is

able to activate other signaling cascades, including the mitogen-activated protein kinase (MAPK), phosphatidylinositol-3'-kinase (PI3K) and mammalian target of rapamycin (mTOR) pathway (reviewed in [116, 117]).

JAK mutations are frequently associated with an overexpression of CRLF2. JAK2 mutants in particular are frequently found in combination with CRLF2 rearrangements, which leads to the overexpression of the CRLF2 cytokine receptor, further activating JAK2 pathways. While JAK2 mutations affecting its kinase or pseudo-kinase domain are found in around 20% of Down Syndrome BCP-ALL patients [118], somatic mutations in JAK1 and JAK3 are present in a minority of T-ALL cases, and are even less common in BCP-ALL [119].

2   Aim of this thesis

The identification of susceptibility loci that predispose to leukemia development helps us to understand the genetic basis of the disease, which is necessary for the implementation of new treatment strategies, e.g. for the emergence of personalized medicine. Although rare, a powerful tool to achieve this goal is the screening of families with more than one case of leukemia for inherited disease associated variants.

The aim of this thesis was to elucidate the inherited genetic basis of a family with three cases of B cell precursor acute lymphoblastic leukemia in childhood according to the following strategy:

1. Identification of the inherited germline mutation, which predisposes to BCP-ALL in the three affected children.

2. Analysis of the mutation's prevalence in a disease-related as well as a healthy cohort.

3. *In vitro* validation of the germline variant for its functional role in leukemia development.

4. Characterization of an *in vivo* mouse model, which mimics the human phenotype with respect to the *in vivo* role of the predisposing mutation and to draw a link back to the human setting.

3 Material

3.1 Patient samples

Patient material was either provided by the Children's Hospital and the Institute of Transplantation Diagnostics and Cell Therapeutics (ITZ) of the University Clinic Duesseldorf under the direction of Dr. med. Julia Hauer, or by Dr. Polina Stepensky from the Hadassah Medical Center in Israel. Donors' informed consent was taken in agreement with the respective Faculties Ethic Committee. To protect the patients' privacy, all personal data were encrypted and concealed. Routine diagnostic assessment of blood samples was carried out in the laboratory of the Children's Hospital of the University Clinic Duesseldorf. Karyotyping of human samples was performed in the Munich Leukemia Laboratory.

3.2 Animals

Animal experiments with heterozygous $Pax5^{+/-}$ mice [94] were carried out in cooperation with Dr. Isidro Sánchez-García's laboratory from the Instituto de Biologia Molecular y Celular del Cancer (IBMCC), CSIC/Universidad de Salamanca (Spain). Mouse housing and handling were conducted according to relevant national and international guidelines and were approved by the Bioethics Committee of the University of Salamanca and by the Bioethics Subcommittee of Consejo Superior de Investigaciones Cientificas (CSIC). Both male and female mice were included in the study. Wildtype (WT) and $Pax5^{+/-}$ mice were bred on a mixed C57BL/6 x CBA background and monitored for disease development up to 24 months, while kept in a specific pathogen-free facility or a conventional facility. Upon first signs of disease, mice were sacrificed and subjected to standard necropsy procedures. All major organs were examined under the dissecting microscope. Differences in Kaplan-Meier survival plots of transgenic and WT mice were analyzed using the log-rank (Mantel-Cox) test.

3.3 Cells

3.3.1 Hek293T

Species: human (*Homo sapiens*)

Cell type: embryonal kidney

Morphology: fibroblastoid cells growing adherently as a monolayer

Subculture: $2\text{-}3 \times 10^6$ cells/75 cm^2; split ratio 1:4 - 1:5 every 2-3 days

Doubling time: circa 24-30 hours

Origen: derivative of the human primary embryonal kidney cell line 293 (ACC 305, DSMZ Heidelberg) with high transfection efficiency. Harbors a plasmid containing the temperature sensitive mutant of SV-40 large T-antigen (tsA1609neo); cells were originally referred to as 293tsA1609neo.

3.3.2 Ba/F3

Species: murine (*Mus musculus*)

Cell type: pro-B cells

Morphology: round, single cells in suspension (occasionally in clumps)

Subculture: $1\text{-}3 \times 10^5$ cells/ml; split ratio 1:10 every 2-3 days

Doubling time: circa 20 hours

Origen: IL-3 dependent murine pro-B cell line with unclear origin. Cells were derived from C3H mice.

3.4 Media, Additives & Buffers

3.4.1 For cell culture

Identifier	Company	Order ID
DMEM GlutaMAX	Gibco	31966-021
RPMI-1640 GlutaMAX	Gibco	61870-010
Fetal bovine serum (heat Inactivated) (FBS)	PAN	P30-1902
L-Glutamine	Gibco	25030-024
Dulbecco's phosphate buffered saline (DPBS)	Sigma-Aldrich	D8537
Trypsin-EDTA	Gibco	25300-054
Penicillin-Streptomycin	Gibco	15140-122
Gentamicin	Gibco	15710-049
Hygromycin B	Invitrogen	10687010
DMSO (Dimethylsulfoxide)	Sigma-Aldrich	D2650
Recombinant Mouse IL-3	Gibco	PMC0034
ACK Lysing Buffer	Thermo Fisher Scientific	A1049201

3.4.2 For bacterial culture

LB medium

5 g/l Yeast extract, 10 g/l Tryptone, 10 g/l NaCl, dH$_2$O. The media was autoclaved for 20 min at 121 °C and stored at 4 °C.

LB agar plates

5 g/l Yeast extract, 10 g/l Tryptone, 10 g/l NaCl, 15 g/l agar, dH$_2$O. The media was autoclaved for 20 min at 121 °C, cooled to 60 °C before the respective antibiotic was added and the solution was poured in agar plates. After solidification, the plates were stored at 4 °C.

<u>Terrific Broth medium</u>

12 g/l Tryptone, 24 g/l Yeast extract, 9.4 g/l K_2HPO_4, 2.2 g/l KH_2PO4, pH 7.2, 0.4% Glycerin. The media was autoclaved for 20 min at 121 °C and stored at 4 °C.

<u>Buffers for chemically competent bacteria</u>

RF1 buffer: 100 mM RbCl, 50 mM $MnCl_2$ x 2 H_2O, 30 mM CH_3COOK, 10 mM $CaCl_2$ x 2 H_2O, 15% w/v Glycerin. Ad 1 l dH_2O and adjust pH to 5.8. Sterilize by filtration.

RF2 buffer: 10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$ x 2 H_2O, 15% w/v Glycerin. Ad 500 ml dH_2O and sterilize by filtration.

3.4.3 Others

<u>Agarose Gel</u>

1 – 2% agarose dissolved in 1 x Tris/Acetic Acid/EDTA (TAE) buffer (50 x TAE stock)

<u>Immunoblotting</u>

RIPA lysis buffer: 50 mM Tris (pH 8.0), 150 mM NaCl, 0.5% Sodium deoxycholate, 1% Nonidet NP40 Substitute, 0.1% SDS, ad 50 ml dH_2O. Freshly add Protease and Phosphatase inhibitors prior to use.

2 x Laemmli buffer: 125 mM Tris (pH 6.8), 10% Glycerin, 4% SDS, spatula tip bromphenolblue, ad 50 ml dH_2O. Add 200 µl β-Mercaptoethanol to 1.8 ml buffer freshly prior to use.

Separation gel buffer: 30% Acryl-Bisacrylamide (19:1), 378 mM Tris-HCl (pH 8.0), 0.1% SDS, 0.1% APS, 0.06 % TEMED

Stacking gel buffer:	30% Acryl-Bisacrylamide (19:1), 126 mM Tris-HCl (pH 6.8), 0.1% SDS, 0.1% APS, 0.1% TEMED
10 x SDS running buffer:	25 mM Tris, 190 mM Glycine, 0.1% SDS
10 x Transfer buffer:	25 mM Tris, 190 mM Glycine
TBS-T buffer:	20 mM Tris (pH 7.5), 150 mM NaCl, 0.1% Tween 20
Blocking buffer:	5% Bovine serum albumin (BSA) or powdered milk dissolved in TBS-T
Stripping buffer:	1 x Re-Blot Plus Mild Antibody Stripping Solution (Merck)

3.5 Chemicals

Identifier	Company	Order ID
2-Propanol	Merck	67-63-0
50 x TAE Buffer	Bio-Rad	161-0743
Agarose	Biozym	840004
Albumin fraction V (BSA)	Roth	8076.3
Ammonium Persulfate (APS)	Sigma-Aldrich	A-3678
Bis-Acrylamide 30% (19:1)	Bio-Rad	161-0154
Bromphenol blue	Sigma-Aldrich	B0126
$CaCl_2$ x 2 H_2O (Calcium chloride dihydrate)	Sigma-Aldrich	C3306
CH_3COOK (Potassium acetate)	Sigma-Aldrich	P1190
Ethidiumbromide Solution	Sigma-Aldrich	E1510
EtOH (Ethanol) absolute	Merck	1009832511
Ficoll-Paque Plus	GE Healthcare	17-1440-02
Glycerin	Merck	1040921000

Glycine	Merck	1042011000
HCl (Hydrochloric acid)	Merck	1099110001
$MnCl_2$ x 2 H_2O (Manganese(II)chloride dihydrate)	Merck	1059340100
MOPS (3-[N-Morpholino]-propanesulfonic acid)	Ambion	AM9570
NaCl (Sodium chloride)	Sigma-Aldrich	S9888
NaOH (Sodiumhydroxide)	Merck	106498100
Nonidet P40 Substitute	Fluka Biochemica	74385
Polyethylenglycol (PEG)	Sigma-Aldrich	P1458
Powdered milk	Roth	T145.3
RbCl (Rubidium chloride)	Sigma-Aldrich	R2252
Sodium dodecyl sulfate 20% (SDS)	Ambion	AM9820
Tetramethylethylendiamin (TEMED)	Merck	1107320100
Tris	Roth	5429.3
Tryphan Blue	Sigma-Aldrich	T8154
Tween 20	Roth	9127.1
β-Mercaptoethanol	Merck	154330100

3.6 Specific reagents, size markers and other materials

Identifier	Company	Order ID
Agencourt AMPure XP	Beckman Coulter	A63881
Complete EDTA-free Protease Inhibitor Cocktail	Roche Diagnostics	11697498001
FuGene 6 Transfection Reagent	Promega	E2691
GeneRuler 1 kb Plus DNA Ladder	Thermo Fisher Scientific	SM1331
Page Ruler Prestained Protein Ladder	Fermentas	SM0671
GeneRuler 100 bp Plus DNA Ladder	Thermo Fisher Scientific	SM0321
Human660W-Quad_v1 BeadChip	Illumina	WG3111501
MyOne Streptavidin T1 Dynabeads	Life Technologies	65601
Nuclease free water	Ambion	AM9937
PhosSTOP Phosphatase Inhibitor Cocktail	Roche Diagnostics	4906845001
Propidium Iodide Staining Solution	BD Biosciences	556463
SuperSignal West Pico Luminescent Substrate	Thermo Fisher Scientific	34077

3.7 Plasmids

Identifier	Provider	Order ID
pDONR223-JAK3	Addgene [120]	23944
luc-CD19 construct	M. Busslinger [121]	
Pax5 (pCCL-cppt-PGK-WPRE)	Addgene [122]	35003
PAX5 (Myc-DDK-tagged) cDNA ORF clone	Origene	RC222785
pMC3 vector	R. Linka [123]	
pRL-TK Vector	Promega	E2241

3.8 Enzymes

Identifier	Provider	Order ID
ApaI	New England Biolabs	R0114S
EcoRI	New England Biolabs	R0101S
MluI	New England Biolabs	R0198S
SacII	New England Biolabs	R0157S
SalI	New England Biolabs	R0138S
SpeI	New England Biolabs	R0133S

3.9 Antibodies

Identifier	Provider	Order ID
Anti-PAX5 antibody	Abcam	ab109443
Anti-β-Actin (AC-74)	Sigma-Aldrich	A2228
FITC Rat Anti-Mouse CD19 (1D3)	BD Biosciences	553785
HSP90 (C45G5)	Cell Signaling	4877
Lamin A (C-20)	Santa Cruz Biotechnology	sc-6214
PE Rat Anti-Mouse IgM (R6-60.2)	BD Biosciences	553409
PerCP Rat Anti-Mouse CD45R/B220 (RA3-6B2)	BD Biosciences	553093
Phospho-Stat5 (Tyr694) (C11C5)	Cell Signaling	9359
Stat5 (3H7) Rabbit	Cell Signaling	9358

3.10 Oligonucleotides

Gene	Primer	Sequence 5'→ 3'	Annealing	Product size
JAK1	hJak1_Exon22_F hJak1_Exon22_R	AAG ACA TCA GGC ACT CTG TCC CGG TCA TCC TTG ACG GTG TAA	60 °C	485 bp
Jak1	mJak1_Ex14_F mJak1_Ex14_R	CCA GAC AGC CAG GAG AAC AG CGT CTG CAT AGT ACC CAC CC	60 °C	277 bp
Jak3	mJak3_genomic_F mJak3_genomic_R	CGG GAT GTG GGG CTT TAA CT GCA GAC ACG GGG TAT AGT GG	60 °C	406 bp
NRAS	hNRAS_Exon2_F hNRAS_Exon2_R	TGG CTC GCC AAT TAA CCC TG GAG ACA GGA TCA GGT CAG CG	60 °C	205 bp
PAX5	hPax5_gDNA_for hPax5_gDNA_rev	CTG ACC GCC CGT CTT TCT C ACT CGC TCC TCT GCA GGT AA	60 °C	275 bp
Pax5	mPax5_Ex3_F mPax5_Ex3_R	CTC GTA CAT GCA CGG AGA CA GGA CCC TTC AGT ACA CCA GC	60 °C	468 bp
Pax5	mPax5_5UTR_F mPax5_3UTR_R	GGA AAC TTT TCC TCG CTG TCC GAA GCT CAT CAA GCG ACC CT	60 °C	1440 bp

Primer	Sequence 5'→ 3'	Size
mJak3_R653H_1_F	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCC CTG TTC CCT CCT GTA ACA C	264 bp
mJak3_R653H_1_R	CAA GCA GAA GAC GGC ATA CGA GAT CGT GAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TAG TGG GAC TGA CAC CAG GAT	
mJak3_R653H_2_F	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCC CTG TTC CCT CCT GTA ACA C	264 bp
mJak3_R653H_2_R	CAA GCA GAA GAC GGC ATA CGA GAT ACA TCG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TAG TGG GAC TGA CAC CAG GAT	
mJak3_V670A_3_F	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAA CGT CTC AGC ACG GAA GG	340 bp
mJak3_V670A_3_R	CAA GCA GAA GAC GGC ATA CGA GAT GCC TAA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TCT GTC ATG GTC ACC TTT GCA C	

Gene	Primer	Sequence 5'→ 3'	Annealing	Product size
PAX5	Pax5_SSP1_F1	CTC GTC GTA CTC CAT CAG CG	65 °C	588 bp
	Pax5_SSP1_R1	CCA ACA CCA TCA TCA GGG CT		
PAX5	Pax5_SSP1_F2	CTC GTC GTA CTC CAT CAG CA	65 °C	794 bp
	Pax5_SSP1_R2	GAG CCC CAG TTA AGG TGT CC		
RHCE	Rh-5e-ds	AGG GCT CAG CTC CAT TCT TT	65 °C	537 bp
	Rh-5e-dR	CAT CAC CCC TCT CTC CAA CA		

Primer	Sequence 5'→ 3'
V$_H$J558	CGA GCT CTC CAR CAC AGC CTW CAT GCA RCT CAR C
	GTC TAG ATT CTC ACA AGA GTC CGA TAG ACC CTG G
V$_H$7183	CGG TAC CAA GAA SAM CCT GTW CCT GCA AAT GAS C
	GTC TAG ATT CTC ACA AGA GTC CGA TAG ACC CTG G
V$_H$Q52	CGG TAC CAG ACT GAR CAT CAS CAA GGA CAA YTC C
	GTC TAG ATT CTC ACA AGA GTC CGA TAG ACC CTG G
DH	TTC AAA GCA CAA TGC CTG GCT
	GTC TAG ATT CTC ACA AGA GTC CGA TAG ACC CTG G
Cμ	TGG CCA TGG GCT GCC TAG CCC GGG ACT T
	GCC TGA CTG AGC TCA CAC AAG GAG GA

Gene	Primer	Sequence 5'→ 3'
Pax5	mPax5_SalI_F	GGC GCG TCG ACG CCA CCA TGG ATT TAG AGA AAA ATT ACC C
	mPax5_SpeI_R	CCT CGA CTA GTT CAG TGA CGG TCA TAG GCG GTG GCT GCG G
	mPax5_P80R_F	GAG ACA GGA AGC ATC AAG CgG GGG GTG ATT GGA GGA TCC
	mPax5_P80R_R	GGA TCC TCC AAT CAC CCC CcG CTT GAT GCT TCC TGT CTC
	mPax5_P80L_F	GAG ACA GGA AGC ATC AAG CtG GGG GTG ATT GGA GGA TCC
	mPax5_P80L_R	GGA TCC TCC AAT CAC CCC CaG CTT GAT GCT TCC TGT CTC
	mPax5_2	TTG CCC ATC AAG GTG TCA GG
	mPax5_3	GCG TCA GCT CCA TCA ACA GG
	mPax5_4	TCA CCA CCA CGG AAC CCA TC
	mPax5_5	CAT GGC TTC ACT GGC TGG AG

Gene	Primer	Sequence 5'→ 3'
	hPax5_SalI_F	GGC GCG TCG ACG CCA CCA TGG ATT TAG AGA AAA ATT ATC C
	hPax5_SpeI_R	CCT CGA CTA GTT CAG TGA CGG TCA TAG GCA GTG GCG GCTG
	hPax5_V26G_F	GTG AAT CAG CTT GGG GGG GgT TTT GTG AAT GGA CGG CCAC
	hPax5_V26G_R	GTG GCC GTC CAT TCA CAA AAc CCC CCC CAA GCT GAT TCA C
	hPax5_G183S_F	CTC GTC GTA CTC CAT CAG CaG CAT CCT GGG CAT CAC GTC C
PAX5	hPax5_G183S_R	GGA CGT GAT GCC CAG GAT GCt GCT GAT GGA GTA CGA CGAG
	pMC3_for	CTC CAT AGA AGA CAC CGG GAC
	hPax5_2	CGA CAT CTC CAG GCA GCT TC
	hPax5_3	CAA TGA CAC CGT GCC TAG CG
	hPax5_4	CAC CAA CAA GCG CAA GAG AG
	hPax5_5	CTT GTT CAC ACA GCA GCA GC
	mJak3_MluI_F	GGC GCA CGC GTA TGG CAC CTC CAA GTG AGG
	mJak3_ApaI_R	CCT CGG GGC CCT CCG GGT CTT CCA CGC CAC
	mJak3_R653H_F	CGG AAG GTG CTC CTG GCT CaT GAG GGG GGT GAT GGG AAT C
	mJak3_R653H_R	GAT TCC CAT CAC CCC CCT Cat GAG CCA GGA GCA CCT TCC G
	mJak3_V670A_F	GCT GAG TGA TCC TGG TGc CAG TCC CAC TGT GCT GAG CCT G
Jak3	mJak3_V670A_R	CAG GCT CAG CAC AGT GGG ACT GgC ACC AGG ATC ACT CAGC
	mJak3_320_F	CTG ACT GGT TTG GGC TGG AG
	mJak3_717_F	TTT GGA CCT GGA GCG GCT AC
	mJak3_1036_F	TTC CGC CTG ATC TGC GAC TC
	mJak3_1401_F	CCT AAC ATC CTG CTG CGC TC
	mJak3_1764_F	GGC TGG AGA CAG CAT CAT GG
	mJak3_2085_F	TCA GAC ACT CGG CTT GGA GG
	mJak3_2470_F	CTG GGC AAG GGC AAC TTT GG
	mJak3_2858_F	TGG AGA GCG AGG CTC ATG TG

3.11 Commercially available Kits

Identifier	Provider	Order ID
AllPrep DNA/RNA Mini Kit	Qiagen	80204
BigDye Terminator v3.1 Cycle Sequencing Kit	Applied Biosystems	4337456
Dual-Glo Luciferase Assay System	Promega	E2920
DyeEx 2.0 Spin Kit	Qiagen	63204
KAPA2G Fast Ready Mix PCR Kit	Peqlab	KK5103
NE-PER Nuclear and Cytoplasmic Extraction Kit	Thermo Fisher Scientific	78833
NucleoBond Xtra Maxi EF	Macherey Nagel	740424.50
Phusion High-Fidelity PCR Master Mix	New England Biolabs	M0531S
QIAquick Gel Extraction Kit	Qiagen	28706
QIAquick PCR Purification Kit	Qiagen	28106
QuantiTect Reverse Transcription Kit	Qiagen	205313
RNase-Free DNase Set	Qiagen	79254
Roche Rapid DNA Ligation Kit	Roche	11635379001
SureSelectXT Human All Exon V5+UTRs	Agilent	5190-6213
SureSelectXT Mouse all Exon kit	Agilent	5190-4641
Taq PCR Core Kit	Qiagen	201223
TruSeq SBS Kit v3	Illumina	FC-401-3001

3.12 Consumables and Devices

Identifier (Company)

3130 Genetic Analyzer (Applied Biosystems)

Agilent 2100 Bioanalyzer (Agilent Technologies)

Amersham Protran Supported 0.45 NC (GE Healthcare)

BD AccuriC6 Cytometer (BD Biosciences)

Blotting paper, extra thick (Bio-Rad)

Cell culture flasks (25 cm^2, 75 cm^2, 175 cm^2) (Corning)

Cell culture plates (96-well, 48-well, 24-well, 6-well) (Corning)

Centrifuge 5403 (Eppendorf)

Centrifuge 5417R (Eppendorf)

Centrifuge tubes (1.5 & 2 ml) (Eppendorf)

CO_2 incubator for cell culture (Thermo Fisher Scientific)

Cryo freezing container (Nalgene)

Dynabeads MPC-S (Magnetic Particle Concentrator) (Thermo Fisher Scientific)

Falcon tubes (15 ml & 50 ml) (Greiner Bio-One)

HeroLab UVT 2035 UV lamp (HeroLab)

HiSeq 2500 (Illumina)

LAS-3000 mini 2UV Transilluminator (Fujifilm)

Mastercycler gradient (Eppendorf)

Milli-Q Integral 15 (Millipore)

NanoDrop Spectrophotometer ND-1000 (PeqLab)

PH meter 7310 (Inolab)

Spark 10M (Tecan)

Thermomixer Comfort (Eppendorf)

Transfer Blot Mini Trans Blot Cell (Bio-Rad)

UVC/T-M-AR, DNA/RNA UV-cleaner box (Kisker)

Vi-cell XR (Beckman Coulter)

Vortex2 (Genie Scientific Industries)

3.13 Software

Identifier (Website)

Control-FREEC (http://boevalab.com/FREEC/index.html#downloads)

Flowjo - FACS analysis (http://www.flowjo.com/)

Graphpad Prism (http://www.graphpad.com/scientific-software/prism/)

Illumina GenomeStudio v2011.1 (http://support.illumina.com/downloads.ilmn)

MuTect (http://archive.broadinstitute.org/cancer/cga/mutect)

PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/dokuwiki/downloads)

SIFT (http://sift.jcvi.org/)

Snupy (https://snupy.bio.inf.h-brs.de/)

4 Methods

4.1 Cell cultivation

Hek293T cells [1-3] were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmBH (DSMZ) (ACC 635) and cultivated in DMEM Medium GlutaMAX (Gibco), supplemented with 10% (v/v) heat inactivated fetal calf serum (PAN) and Penicillin-Streptomycin (100 units/mL of penicillin and 100 µg/mL of streptomycin, Gibco), at 37 °C and 5% CO_2.

Ba/F3 cells [1, 2] were obtained from DSMZ (ACC 300) and cultivated in RPMI 1640 Medium GlutaMAX (Gibco), supplemented with 10% (v/v) heat inactivated fetal bovine serum (PAN), Gentamicin (50 µg/ml) and recombinant mouse Interleukin-3 (10 µg/ml) (Gibco), at 37 °C and 5% CO_2. The cell density was kept $<2 \times 10^6$ cells/ml to avoid the outgrowth of a cytokine-independent subclone. The total number of viable cells was determined with either a Neubauer counting chamber or on a Vi-cell XR cell viability analyzer (Beckman Coulter).

4.2 Cryopreservation of cells

Cells were removed from the culture flask with (adherent culture) or without (suspension culture) trypsin and pelleted for 5 min at 400 x g. Either 2.5×10^6 cells/ml (adherent cells) or 5×10^6 cells/ml (suspension cells) were resuspended in 1.8 ml freezing medium (90% FBS and 10% DMSO) and transferred to a cryotube. The cryotubes were frozen in a cryobox filled with isopropanol at -80 °C for at least 24 h. For long term storage, the cells were kept in the gas phase of liquid nitrogen. For recultivation, the cryotube was thawed in a water bath at 37 °C. In order to completely remove the freezing media, the cells were centrifuged at 400 x g for 5 min and the pellet resuspended in fresh culture medium.

4.3 Flow Cytometric analysis

Nucleated cells were obtained from total mouse bone marrow (BM) (by flushing from the long bones), peripheral blood (PB) or spleen. Contaminating red blood cells were lysed with ACK Lysing Buffer (Thermo Fisher Scientific) and the remaining cells washed in PBS with 1% FBS. After staining, all cells were washed once in PBS and were resuspended in PBS with 1% FBS containing 2 mg/mL propidium iodide (PI), in order to allow the exclusion of dead cells. The data were acquired on an AccuriC6 Flow Cytometer (BD Biosciences) and analyzed using the Flowjo software. Specific fluorescence of FITC/PE and APC excited at 488 nm and 633 nm, respectively, as well as known forward and orthogonal light scattering properties of mouse cells were used to set gates. For each analysis, a total of at least 50,000 viable cells (minus PI) were assessed. Antibodies were used at a 1:100 dilution. Flow cytometric analysis was carried out in cooperation with Dr. Isidro Sánchez-García's laboratory from the Instituto de Biologia Molecular y Celular del Cancer (IBMCC), CSIC/Universidad de Salamanca (Spain).

4.4 DNA/RNA preparation

4.4.1 From whole blood

For the isolation of lymphocytes from whole blood samples, a co-polymer of sucrose-epichlorohydrin, commonly known under its commercial brand name "Ficoll", was used (Ficoll-Paque Plus, GE Healthcare). It is based on the principle that peripheral blood mononuclear cells (PBMCs) (lymphocytes & monocytes) show a lower buoyant density in solution, compared to polymorphonuclear cells (PMNCs) (granulocytes). The ficoll has a density of 1.077 g/ml, and can therefore serve as an isosmotic medium, allowing the heavier erythrocytes and PMNCs to sediment through, while retaining the majority of PBMCs (density below 1.077 g/ml) at the interface between plasma (top) and ficoll (bottom) layers (**Figure 8**). First, whole blood was diluted 1:1 with PBS, before being

layered over 15 ml of ficoll in a 50 ml falcon tube. The sample was centrifuged for 20 min at 1003 x g at RT with no active brake. The white layer containing PBMCs (also referred to as buffy coat) was carefully transferred to a new falcon tube and PBS was added up to a volume of 50 ml. The solution was centrifuged for 10 min at 600 x g, the supernatant discarded and the resulting pellet washed twice with PBS. Afterwards, the pellet was resuspended in 10 ml PBS and centrifuged for 10 min at 200 x g. The supernatant containing thrombocytes was discarded and the cell pellet containing the freshly isolated PBMCs was either directly used for DNA/RNA preparation or resuspended in freezing media (90% FBS, 10% DMSO) and frozen at -80 °C.
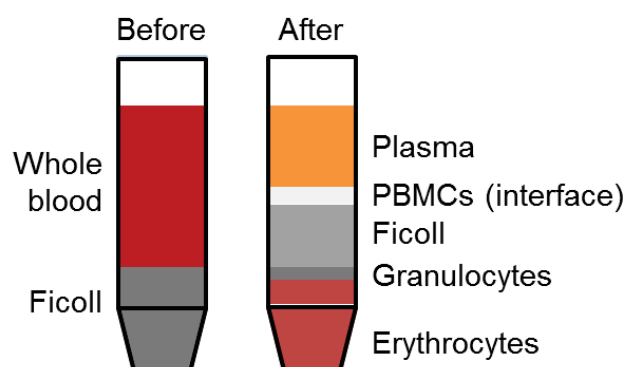
Figure 8: Schematic drawing of ficoll density gradient. After centrifugation, the PBMCs can be withdrawn from the interphase between plasma and ficoll. Adapted from [124].

4.4.2 From cultured cells

5×10^6 cells were centrifuged (400 x g, 5 min) and washed once with PBS, before they were resuspended in 600 µl RLT buffer and processed for DNA/RNA preparation using the AllPrep DNA/RNA Mini Kit (Qiagen) according to manufacturer's instructions. The optional DNase digestion was performed for RNA preparations in order to clear the RNA of possibly contaminating DNA residues. Eluted DNA was stored at -20 °C and RNA at -80 °C until use.

4.4.3 DNA/RNA concentration and purity

To determine the DNA or RNA concentration, the respective sample was assessed on a NanoDrop spectrophotometer (PeqLab) by measuring its absorbance at a

wavelength of 260 nm. Samples with an E260/E280 ratio above 1.8 were considered as pure solutions and used for further experiments.

4.4.4 Complementary DNA synthesis

Complementary DNA (cDNA) was synthesized from RNA using the QuantiTect Reverse Transcription Kit (Qiagen) according to the manufacturer's protocol. To achieve equal expression levels between different samples, all RNAs were diluted to a concentration of 100 ng/µl, before equal amounts were used for cDNA synthesis.

4.5 DNA Amplification

4.5.1 Standard Polymerase Chain Reaction

Standard Polymerase Chain Reaction (PCR) is used to amplify a specific, short DNA region to create millions of identical copies. This method enables the detection of target gene regions on an agarose gel, or is the start point for further downstream processing, like sequencing or cloning. In addition to the template DNA and unique single stranded DNA oligonucleotide primers, free deoxynucleotide triphosphates (dNTPs) are added to the reaction, which serve as the essential building blocks for the newly synthesized DNA strands. Furthermore, a heat resistant DNA polymerase from the bacterium *Thermus aquaticus* (Taq) catalyzes the DNA synthesis in a series of heating (DNA denaturation) and cooling (DNA annealing and elongation) steps. Standard PCR was performed with the following PCR mix and program (**Table 3**) using the Taq PCR Core Kit (Qiagen):

PCR Mix	
10 x Buffer	2.5 µl
Forward Primer (10 µM)	0.5 µl
Reverse Primer (10 µM)	0.5 µl
dNTPs (10 mM)	0.5 µl
Taq Polymerase	0.1 µl
20-50 ng DNA	x
dH$_2$O	up to 25 µl

Step	Temperature	Time	
Initial Denaturation	94 °C	2 min	
Denaturation	94 °C	30 sec	
Annealing	60 °C	30 sec	35 cycles
Elongation	72 °C	1 min	
Final Elongation	72 °C	7 min	

Table 3: Standard polymerase chain reaction components and program.

4.5.2 Visualization and cleanup

The final PCR product was visualized using gel electrophoresis. It is based on the principle that, when loaded onto an agarose gel, negatively charged DNA will travel towards a positively charged anode at a speed that is inversely related to their length when an electrical field is applied. Separated DNA molecules were detected using Ethidium bromide as a DNA intercalating dye, and their length evaluated with a suitable DNA ladder (Thermo Scientific), composed of molecules of defined lengths. In order to purify the final PCR product, the DNA was prepared directly using either the QIAquick PCR purification Kit (Qiagen) or after extraction from the agarose gel, using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's protocol.

4.6 V(D)J clonality PCR assay

In order to verify a clonal origin of blast cells, immunoglobulin rearrangements from murine leukemic samples were amplified by PCR using the primers VHJ558, VH7183, VHQ52 DH and Cµ and following cycling conditions (**Table 4**):

Step	Temperature	Time	
Initial Denaturation	95 °C	30 sec	
Denaturation	95 °C	1 min	
Annealing	65 °C	1 min	35 cycles
Elongation	72 °C	45 sec	
Final Elongation	72 °C	10 min	

Table 4: PCR program for V(D)J clonality assay

Ambiguous positions in the DNA sequence were marked with K (G or T), M (A or C), S (C or G), R (A or G), W (A or T) and Y (C or T) in the primer sequence.

4.7 Complementary DNA deletion screening

1 µg BM RNA from leukemic $Pax5^{+/-}$ mice was reverse transcribed in cDNA as explained in 4.4.4. To amplify the complete $Pax5$ gene, a standard PCR (2 min elongation) was carried out with primers located in the gene's 3' and 5' untranslated region (UTR) (mPax5_5UTR_F/mPax5_3UTR_R). The PCR products were visualized on an agarose gel, gel extracted and each band was identified by Sanger sequencing.

4.8 Next generation sequencing

In general, sequencing technology relies on the principle that fluorescent labeled dNTPs are incorporated into a DNA template strand during repeated cycles of DNA synthesis. In each cycle at the point of incorporation, they can be identified by fluorophore excitation, thereby revealing the exact nucleotide sequence of the DNA molecule. Next generation sequencing (NGS) allows this process to take place in a

massively parallel manner across millions of fragments simultaneously. In this study, next generation sequencing technology from Illumina was used (**Figure 9)**. Tumor (C3L, C6L and C7L) and germline (B1, B2, B4, B5, C1, C2, C3, C4, C5, C6, C7) DNA was extracted and purified using the AllPrep DNA/RNA Mini Kit (Qiagen) according to the manufacturer's instructions.
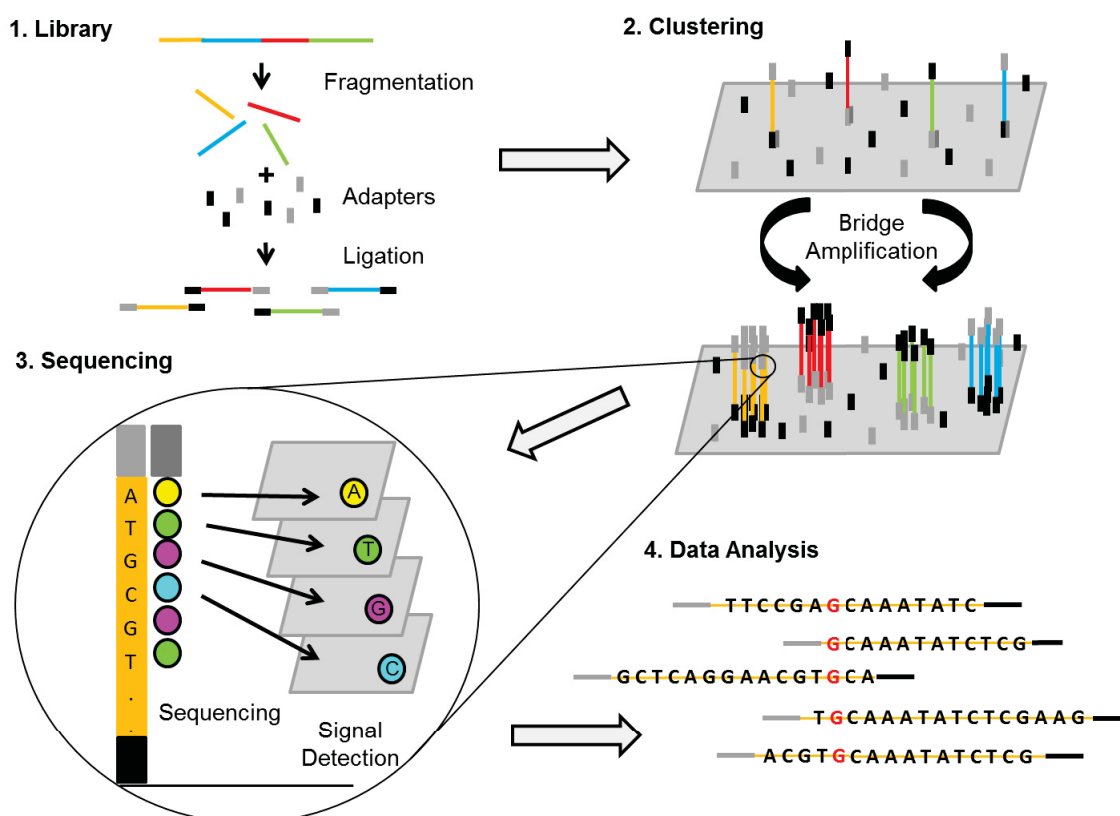
Figure 9: Ilumina whole exome sequencing workflow. (1) DNA samples to be analyzed are mechanically fragmented and ligated to adapters in order to obtain the sequencing library. (2) This library is further loaded onto a flow cell, where it is used as a template for subsequent PCR amplification, which results in clusters with identical DNA fragments. (3) Subsequent sequencing-by-synthesis (SBS) using florescence-tagged molecules is carried out to identify the DNA sequence of each cluster. (4) The sequenced fragments are aligned to obtain the full length sequence of the analyzed DNA sample. Adapted from Illumina's "An Introduction to Next-Generation Sequencing Technology".

4.8.1 Library preparation

In order to obtain a sequencing library, the DNA samples were randomly fragmented by shearing, before adapters were ligated to the 5' and 3' end. Exome library preparation was performed using the SureSelectXT Human or Mouse All Exon kit (Agilent) with modifications adapted from Fisher *et al.*, 2011 [125]. Briefly, SPRI beads were added to the original protocol and the size of the reaction reduced to 0.5 µl. Furthermore, the

volume for washing was reduced and freshly prepared 20% PEG/2.5 M NaCl was added instead of elution of the samples from the SPRI beads.

4.8.2 Cluster generation

In order to generate clusters to serve as a template for sequencing, the whole exome library was loaded on a flow cell. During this step the fragmented DNA is covalently hybridized to surface-bound oligos, which are complementary to the library adapters. Clusters from each single fragment are generated via a PCR-based approach (Bridge-amplification). The targeted capture was performed according to the manufacturer's protocol. Purification and enrichment after hybridization was carried out using MyOne Streptavidin T1 Dynabeads (Life Technologies) and off-bead PCR amplification in the linear range.

4.8.3 Sequencing

The sequencing reaction was carried out stepwise with the sequencing-by-synthesis (SBS) method, using reversible termination chemistry and florescence tagged molecules. In each sequencing cycle only the reversible terminator-bound dNTPs, which are complementary to one of the template DNA clusters, are incorporated. Afterwards, the fluorescence group is cleaved, its specific light signal detected and the terminator group removed in order to start a new sequencing cycle. Illumina's paired-end (PE) sequencing further allows sequencing of both ends of the DNA fragment with read lengths between 100-250 bp. These sequences are aligned as forward and reverse reads as so-called read pairs, which increase accuracy and enable the detection of insertion or deletions (indels). Here, 2 x 100 bp sequencing with an either 6 or 8 bp index read was performed using the TruSeq SBS Kit v3 on a HiSeq 2500 platform (Illumina).

4.8.4   Data analysis

To analyze the obtained data, the newly created sequence reads were aligned to a reference genome from an online database in order to identify DNA variants, like single nucleotide polymorphisms (SNP) or indels. First, Fastq files were generated using BcltoFastq 1.8.4 (Illumina). The BWA version 0.7.4. [126] was used to align the sequence data to either the human (GRCh37.p13) or mouse (GRCm38.p4) reference genome [127]. Conversion steps were performed using Samtools [128, 129], followed by the removal of duplicate reads. Local realignment around indels, SNP-calling, annotation and recalibration was facilitated by GATK 2.4.9 [130]. Resulting variation calls were annotated by the Variant Effect Predictor [131] using the Ensembl database (v84) [132]. Afterwards, the data was imported to our in-house MySQL database, called Single NUcleotide PolYmorphism (SNUPY), to facilitate automatic and manual annotation, reconciliation and data analysis by complex database queries. Predictions from the online prediction tools SIFT [133] and Polyphen-2 [134] were integrated into the SNUPY data output. Somatic calls were analyzed using MuTect [135]. Only entries with at least 9% difference in allele frequency between tumor and germline sample were kept for further analysis. Cancer-related genes were retrieved from the cancer gene census list (COSMIC [136]) using ENSEMBL's biomart [137]. For the detection of copy number alterations, the alignments were passed to Control-FREEC (version 7.2) [138], which was configured to use overlapping 50 kb windows in 10 kb steps and to normalize read counts within each window by a paired normal control sample.

4.9   Amplicon sequencing

Amplicon sequencing is a highly specific, ultra-deep sequencing technique, which allows the identification of a variant occurring even in very low copy numbers. Therefore PCR products of targeted regions of interest are sequenced by NGS, enabling a read depth up to 2 million reads per sample. Using this method, an identified somatic mutation in the tumor sample can be traced back to pinpoint, for example, its

first appearance in the blood. For deep sequencing, product-specific primers (mJak3_R653H_1_F/R, mJak3_R653H_2_F/R, mJak3_V670A_3_F/R) containing Illumina TruSeq Adapters were used in a standard PCR Mix with the following PCR program (**Table 5**):

Step	Temperature	Time	
Initial Denaturation	94 °C	2 min	
Denaturation	94 °C	30 sec	
Annealing	65 °C, down 1 °C per cycle	30 sec	10 cycles
Elongation	72 °C	20 sec	
Denaturation	94 °C	30 sec	
Annealing	55 °C	30 sec	25 cycles
Elongation	72 °C	20 sec	

Table 5: PCR program for amplicon sequencing.

Resulting PCR products were visualized on an agarose gel and purified using the QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's instructions, followed by a short fragment removal (see 4.9.1). All steps were carried out using a DNA/RNA UV-cleaner box (Kisker) to avoid cross contamination. The final samples were checked for purity and adjusted to a concentration of 2 M on a Agilent 2100 Bioanalyzer (Agilent Technologies), before they were sequenced on a HiSeq 2500 (Illumina).

4.9.1 Short fragment removal

To purify the amplicon for later sequencing, fragments below 100 bps were removed using Agencourt AMPure XP Beads (Beckman Coulter). Therefore, the DNA sample was mixed 1:1 with AMPure Beads in a 1.5 ml eppendorf tube, vortexed and incubated at RT for 10 min. Next, the tube containing the sample was put on a magnetic particle concentrator (MPC) (Thermo Fisher Scientific) for 5 min. The supernatant was discarded and the pellet washed with 500 µl of 70% ethanol on the MPC. Afterwards,

the supernatant was removed completely and the pellet dried on the MPC for 10 min at RT. For elution, the pellet was resuspended in 50 µl EB buffer (PCR Purification Kit – Qiagen) and incubated at RT for 1 min. The sample was put on the MPC and the supernatant containing the purified DNA was transferred to a new microcentrifuge tube.

4.10 Sanger sequencing

All candidate variants identified by whole exome sequencing were verified using Sanger sequencing. Amplified regions around the variant were obtained by standard PCR with the respective primer pair, and purified using the QIAquick PCR Purification Kit (Qiagen). Purified PCR products were prepared for sequencing using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) according to manufacturer's instructions. To remove unincorporated dye terminators, the obtained products were purified using the Dye Ex 2.0 Spin Kit (Qiagen). Sanger sequencing was performed on a 3130 Genetic Analyzer (Applied Biosystems). Primer pairs hJak1_Exon22_F/R, hNRAS_Exon2_F/R and hPax5_gDNA_for/rev were used for human validation and mJak1_Ex14_F/R, mJak3_genomic_F/R and mPax5_Ex3_F/R for mouse validation.

4.11 Single nucleotide polymorphism genotyping

Single Nucleotide Polymorphism (SNP) genotyping can be utilized with microarray technology, to detect genetic variations of SNPs between different individuals of the same species. Therefore, it provides a genetic fingerprint, which can for example be used to assess the inheritance of different alleles in a family pedigree. Moreover the microarray data can be employed to detect copy number variants (CNVs) and other structural aberrations. Therefor a whole genome amplified and fragmented DNA sample is hybridized to the complementary locus-specific 50-mers, which are covalently linked to one of up to one million bead types. The captured DNA is then used as a template for single-base extension of the oligos on the BeadChip, thereby

incorporating a detectable label on the BeadChip, which determines the genotype call for the DNA sample. SNP genotyping was performed with germline (C3, C6 and C7) and leukemia samples (C3L, C6L and C7L) of the affected children. The Illumina Human660W-Quad_v1 BeadChip (Illumina) was used, which offers 559,348 SNPs and 95,866 Copy Number probes. Genotype calling and CNV analysis was performed with the Illumina GenomeStudio v2011.1. SNP genotyping and data analysis was carried out in the Max-Delbrueck-Center for Molecular Medicine (MDC) (Berlin-Buch, Germany) by the group of Franz Rüschendorf.

4.12 Sequence specific primer PCR

The sequence specific primer (SSP) PCR was originally designed as a molecular typing method to define human leukocyte antigen (HLA) class I & II alleles [139]. The technique relies on the design of allele sequence-specific primers, which selectively amplify a target sequence. It is based on the principle that, under controlled PCR conditions, the base on the 3 prime end of a primer needs to be matched to the target sequence in order to ensure the amplification of a specific product. If a mismatch occurs at this position, no amplicons are generated when high annealing temperatures are applied. Therefore, the primers are designed in a way that the target mutation is positioned as the last base of the primers 3' end. The PCR reaction is carried out using SSP primers for the WT, as well as the mutant allele. Depending on which PCR reaction yields a positive product, conclusions can be drawn regarding the existence of a specific mutation in the genomic DNA. In addition to the SSPs, an internal control primer pair is added, which amplifies a conserved region of DNA and therefore ensures the integrity of the DNA sample.

For the identification of human *PAX5* c.547G>A, SSP-PCR was carried out using the KAPA2G Fast ReadyMix PCR Kit (Peqlab) with the following PCR conditions (**Table 6**):

| Step	Temperature	Time	
Initial Denaturation	95 °C	3 min	
Denaturation	95 °C	15 sec	
Annealing	65 °C	15 sec	35 cycles
Elongation	72 °C	15 sec	
Final Elongation	72 °C	1 min	

Table 6: PCR program for Sequencing specific primer PCR.

DNA samples were obtained from buccal swab samples taken from mixed male and female donors between the age of 18 and 55 (kindly provided by the Institute of Transplantation Diagnostics and Cell Therapeutics (ITZ); University Clinic Duesseldorf). In two separate PCR reactions, each DNA was amplified with the SSP1_F1/R1 primer for the WT, as well as the SSP1-F2/R2 primer pair for the mutant allele. The primer pair Rh-5e-ds/dR, containing a conserved region of the Rhesus (Rh) factor gene was used as a control in the mutant PCR in order to always get a positive signal. PCR products were separated on a 2% agarose gel.

4.13 Cloning

4.13.1 Pax5 constructs

The coding sequence (CDS) of the murine or human Pax5 WT was obtained from the Pax5 Plasmid from Addgene [57] (#35003) or from Origene (RC222785), respectively, via PCR using the Phusion High-Fidelity DNA Polymerase (Thermo Scientific) with the primer pair Pax5_SalI_F/Pax5_SpeI_R. The mutant sequences for murine *Pax5* (p.P80R and p.P80L) and human *PAX5* (p.G183S and p.V26G) were created by site directed mutagenesis by PCR (**Figure 10**).
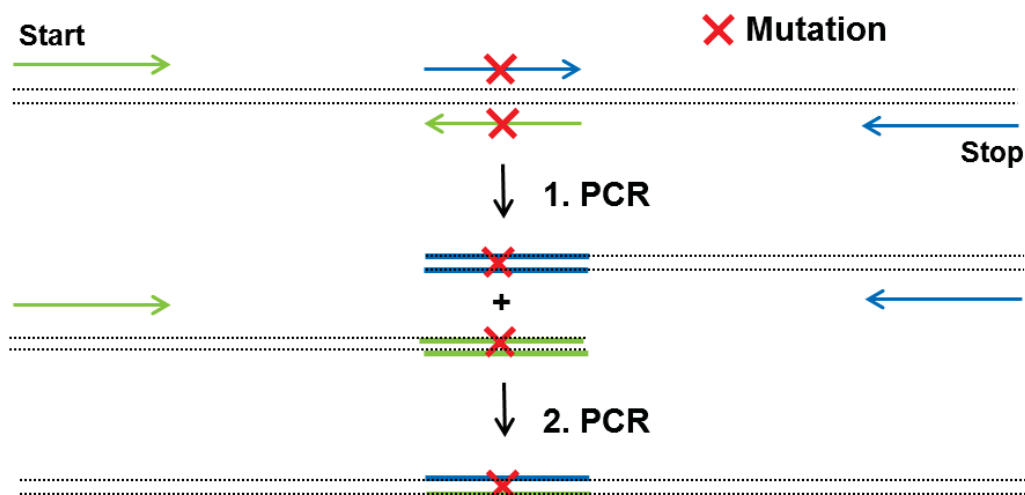
Figure 10: Schema describing the site directed mutagenesis by PCR method. Forward and reverse primers containing the target mutation are used in combination with primers located on the end and start of the coding sequence respectively. The two obtained "left" and "right" fragments are then used as templates in a subsequent PCR reaction with both outer primers, thereby amplifying the whole coding region with the target mutation.

Therefore, specific *forward* and *reverse* oligos were designed containing the target region with the respective mutation (mPax5_P80R_F/R, mPax5_P80L_F/R, hPax5_V26G_F/R and hPax5_G183S_F/R). Next, the *forward* oligo was used together with Pax5_SpeI_R to amplify the coding region downstream of the mutation, while the *reverse* oligo was combined with Pax5_SalI_F to amplify the upstream coding sequence respectively. These two PCR products then served as templates in a subsequent PCR reaction with the outer primers Pax5_SalI_F and Pax5_SpeI_R in order to combine both regions and to obtain the complete coding sequence harboring the mutation. For site directed mutagenesis PCR conditions according to (**Table 7**) were used. The restriction enzymes SalI/SpeI (New England Biolabs) and Rapid DNA Ligation Kit (Roche) were utilized according to manufacturer's protocol to clone the obtained *Pax5* sequences into a derivative of a bicistronic expression vector (pMC3), which was used previously for stable expression of genes in cell lines [58]. Here, the vector was modified to encode the hygromycin resistance gene as the second cistron (pMC3-Pax5.Hygro). The identity of the respective sequences was confirmed by control digest using EcoRI and Sanger sequencing with the primers pMC3_for, Pax5_2, Pax5_3, Pax5_4 and Pax5_5.

PCR Mix (1)	
Phusion PCR Master Mix	25 µl
Forward Primer (10 µM)	2.5 µl
Reverse Primer (10 µM)	2.5 µl
20-50 ng DNA	x
dH2O	up to 50 µl

PCR Mix (2)	
Phusion PCR Master Mix	25 µl
Forward Primer (10 µM)	2.5 µl
Reverse Primer (10 µM)	2.5 µl
PCR product upstream	5 µl
PCR product downstream	5 µl
dH2O	10 µl

Step	Temperature	Time	
Initial Denaturation	98 °C	30 sec	
Denaturation	98 °C	7 sec	
Annealing	67 °C	23 sec	35 cycles
Elongation	72 °C	1 min	
Final Elongation	72 °C	7 min	

Table 7: PCR components and program for cloning.

4.13.2 Jak3 constructs

The CDS of murine *Jak3* WT was obtained from the Riken Full Length cDNA Clone A130091E14 (Source Bioscience). WT sequence and p.R653H and p.V670A mutant sequences were obtained by site directed mutagenesis by PCR as described above using the outer primers mJak3_MluI_F, mJak3_ApaI_R in combination with the mutant-specific primers mJak3_R653H_F/R and mJak3_V670A_F/R. Next, the coding sequences were cloned into the pMC3 expression vector containing a Hygromycin resistance (pMC3-Jak3.Hygro) using the restriction enzymes MluI and ApaI (New

England Biolabs) and the respective cDNAs were confirmed by control digest with SacII and Sanger sequencing (primers: pMC3_for, mJak3_320_F, mJak3_717_F, mJak3_1036_F, mJak3_1401_F, mJak3_1764_F, mJak3_2085_F, mJak3_2470_F, mJak3_2858_F).

4.13.3 Preparation of chemically competent *E.coli*

For the preparation of chemically competent bacteria, the *Escherichia coli* (*E.coli*) strain DH5α was used following the protocol of Sambrook & Russell [140]. In short, a culture of DH5α was grown while shaking at 37 °C in LB media until an OD_{600} of 0.4 - 0.6, centrifuged (5000 x g, 20 min at 4 °C) and the pellet resuspended with buffer RF1 in 1/3 of the starting volume. After incubation time of 90 min at 4 °C the solution was centrifuged again and the resulting pellet was resuspended in 1/12.5 of the starting volume with RF2 buffer. Subsequently, the chemically competent *E.coli* suspension was aliquoted in 75 µl portions, while constantly kept on ice in a 4 °C room. Aliquots were shock frozen in liquid nitrogen and stored for later use at -80 °C

4.13.4 Transformation of chemically competent *E. coli*

Chemically competent *E.coli* were thawed on ice, before 3 µl of the ligation reaction was added. The solution was mixed gently and put on ice for 30 min. Afterwards the sample was heat-shocked at 42 °C for 30 sec, followed by a 2 min recovery on ice. Carefully the mix was resuspended in 250 µl LB medium (without antibiotic) and incubated at 37° C for 45 min while shaking. Transformed *E.coli* cells were plated onto LB agar plates containing 100 µg/ml ampicillin and incubated overnight at 37 °C.

4.13.5 Maxi Preparation

Prior to plasmid preparation, 3 ml LB medium with 100 mg/ml Ampicillin were inoculated with a single bacteria colony containing the plasmid of interest and grown at 37 °C while shaking for 6-8 h. 200 µl of the bacterial culture were further diluted in 200 ml LB medium with the respective antibiotic and grown at 37 °C while shaking

overnight. In order to extract and purify the plasmid DNA, the overnight culture was prepared using the NucleoBond Xtra Maxi EF Kit (Macherey-Nagel) according to the manufacturer's protocol. Resulting plasmid DNA was checked for concentration and purity on a NanoDrop spectrophotometer and stored at -20 °C until further use.

4.13.6 Transfection of cultured cells

4.13.6.1 Attractene

Hek293T cells were transfected using the Attractene Transfection Reagent (Qiagen), according to manufacturer's protocol. Briefly, 1.2 µg of each pMC3-Pax5.Hygro construct or the empty vector were diluted in 100 µl serum-free media, before 4.5 µl of Attractene was added. After an incubation period of 15 minutes the DNA/Attractene mix was added to Hek293T cells (6×10^5) in suspension and plated on a 6-well plate. Cells harboring the plasmids were selected using 200 µg/ml Hygromycin B (Life Technologies), which was first administered 48 h after transfection.

4.13.6.2 Nucleofection

Ba/F3 cells were transfected using the Amaxa Nucleofector Technology (Lonza), according to manufacturer's protocol. Briefly, Ba/F3 cells (5×10^6) were resuspended in 100 µl Amaxa Nucleofector solution V, containing either 5 µg of pMC3-Jak3.HygroWT, pMC3-Jak3.HygroR653H, pMC3-Jak3.HygroV670A or the empty vector and electroporated using the Nucleofector program X-01. Cells harboring the plasmids were selected using 600 µg/ml Hygromycin B (Life Technologies), which was first added 48 h after transfection.

4.14 Luciferase reporter assay

Luciferase reporter assays are used to study the gene expression of a target gene with or without the variant of interest at the transcriptional level. Here, firefly luciferase was used as a reporter gene, which is able to release light via the conversion of luciferin to oxyluciferin. The firefly was cloned downstream of the regulatory elements of the target

gene and could therefore function as a "reporter" of the genes transcriptional status, when transfected into a cell. The Renilla luciferase (also known as sea pansy luciferase) catalyzes a similar reaction like the firefly luciferase, using coelenterazine as a substrate, but generates light of a shorter wavelength. A vector with constitutive Renilla expression served as an internal control, which was used for normalization of the firefly luciferase reporter gene. Here, a reporter construct containing a CD19 promoter derived high affinity binding site (luc-CD19), which is a target of the *PAX5* gene, was used. Hek293T cells expressing pMC3-Pax5.HygroWT, pMC3-Pax5.HygroEmpty or pMC3-Pax5.HygroMutant were transfected with 2 µg luc-CD19 construct (kindly provided by M. Busslinger) [121] and 100 ng pRL-TK Renilla luciferase plasmid DNA (Promega) using FuGene 6 (Promega). 48 h after transfection, cell lysis and subsequent firefly and Renilla luciferase activity measurement was performed on a Spark 10M multimode microplate reader (Tecan) using the Dual-Glo Luciferase Assay System (Promega), according to manufacturer's instructions. All transfections were carried out in triplicate in at least 3 independent experiments. The firefly luciferase activity was normalized according to the corresponding Renilla luciferase activity.

4.15 Subcellular fractionation

Cellular and nuclear protein fractions of Pax5.WT as well as Pax5.Mutant proteins were obtained by using stable transfected Hek293T cells that were separated in cellular and nuclear protein fractions using NE-PER Nuclear and Cytoplasmic Extraction Reagent (Thermo Fisher Scientific), according to the manufacturer's instructions.

4.16 Immunoblotting

Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) separates proteins according to the relative size of pores formed within a polyacrylamide gel. To cleave disulfide bonds, Laemmli buffer containing 2-mercaptoethanol is added. The

anionic detergent SDS in the Laemmli and running buffer also breaks down secondary and tertiary structures (without disulfide bonds) and moreover applies a negative charge to each protein in proportion to its mass. When a voltage is impressed on the polyacrylamide gel harboring the protein samples, the negative charged proteins migrate through the gel matrix toward the positively charged anode, thereby fractionating themselves. To increase the resolution during separation, a stacking gel is used. Afterwards, the protein enters a resolving gel, where the protein is separated based on its size. Whole cell extracts from Ba/F3 cells were either obtained as previously described [141] or by subcellular fractionation, as described above. Spleen samples were prepared using a 100 μm cell strainer and depleted of erythrocytes using NH4Cl, before lymphocytes were lysed in RIPA buffer, containing protease and phosphatase inhibitors (Roche Diagnostics). 20 μg of whole protein was separated on SDS-PAGE (20 mA per gel) and transferred to Hybond-C Extra membranes (Amersham Biosciences) (200 mA, 2 h). Next, the blotted membrane was saturated by 1 h incubation in 5% milk powder dissolved in T-BST. After 3 washing steps in T-BST, immunoblotting was carried out overnight in 5% BSA (dissolved in T-BST) using the following antibodies and dilutions: anti-phospho-STAT5 1:1000, anti-STAT5 1:1000, anti-HSP90 1:2500 (Cell Signaling), anti-PAX5 ab109443 1:5000 (Abcam), anti-Lamin A 1:500 (Cell Signaling) and anti-β-Actin clone AC-74 1:10000 (Sigma-Aldrich). Detection was achieved using anti-rabbit, anti-mouse or anti-goat horseradish peroxidase conjugates (Santa Cruz Biotechnology), respectively, with an enhanced chemiluminescent HRP substrate (Thermo Fisher Scientific).

4.17 IL-3 depletion assay

Ba/F3 cells are a murine pro-B cell line the survival of which is dependent on the presence of the cytokine IL-3 in the growth media. In case of IL-3 withdrawal the cells are only able to survive when they express a constitutively active tyrosine kinase or similar oncogenes. Therefore, the Ba/F3 cell system can be used to test potentially

activating mutations in the respective genes. To assess the *Jak3* mutants p.R653H and p.V670A, 2 x 10^5 Ba/F3 cells expressing either murine Jak3.WT, Jak3.R653H, Jak3.V670A or the empty vector control, were washed twice with medium without IL-3, before they were cultured in the absence of IL-3 for 10 days. Proliferation was measured by counting the cells every day using Trypan Blue (Sigma-Aldrich).

4.18 Statistical analysis

Statistical data analysis was performed with the Graphpad Prism software with the appropriate statistical test. The level of significance is indicated according to the following nomenclature:

ns = not significant; p > 0.05

* = p ≤ 0.05

** = p ≤ 0.01

*** = p ≤ 0.001

5 Results

5.1 Family with three cases of pediatric B cell precursor ALL

Although it is well established that somatic mutations are key events in the development and progression of childhood B cell precursor acute lymphoblastic leukemia (BCP-ALL) [45], the relevance of inherited predisposition to disease development is only rudimentarily understood. In order to increase the chances of identifying low penetrance susceptibility loci or genomic variations that functionally contribute to BCP-ALL leukemogenesis, families with recurrent leukemia cases can be screened for common disease-causing variants. We characterized a family from Ashkenazi Jewish origin, with three cases of BCP-ALL in childhood (**Figure 11**).
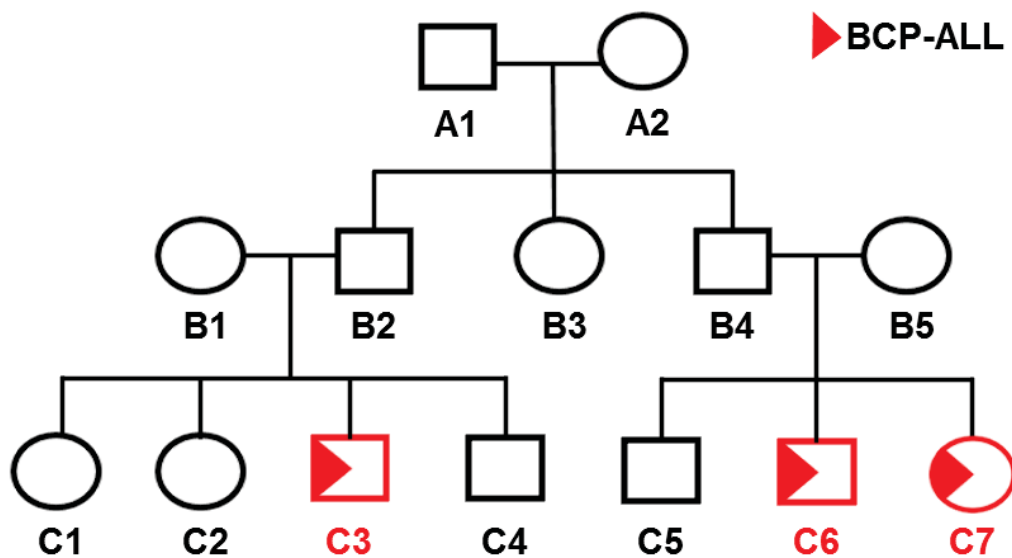
Figure 11: Family pedigree with three cases of BCP-ALL in childhood. Red triangles mark BCP-ALL cases, which all occurred in generation C of the family (C3, C6 and C7).

Two of the affected children are siblings (C6 and C7), while one is their cousin (C3). The leukemias occurred at the ages of 55 months (C3), 16 months (C6) and 21 months (C7), with C3 suffering from a relapse 8 years after the initial diagnosis. The immunophenotype of the leukemias was analyzed as CD19, CD10 and CD20 positive (**Table 8**).

ID	Age at Diagnosis (years)	Immunophenotype	Karyotype
C3	4.5	CD19$^+$, 10$^+$, 22$^+$	Dic(9;12)(p11-13;p11-13)
C3 relapse	12.5	CD19$^+$, 10$^+$, 20$^+$, 22$^+$	Dic(9;12)(p11-13;p11)
C6	1.3	CD19$^+$, 10$^+$, 22$^+$, 20$^+$	Del(9)(p13;p22)
C7	1.75	CD19$^+$, 10$^+$, 20$^+$	i(9)(q10)

Table 8: Diagnostic results of BCP-ALL samples from all three affected children, including the age at diagnosis, the immunophenotype and the karyotype. All leukemias display a B cell precursor phenotype with aberrations affecting chromosome 9. C3 suffered from a relapse 8 years after the initial diagnosis.

CD10 is expressed on B cell precursors in the bone marrow (BM), including pro-B and pre-B cells [142]. Therefore, the leukemias were classified as B cell precursor ALLs. Additionally to BCP-ALL in C3, C6 and C7, gastric cancers appeared in the broader family history, while B2 suffered from Diabetes Mellitus Type I and C3 additionally developed bronchiolitis obliterans. Although no common translocations (ETV6-RUNX1, MLL or BCR-ABL1) were detected in the BCP-ALLs of the affected children, all leukemic samples displayed an abnormal karyotype regarding the chromosomal region 9p. These aberrations manifested with the presence of dicentric chromosomes in the initial and relapsed karyotype of C3, 9p deletions in C6 and an iso-chromosome of 9p in C7 (**Table 8**). Abnormalities in this chromosomal area occur in around 35–60% of childhood ALL cases, making it the most common known molecular aberration [45, 143, 144]. Although inherited predispositions in pediatric ALL are rare [51], the high incidence of BCP-ALL in this family indicates a familial germline susceptibility to the disease.

5.2 Whole exome sequencing identifies *PAX5* c.547G>A

In order to elucidate a possible inherited susceptibility that caused the leukemia predisposition in the family, whole exome sequencing was performed for germline samples of generation B (B1, B2, B4, B5) and C (C1, C2, C3, C4, C5, C6, C7) as well as tumor samples from the affected children (C3L, C6L and C7L). We assumed an

autosomal dominant mendelian inheritance via the fathers, since the mothers were neither related to the fathers nor to each other nor originated from the same ethnic backgrounds. Since incomplete penetrance had to be taken into account, variants of interest were narrowed down to their presence in both fathers as well as the affected children. This analysis revealed 518 heterozygous single nucleotide variants (SNVs), which were shared in the germline probes of the affected children and fathers, but also by non-affected family members in generation C. Thereby, we identified a heterozygous mutation affecting the coding region of the Paired Box 5 (*PAX5*) gene at position 547, with a G>A substitution (Chromosome 9:37002702). Throughout B cell development, PAX5 is one of the main B cell specific transcription factors controlling B cell identity, commitment and function. Aberrations of the PAX5 gene are therefore often implicated in B cell tumorigenesis (reviewed in [96]). *PAX5* c.547G>A translates into an amino acid exchange from glycine to serine at position 183 of the PAX5 protein (p.G183S). Databank research validated the variant as extremely rare, since it could be detected neither in the 1000 genomes project [145], nor in the ExAC Browser, which harbors more than 60,000 sequenced exomes [146]. *PAX5* c.547G>A was predicted to be "deleterious" by the online available prediction tool SIFT (score=0.03; range 0 to 1; deleterious if score is <0.05) [133] and "probably damaging" by PolyPhen-2 (score=0.989; range 0 to 1; 1 is most confident for damaging). Nevertheless, due to their low specificity, these predictions tools should be used with caution [147]. The *PAX5* c.547G>A variant could further be validated by Sanger sequencing, which excludes the possibility of a false positive call from whole exome sequencing (**Figure 12**). In addition to both fathers and the affected children, two healthy children (C2 and C4) carried the mutation, indicating that certain environmental factors or additional secondary genomic alterations need to be present for disease development. Taken together, *PAX5* c.547G>A was transmitted from the grandparents to the fathers and subsequently to 5 out of 7 children, three of whom developed BCP-ALL. Hence, the

germline variant was passed on via autosomal dominant mendelian inheritance and the

phenotype developed with incomplete penetrance (**Figure 12**).
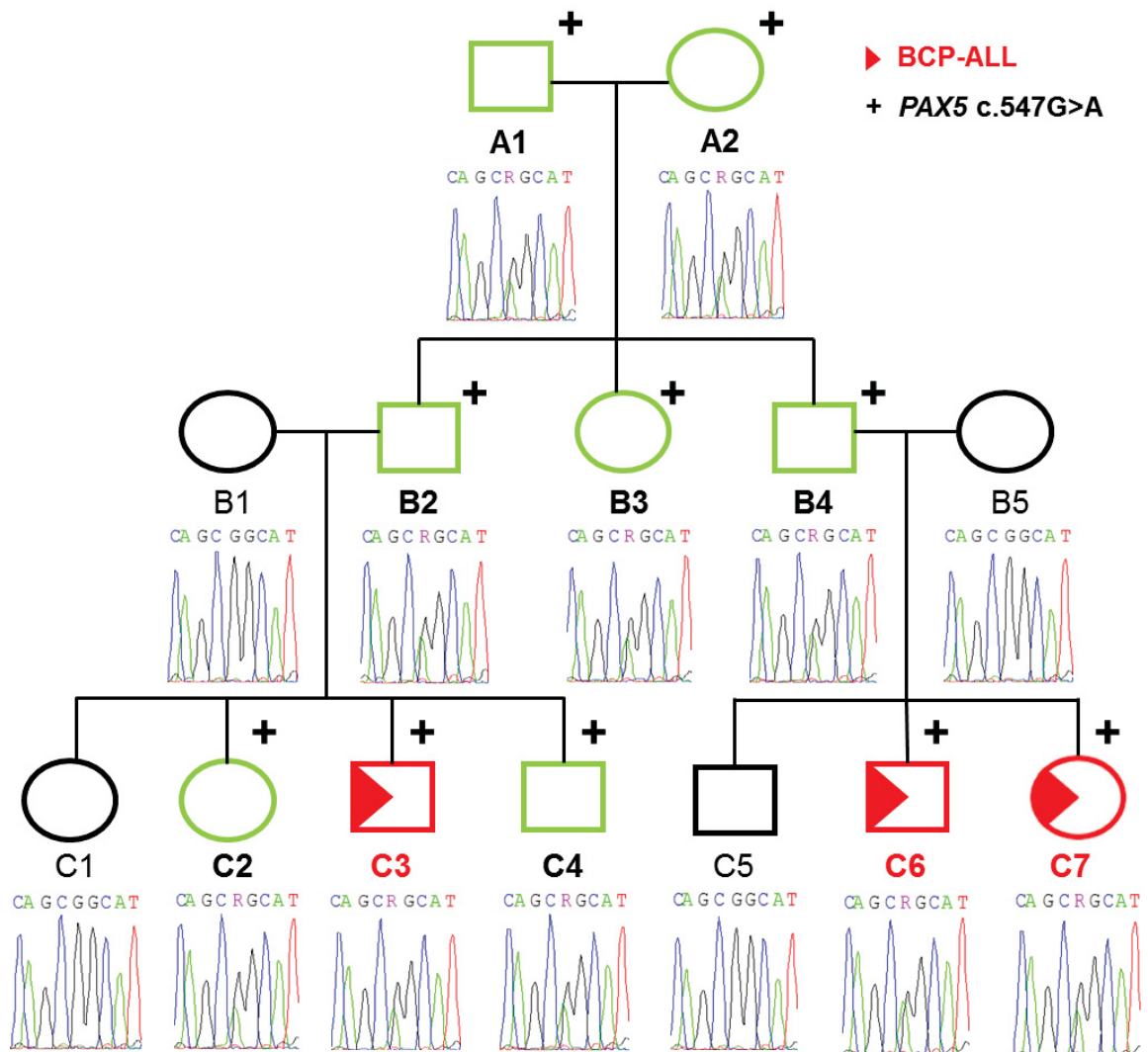

Figure 12: Sanger sequencing of the family validates a recurrent heterozygous *PAX5* c.547G>A mutation. "+" marks family members with *PAX5* c.547G>A positive carrier-status. The variant was transmitted from the grandparents (A1, A2) to the fathers (B2, B4) and aunt (B3) and subsequently to five (C2, C3, C4, C6, C7) out of seven children, three of whom developed BCP-ALL (marked in red). The PAX5 mutation displays incomplete penetrance, since 7 out of 10 family members were not phenotypically affected (marked in green).

5.2.1 Somatic makeup of BCP-ALLs

We next aimed to identify the disease-driving mechanism in the BCP-ALL cases. Therefore, we screened the whole exome sequencing data for tumor-specific somatic mutations. For this purpose, the bioinformatics tool MuTect was used, which is designed to calculate the probability of a tumor-specific point mutation by filtering germline SNVs from the tumor sequence data [135]. From 30548 to 59832 SNVs that were detected exome-wide in germline and tumor samples of C3, C6 and C7, only 11 to 20 variants were deemed tumor-specific by MuTect. 1 to 2 of these somatic calls per sample were indicated to be involved in cancer development by the COSMIC cancer gene census database (http://cancer.sanger.ac.uk/census/) [136], including mutations in NRAS, DDX5, JAK1 and MLL3 (**Figure 13**). Since both NRAS and JAK1 are involved in B cell survival and proliferation, the identified mutations were chosen for further validation.

A recurrent mutation in the *NRAS* gene was validated for C3L (p.G12D) and C6L (p.G12S). The respective RAS-RAF-MEK-ERK pathway mediates mitogenic and antiapoptotic responses. Somatic mutations that deregulate this signaling make up one of the most common genetic aberrations in childhood ALL [148]. Although no *NRAS* mutation was found in C7L, the tumor displayed a point mutation (p.V1009I) followed by a 3 bp inframe insertion, leading to the addition of an extra leucine at amino acid position 1010 (p.L1010LL), in the Janus kinase 1 (JAK1) (**Figure 13**). Mutations affecting the tyrosine kinase JAK1 are able to confer constitutive active JAK-STAT signaling and have been reported in 3% of B lineage ALL patients with poor prognosis. Moreover, the presence of JAK mutations in general was shown to be significantly associated with deletion of the CDKN2A/B locus [149].
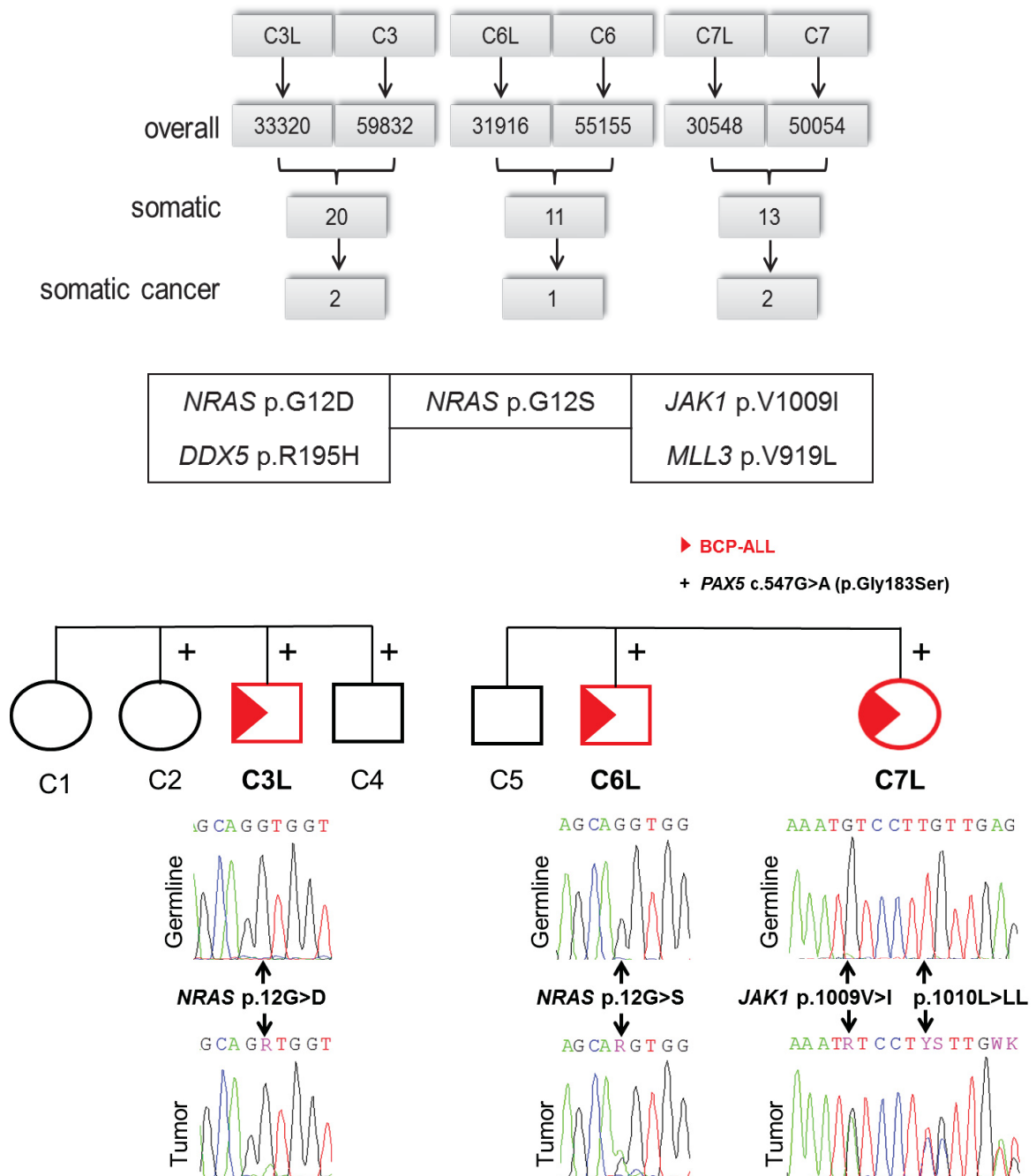
Figure 13: Whole exome sequencing identifies somatic mutations in cancer related genes. Overall between 31916 and 59832 mutations were detected exome wide by next generation sequencing. Of these, 11–20 were determined tumor-specific by MuTect analysis, with 1 or 2 per sample being potentially involved in cancer development according to the COSMICs cancer gene census database. The NRAS gene is recurrently mutated in C3L and C6L. The identified *NRAS* (p.G12D/S) and *JAK1* (p.V1009I and p.L1010LL) mutations could be further validated by Sanger sequencing.

CDKN2A/B is located on chromosome 9p, which had already been identified to display

multiple aberrations in all three children by karyotype analysis (**Table 8**). Therefore,

single nucleotide polymorphism (SNP) genotyping of leukemic and germline DNA from

C3, C6 and C7 was performed in addition to whole exome sequencing on an Illumina Human660W-Quad_v1 Bead Chip (Illumina), covering 590,000 common SNPs and 65,000 common copy number variation (CNV) probes. In agreement with the cytogenetic findings, the 9p region showed a homozygous loss affecting the tumor suppressor locus CDKN2A in all tumor samples (C3L Chr.9 19.995.895-22.228.756, C6L Chr.9 20.744.371–28.078.501 and C7L Chr. 9 21.825.996-22.008.026) (**Figure 14**). Moreover heterozygous deletions spanning almost the entire 9p region were observed in the leukemic samples (C3L, C6L and C7L), which resulted in loss of the wildtype (WT) *PAX5* allele. Subsequently, this loss of heterozygosity (LOH) caused retention of the mutated germline allele encoding *PAX5* c.547G>A.
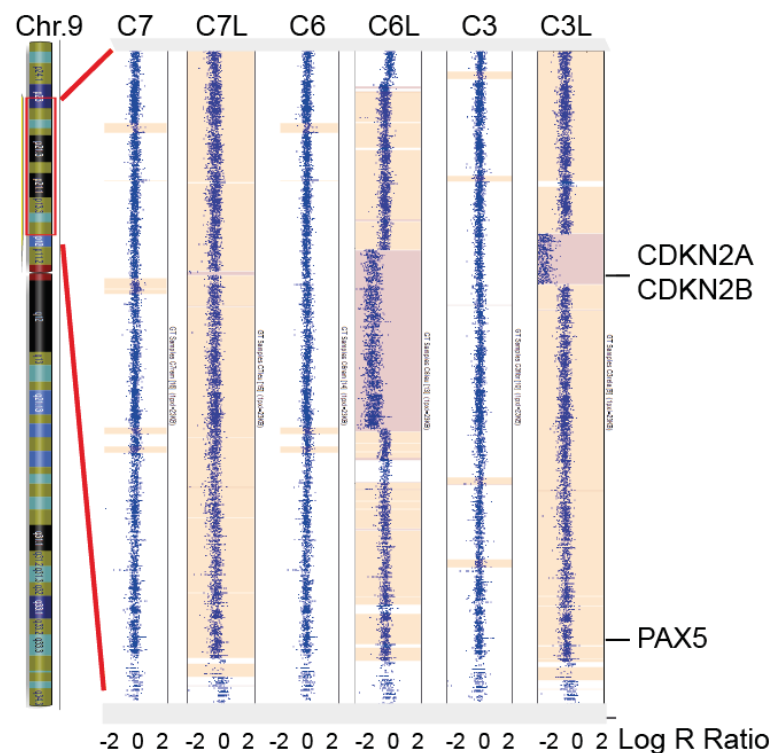
Figure 14: SNP genotyping reveals leukemia specific aberrations on chromosome 9. Indicated are structural variations on the chromosomal region 9p comparing the Log R ratio of leukemic (C3L, C6L and C7L) and germline (C3, C6 and C7) cells of the BCP-ALL affected children. All BCP-ALL samples display a homozygous loss affecting the CDKN2A/B locus (dark red coloring) in combination with a loss of heterozygosity for the *PAX5* gene (light orange coloring). Genotype calling and copy number variation (CNV) analysis was done with the Illumina GenomeStudio v2011.1.

5.2.2 *PAX5* variant c.547G>A confers reduced transcriptional activity

PAX5 somatic mutations commonly result in a marked reduction of PAX5's transcriptional activity [45]. Therefore the activity of the mutant protein p.G183S, which originates from the identified germline *PAX5* variant c.547G>A, was assessed using a PAX5-dependent reporter gene assay. The reporter contained copies of a CD19 promoter-derived high-affinity PAX5 binding site and allowed a luciferase based readout. The inactivating PAX5 p.V26G mutation served as a positive control. PAX5 V26G is already known from the literature to confer alterations in the protein structure, since it is part of the second beta-sheet and the valine to glycine substitution increases the protein backbone flexibility and affects the adjacent DNA contact residue p.F27 [45]. In the luciferase assay, p.G183S showed significantly reduced transcriptional activity compared to the WT PAX5 protein, which even exceeded the loss of function capacity of the known inactivating mutation p.V26G (Dunnett's test, p<0.0004) (**Figure 15**).
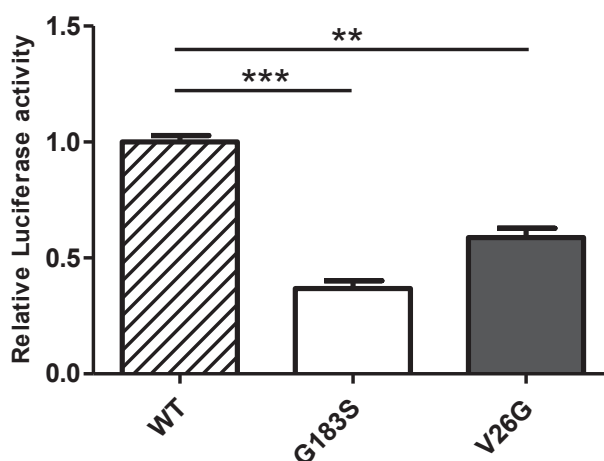
Figure 15: The PAX5^{mutant} protein G183S shows reduced transcriptional activity. Using a PAX5-dependent reporter gene assay, the relative luciferase activity was measured in Hek293T cells overexpressing wildtype (WT) PAX5, PAX5 p.G183S or PAX5 p.V26G, which served as control for reduced PAX5 activity. Bars show mean luciferase activity from three individual experiments with triplicate measurements. Background luciferase activity measured from the empty vector transfection was subtracted from all measurements and the WT was set to 1. Asterisks indicate significant differences as calculated by Dunnett's test (p<0.0004).

Moreover, both PAX5 mutations displayed no detectable differences in subcellular localization compared to the WT protein, as shown by western blot (**Figure 16**). These results indicate that PAX5 p.G183S is functional, but only with reduced activity.
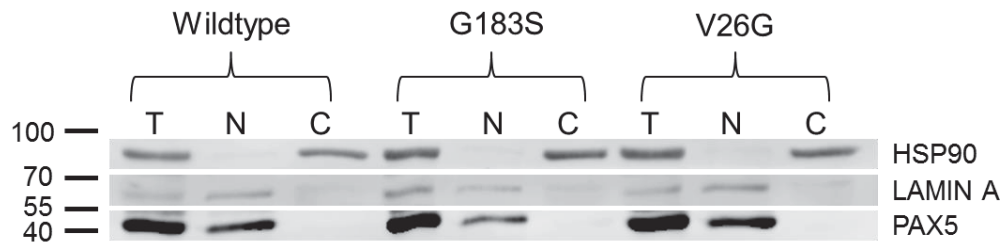

Figure 16: The Pax5 G183S mutant shows no differences in subcellular distribution compared to the wildtype (WT). Immunoblot displaying total (T), nuclear (N) as well as cytoplasmic (C) cell extracts from HEK293T cells either overexpressing WT or mutant (p.G183S and p.V26G) PAX5 proteins. HSP90 serves as the cytoplasmic control, while LAMIN A only detects nuclear-specific proteins. Likewise the WT PAX5 protein, both PAX5 mutations are only detected in the nuclear extracts (n=2).

5.2.3 *PAX5* c.547G>A prevalence

So far, *PAX5* c.547G>A was not officially described in a public online database. Therefore, we decided to further assess the frequency of the mutation using our own available cohorts. Since all three affected children showed aberrations of the chromosomal region 9p, which resulted in LOH of the WT allele, the co-occurrence of *PAX5* mutations with abnormalities on chromosome 9 seemed possible. Hence, we sequenced the gene region containing *PAX5* c.547G>A in a cohort of 23 patients with sporadic BCP-ALL, which showed aberrations on chromosome 9p in the tumors (**Supplementary Table 1**). In all patients, neither *PAX5* c.547G>A nor other *PAX5* mutations affecting this region could be validated. In order to assess the general frequency of *PAX5* c.547G>A and whether it is rather restricted to occurring in distinct families with an inherited susceptibility, we screened an additional cohort of 1100 healthy individuals for its appearance. The individuals were of mixed gender and ethnic background, and aged 18 to 55. For high throughput mutation screening, sequence-specific primer PCR (SSP-PCR) was used. This technique utilizes unique primers that harbor the mutation site at their 3' end together with high annealing temperatures, making the PCR reaction highly specific for the correct product. By applying SSP-PCR,

1100 samples could be tested in parallel and their mutation status visualized on an agarose gel (**Figure 17**).
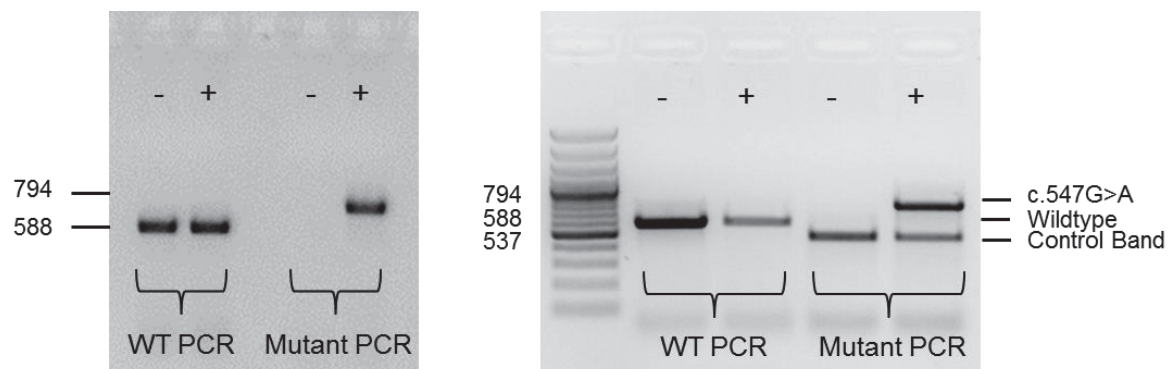
Figure 17: Sequence specific primer (SSP) PCR for the detection of heterozygous *PAX5* c.547G>A. Two PCR reactions are performed that specifically amplify either the wildtype (WT) or the mutant *PAX5* sequence. Since *PAX5* c.547G>A was only found heterozygous in patient samples the remaining PAX5 WT-allele is detected by the WT PCR in an affected individual (+), while samples without *PAX5* c.547G>A (-) will show no positive signal in the Mutant PCR (left picture). In order to always display a band to ensure DNA integrity, a control primer was added to the Mutant PCR (right picture).

The complete cohort was tested negative for *PAX5* c.547G>A, which highlights the rarity of the mutation and lowers the chances of a sporadic appearance without a familial predisposition, as it occurred in our family. This hypothesis was further supported by a Nature publication, which was published during the preparation of this work, presenting two additional families harboring *PAX5* c.547G>A [150]. The data presented by Shah *et al.*, are in agreement with our findings in regard to the leukemia phenotype, the variable penetrance and the LOH for the *PAX5* WT allele in combination with CDKN2A loss on chromosome 9p in the tumors. Altogether, we showed that, although rare, the heterozygous germline *PAX5* c.547G>A variant confers reduced PAX5 transcriptional activity, which can lead to substantial decrease of PAX5 expression in tumors with additional LOH of the PAX5 WT allele though aberrations on chromosome 9.

5.3 *Pax5*⁺/⁻ mice develop BCP-ALL when exposed to infection

For almost a century, infections have been proposed as the most likely cause of childhood leukemia [151]. The leukemia phenotype analyzed in the family presented

here displayed incomplete penetrance for *PAX5* c.547G>A, with a majority of family members being mutation carriers without any phenotypic signs of illness (**Figure 12**). Infection as an environmental factor influencing BCP-ALL development could therefore have an effect on disease penetrance. In order to elucidate a potential role of infection exposure and *PAX5* susceptibility, we used *Pax5* heterozygous (*Pax5*$^{+/-}$) mice, which harbor only one functioning *Pax5* allele, as a readout model. Although Pax5 is required for normal B cell development, *Pax5*$^{+/-}$ mice housed in a specific pathogen-free (SPF) facility are known to display the same growth, viability and behavior as their WT littermates and never spontaneously develop B-ALL [94, 152]. However, the SPF environment is sterile and monitored for the absence of common viruses, bacteria and parasites. Therefore, we tested whether exposure to common infection can provoke BCP-ALL development in heterozygous *Pax5* mice, when they are exposed to an environment with common pathogens, after being born in sterile surroundings. This experimental design was based on the "delayed infection hypothesis" proposed by Mel Greaves, who postulated that infection as the second hit can promote leukemogenesis in the scenario of an underdeveloped immune system [70].

For the experiments, two mouse cohorts were monitored. Both groups were composed of *Pax5*$^{+/-}$ mice as well as littermate WT controls and were observed for an experimental period of two years. Cohort 1 was born and housed in a SPF facility, while cohort 2 was born in the SPF facility but moved to a conventional facility between two and five months after birth, where the mice encountered a common infectious environment. The microbiological status of the mice in both facilities was tested and monitored for the entire experimental duration (**Supplementary Table 2**). In the conventional animal facility the mice were exposed to a variety of pathogens including mouse hepatitis virus (MHV), murine norovirus (MNV), helicobacter species and trichomonas muris, which indicates a significant immune response in the second cohort. In line with the infection hypothesis, BCP-ALL development was observed in

22% (9 out of 41) $Pax5^{+/-}$ mice from the pathogen exposed cohort 2, while the control WT mice from the same cohort and all mice from cohort 1, which were housed in the SPF environment, stayed healthy (**Figure 18**).
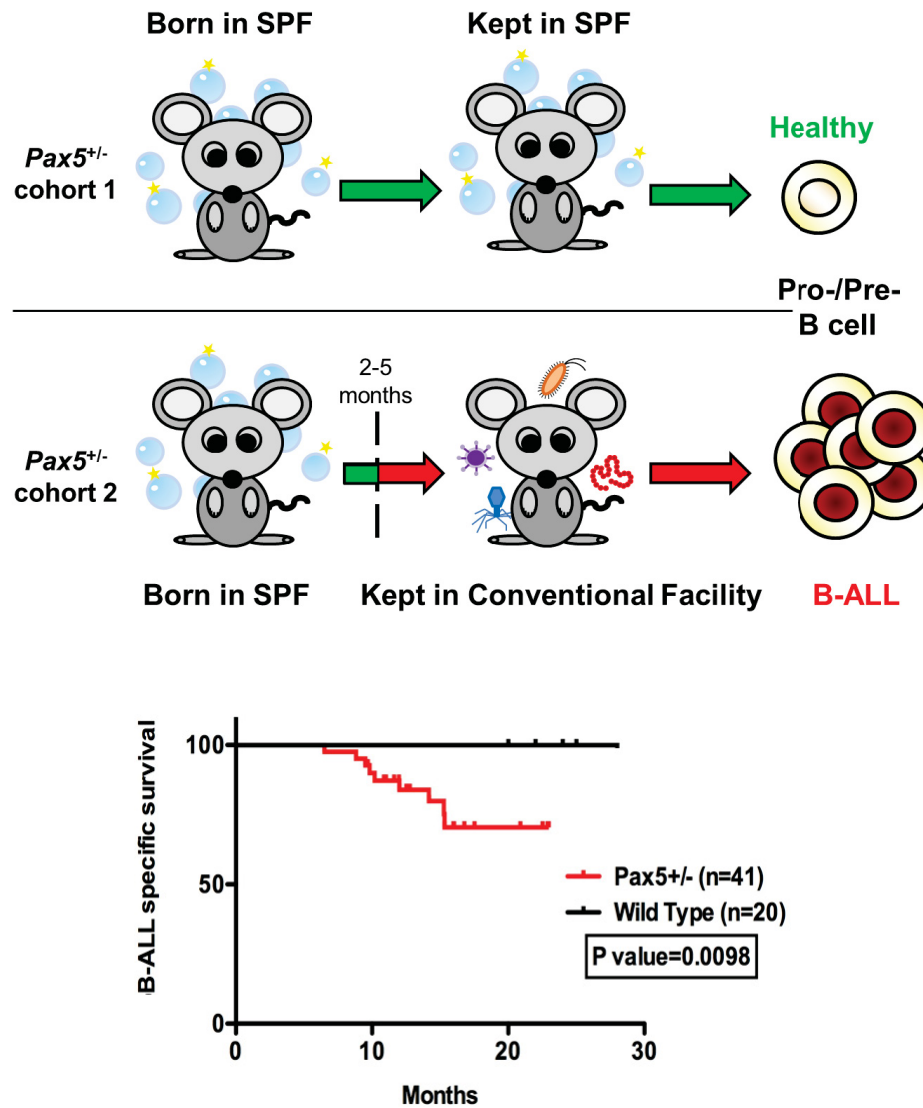
Figure 18: BCP-ALL development in infection exposed $Pax5^{+/-}$ mice. $Pax5^{+/-}$ cohort 1 was born and kept in a specific pathogen-free (SPF) facility, while mice from $Pax5^{+/-}$ cohort 2 were born in a SPF facility and moved to a conventional facility after 2-5 months, where they were exposed to common pathogens, leading to BCP-ALL development in 22% of the mice. $Pax5^{+/-}$ mice housed in a conventional facility show a significant shortened life span (log-rank Test; p=0.0098) compared to respective wildtype mice. Mouse experiments were carried out in cooperation with Dr. Isidro Sánchez-García's laboratory (IBMCC).

The leukemias showed a precursor B cell surface phenotype of $CD19^{+/-}B220^{+}IgM^{-}cKit^{+/-}CD25^{+/-}$, which manifested with splenomegaly and disruption of splenic and lymph node architecture (**Figure 19**).
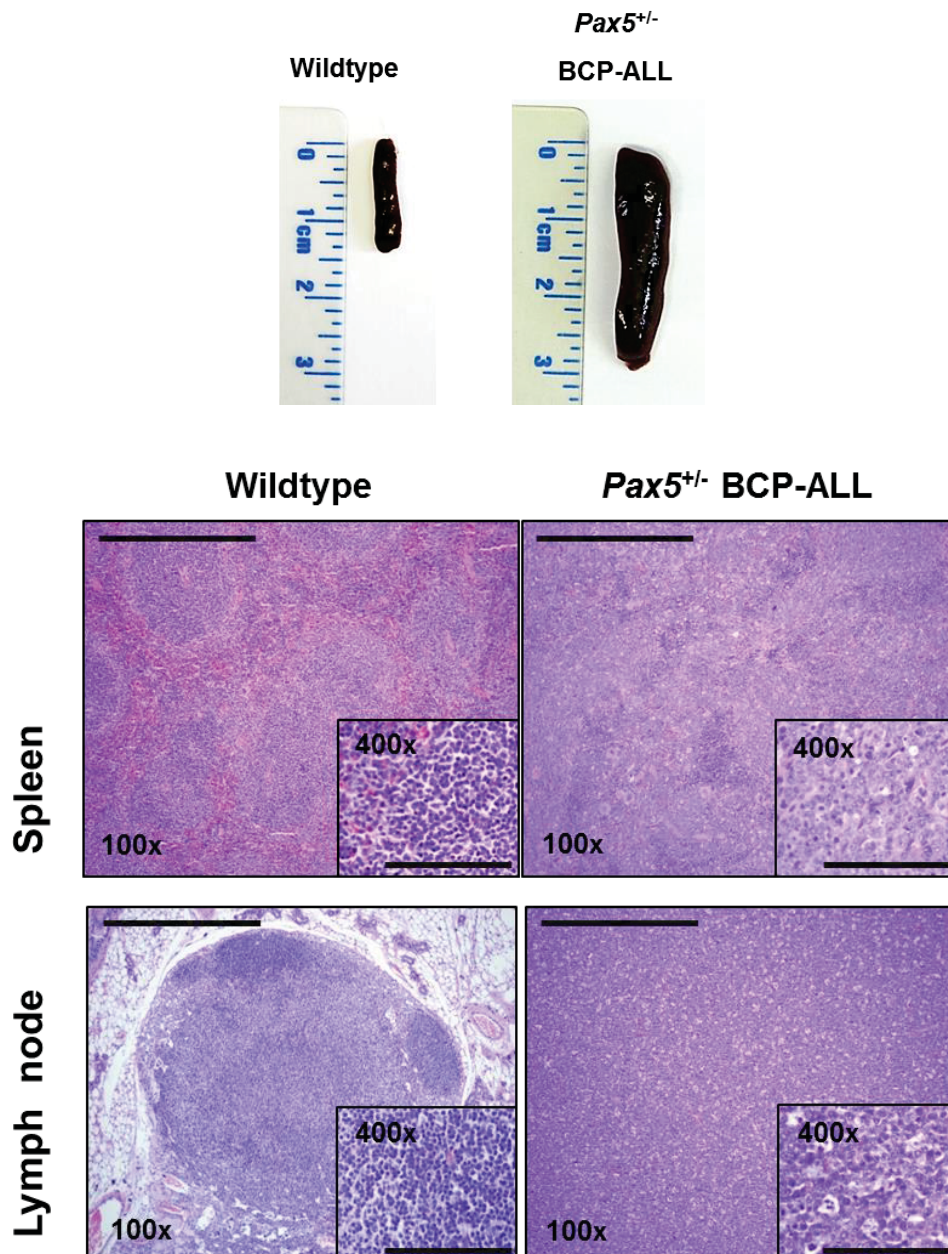
Figure 19: Representative blast infiltration of secondary lymphoid organs in *Pax5*^+/- BCP-ALLs. Top: Picture of splenomegaly that was observed in 55% (5 out of 9) of *Pax5*^+/- BCP-ALLs. Wildtype (WT) mouse spleen serves as reference for a healthy spleen size. Bottom: Haematoxylin and Eosin staining of WT mice and *Pax5*^+/- BCP-ALLs. Staining depicts infiltrating blast cells in spleen and lymph nodes, showing loss of structural architecture resulting from effacement with cells morphologically resembling lymphoblasts in the leukemias. Scale bar represent 500 µm for 100x enlargement and 100 µm for 400x enlarged images. Experiments were performed in cooperation with Dr. Isidro Sánchez García's group (IBMCC).

Moreover, we observed dissemination of blast cells in the BM, peripheral blood (PB), spleen and lymph nodes, as well as the infiltration of non-lymphoid tissues like liver, lung and kidney [153]. All BCP-ALLs displayed clonal blast populations, since only

clonal immature BCR rearrangement were detected in the tumor samples by V(D)J clonality PCR assay (**Figure 20**).
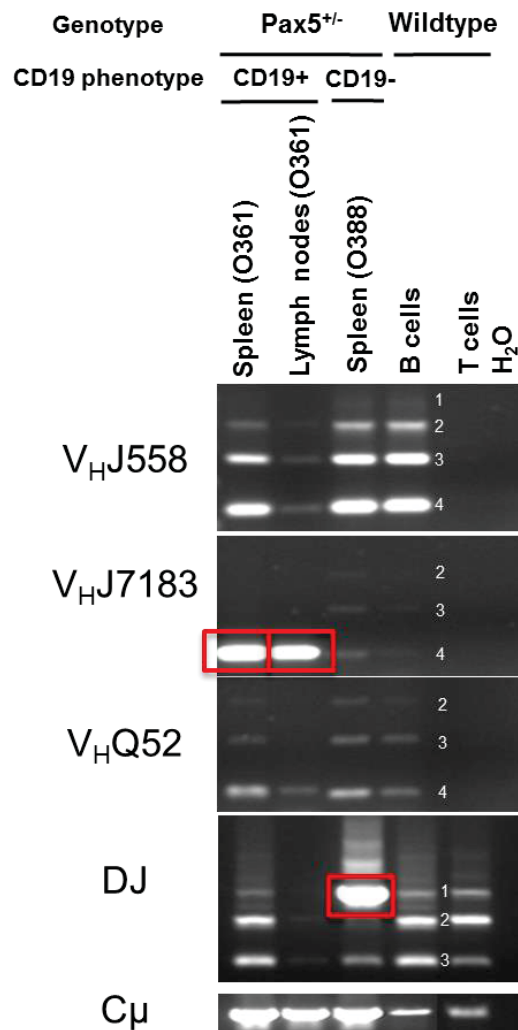

Figure 20: Blast cells of *Pax5*[+/-] mice are clonal. PCR analysis of immunoglobulin heavy-chain gene rearrangements in infiltrated spleens and lymph nodes of BCP-ALLs from *Pax5* heterozygous mice. T cells served as a negative control while sorted CD19[+] B cells from the spleens of healthy mice were included as positive control for polyclonal rearrangements within the mature B cell population (indicated by numbers, 1-4). Infiltrated tissues show increased clonality within their immunoglobulin repertoire (red squares). The experimental data was verified in 2 independent experiments. Analysis was performed by Isidro Sánchez García's group (IBMCC).

Taken together, murine BCP-ALLs closely resemble the leukemia phenotype observed in humans. Moreover, the low frequency of disease incidence of BCP-ALL development in infection-exposed *Pax5* heterozygous mice reproduces the low penetrance of BCP-ALL development in the family with the heterozygous *PAX5* c.547G>A mutation.

5.3.1   Additional alterations in the *Pax5* WT allele reduce Pax5 activity

The disease phenotype of *Pax5*$^{+/-}$ BCP-ALLs (B220$^+$, IgM$^-$, CD19$^{+/-}$) revealed a potential

loss of Pax5 transcriptional activity from the remaining *Pax5* WT allele, since some

leukemias displayed absence of CD19 expression, which is one of the most important

downstream targets of Pax5 (**Figure 21**).
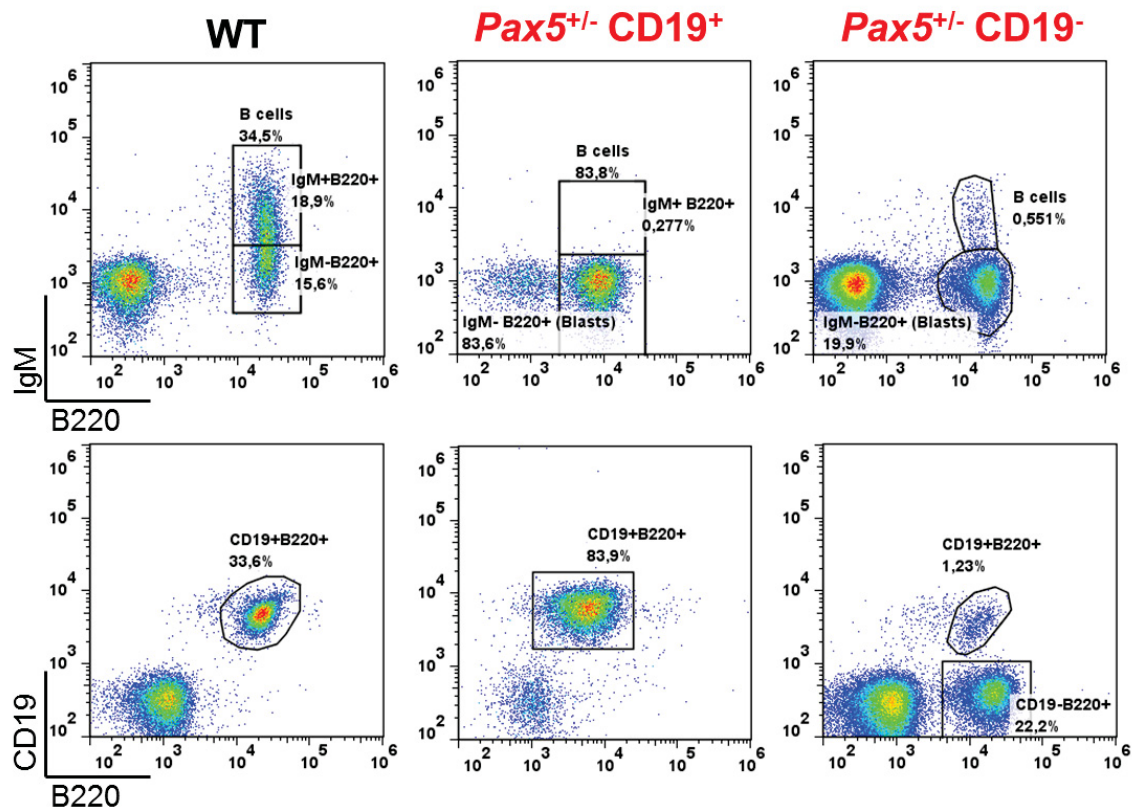
Figure 21: Flow cytometric analysis of hematopoietic subsets in *Pax5*$^{+/-}$ mice with BCP-ALL.
Representative plots of cell subsets from the peripheral blood are shown. These display accumulation of
IgM negative, B220 positive blast cells of the B cell precursor lineage in diseased *Pax5*$^{+/-}$ mice (n=9)
compared to wildtype (WT) mice (n=4). The cell surface phenotype of the blast cells was either CD19
positive or negative.

Therefore, whole exome sequencing was performed with 7 out of 9 murine BCP-ALLs,

with 2 samples being excluded due to blast counts below 50%. Sequencing analysis of

the *Pax5* gene revealed somatic Pax5 point mutations (p.P80R and p.P80L) in two

mice (S665 and S748), which could be positively validated by Sanger sequencing
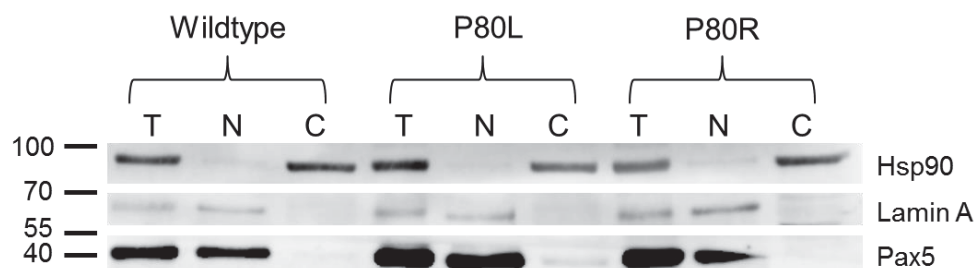
(**Figure 22**).
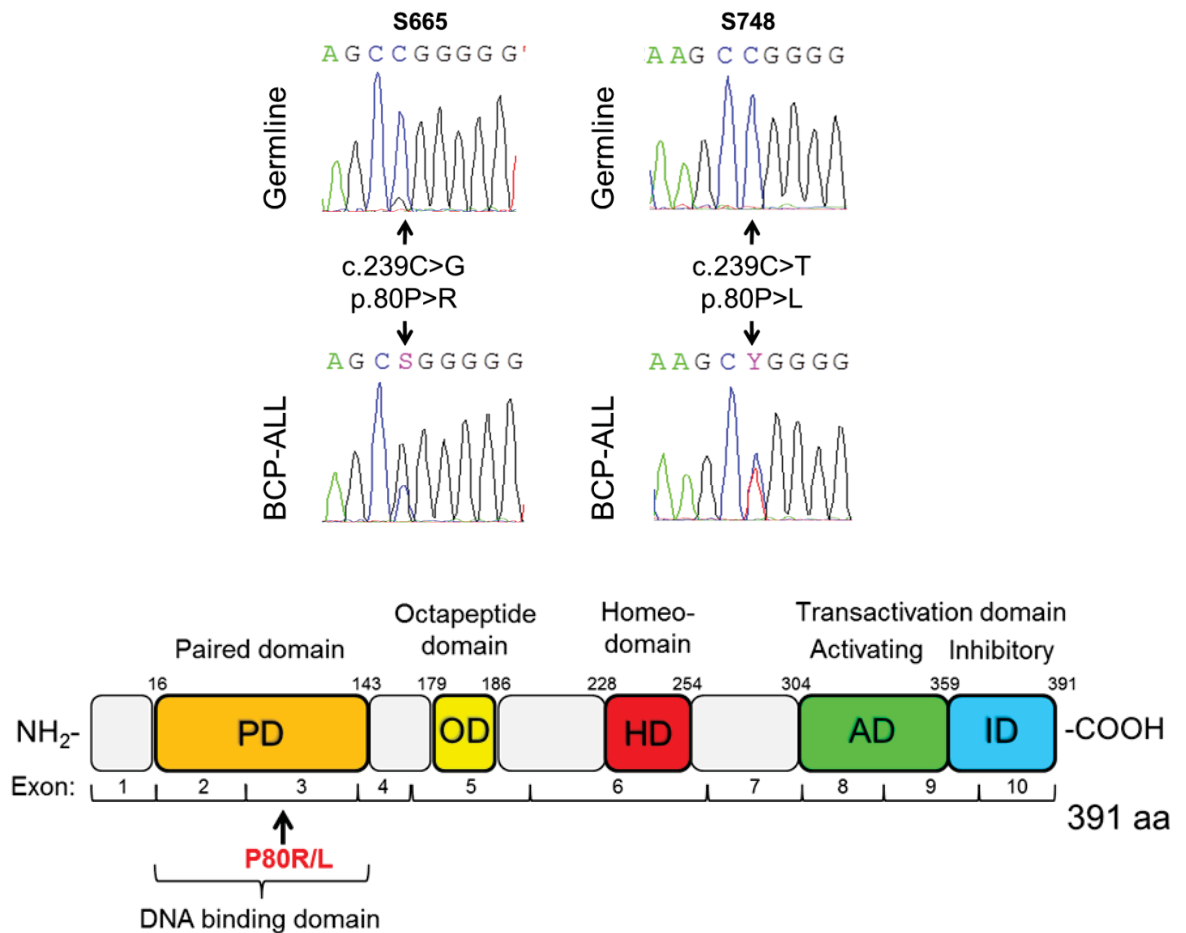
Figure 22: Pax5 variants p.P80R and p.P80L. Top: Sanger sequencing validates two somatic Pax5 variants p.P80R and p.P80L in BCP-ALLs from the heterozygous *Pax5* animals S665 and S748 respectively. Bottom: schematic of the *Pax5* gene domains, displaying the location of the single nucleotide variants Pax5 p.P80R and p.P80L. Both variants are located in the gene's DNA binding domain.

The variants were found as heterozygous mutations, which can be explained by the construction of the *Pax5*^+/- background. In order to generate *Pax5* heterozygous mice, exon 2 of the *Pax5* gene was replaced by *Escherichia colia* lacZ and neomycin resistance genes through homologous recombination [94]. Since exon 2 encodes the n-terminal half of the paired domain, which is indispensable for DNA binding [154], this targeted approach inactivated the respective *Pax5* allele. Nevertheless, Pax5 p.P80R and p.P80L, which are located in exon 3 on the c-terminal half of the DNA binding domain (**Figure 22**), are still present as WT in the inactivated allele, therefore, displaying as heterozygous mutations in the sequencing. Crystal structure modelling of *PAX5* p.P80R proposed structural consequences through the amino acid exchange,

since the proline contacts the phosphate backbone of the DNA, which might be disrupted by the arginine replacement [45]. To assess whether the mutations influence the transcriptional activation mediated by Pax5, the Pax5-dependent reporter gene assay, which was already employed for the human PAX5 p.G183S variant, was used. Again, both mutations resulted in a marked reduction of Pax5 transcriptional activity (Dunnett's test, P<0.0001) compared to the WT Pax5 protein, while their subcellular localization remained unaffected (**Figure 23 and Figure 24**).
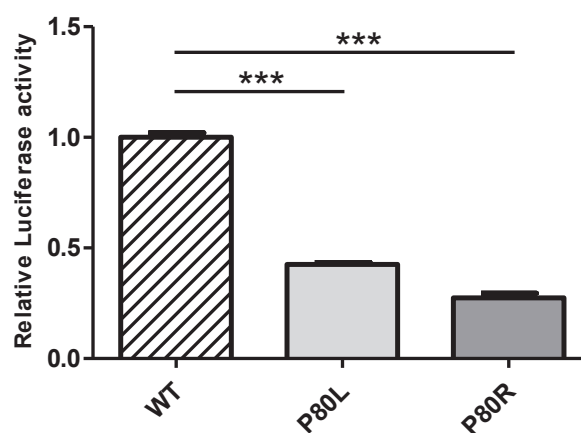

Figure 23: Attenuated transcriptional activity of Pax5 mutant proteins p.P80L and p.P80R. Relative luciferase activity is indicated for Pax5.WT as well as Pax5.mutant proteins, as measured by a Pax5-dependent reporter gene assay in Hek293T cells, which were stable transfected with the different Pax5 overexpressing constructs. Bars show mean luciferase activity from three individual experiments with triplicate measurements. Luciferase activity measured from the empty vector transfection was subtracted from all measurements and the wildtype (WT) was set to 1. Asterisks indicate significant differences as calculated by Dunnett's test (p<0.0001).

Figure 24: Pax5 p.P80R and p.P80L show the same nuclear distribution like the Pax5 wildtype protein. Hek293T cells were transfected with pMC3-Pax5.WT or pMC3-Pax5.mutant and hygromycin-resistant cells were selected. Cell lysates of total protein (T) nuclear (N) and cytosolic (C) fractions were subjected to immunoblot analysis for Pax5, Hsp90 as cellular control, as well as LaminA as nuclear control (n=2).

In the murine tumors, p.P80R conferred absence of CD19 expression, while the leukemic sample harboring p.P80L still showed CD19 on the cellular surface, suggesting that the mutation was not able to completely abolish Pax5 signaling downstream. This observation is in line with the reporter gene assay results, which detected significantly less Pax5 transcriptional activity for p.P80R when compared to p.P80L (student's t-test; $p < 0.0231$).

To expand the screen to include potential deletions in the *Pax5* gene, we further performed cDNA deletion PCR analysis. Therefore, the whole *Pax5* gene transcript was amplified from murine BCP-ALL samples using primers localized in the gene's 5 and 3 prime untranslated region. The PCR products were visualized on an agarose gel and the identity of the respective bands was determined by Sanger sequencing. The PCR was optimized to detect Pax5 isoforms a (1175 bp) and b (1011 bp), while the smaller isoforms Pax5d (735 bp) and Pax5e (570 bp) [155] were not preferentially amplified. This analysis revealed the presence of truncated Pax5a and b transcripts in mouse S767 (**Figure 25**). Sanger sequencing identified deletion of Ex7 and 8 in the respective isoforms, affecting part of the gene's activating transactivation domain (**Figure 22**). Unfortunately, breakpoint detection on a genomic level was not possible due to large intronic regions and limited sample amounts. Phenotypically, the blast cells of mouse S767 were classified as CD19 positive by FACS analysis, which points towards a mild effect of the alteration. The observation that 4 out of 5 BCP-ALLs with absence of CD19 expression did not show mutations or deletions in the *Pax5* gene (**Table 9**), suggests that other mechanisms, which influence Pax5 activity, took place in these leukemias. Nevertheless, the data reproduce the observation made in the human family, which showed LOH of the WT Pax5 allele in combination with a loss of function mutation in the remaining Pax5 allele in the tumors.
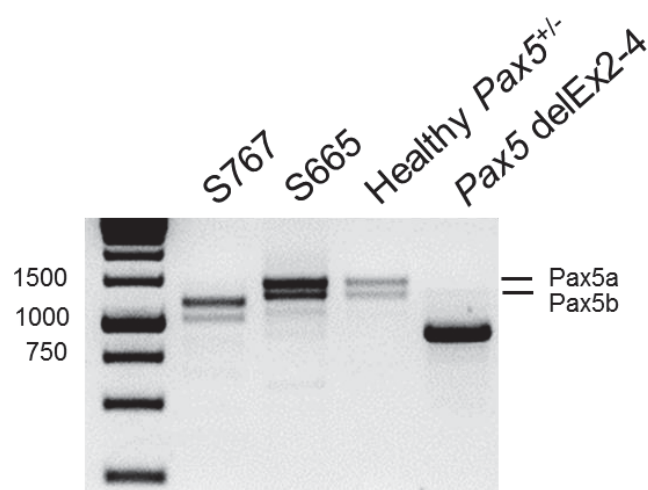
Figure 25: CDNA deletion PCR identifies aberrant Pax5 transcript in mouse S767. The complete Pax5 coding region of bone marrow tumor cDNA from $Pax5^{+/-}$ mice was amplified via PCR and visualized on an agarose gel. Mouse S767 displays truncated Pax5 transcripts, while mouse S665 (harboring Pax5 p.P80R) shows distinct bands for Pax5 isoforms a and b. RNA from a healthy $Pax5$ heterozygous mouse served as normal control. Tumor cDNA with loss of Ex2-4 of the Pax5 gene served as control for aberrant Pax5 expression.

Mouse ID	CD19 phenotype	Pax5 alteration
O361	CD19 pos	-
O388	CD19 neg	-
S665	CD19 neg	p.P80R
O332	CD19 neg	-
W495	CD19 neg	-
S748	CD19 pos	p.P80L
W893	CD19 neg	-
S767	CD19 pos	Pax5 del Ex7-8
W634	CD19 pos	-

Table 9: CD19 phenotype with identified Pax5 alterations for all murine BCP-ALLs analyzed. 4 out of 9 tumors are CD19 positive with the remaining 5 displaying no CD19 expression. Pax5 alterations could only be identified in one CD19 negative leukemia (S665), while 2 of the BCP-ALLs expressing CD19 harbor mutations/deletions in the Pax5 gene (S748, S767).

In the human scenario, this LOH for Pax5 was accompanied by deletions in the tumor suppressor locus *CDKN2A*. To further validate the mouse model in this regard, we screened the tumor whole exome data for possible CNVs in the *CDKN2A* region. Likewise the somatic makeup of BCP-ALLs in the family, deletions affecting the *Cdkn2A* locus were identified but only in 1 out of 9 murine leukemias (**Figure 26**).
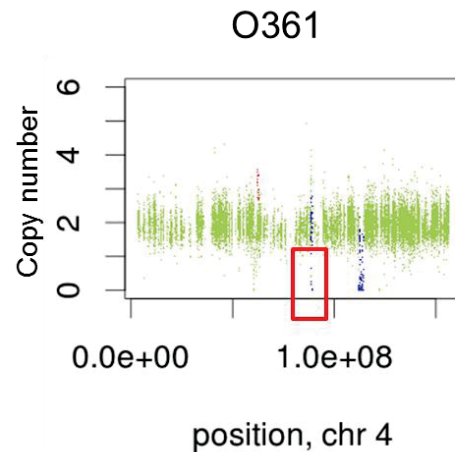
Figure 26: Loss of *Cdkn2a* locus in mouse O361. Copy number variation (CNV) profiles were calculated from tumor/reference pairs using Control-FREEC1. Copy number profiles displayed a loss of heterozygosity at 4:89274500-89276999 in mouse O361 (red square), correlating with the absence of p19Arf expression in this animal.

The low incidence of *Cdkn2A* aberrations in the mouse might be explained by different chromosomal distances of the *CDKN2A* locus and the *PAX5* gene between humans and mice. In humans, both *PAX5* and *CDKN2A* are located in close proximity to each other being merely 15 kb apart on chromosome 9, while they are separated by around 45 kb on chromosome 4 in mice. Together, these data indicate a high resemblance between the BCP-ALLs developed in the family harboring an inherited *PAX5* variant and the heterozygous *Pax5* mouse model.

5.3.2 *Pax5* heterozygosity creates an aberrant B cell compartment in the BM

Next, we aimed to explain the mechanism behind the specific BCP-ALL susceptibility, which is conferred by *Pax5* heterozygosity under infection exposure. Therefore, the different B cell developmental stages were analyzed in healthy *Pax5*$^{+/-}$ as well as WT littermates of the same breeding, which were housed in the conventional facility. In the PB, total B cell numbers (B220$^+$IgM$^{+/-}$) were significantly reduced in *Pax5*$^{+/-}$ mice at 3, 6 and 9 months of age, when compared to WT littermates. On the other hand, we observed significantly higher amounts of pro-/pre-B and immature B cells in the BM of *Pax5* heterozygous mice as compared to age matched WT mice (**Figure 27**).
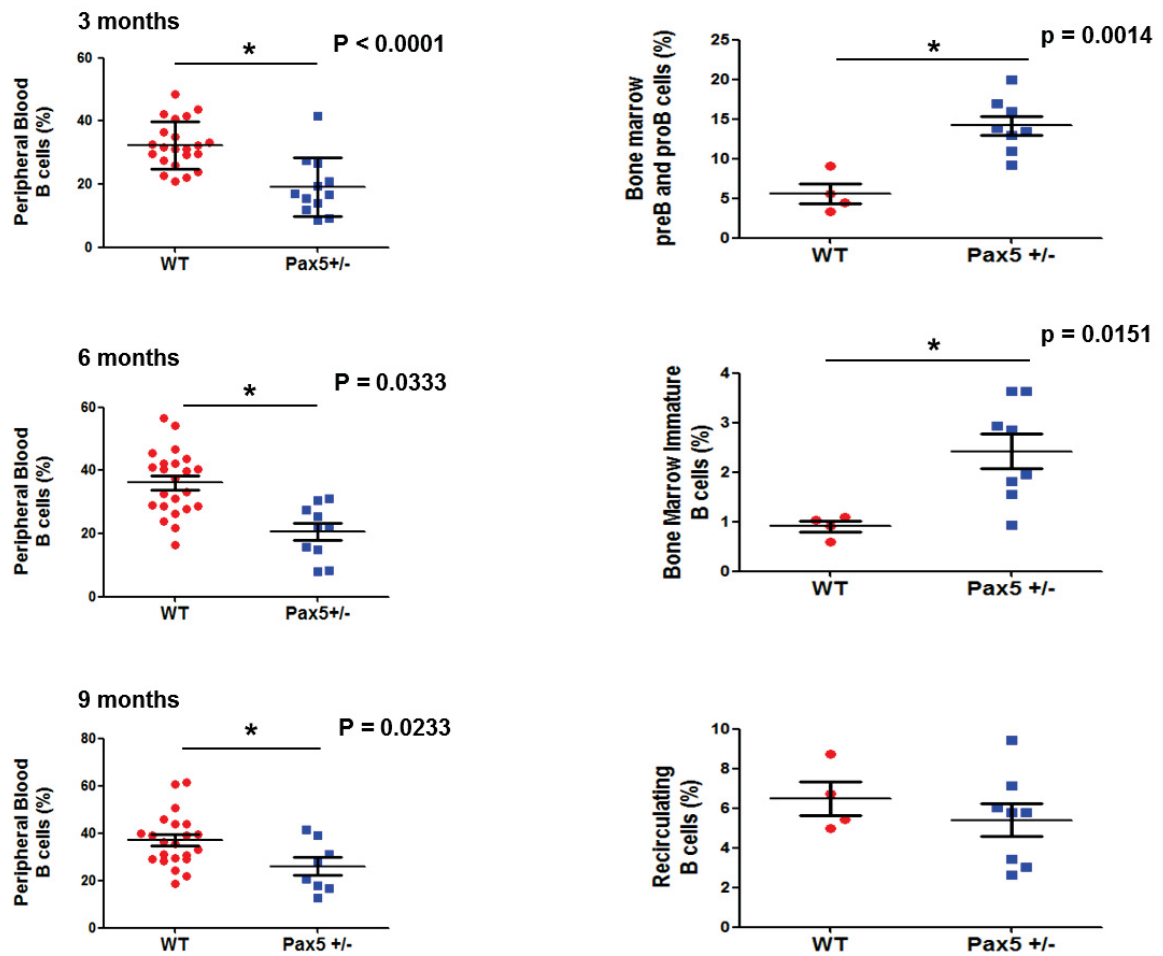
Figure 27: Healthy *Pax5*^+/- mice display a B cell decrease in the peripheral blood (PB) and an increase in precursor B cell populations in the bone marrow (BM). Left: Percentage of PB B cells (B220$^+$IgM$^{+/-}$) at different time points in *Pax5*^+/- mice (n=12) compared to wildtype (WT) mice (n=22) as analyzed by flow cytometry. A significant decrease in PB B cells can be observed in the mice aged 3, 6 and 9 months. Error bars represent the standard deviation. Indicated p-values were calculated by student's t-test. Right: 4 month old heterozygous *Pax5* mice (n=8) show increased numbers of pro-/pre-B cells (B220^{low}IgM$^-$) as well as immature B cells (B220^{low}IgM$^+$) in the BM compared to age-matched WT mice (n=4). No differences were observed in recirculating B cells (B220^{++}IgM$^+$). Unpaired t-test p-values are indicated. FACS analysis was performed in cooperation with Dr. Isidro Sánchez García's group in Salamanca (Spain).

These data indicate that *Pax5*-related susceptibility to BCP-ALL is associated with a decrease of PB B cells *in vivo* due to accumulation of B cell precursors in the BM. Therefore *Pax5* heterozygosity in the mouse accounts for the appearance of an aberrant B cell precursor compartment in the BM, which is susceptible to malignant transformation by disease-driving somatic mutations.

5.4   Somatic *Jak3* variants drive tumor progression in BCP-ALLs of *Pax5* heterozygous mice

5.4.1   Identification of recurrent somatic *Jak3* variants

Since loss of Pax5 expression as a single event is not sufficient to produce a disease phenotype [45] and no recurrent CDKN2A loss could be validated in the murine leukemias, BCP-ALLs of *Pax5* heterozygous mice were further screened for second hits, driving leukemogenesis. Whole exome sequencing analysis identified between 25573 and 32724 SNVs in tumor samples and between 27141 and 33402 SNVs in germline samples. 8 to 173 variants were determined to be tumor specific by MuTect analysis with recurrent somatic cancer related mutations detected in the Janus kinase 3 (*Jak3*) gene (p.R653H/C and p.V670A) in 6 out of 7 murine tumors (**Figure 28**).
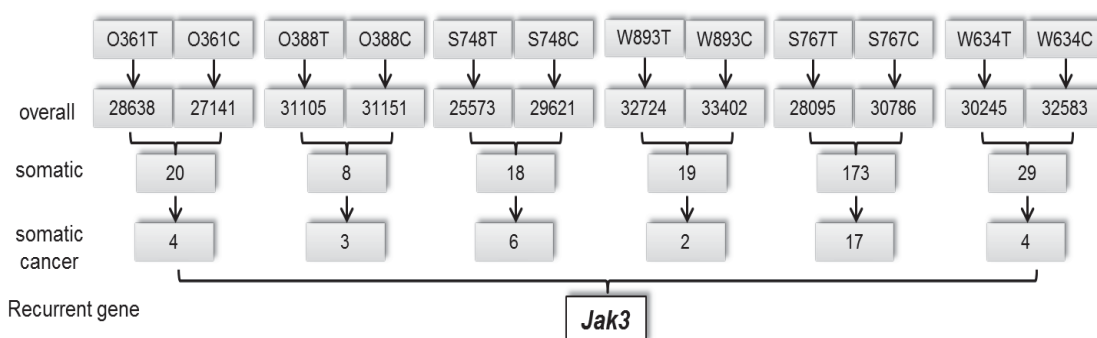
Figure 28: *Pax5*^+/- BCP-ALLs harbor recurrent Janus kinase 3 (*Jak3*) variants. Whole exome sequencing analysis yielded around 30000 SNVs exome-wide. 2-17 were deemed tumor-specific by MUTECT analysis with recurrent variants in the gene *Jak3*, which were subject for further studies.

All variants displayed as heterozygous non-synonymous somatic SNVs, which are located in the pseudokinase domaine (JH2) of the *Jak3* gene and are conserved among a variety of species (**Figure 29**). The corresponding human homologues (JAK3 p.R657Q and p.V674A) are recorded in the cancer gene census data base. In humans, p.R657Q and p.V674A are already known to increase accessibility of the catalytic loop from 3D structure modelling [156] and to confer constitutive kinase activity of JAK3 from *in vitro* assays [157].
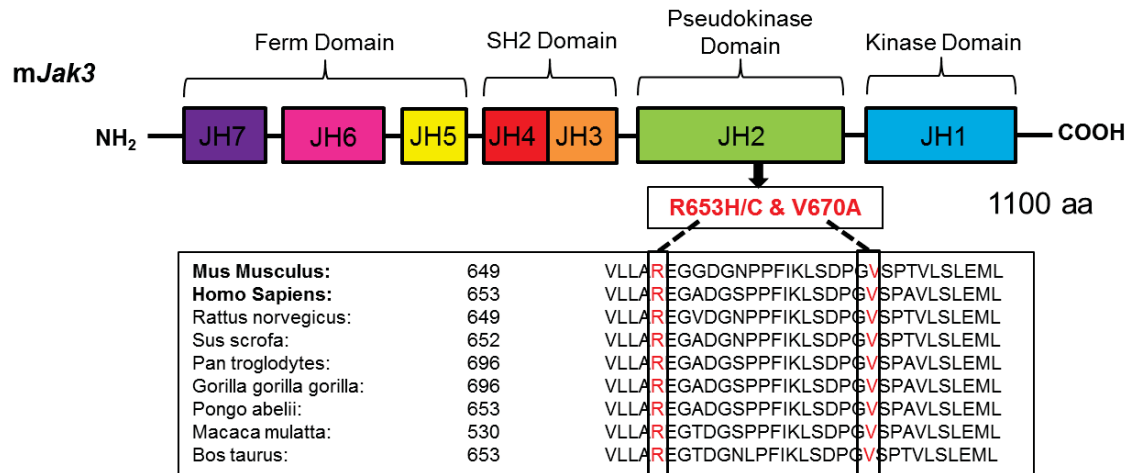
Figure 29: Domain organization of murine *Jak3*. Corresponding mutations p.R653H/C and p.V670A in BCP-ALLs of *Pax5*+/- mice are highlighted. Additionally, homology analysis among 9 species for the depicted region in the JH2 domain of *Jak3* is shown. Boxes indicate the conserved amino acid residues affected by the identified variants.

Additionally, 2 out of 7 mice displayed heterozygous mutations in the *Jak1* gene (p.S645P and p.V657F). All mutations could be positively validated by Sanger sequencing (**Figure 30**). The high frequency of recurrent *Jak* mutations and their appearance in highly conserved regions across species indicate that the variants might be relevant for the BCP-ALL development in infection exposed *Pax5* heterozygous mice.
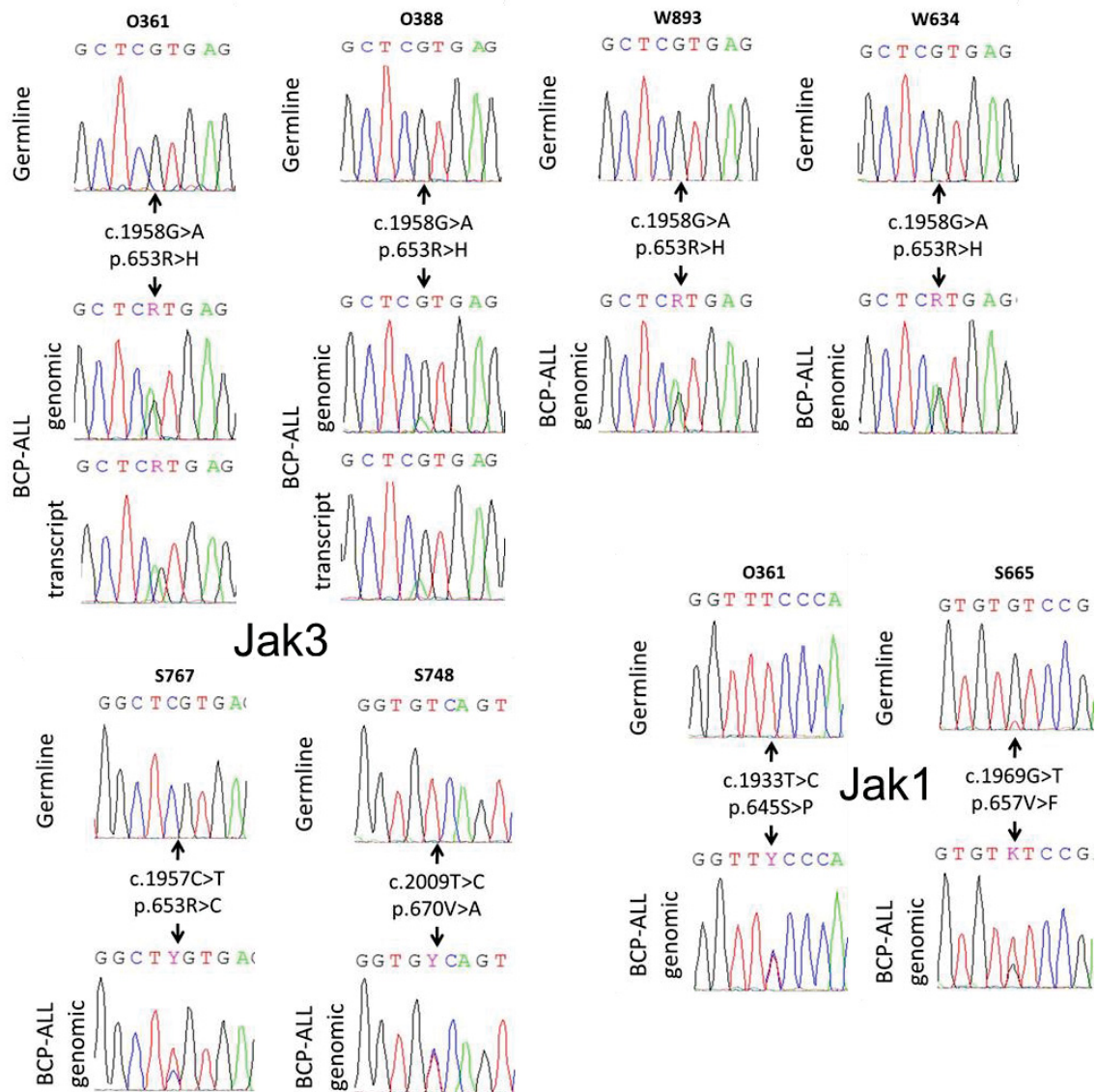
Figure 30: Validation of *Jak3* and *Jak1* mutations by Sanger Sequencing. Displayed are sequencing results of germline versus tumor DNA on genomic and for mice O361 and O388 additional on transcript level. *Jak3* p.R653H was found in 4 out of 9 mice, while p.R653C and p.V670A were only found once in mouse S767 and 748, respectively. Additionally, mouse O361 harbors the *Jak1* mutation p.S645P and mouse S665 the *Jak1* mutation p.V657F.

5.4.2 First appearance and frequency of *Jak3* variants

To monitor the frequency of the identified Jak3 mutations in the bulk of leukemic blast cells, we performed deep sequencing. This next-generation-sequencing-based technique allows the sequencing of a PCR-amplified short region containing the mutation site, achieving a read depth between 600,000 and 2.5×10^6 per *Jak3* SNV. With this approach, we were able to identify the presence of *Jak3* variants in the BM of 7 out of the 9 BCP-ALLs analyzed. *Jak3* mutant detection frequencies in the tumor

samples ranged from 2% up to 63%, while Jak3 mutations were absent in the corresponding germline samples as well as in BM cells of healthy *Pax5*[+/-] or WT mice (**Figure 31**).
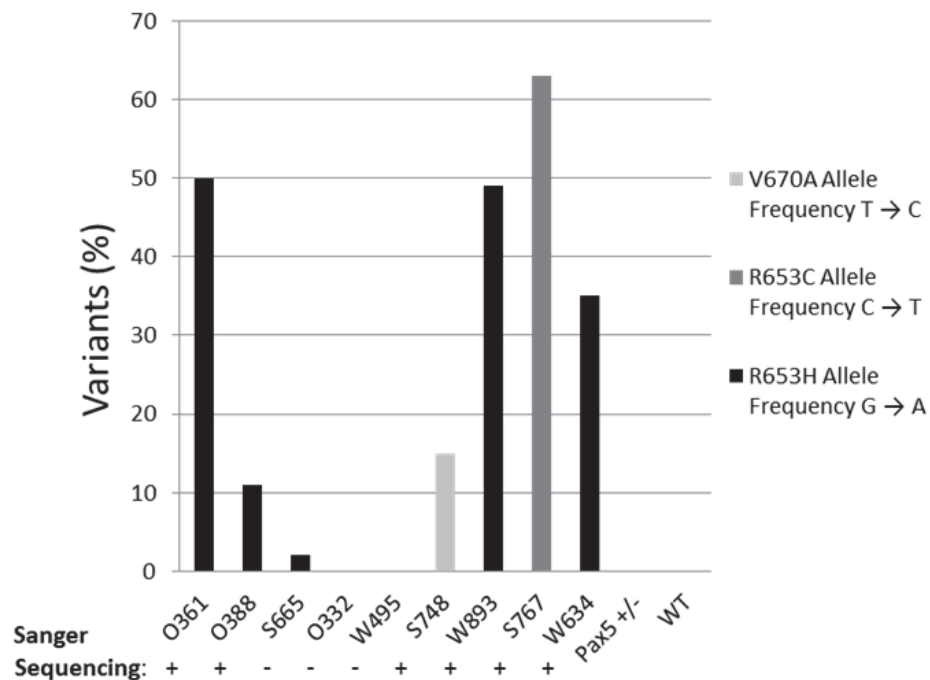
Figure 31: Deep sequencing of *Pax5*[+/-] tumor samples revealed *Jak3* variants in the BM of 7 out of 9 murine leukemias. The allele frequency of the *Jak3* variants p.V670A, p.R653C and p.R653H is shown for tumor DNA of 9 *Pax5* heterozygous mice. Healthy *Pax5* heterozygous and wildtype mice, for each of which three mice were pooled, serve as negative controls. Samples in which the variants were confirmed by Sanger sequencing are marked +.

Additionally, in mice W634 (*Jak3* p.R653H), S748 (*Jak3* p.V670A) and S767 (*Jak3* p.R653C), we were able to monitor leukemia evolution over time by analyzing extracted DNA from PB that was collected at routine intervals of every 2 months after birth. Sanger and deep sequencing at the same read depth as described above verified that all *Jak3* variants were first detectable when the mouse appeared clinically ill with BCP-ALL (**Figure 32**), which suggests a critical involvement of the mutations in disease progression.
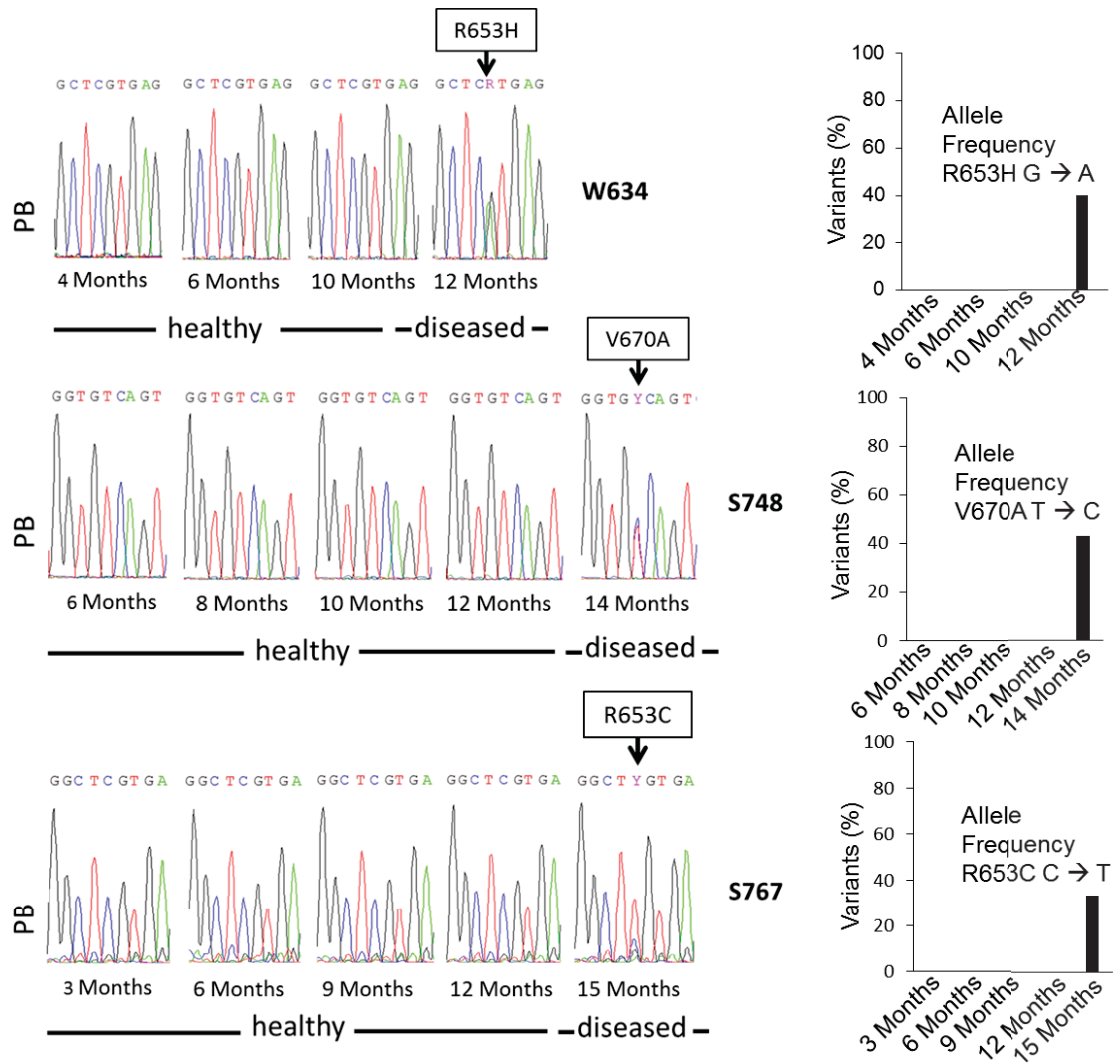
Figure 32: Backtracking validates the presence of *Jak3* mutations only in diseased mice. Displayed are backtrackings of the murine *Jak3* p.R653H, p.V670A and p.R653C mutations by Sanger sequencing and deep sequencing, using genomic DNA of peripheral blood (PB) from *Pax5*^+/- mice W634, S748 and S665. The leukemic clone is only detectable after BCP-ALL manifestation at 12, 14 and 15 months of age, respectively.

5.4.3 Jak3 mutant proteins induce constitutive active downstream signaling

Activating mutations in Jak3 confer a constitutive active function of the tyrosine kinase downstream, mediating constant survival and proliferation signals to the affected cell via Stat5 phosphorylation [116]. To assess whether the recurrent murine Jak3 mutations p.R653H and p.V670A are activating mutations, we used the IL-3 dependent murine pro-B cell line Ba/F3 as a readout model. Using this approach, Ba/F3 cells were only able to survive with constitutive activated Jak3-Stat signaling when IL-3 was

depleted from the culture medium. The overexpression of both tested Jak3 mutants was able to transform Ba/F3 cells to cytokine independent growth, while cells harboring Jak3 WT or the empty vector control died shortly after IL-3 depletion (**Figure 33**).
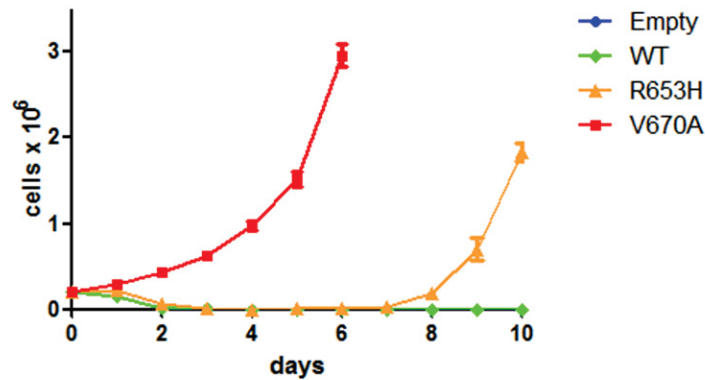
Figure 33: Murine Jak3 p.R653H and p.V670A transform Ba/F3 cells to IL-3 independent growth. Ba/F3 cells were transfected with pMC3 empty vector, pMC3-JAK3.WT, pMC3-JAK3.R653H or pMC3-JAK3.V670A and hygromycin-resistant cells were selected. Cells were cultured in media without IL-3 and their proliferation measured every day using Trypan Blue. Values represent the mean out of four replicates with essentially identical datasets.

Tumorigenic Jak3 signaling downstream could further be shown to be mediated by p.R653H in Ba/F3 cells as well as by p.V670A in a leukemic spleen sample by western blot analysis, since pStat5 was highly phosphorylated in the transformed samples (**Figure 34**).
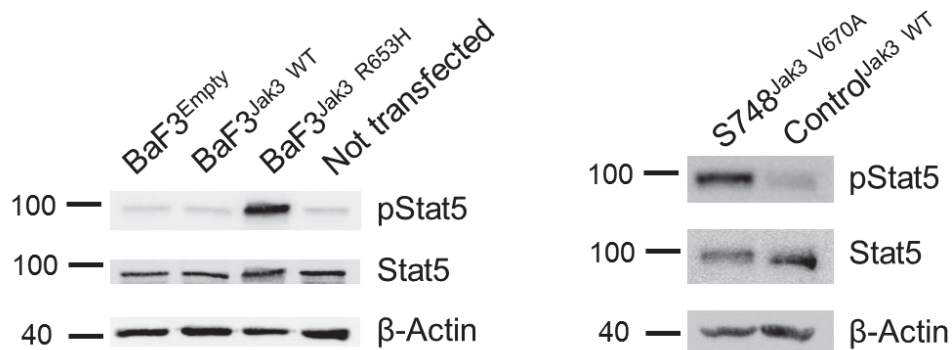
Figure 34: Murine Jak3 p.R653H and p.V670A cause constitutive active Jak3 signaling downstream. Left: Ba/F3 cells harboring pMC3 empty vector, pMC3-JAK3.WT or pMC3-JAK3.R653H were cultivated in media without IL-3 and serum for 4h, before being subjected to immunoblot analysis for phospho-Stat5, as well as total Stat5 and beta-Actin as loading control (n=3). Stat5 phosphorylation is only detected in cells transfected with Jak3 p.R653H. Right: Lysate of spleen tumor cells from mouse S748, harboring Jak3 p.V670A, was subjected to immunoblot analysis for phospho-Stat5, total Stat5 and beta-Actin. Wildtype (WT) spleen cells served as negative control. Compared to the control, the tumor cells harboring Jak3 p.V670A show constitutive pStat5 activation (n=2).

Taken together, these findings suggest that, in precursor B cells of *Pax5* heterozygous mice, exposure to infection as an environmental event leads to the acquisition of additional somatic aberrations of the *Pax5* WT allele as well as activating *Jak3* variants, which are first detectable in the PB shortly before leukemia outgrowth.

6 Discussion

6.1 *PAX5* c.547G>A confers susceptibility to BCP-ALL

6.1.1 Pediatric ALL: Progress through collaboration

The treatment of acute lymphoblastic leukemia (ALL) in children is an ideal example of how the collaborative effort of many different study groups in combination with newly evolved technologies can achieve a common goal. With survival rates up to 90% [34, 158] and a variety of promising new drugs on the way, the research focus can be directed more and more towards the elucidation of underlying disease mechanisms. Increasing our knowledge related to the molecular basis of ALL is particularly important, since, although the refinement and improvement of prognostic factors and chemotherapy have tremendously increased chances of cure, their optimization potential will soon be exhausted. The pediatric cancer genome project is one ongoing study to resolve this issue, utilizing whole genome sequencing to define the landscape of somatic alterations that drive tumorigenesis in different pediatric cancers [159]. The identification of tumor-specific somatic mutations is invaluable in developing new drug targets, e.g. the most prominent example being the application of the ABL tyrosine kinase inhibitor imatinib in combination with intensive chemotherapy, which improved the 5-year disease-free survival of Ph$^+$ ALL patients from 30 to 70% [43]. Nevertheless, it is also important to uncover the genetic susceptibility that constitutes the basis for disease development. Regarding this issue, considerable progress has been made over the last two decades through the introduction of genome-wide association studies (GWASs), which are designed to reveal common inherited genetic variants of leukemogenesis (reviewed in [48]). Important ALL susceptibility loci that could be identified using GWASs include single nucleotide polymorphisms (SNPs) in e.g. *ARID5B*, *IKZF1* and *CDKN2A/CDKN2B* [53-55, 61]. Another strategy to elucidate inherited leukemia predisposing variants is the screening of rare cases of familial ALL.

So far, recognized alterations in genes responsible for autosomal dominant transmission of childhood leukemia include *ETV6* [66], *TP53* [65], *CEBPA* [160] and *GATA-2* [161].

6.1.2   Impact of *PAX5* c.547G>A on leukemia development

Here, we report a new predisposing variant in the *PAX5* gene, which was recurrently present in three BCP-ALL affected children of a familial ALL case. The identified heterozygous *PAX5* mutation c.547G>A followed an autosomal dominant mendelian transmission through the fathers and reduces the transcriptional activity of *PAX5* compared to its wildtype (WT) counterpart. During physiologic hematopoiesis, PAX5 is one of the master regulators for B cell development and function (reviewed in [96]), since absence of Pax5 expression leads to B cell arrests at an early precursor stage [94]. Moreover, *PAX5* is recurrently mutated in ALL, being somatically lesioned in up to 40% of BCP-ALL cases [36, 45]. It is important to note that in the family presented here, *PAX5* c.547G>A displayed incomplete penetrance, since the single nucleotide variant (SNV) was also positively validated in healthy relatives. This observation implies that the identified inherited mutation itself poses only a mild predisposition in the sense of defective B cell differentiation and requires synergistic events for a leukemia phenotype to evolve. This is in line with the observation that *Pax5* heterozygous (*Pax5*$^{+/-}$) mice, which display a genetically engineered disruption of one *Pax5* allele, develop normally and stay healthy in specific pathogen-free facilities [94]. In contrast, heterozygous disruptions of other genes from the *PAX5* family have severe consequences, e.g. renal-coloboma, which is caused by mono-allelic deletions of PAX2 [162].

6.1.3   Somatic events cooperating with *PAX5* c.547G>A susceptibility

BCP-ALL affected children in the analyzed family displayed somatic loss of large regions on chromosome 9p, which not only encompassed the tumor suppressor locus CDKN2A, but also the *PAX5* gene itself. This recurrent tumor-specific fingerprint of all

three BCP-ALLs revealed the mechanism behind the *PAX5* c.547G>A mediated leukemia susceptibility, since the mutant protein conferred insufficient transcriptional activity after the somatic loss of heterozygosity (LOH) of the WT *PAX5* allele in the tumor cells. We further propose that the remarkably reduced PAX5 expression in this scenario created an aberrant B cell precursor compartment, which was susceptible to leukemic transformation. One important protein encoded by the *CDKN2A* locus is p14ARF, which stabilizes the tumor suppressor protein p53 by antagonizing MDM2. Stable p53 can counteract the propagation of transformed cells through the induction of downstream signaling that lead to cell-cycle arrest, senescence and/or apoptosis (reviewed in [163]). Currently, there are several clinical trials being carried out with small-molecule MDM2 antagonists, which aim for the pharmacological activation of the WT p53. This is a promising approach to restoring a physiological function of p53 in the case of CDKN2A loss [164].

Next to deletions of the CDKN2A locus, a detailed somatic mutation screen of the BCP-ALL blast cells from the affected children revealed additional tumor-specific mutations in *NRAS* (G12D/S) and *JAK1* (p.V1009I and p.L1010LL). Ras proteins, including HRAS, NRAS and KRAS, can activate multiple downstream signaling pathways, regulating cell survival, proliferation and differentiation. Secondary activating mutations affecting RAS genes have been identified in virtually all types of human cancers, with alterations in the amino acids 12, 13 and 61 being the most oncogenic (reviewed in [165]). A recent study also identified recurrent activating mutations in JAK1 in a cohort of BCR-ABL1-negative, high-risk pediatric ALL cases. Interestingly, additional deletions affecting the CDKN2A/B locus were detected in 70% of these JAK1 mutated cases [149]. Here, the identified somatic JAK1 mutations (p.V1009I and p.L1010LL) were located in the protein's kinase domain, which may lead to enhanced kinase activity, since it has been suggested that transformation of progenitor B cells requires very high levels of JAK/STAT5 activation. This was demonstrated by a mutagenesis screen in

Stat5 overexpressing mice with a B-cell- specific sleeping beauty transposon, showing that, in this scenario Jak1, is frequently targeted (26%), which increased Jak1 protein levels and resulted in B-ALL development [166]. In conclusion, we show how an inherited *PAX5* c.547G>A variant can predispose to BCP-ALL development in the scenario of secondary concomitant LOH of the WT PAX5 allele. Furthermore, the BCP-ALLs displayed disruption of the tumor suppressor locus CDKN2A and potentially activating mutations in signaling pathways affecting cell proliferation and survival (**Figure 35**).
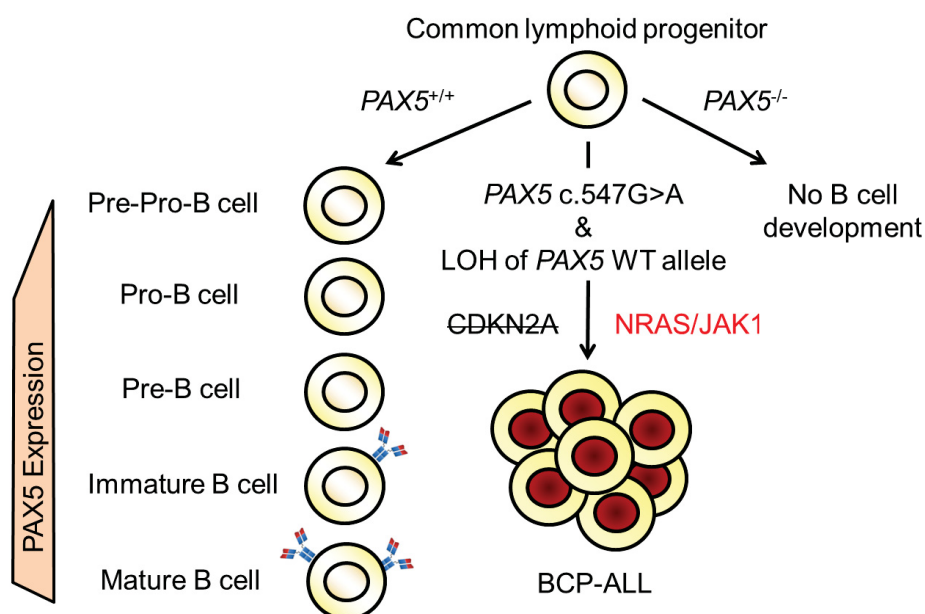
Figure 35: Normal and aberrant PAX5 expression during B cell development. Physiologically PAX5 expression is induced in early pro-B cells and is maintained throughout the life of a B cell. In case of complete PAX5 loss, a block in B cell development occurs, while B cell malignancies can arise in the scenario of aberrant PAX5 expression. Here, a germline *PAX5* predisposition (c.547G>A) conferring reduced transcriptional activity was combined with a loss of heterozygosity (LOH) in the remaining *PAX5* wildtype (WT) allele in the leukemic cells. This was accompanied by concomitant disruption of the tumor suppressor locus CDKN2A and activation of pathways involved in cell proliferation and survival (NRAS/JAK1).

The same results as obtained here could be further validated in two additional families, from an independent research group that published their data during the preparation of this work [150]. The authors could reproduce our findings in two unrelated kindreds from Puerto Rican and African American ancestry, with autosomal dominant segregation of *PAX5* c.547G>A, which also displayed variable penetrance and only

manifested with BCP-ALL in individuals with deletions on chromosome 9p. A mutational screen of 44 sporadic BCP-ALL cases with additional 9p aberrations revealed a low frequency of two *PAX5* mutations in this cohort, affecting the amino acid 183, which is translated from c.547 [150]. In our work we could not reproduce this finding by screening 23 cases with the respective phenotype negative for a PAX5 mutation affecting the protein position 183, which could be due to the small cohort size. We could further expand this screen to 1100 healthy individuals, which all tested negative for *PAX5* c.547G>A. This, and the fact that the mutation was so far only recurrently validated in three isolated families from all over the world, suggests that the susceptibility conferred through *PAX5* c.547G>A is confined to distinct familial backgrounds. The identification of the inherited susceptibility to BCP-ALL caused by *PAX5* c.547G>A can therefore have clinical implications, e.g. for pre-implantation genetic diagnosis. Nevertheless, in order to define the complete spectrum of how inherited predisposing variants contribute to leukemogenesis, further familial ALL cases need to be sequenced and assessed regarding their underlying genetic makeup.

6.2 Infection exposure as a causal factor for BCP-ALL development

6.2.1 Infection exposure in *Pax5*$^{+/-}$ mice

Although we could show that *PAX5* c.547G>A confers susceptibility to BCP-ALL, only 1/3 of the family members carrying the mutations eventually developed a clinical disease manifestation. This incomplete penetrance suggests that the inherited PAX5 variant only mediated the appearance of an aberrant precursor B cell compartment, comprised of immunologically hidden but persistent pre-leukemic clones that were prone to acquiring additional genetic hits through external influences. However, by solely analyzing the family we are able to characterize the phenotypic endpoint of the leukemia, but the mechanisms responsible for the conversion of the pre-leukemic clone into full-blown BCP-ALL remain unclear. Next to radiation and chemical exposures [72,

76], infections have been regarded as one of the most likely causes of childhood leukemia over the past century [80-82]. Infection exposure as causal factor for the development of childhood ALL has further been supported by descriptive epidemiologic studies (reviewed in [89]), although biological proof of its involvement as a second hit, related to the natural history of the disease with prenatal initiation, was still missing. It is known that *Pax5* heterozygous mice never spontaneously develop leukemia when housed in a pathogen-free environment [94, 153]. However, we could show that the exposure of initially pathogen-free *Pax5$^{+/-}$* mice to an infectious environment at the age of two to five months represented the necessary oncogenic trigger to initiate BCP-ALL development in around 1/4 of the mice (**Figure 36**).
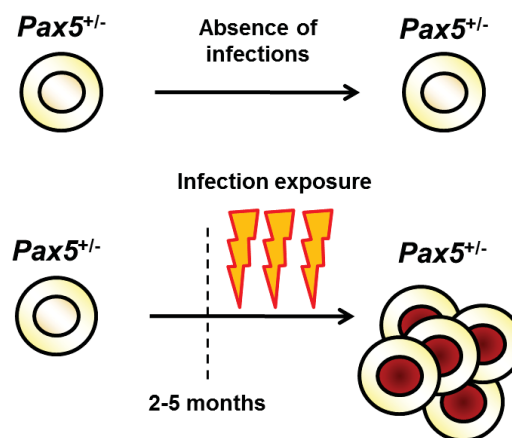
Figure 36: In mice, exposure to infection is a causal factor for the development of BCP-ALL as a result of *Pax5* inherited susceptibility. While *Pax5$^{+/-}$* mice stay healthy in the absence of infections, infection exposure between 2-5 months after birth can lead to leukemia development in *Pax5$^{+/-}$* mice.

This closely mimics the human situation, since children encounter with a variety of different infectious agents in early life, which shape the immune system and the later immune response. Nevertheless, all attempts to link a single consistent infectious agent with BCP-ALL development in childhood have failed (reviewed in [89]). This would suggest that for disease occurrence, a sequence of infections and broad immune responses in an unprimed immune system are necessary, rather than exposure to a single distinct pathogen. Likewise, around 5 months prior to ALL manifestation, children display marked deterioration of their immune response and

profound liability to a variety of infections [167]. The results presented here are in support of the different infection hypotheses, although in our setup, we were not able to pinpoint distinct infectious agents that trigger disease development or to define the exact time point/s of infection that is/are necessary to catch the immune system off-guard. In order to elucidate underlying mechanisms, further experiments utilizing *Pax5*$^{+/-}$ mice have to be conducted, including infection exposures at various time points, artificial bacterial/viral stimulations, microbiome analyses, antibiotic treatments or the backcross with mice deficient for certain infection specific receptors.

6.2.2 Somatic Pax5 and Jak3 mutations in murine *Pax5*$^{+/-}$ BCP-ALLs

The somatic mutations identified in BCP-ALLs of Pax5 heterozygous mice are summarized in (**Table 10**).

Mouse ID	Blast count	CD19 phenotype	Jak3 Mutation*	Mutation rate#	Human homolog	other Mutations*
O361	98.5%	CD19+	R653H	50%	R657Q	Jak1 S645P
O388	66.9%	CD19-	R653H	11%	R657Q	-
S665	97.9%	CD19-	-	2%	R657Q	Jak1 V657F Pax5 P80R
O332	13.4%	CD19-	-	-	-	-
W495	27.6%	CD19-	-	-	-	-
S748	95.1%	CD19+	V670A	16%	V674A	Pax5 P80L
W893	73.6%	CD19-	R653H	49%	R657Q	-
S767	74.7%	CD19+	R653C	63%	R657Q	Pax5 delEx7-8
W634	93.8%	CD19+	R653H	35%	R657Q	-

Table 10: Genotype and phenotype of *Pax5*$^{+/-}$ BCP-ALLs. (*) marks mutations confirmed by Sanger sequencing, while (#) refers to the mutation rate estimated through amplicon sequencing.

Regarding the tumor-specific molecular fingerprint of *Pax5*$^{+/-}$ BCP-ALLs, we were able to identify secondary aberrations affecting the WT Pax5 allele in 3 out of 9 mice. These alterations did not correspond to a complete absence of Pax5 functionality in the

tumors, since 2 out of 3 tumors still displayed CD19 expression, one of the most important downstream targets of Pax5. This phenomenon suggests that the identified mutations/deletion only confer a mild reduction of Pax5 activity. Likewise, in human BCP-ALL cases, no correlation between PAX5 mutation status and expression of the PAX5 target genes CD19 and CD79A could be found [45, 168, 169]. Nevertheless, gene expression profiling demonstrated that the existence of PAX5 mutations can have major effects on the intracellular transcriptional network within primary leukemic cells [45]. However, in our cohort, 4 out of 9 BCP-ALLs displayed absence of CD19 without the detection of additional aberrations affecting Pax5. One possible explanation is post-transcriptional silencing of Pax5, e.g. epigenetic mechanisms, which can only be speculated, since a study only detected minimal levels of *PAX5* promoter methylation in human BCP-ALL, irrespective of the mutational status of *PAX5* [45].

In addition to inactivating Pax5 aberrations affecting the remaining WT Pax5 allele, activating mutations of Jak3 were recurrently detected in the murine BCP-ALLs. Therefore we propose that disruption of the *Pax5* gene promotes leukemogenesis by creating an aberrant Interleukin-7 (IL-7) sensitive progenitor compartment in the murine bone marrow. The cytokine IL-7 is produced by stromal cells in the BM and is essential for B cell development, since IL-7 or IL-7-Receptor (IL-7R) knockout mice show a B cell arrest at the pro-B cell stage [170, 171]. IL-7 is especially important for B cell precursors, orchestrating their proliferation, differentiation and survival (reviewed in [172]), through active Jak-Stat signaling. Cell survival of early B cell precursors is one key function of the IL-7R signaling, through the regulating of anti-apoptotic (e.g. Bcl-2 and Mcl-1) and pro-apoptotic (e.g. Bax and Bim) factors [173, 174]. IL-7 withdrawal leads to the death of B cell precursors, while the assembly of a functioning B cell receptor (BCR) renders the cells IL-7 unresponsive (reviewed in [175]). Our data demonstrates the propagation of secondary Jak3 mutations in BCP-ALLs of *Pax5* heterozygous mice, which constitutively activate the IL-7R signaling. This could be

validated by increased Stat5 phosphorylation, which is the predominant Stat protein activated by IL-7 (reviewed in [176]). We therefore suggest that *Pax5*[+/-] precursor B cells that are recurrently triggered by infectious agents in the conventional facility are susceptible to malignant transformation through the acquisition of somatic Jak3 mutations. These somatic Jak3 mutations promote leukemia development with a phenotype very similar to human BCP-ALL, with respect to clinical, immune-phenotypic and molecular genetic characteristics. The mutations were first detectable in the peripheral blood immediately before leukemia outgrowth, and further *in vivo* transplantation experiments demonstrated that the activating Jak3 mutations per se are sufficient to drive leukemia in recipient mice [153]. Moreover, the *in vitro* treatment of tumor pro-B cells harboring Jak3 p.V670A shows sensitivity to Jak1/3 as well as Jak1/2 inhibitors [153]. The responsiveness to Jak1 inhibitors in this case can be explained by the fact that Jak3 mutants are dependent on a functional Jak1/Jak3 cytokine receptor complex for signaling [177]. Thus, targeting the deregulated Jak/Stat pathway could be a promising therapy for this disease phenotype, although the selection and dosing of the inhibitor is a difficult task and further *in vivo* studies are required. Nevertheless, our findings are important not only for endorsing the credibility of the infection hypotheses, but also for the development of novel therapeutic targets, which could help to prevent as well as effectively treat a significant proportion of childhood leukemias.

6.2.3 Infection exposure in other murine leukemia models

Since inherited defects in the *Pax5* gene only pose a small proportion of BCP-ALLs, further ongoing studies aim to reproduce these findings in more prevalent ALL subtypes. The ETV6-RUNX1 fusion gene is the most common chromosomal alteration in pediatric cancer and occurs in approximately 25% of childhood BCP-ALL [178]. In agreement with our hypothesis, mice which express the human ETV6-RUNX1 fusion gene under the control of the Sca1 promoter in the hematopoietic system develop BCP-ALLs at low frequencies of 9% only when they were exposed to common

pathogens like *Pax5*[+/-] mice (Manuscript in preparation) (**Figure 37**). This low disease penetrance mirrors the human scenario, since RUNX1 fusion genes are detectable in cord bloods at a frequency that is 100-fold greater than the risk of the corresponding leukemia [179]. Nevertheless, an infectious environment as leukemia trigger is only needed in combination with mild B cell developmental defects, since expression of the human BCR-ABL[p190] in murine hematopoietic stem cells, which is an oncogene conferring active kinase signaling, is sufficient to induce BCP-ALL development at a frequency of 13% in infection-unexposed mice. This disease incidence could be accelerated to 90% when an additional *Pax5* heterozygosity was introduced in the genetic background, which again highlights the essential role of Pax5 in regard to B-ALL manifestation (Manuscript submitted) (**Figure 37**).
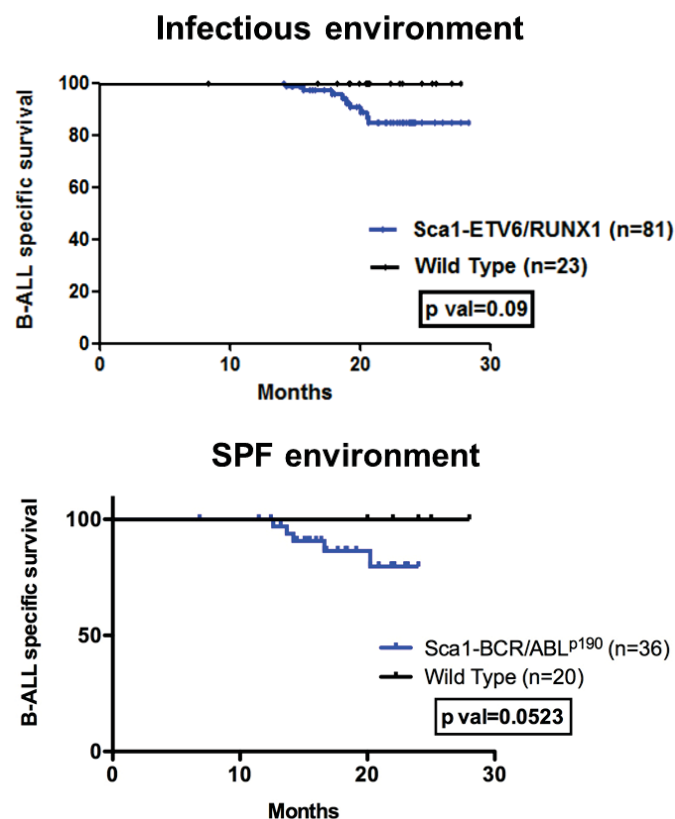
Figure 37: Relevance of infection exposure for other B-ALL subtypes. Preliminary data showing that mice expressing the human ETV6-RUNX1 fusion in hematopoietic stem cells (HSCs) develop B-ALL only when exposed to common pathogens (top graph), while HSC restricted expression of the human BCR-ABL[p190] fusion is sufficient to induce B-ALL in mice housed in a specific pathogen free (SPF) facility (bottom graph).

6.3 Pax5 inherited susceptibility - from humans to mice to humans

6.3.1 Relevance of mouse models for human leukemia research

The implementation of new technologies, e.g. next generation sequencing, has opened up new possibilities to elucidate disease-causing mechanisms according to their genetic basis. Nevertheless, whether screening large patient cohorts or distinct familial backgrounds, it still comes down to the analysis of large data sets, which are subsequently filtered and narrowed down to a handful of potential target variants. The consequences of these variants can be predicted by structural modelling of the corresponding protein, or their biological effects can be assessed using cell culture systems *in vitro*. However, in order to provide credible proof that the identified genetic aberration is indeed associated with a certain disease phenotype, *in vivo* modelling of the alteration is inevitable. So far, a variety of experimental *in vivo* models have been utilized to characterize the hematopoietic system. Drosophila, for example, is a well-established model system for hematopoiesis and a powerful tool to study the molecular interactions among key genes involved in leukemia (reviewed in [180]). In addition, zebrafish *in vivo* models have been useful to reveal potential partners of initial leukemia lesions, through forward genetics mutagenesis screens (reviewed in [181]). Nevertheless, the power of these models to recapitulate all aspects of the human disease is limited. For mice, on the other hand, a high degree of conservation between the human and mouse genome has been reported, with merely around 300 genes being unique to one or the other [182]. Furthermore, in mice, striking similarities to humans regarding the mechanisms and key events in B cell development have been found [183]. The different B cell developmental stages in humans closely resemble the corresponding murine counterparts, with some differences in surface marker expression and growth requirements [183, 184]. This analogy could be utilized to engineer a variety of different mouse models to study the underlying genetic mechanisms of BCP-ALL [185]. Therefore, over the past century, mouse models have

been the animal system of choice to validate candidate genes involved in leukemia development. So far, murine models which tried to uncover the role of aberrant PAX5 expression in childhood BCP-ALL have been rare. Since *Pax5* homozygous knockout mice show a severe block in B cell development [97], most existing genetically engineered mouse models (GEMMs) utilize the *Pax5*$^{+/-}$ strain. Moreover, *Pax5*$^{+/-}$ mice are a useful model to study the relevance of Pax5 due to amino acid sequence similarities of 99% between the human and murine Pax5 gene.

6.3.2 Comparison between Pax5 susceptibility in humans and mice

We were able to show that a heterozygous germline mutation in *PAX5* c.547G>A poses an increased risk of BCP-ALL development in humans. In *Pax5* heterozygous mice, an additional exposure to infection was necessary in order to reproduce these findings. Both scenarios display high resemblances in regard to phenotype, penetrance and somatic make-up of the leukemias (**Table 11**). Therefore, the following paragraphs will focus on similarities as well as disparities between the *Pax5* mediated BCP-ALLs in humans and mice, as well as highlight what we could learn from the mouse model to close the circle back to the human situation.

	Penetrance	Environmental trigger	Phenotype	Pax5 activity	Somatic molecular fingerprint
Human *PAX5* c.547G>A	Incomplete	Unknown	Pre-B CD19+	Reduced	LOH *PAX5* WT, CDKN2A, JAK1/NRAS
Pax5+/-	Incomplete	Infection	Pro/Pre-B CD19+/-	Reduced	Pax5 aberrations, Cdkn2a, Jak1/Jak3

Table 11: Differences and similarities of BCP-ALLs from humans carrying *PAX5* c.547G>A and *Pax5* heterozygous mice.

6.3.2.1 Penetrance and environmental trigger

Although *PAX5* c.547G>A could be positively validated in 9 family members, only 3 of these eventually developed BCP-ALL [186]. This reduced penetrance could be reproduced in the *Pax5*$^{+/-}$ murine model, which displayed a leukemia incidence of 22%. Incomplete or reduced penetrance in general is found in a variety of diseases, while the parameters affecting penetrance are numerous and often hard to assess, since they can include genetic as well as environmental factors [187]. The fact that infection exposure was able to induce leukemia in *Pax5*$^{+/-}$ mice is one potential explanation for an environmental factor that might have also played a role in the human scenario, since the etiology of leukemia in childhood has frequently been linked to infectious stimuli [69, 70]. Nevertheless, the sole presence of pathogens was not sufficient to confer complete penetrance in the mouse model. Hence, further studies which aim to increase the leukemia incidence in *Pax5*$^{+/-}$ mice in the context of infection exposure will be valuable to reveal the leukemia-driving events that took place in the BCP-ALL affected children, in contrast to their healthy relatives susceptible to the same phenotype.

6.3.2.2 Phenotype and reduced Pax5 activity

The phenotype of the BCP-ALLs identified in the human family and in the *Pax5*$^{+/-}$ mice show a high resemblance, since in both leukemias precursor B cell populations were targets of the malignant transformation. Nevertheless, differences occurred in the expression of the Pax5 target gene *CD19*, which was absent on the cell surface of 5 out of 9 murine BCP-ALLs analyzed, while the identified human leukemias were uniformly CD19 positive. This shows that for BCP-ALL manifestation, a complete disruption of Pax5 function is possible but not necessary. Likewise, somatic PAX5 mutations can have major implications on the intracellular transcriptional network in human leukemias, while they still express CD19 [45]. However, the fact that *Pax5*$^{+/-}$ mice are more prone to a complete loss of Pax5 expression in the tumors can be

explained by their genotype, since the inbred heterozygous disruption of one Pax5 allele renders the remaining WT allele more susceptible to acquiring mutations. Again, in both scenarios, a reduced Pax5 activity could be validated in the tumors. In the family *PAX5* c.547G>A confers reduced PAX5 transcriptional activity *in vitro* [150, 186], which exerted functional consequences after a LOH of the *PAX5* WT allele in the BCP-ALL blast cells. Likewise, the majority of the murine BCP-ALLs either lacked the expression of Pax5 target genes (e.g. CD19) or displayed somatically acquired *Pax5* aberrations, influencing Pax5 functionality (Pax5 p.P80R/L, Pax5 delEx7-8).

6.3.2.3 Somatic molecular fingerprint

Analyzing the tumor-specific fingerprints that arose in *Pax5*[+/-] BCP-ALLs is particularly important, to identify potential secondary genetic hits which cooperate with *Pax5* susceptibility. These could in turn serve as potential targets for directed therapy in the human setting. We were able to show that BCP-ALL-affected patients harboring *PAX5* c.547G>A lost the remaining *PAX5* WT allele through secondary structural aberrations of chromosome 9p. In *Pax5*[+/-] mice we face the same scenario, but the sequence is reverse, since the mice display one disrupted *Pax5* allele from birth and somatically acquire additional Pax5 mutations or deletions on the remaining WT allele. Nevertheless, the bottom line in both species is a somatic reduction of Pax5 activity. In this regard, proper Pax5 dosage plays a key role, since a transgenic RNA-Interferenz based B-ALL mouse model, which allowed inducible suppression and restoration of endogenous Pax5 levels showed that restoration of Pax5 not only reengaged B-lineage differentiation, but also led to disease remission and long term survival [188]. Therefore, restoring endogenous Pax5 expression through e.g. gene editing in the human setting could be a valuable approach in the future to directly target inherited genetic lesions of PAX5.

The tumor-specific inconsistencies between the identified somatic mutational pattern in the human and murine scenario can mostly be explained through species differences.

While all three BCP-ALL-affected children displayed deletions encompassing the tumor suppressor locus CDKN2A, only 1 out of 9 mice showed similar findings. However, since the close genomic proximity between CDKN2A and PAX5 in humans cannot be translated to the murine genome, we would expect a disruption of other tumor suppressor loci or proto-oncogenes in the mouse. Therefore, instead of deletions affecting CDKN2A, murine BCP-ALLs displayed recurrent mutations in Jak3, which rendered the tyrosine kinase constitutively active and conferred continuous proliferation and survival signals to the leukemic cell. Murine precursors B cells are particularly sensitive to IL-7 withdrawal and are dependent on an active IL-7R receptor signaling in order to survive and proliferate [171]. The human B cell development, on the other hand, is less dependent on the IL-7R signaling cascade, including JAK3 and STAT5, since several patient-derived mutations affecting the IL-7Ralpha gene are associated with normal numbers of peripheral CD19$^+$ B cells [189, 190]. Therefore, in humans, an alternative mechanism that can bypass the IL-7 signaling seems possible, while deregulation of the IL-7R severely impairs B cell lymphopoiesis in mice [171]. In agreement, the leukemia incidence of mice with constitutive active expression of Stat5 (Stat5-CA) is accelerated from 1 to 100% when these mice are crossed back on a $Pax5^{+/-}$ background. Consistently, these GEMMs fail to develop leukemia when backcrossed on an IL-7 deficient background, which again highlights the dependence of murine B cell precursors on the IL-7/Jak3 signaling pathway [100, 191]. Although the authors did not analyze the mutational spectrum of secondary alterations in these mice, one can speculate that activating mutations of the IL-7R signaling, e.g. Jak3, could have been present in Stat5-CA x $Pax5^{+/-}$ leukemias. This speculation is further supported by a study which exposed thymectomized $Pax5^{+/-}$ mice to chemical and viral mutagens and identified somatic mutations affecting Jak kinases in the resulting B cell leukemias [192]. Altogether, the recurrently mutated Jak3 gene in the murine model might be due to the species specific B cell dependency on IL-7, since no aberrations affecting the Jak3 gene could be detected in the BCP-ALL-affected children.

Nevertheless, the connection between Pax5 defects and active IL-7 signaling in the mouse can still be valuable in regard to the human scenario. The successful implementation of tyrosine kinase inhibitors in the treatment of BCR-ABL is only one example of how the targeting of this signaling pathway is essential for leukemia progression in humans (reviewed in [193]).

7   Concluding remarks and impact of this work

We were able to identify an inherited leukemia predisposing variant in PAX5 in humans, and further link Pax5 aberrations and BCP-ALL development to infection exposure in mice. However, several challenges remain that need to be addressed. The application of next generation sequencing techniques offers huge benefits in terms of data acquisition, but at the same time poses a risk of overlooking essential targets in the large data output. Tumor heterogeneity is an important issue in this regard, since the analyzed cell population is normally composed of a mixture of cancer cells intermingled with healthy support and immune cells [159]. In addition, it has been suggested that genetic lesions, which initiate the cancer process, might be dispensable for tumor progression and maintenance, implying that the vast majority of cancer cells are not necessarily the source of leukemia development and relapse [194]. Likewise, mutations which are present in as little as 1% of the leukemic population are often founders of metastatic dissemination and relapse [195, 196]. Another important point to address is that the major focus in the analysis of sequencing data is put on the identification of non-silent mutations in annotated genes, while aberrations in regulatory areas, e.g. promoter and enhancer regions, are often neglected [159]. In order to fully grasp the scope of these mutations, further algorithms that integrate genomic, expression and epigenetic data have to be implemented. Together, this highlights how easily the interpretation of sequencing data can remain incomplete, due to low sensitivity or insufficient knowledge of specific regulatory regions.

Nevertheless, by analyzing the familial ALL discussed in this work, we were able to pinpoint *PAX5* c.547G>A as one important inherited factor in the development of BCP-ALL in this particular family background. Although the variant rarely occurs in the general population, we were able to employ our findings in the analysis of an additional child who was born in the family during the preparation of this work. Fortunately, the child tested negative for the *PAX5* c.547G>A variant, which enabled the family a

normal life without the fear of a leukemia predisposition in their newborn. Furthermore, the *Pax5*$^{+/-}$ mouse model helped us to reveal secondary events that cooperate with inherited Pax5 susceptibility, which generates a basis for future experimental designs based on infection exposures, Jak3 inhibitors and gene editing approaches for Pax5.

8 References

1. Howlader, N., et al., eds, *SEER cancer statistics review, 1975-2010.* Bethesda, Md: National Cancer Institute, 2013.
2. Pizzo, P.A., Poplack, D.G., eds., *Biologcial basis of childhood cancer*, in *Principles and practice of pediatric oncology. 7th ed.*: Philadelphia, Pa: Lippincott Williams and Wilkins, 2015, p. 1-98.
3. Winslow, T. 2007. http://www.cancer.gov.
4. Ries, L.A., et al., *Cancer incidence and survival among children and adolescents: United States SEER Program 1975-1995.* National Cancer Institute, SEER Program. NIH Pub. No. 99-4649. Bethesda, MD, 1999.
5. Linabery, A.M. and J.A. Ross, *Trends in childhood cancer incidence in the U.S. (1992-2004).* Cancer, 2008. **112**(2): p. 416-32.
6. Lim, J.Y., et al., *Genomics of racial and ethnic disparities in childhood acute lymphoblastic leukemia.* Cancer, 2014. **120**(7): p. 955-62.
7. Pizzo, P.A., Poplack, D.G., eds., *Acute lymphoblastic leukemia*, in *Principles and practice of pediatric oncology. 7th ed.*: Philadelphia, Pa: Lippincott Williams and Wilkins, 2015, p. 463-97.
8. Behm, F., *Classification of acute leukemias: perspective 2*, in *Treatment of acute leukemias: new directions for clincial research*. Totowa, NJ: Humana Press, 2003.
9. Peters, J.M. and M.Q. Ansari, *Multiparameter flow cytometry in the diagnosis and management of acute leukemia.* Arch Pathol Lab Med, 2011. **135**(1): p. 44-54.
10. Chiaretti, S., G. Zini, and R. Bassan, *Diagnosis and subclassification of acute lymphoblastic leukemia.* Mediterr J Hematol Infect Dis, 2014. **6**(1): p. e2014073.
11. Bene, M.C., et al., *Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL).* Leukemia, 1995. **9**(10): p. 1783-6.
12. Mullighan, C.G., *Molecular genetics of B-precursor acute lymphoblastic leukemia.* J Clin Invest, 2012. **122**(10): p. 3407-15.
13. Hunger, S.P. and C.G. Mullighan, *Acute Lymphoblastic Leukemia in Children.* N Engl J Med, 2015. **373**(16): p. 1541-52.
14. Harrison, C.J., *Cytogenetics of paediatric and adolescent acute lymphoblastic leukaemia.* Br J Haematol, 2009. **144**(2): p. 147-56.
15. Mrozek, K., D.P. Harper, and P.D. Aplan, *Cytogenetics and molecular genetics of acute lymphoblastic leukemia.* Hematol Oncol Clin North Am, 2009. **23**(5): p. 991-1010, v.
16. Zhang, J., et al., *Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group.* Blood, 2011. **118**(11): p. 3080-7.
17. Secker-Walker, L.M., S.D. Lawler, and R.M. Hardisty, *Prognostic implications of chromosomal findings in acute lymphoblastic leukaemia at diagnosis.* Br Med J, 1978. **2**(6151): p. 1529-30.
18. Gibbons, B., et al., *Near haploid acute lymphoblastic leukemia: seven new cases and a review of the literature.* Leukemia, 1991. **5**(9): p. 738-43.
19. Hann, I., et al., *Determinants of outcome after intensified therapy of childhood lymphoblastic leukaemia: results from Medical Research Council United Kingdom acute lymphoblastic leukaemia XI protocol.* Br J Haematol, 2001. **113**(1): p. 103-14.
20. Heerema, N.A., et al., *Prognostic impact of trisomies of chromosomes 10, 17, and 5 among children with acute lymphoblastic leukemia and high hyperdiploidy (> 50 chromosomes).* J Clin Oncol, 2000. **18**(9): p. 1876-87.
21. Mitelman, F., B. Johansson, and F. Mertens, *The impact of translocations and gene fusions on cancer causation.* Nat Rev Cancer, 2007. **7**(4): p. 233-45.
22. Wiemels, J.L., et al., *Prenatal origin of acute lymphoblastic leukaemia in children.* Lancet, 1999. **354**(9189): p. 1499-503.
23. Ma, X., et al., *Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia.* Nat Commun, 2015. **6**: p. 6604.
24. Dyer, M.J., et al., *Immunoglobulin heavy chain locus chromosomal translocations in B-cell precursor acute lymphoblastic leukemia: rare clinical curios or potent genetic drivers?* Blood, 2010. **115**(8): p. 1490-9.

25. Shurtleff, S.A., et al., *TEL/AML1 fusion resulting from a cryptic t(12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis.* Leukemia, 1995. **9**(12): p. 1985-9.

26. Rowley, J.D., *Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining.* Nature, 1973. **243**(5405): p. 290-3.

27. Kurzrock, R., et al., *A novel c-abl protein product in Philadelphia-positive acute lymphoblastic leukaemia.* Nature, 1987. **325**(6105): p. 631-5.

28. Meyer, C., et al., *The MLL recombinome of acute leukemias in 2013.* Leukemia, 2013. **27**(11): p. 2165-76.

29. Mellentin, J.D., et al., *The gene for enhancer binding proteins E12/E47 lies at the t(1;19) breakpoint in acute leukemias.* Science, 1989. **246**(4928): p. 379-82.

30. Harrison, C.J., *Blood Spotlight on iAMP21 acute lymphoblastic leukemia (ALL), a high-risk pediatric disease.* Blood, 2015. **125**(9): p. 1383-6.

31. Roberts, K.G., et al., *Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia.* N Engl J Med, 2014. **371**(11): p. 1005-15.

32. Riehm, H., et al., *The Berlin Childhood Acute Lymphoblastic-Leukemia Therapy Study, 1970-1976.* American Journal of Pediatric Hematology Oncology, 1980. **2**(4): p. 299-306.

33. Gaynon, P.S., et al., *Intensive Therapy for Children with Acute Lymphoblastic-Leukemia and Unfavorable Presenting Features - Early Conclusions of Study Ccg-106 by the Childrens-Cancer-Study-Group.* Lancet, 1988. **2**(8617): p. 921-924.

34. Hunger, S.P., et al., *Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the children's oncology group.* J Clin Oncol, 2012. **30**(14): p. 1663-9.

35. Wassmann, B., et al., *Alternating versus concurrent schedules of imatinib and chemotherapy as front-line therapy for Philadelphia-positive acute lymphoblastic leukemia (Ph+ ALL).* Blood, 2006. **108**(5): p. 1469-77.

36. Mullighan, C.G., et al., *Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia.* N Engl J Med, 2009. **360**(5): p. 470-80.

37. Mann, G., et al., *Improved outcome with hematopoietic stem cell transplantation in a poor prognostic subgroup of infants with mixed-lineage-leukemia (MLL)-rearranged acute lymphoblastic leukemia: results from the Interfant-99 Study.* Blood, 2010. **116**(15): p. 2644-50.

38. Harrison, C.J., et al., *An international study of intrachromosomal amplification of chromosome 21 (iAMP21): cytogenetic characterization and outcome.* Leukemia, 2014. **28**(5): p. 1015-21.

39. Moorman, A.V., et al., *Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial.* Lancet Oncol, 2010. **11**(5): p. 429-38.

40. Collins, F.S. and H. Varmus, *A new initiative on precision medicine.* N Engl J Med, 2015. **372**(9): p. 793-5.

41. National Research Council (US) Committee, in *Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease.* 2011: Washington (DC).

42. Ross, D.M., et al., *Safety and efficacy of imatinib cessation for CML patients with stable undetectable minimal residual disease: results from the TWISTER study.* Blood, 2013. **122**(4): p. 515-22.

43. Schultz, K.R., et al., *Long-term follow-up of imatinib in pediatric Philadelphia chromosome-positive acute lymphoblastic leukemia: Children's Oncology Group study AALL0031.* Leukemia, 2014. **28**(7): p. 1467-71.

44. Chiaretti, S., et al., *Advances in the Genetics and Therapy of Acute Lymphoblastic Leukemia.* Am Soc Clin Oncol Educ Book, 2016. **35**: p. e314-22.

45. Mullighan, C.G., et al., *Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia.* Nature, 2007. **446**(7137): p. 758-64.

46. Mullighan, C.G., *Genomic characterization of childhood acute lymphoblastic leukemia.* Semin Hematol, 2013. **50**(4): p. 314-24.

47. McCarthy, M.I., et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges.* Nat Rev Genet, 2008. **9**(5): p. 356-69.

48. Pui, C.H., et al., *Childhood Acute Lymphoblastic Leukemia: Progress Through Collaboration.* J Clin Oncol, 2015. **33**(27): p. 2938-48.

49. Kharazmi, E., et al., *Familial risks for childhood acute lymphocytic leukaemia in Sweden and Finland: far exceeding the effects of known germline variants.* Br J Haematol, 2012. **159**(5): p. 585-8.

50. Greaves, M.F., et al., *Leukemia in twins: lessons in natural history.* Blood, 2003. **102**(7): p. 2321-33.

51. Zhang, J., K.E. Nichols, and J.R. Downing, *Germline Mutations in Predisposition Genes in Pediatric Cancer.* N Engl J Med, 2016. **374**(14): p. 1391.

52. Moriyama, T., M.V. Relling, and J.J. Yang, *Inherited genetic variation in childhood acute lymphoblastic leukemia.* Blood, 2015. **125**(26): p. 3988-95.

53. Trevino, L.R., et al., *Germline genomic variants associated with childhood acute lymphoblastic leukemia.* Nat Genet, 2009. **41**(9): p. 1001-5.

54. Papaemmanuil, E., et al., *Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia.* Nat Genet, 2009. **41**(9): p. 1006-10.

55. Sherborne, A.L., et al., *Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk.* Nat Genet, 2010. **42**(6): p. 492-4.

56. Patsialou, A., D. Wilsker, and E. Moran, *DNA-binding properties of ARID family proteins.* Nucleic Acids Res, 2005. **33**(1): p. 66-80.

57. Akagi, T., et al., *In vivo deficiency of both C/EBPbeta and C/EBPepsilon results in highly defective myeloid differentiation and lack of cytokine response.* PLoS One, 2010. **5**(11): p. e15419.

58. Molnar, A. and K. Georgopoulos, *The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins.* Mol Cell Biol, 1994. **14**(12): p. 8292-303.

59. Serrano, M., G.J. Hannon, and D. Beach, *A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4.* Nature, 1993. **366**(6456): p. 704-7.

60. Kamb, A., et al., *A cell cycle regulator potentially involved in genesis of many tumor types.* Science, 1994. **264**(5157): p. 436-40.

61. Xu, H., et al., *Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations.* J Natl Cancer Inst, 2013. **105**(10): p. 733-42.

62. Migliorini, G., et al., *Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype.* Blood, 2013. **122**(19): p. 3298-307.

63. Perez-Andreu, V., et al., *Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse.* Nat Genet, 2013. **45**(12): p. 1494-8.

64. Ellinghaus, E., et al., *Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia.* Leukemia, 2012. **26**(5): p. 902-9.

65. Powell, B.C., et al., *Identification of TP53 as an acute lymphocytic leukemia susceptibility gene through exome sequencing.* Pediatr Blood Cancer, 2013. **60**(6): p. E1-3.

66. Topka, S., et al., *Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia.* PLoS Genet, 2015. **11**(6): p. e1005262.

67. Lee, P., et al., *The biology, pathogenesis and clinical aspects of acute lymphoblastic leukemia in children with Down syndrome.* Leukemia, 2016.

68. Yang, W., et al., *ARID5B SNP rs10821936 is associated with risk of childhood acute lymphoblastic leukemia in blacks and contributes to racial differences in leukemia incidence.* Leukemia, 2010. **24**(4): p. 894-6.

69. Eden, T., *Aetiology of childhood leukaemia.* Cancer Treat Rev, 2010. **36**(4): p. 286-97.

70. Greaves, M., *Infection, immune responses and the aetiology of childhood leukaemia.* Nat Rev Cancer, 2006. **6**(3): p. 193-203.

71. Ichimaru, M., *[Atomic bomb and leukemia].* Rinsho Ketsueki, 1979. **20**(6): p. 574-89.

72. Tomonaga, M., *Leukaemia in Nagasaki atomic bomb survivors from 1945 through 1959.* Bull World Health Organ, 1962. **26**: p. 619-31.

73. Investigators, U.K.C.C.S., *The United Kingdom Childhood Cancer Study of exposure to domestic sources of ionising radiation: 1: radon gas.* Br J Cancer, 2002. **86**(11): p. 1721-6.

74. Investigators, U.K.C.C.S., *The United Kingdom Childhood Cancer Study of exposure to domestic sources of ionising radiation: 2: gamma radiation.* Br J Cancer, 2002. **86**(11): p. 1727-31.

75. Savitz, D.A. and K.W. Andrews, *Review of epidemiologic evidence on benzene and lymphatic and hematopoietic cancers.* Am J Ind Med, 1997. **31**(3): p. 287-95.

76. Meinert, R., et al., *Leukemia and non-Hodgkin's lymphoma in childhood and exposure to pesticides: results of a register-based case-control study in Germany.* Am J Epidemiol, 2000. **151**(7): p. 639-46; discussion 647-50.

77. Greaves, M.F., et al., *Geographical distribution of acute lymphoblastic leukaemia subtypes: second report of the collaborative group study.* Leukemia, 1993. **7**(1): p. 27-34.

78. Hrusak, O., et al., *Acute lymphoblastic leukemia incidence during socioeconomic transition: selective increase in children from 1 to 4 years.* Leukemia, 2002. **16**(4): p. 720-5.

79. Pang, D., et al., *Parental smoking and childhood cancer: results from the United Kingdom Childhood Cancer Study.* British Journal of Cancer, 2003. **88**(3): p. 373-381.

80. Kinlen, L.J., *Infective cause of childhood leukaemia.* Lancet, 1989. **1**(8634): p. 378-9.

81. Greaves, M.F., *Speculations on the cause of childhood acute lymphoblastic leukemia.* Leukemia, 1988. **2**(2): p. 120-5.

82. Smith, M., *Considerations on a possible viral etiology for B-precursor acute lymphoblastic leukemia of childhood.* J Immunother, 1997. **20**(2): p. 89-100.

83. Smith, M.A., et al., *Evidence that childhood acute lymphoblastic leukemia is associated with an infectious agent linked to hygiene conditions.* Cancer Causes Control, 1998. **9**(3): p. 285-98.

84. Yazdanbakhsh, M., P.G. Kremsner, and R. van Ree, *Allergy, parasites, and the hygiene hypothesis.* Science, 2002. **296**(5567): p. 490-4.

85. Kramer, U., et al., *Age of entry to day nursery and allergy in later childhood.* Lancet, 1999. **353**(9151): p. 450-4.

86. Gilham, C., et al., *Day care in infancy and risk of childhood acute lymphoblastic leukaemia: findings from UK case-control study.* BMJ, 2005. **330**(7503): p. 1294.

87. Kamper-Jorgensen, M., et al., *Childcare in the first 2 years of life reduces the risk of childhood acute lymphoblastic leukemia.* Leukemia, 2008. **22**(1): p. 189-93.

88. Schmiegelow, K., et al., *Etiology of common childhood acute lymphoblastic leukemia: the adrenal hypothesis.* Leukemia, 2008. **22**(12): p. 2137-41.

89. McNally, R.J. and T.O. Eden, *An infectious aetiology for childhood acute leukaemia: a review of the evidence.* Br J Haematol, 2004. **127**(3): p. 243-63.

90. O'Neil, J. and A.T. Look, *Mechanisms of transcription factor deregulation in lymphoid cell transformation.* Oncogene, 2007. **26**(47): p. 6838-49.

91. Georgopoulos, K., et al., *The Ikaros gene is required for the development of all lymphoid lineages.* Cell, 1994. **79**(1): p. 143-56.

92. Mullighan, C.G., et al., *BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros.* Nature, 2008. **453**(7191): p. 110-4.

93. Blake, J.A. and M.R. Ziman, *Pax genes: regulators of lineage specification and progenitor cell maintenance.* Development, 2014. **141**(4): p. 737-51.

94. Urbanek, P., et al., *Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP.* Cell, 1994. **79**(5): p. 901-12.

95. Adams, B., et al., *Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis.* Genes Dev, 1992. **6**(9): p. 1589-607.

96. Cobaleda, C., et al., *Pax5: the guardian of B cell identity and function.* Nat Immunol, 2007. **8**(5): p. 463-70.

97. Nutt, S.L., et al., *Commitment to the B-lymphoid lineage depends on the transcription factor Pax5.* Nature, 1999. **401**(6753): p. 556-62.

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

99. Matthias, P. and A.G. Rolink, *Transcriptional networks in developing and mature B cells.* Nat Rev Immunol, 2005. **5**(6): p. 497-508.

100. Rolink, A.G., et al., *In vitro and in vivo plasticity of Pax5-deficient pre-B I cells.* Immunol Lett, 2002. **82**(1-2): p. 35-40.

101. Fuxa, M. and M. Busslinger, *Reporter gene insertions reveal a strictly B lymphoid-specific expression pattern of Pax5 in support of its B cell identity function.* J Immunol, 2007. **178**(5): p. 3031-7.

102. O'Riordan, M. and R. Grosschedl, *Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A.* Immunity, 1999. **11**(1): p. 21-31.

103. Holmes, M.L., C. Pridans, and S.L. Nutt, *The regulation of the B-cell gene expression programme by Pax5.* Immunol Cell Biol, 2008. **86**(1): p. 47-53.

104. Cobaleda, C., W. Jochum, and M. Busslinger, *Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors.* Nature, 2007. **449**(7161): p. 473-7.

105. Pridans, C., et al., *Identification of Pax5 target genes in early B cell differentiation.* J Immunol, 2008. **180**(3): p. 1719-28.

106. Schebesta, A., et al., *Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration, and immune function.* Immunity, 2007. **27**(1): p. 49-63.

107. Iacobucci, I., et al., *The PAX5 gene is frequently rearranged in BCR-ABL1-positive acute lymphoblastic leukemia but is not associated with outcome. A report on behalf of the GIMEMA Acute Leukemia Working Party.* Haematologica, 2010. **95**(10): p. 1683-90.

108. Familiades, J., et al., *PAX5 mutations occur frequently in adult B-cell progenitor acute lymphoblastic leukemia and PAX5 haploinsufficiency is associated with BCR-ABL1 and TCF3-PBX1 fusion genes: a GRAALL study.* Leukemia, 2009. **23**(11): p. 1989-98.

109. Paulsson, K., et al., *Microdeletions are a general feature of adult and adolescent acute lymphoblastic leukemia: Unexpected similarities with pediatric disease.* Proc Natl Acad Sci U S A, 2008. **105**(18): p. 6708-13.

110. Garvie, C.W., J. Hagman, and C. Wolberger, *Structural studies of Ets-1/Pax5 complex formation on DNA.* Mol Cell, 2001. **8**(6): p. 1267-76.

111. Nebral, K., et al., *Incidence and diversity of PAX5 fusion genes in childhood acute lymphoblastic leukemia.* Leukemia, 2009. **23**(1): p. 134-43.

112. Coyaud, E., et al., *Wide diversity of PAX5 alterations in B-ALL: a Groupe Francophone de Cytogenetique Hematologique study.* Blood, 2010. **115**(15): p. 3089-97.

113. Strehl, S., et al., *PAX5/ETV6 fusion defines cytogenetic entity dic(9;12)(p13;p13).* Leukemia, 2003. **17**(6): p. 1121-3.

114. Gupta, S.K., et al., *Gene copy number alteration profile and its clinical correlation in B-cell acute lymphoblastic leukemia.* Leuk Lymphoma, 2016: p. 1-10.

115. Mullighan, C.G., *The genomic landscape of acute lymphoblastic leukemia in children and young adults.* Hematology Am Soc Hematol Educ Program, 2014. **2014**(1): p. 174-80.

116. Babon, J.J., et al., *The molecular regulation of Janus kinase (JAK) activation.* Biochem J, 2014. **462**(1): p. 1-13.

117. Vainchenker, W. and S.N. Constantinescu, *JAK/STAT signaling in hematological malignancies.* Oncogene, 2013. **32**(21): p. 2601-13.

118. Gaikwad, A., et al., *Prevalence and clinical correlates of JAK2 mutations in Down syndrome acute lymphoblastic leukaemia.* Br J Haematol, 2009. **144**(6): p. 930-2.

119. Jeong, E.G., et al., *Somatic mutations of JAK1 and JAK3 in acute leukemias and solid cancers.* Clin Cancer Res, 2008. **14**(12): p. 3716-21.

120. Johannessen, C.M., et al., *COT drives resistance to RAF inhibition through MAP kinase pathway reactivation.* Nature, 2010. **468**(7326): p. 968-72.

121. Czerny, T. and M. Busslinger, *DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5).* Mol Cell Biol, 1995. **15**(5): p. 2858-71.

122. Pfisterer, U., et al., *Direct conversion of human fibroblasts to dopaminergic neurons.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10343-8.

123. Linka, R.M., et al., *Loss-of-function mutations within the IL-2 inducible kinase ITK in patients with EBV-associated lymphoproliferative diseases.* Leukemia, 2012. **26**(5): p. 963-71.

124. Lin, Z., et al., *In vivo antigen-driven plasmablast enrichment in combination with antigen-specific cell sorting to facilitate the isolation of rare monoclonal antibodies from human B cells.* Nat Protoc, 2014. **9**(7): p. 1563-77.

125. Fisher, S., et al., *A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries.* Genome Biol, 2011. **12**(1): p. R1.

126. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform.* Bioinformatics, 2010. **26**(5): p. 589-95.

127. Kersey, P.J., et al., *Ensembl Genomes 2016: more genomes, more complexity.* Nucleic Acids Res, 2016. **44**(D1): p. D574-80.

128. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

129. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

130. DePristo, M.A., et al., *A framework for variation discovery and genotyping using next-generation DNA sequencing data.* Nat Genet, 2011. **43**(5): p. 491-8.

131. McLaren, W., et al., *Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor.* Bioinformatics, 2010. **26**(16): p. 2069-70.

132. Flicek, P., et al., *Ensembl 2014.* Nucleic Acids Res, 2014. **42**(Database issue): p. D749-55.

133. Ng, P.C. and S. Henikoff, *Predicting deleterious amino acid substitutions.* Genome Res, 2001. **11**(5): p. 863-74.

134. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.

135. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.* Nat Biotechnol, 2013. **31**(3): p. 213-9.

136. Forbes, S.A., et al., *COSMIC: exploring the world's knowledge of somatic mutations in human cancer.* Nucleic Acids Res, 2015. **43**(Database issue): p. D805-11.

137. Smedley, D., et al., *The BioMart community portal: an innovative alternative to large, centralized data repositories.* Nucleic Acids Res, 2015. **43**(W1): p. W589-98.

138. Boeva, V., et al., *Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data.* Bioinformatics, 2012. **28**(3): p. 423-5.

139. Olerup, O. and H. Zetterquist, *HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation.* Tissue Antigens, 1992. **39**(5): p. 225-35.

140. Sambrook, J. and D.W. Russell, *Molecular cloning : a laboratory manual*. 3rd ed ed. 2001, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. 3 volumes.

141. Sato, T., et al., *Functional analysis of JAK3 mutations in transient myeloproliferative disorder and acute megakaryoblastic leukaemia accompanying Down syndrome.* Br J Haematol, 2008. **141**(5): p. 681-8.

142. Loken, M.R., et al., *Flow cytometric analysis of normal B lymphoid development.* Pathol Immunopathol Res, 1988. **7**(5): p. 357-70.

143. Chambon-Pautas, C., et al., *High-resolution allelotype analysis of childhood B-lineage acute lymphoblastic leukemia.* Leukemia, 1998. **12**(7): p. 1107-13.

144. Takeuchi, S., et al., *Homozygous deletions at 9p21 in childhood acute lymphoblastic leukemia detected by microsatellite analysis.* Leukemia, 1997. **11**(10): p. 1636-40.

145. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

146. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-291.

147. Flanagan, S.E., A.M. Patch, and S. Ellard, *Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations.* Genet Test Mol Biomarkers, 2010. **14**(4): p. 533-7.

148. Case, M., et al., *Mutation of genes affecting the RAS pathway is common in childhood acute lymphoblastic leukemia.* Cancer Res, 2008. **68**(16): p. 6803-9.

149. Mullighan, C.G., et al., *JAK mutations in high-risk childhood acute lymphoblastic leukemia.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9414-8.

150. Shah, S., et al., *A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia.* Nat Genet, 2013. **45**(10): p. 1226-31.

151. Ward, G., *The infective theory of acute leukaemia.* Br J Child 1917. **Dis 14, 11**.

152. Nutt, S.L., C. Thevenin, and M. Busslinger, *Essential functions of Pax-5 (BSAP) in pro-B cell development.* Immunobiology, 1997. **198**(1-3): p. 227-35.

153. Martin-Lorenzo, A., et al., *Infection Exposure is a Causal Factor in B-cell Precursor Acute Lymphoblastic Leukemia as a Result of Pax5-Inherited Susceptibility.* Cancer Discov, 2015. **5**(12): p. 1328-43.

154. Czerny, T., G. Schaffner, and M. Busslinger, *DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site.* Genes Dev, 1993. **7**(10): p. 2048-61.

155. Zwollo, P., et al., *The Pax-5 gene is alternatively spliced during B-cell development.* J Biol Chem, 1997. **272**(15): p. 10160-8.

156. Bergmann, A.K., et al., *Recurrent mutation of JAK3 in T-cell prolymphocytic leukemia.* Genes Chromosomes Cancer, 2014. **53**(4): p. 309-16.

157. Degryse, S., et al., *JAK3 mutants transform hematopoietic cells through JAK1 activation, causing T-cell acute lymphoblastic leukemia in a mouse model.* Blood, 2014. **124**(20): p. 3092-100.

158. Schmiegelow, K., et al., *Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia.* Leukemia, 2010. **24**(2): p. 345-54.

159. Downing, J.R., et al., *The Pediatric Cancer Genome Project.* Nat Genet, 2012. **44**(6): p. 619-22.

160. Smith, M.L., et al., *Mutation of CEBPA in familial acute myeloid leukemia.* N Engl J Med, 2004. **351**(23): p. 2403-7.

161. Hahn, C.N., et al., *Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia.* Nat Genet, 2011. **43**(10): p. 1012-7.

162. Sanyanusin, P., et al., *Mutation of the PAX2 gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux.* Nat Genet, 1995. **9**(4): p. 358-64.

163. Kojima, K., J. Ishizawa, and M. Andreeff, *Pharmacological activation of wild-type p53 in the therapy of leukemia.* Exp Hematol, 2016. **44**(9): p. 791-8.

164. Vassilev, L.T., et al., *In vivo activation of the p53 pathway by small-molecule antagonists of MDM2.* Science, 2004. **303**(5659): p. 844-8.

165. Bos, J.L., *ras oncogenes in human cancer: a review.* Cancer Res, 1989. **49**(17): p. 4682-9.

166. Heltemes-Harris, L.M., et al., *Sleeping Beauty transposon screen identifies signaling modules that cooperate with STAT5 activation to induce B-cell acute lymphoblastic leukemia.* Oncogene, 2016. **35**(26): p. 3454-64.

167. Crouch, S., et al., *Infectious illness in children subsequently diagnosed with acute lymphoblastic leukemia: modeling the trends from birth to diagnosis.* Am J Epidemiol, 2012. **176**(5): p. 402-8.

168. Ying, H., et al., *Regulation of mouse CD72 gene expression during B lymphocyte development.* J Immunol, 1998. **161**(9): p. 4760-7.

169. Nutt, S.L., et al., *Identification of BSAP (Pax-5) target genes in early B-cell development by loss- and gain-of-function experiments.* EMBO J, 1998. **17**(8): p. 2319-33.

170. von Freeden-Jeffry, U., et al., *Lymphopenia in interleukin (IL)-7 gene-deleted mice identifies IL-7 as a nonredundant cytokine.* J Exp Med, 1995. **181**(4): p. 1519-26.

171. Peschon, J.J., et al., *Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice.* J Exp Med, 1994. **180**(5): p. 1955-60.

172. Corfe, S.A. and C.J. Paige, *The many roles of IL-7 in B cell development; mediator of survival, proliferation and differentiation.* Semin Immunol, 2012. **24**(3): p. 198-208.

173. Lu, L., P. Chaudhury, and D.G. Osmond, *Regulation of cell survival during B lymphopoiesis: apoptosis and Bcl-2/Bax content of precursor B cells in bone marrow of mice with altered expression of IL-7 and recombinase-activating gene-2.* J Immunol, 1999. **162**(4): p. 1931-40.

174. Oliver, P.M., et al., *Loss of Bim allows precursor B cell survival but not precursor B cell differentiation in the absence of interleukin 7.* J Exp Med, 2004. **200**(9): p. 1179-87.

175. Clark, M.R., et al., *Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling.* Nat Rev Immunol, 2014. **14**(2): p. 69-80.

176. Heltemes-Harris, L.M., et al., *The role of STAT5 in the development, function, and transformation of B and T lymphocytes.* Ann N Y Acad Sci, 2011. **1217**: p. 18-31.

177. Losdyck, E., et al., *Distinct Acute Lymphoblastic Leukemia (ALL)-associated Janus Kinase 3 (JAK3) Mutants Exhibit Different Cytokine-Receptor Requirements and JAK Inhibitor Specificities.* J Biol Chem, 2015. **290**(48): p. 29022-34.

178. Borkhardt, A., et al., *Incidence and clinical relevance of TEL/AML1 fusion genes in children with acute lymphoblastic leukemia enrolled in the German and Italian multicenter therapy trials. Associazione Italiana Ematologia Oncologia Pediatrica and the Berlin-Frankfurt-Munster Study Group.* Blood, 1997. **90**(2): p. 571-7.

179. Mori, H., et al., *Chromosome translocations and covert leukemic clones are generated during normal fetal development.* Proc Natl Acad Sci U S A, 2002. **99**(12): p. 8242-7.

180. Jung, S.H., et al., *The Drosophila lymph gland as a developmental model of hematopoiesis.* Development, 2005. **132**(11): p. 2521-33.

181. Teittinen, K.J., et al., *The zebrafish as a tool in leukemia research.* Leuk Res, 2012. **36**(9): p. 1082-8.

182. Mouse Genome Sequencing, C., et al., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.

183. Ghia, P., et al., *B-cell development: a comparison between mouse and man.* Immunol Today, 1998. **19**(10): p. 480-5.

184. Mestas, J. and C.C. Hughes, *Of mice and not men: differences between mouse and human immunology.* J Immunol, 2004. **172**(5): p. 2731-8.

185. Hauer, J., et al., *Genetically engineered mouse models of human B-cell precursor leukemias.* Cell Cycle, 2014. **13**(18): p. 2836-46.

186. Auer, F., et al., *Inherited susceptibility to pre B-ALL caused by germline transmission of PAX5 c.547G>A.* Leukemia, 2014. **28**(5): p. 1136-8.

187. Cooper, D.N., et al., *Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease.* Hum Genet, 2013. **132**(10): p. 1077-130.

188. Liu, G.J., et al., *Pax5 loss imposes a reversible differentiation block in B-progenitor acute lymphoblastic leukemia.* Genes Dev, 2014. **28**(12): p. 1337-50.

189. Puel, A., et al., *Defective IL7R expression in T(-)B(+)NK(+) severe combined immunodeficiency.* Nat Genet, 1998. **20**(4): p. 394-7.

190. Giliani, S., et al., *Interleukin-7 receptor alpha (IL-7Ralpha) deficiency: cellular and molecular bases. Analysis of clinical, immunological, and molecular features in 16 novel patients.* Immunol Rev, 2005. **203**: p. 110-26.

191. Heltemes-Harris, L.M., et al., *Ebf1 or Pax5 haploinsufficiency synergizes with STAT5 activation to initiate acute lymphoblastic leukemia.* J Exp Med, 2011. **208**(6): p. 1135-49.

192. Dang, J., et al., *PAX5 is a tumor suppressor in mouse mutagenesis models of acute lymphoblastic leukemia.* Blood, 2015. **125**(23): p. 3609-17.

193. Bleckmann, K. and M. Schrappe, *Advances in therapy for Philadelphia-positive acute lymphoblastic leukaemia of childhood and adolescence.* Br J Haematol, 2016. **172**(6): p. 855-69.

194. Vicente-Duenas, C., et al., *Tumoral stem cell reprogramming as a driver of cancer: Theory, biological models, implications in cancer therapy.* Semin Cancer Biol, 2015. **32**: p. 3-9.

195. Notta, F., et al., *Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells.* Nature, 2011. **469**(7330): p. 362-7.

196. Mullighan, C.G., et al., *Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia.* Science, 2008. **322**(5906): p. 1377-80.

9 Acknowledgements

First of all I want to thank Prof. Dr. Arndt Borkhardt for giving me the opportunity to complete this PhD thesis after 4 inspiring years of successful research in the KMT lab. I really appreciate your immense support and the passion for research that you communicate to all of us. I am especially grateful for the chance to work in joint projects with outstanding collaborators from all over the world.

Moreover, I would like to thank my second examiner, Prof. Dr. Hermann Aberle for mentoring my PhD thesis. Thank you for always taking the time to discuss my work and for your excellent expertise in the field of biology, which greatly helped me to refine my thesis.

Special thanks go to my group leader, PD Dr. Julia Hauer, for her guidance, ideas and endless support during these years. You were able to confer a comfortable, independent working atmosphere, while never losing our focus. The success of this project was only achieved by your immense input and vision. I could not have imagined having a better PI.

Furthermore, I am very grateful to Prof. Dr. Isidro Sánchez-García and his group, who gave me the opportunity to work on tremendously successful projects and to spend a fantastic time in Salamanca. Your kindliness and support are very much appreciated.

I also want to thank all members of the KMT-Lab, who greatly improved my research and created a magnificent working atmosphere. We are not just colleagues, but friends!

Special thanks go to:

Stewart, Daniel and Julia for being a huge help in the correction of this thesis;

Deborah, for her help in the lab and for her support during the whole time;

Silke, for training us to be independent PhD students ;) and for your hard work;

Julia, for your support and fruitful discussions and for being a really good friend;

Daniel, for your ideas and support (especially in math and german grammar!);

…and to so many more who supported me during all this time (including Cyrill, Daniel, Sanil, Ute, Andrea, Jasmin, Michael, Kati, René…)

Last but not least I want to express the dearest thanks to my parents, who supported me immensely during my whole life and always helped me to achieve this goal. Finally, I am very grateful for my boyfriend Basti, who is my source of emotional support and who is willing to put his own needs last.

10 Appendix

10.1 Abbreviations

ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
AlloHSCT	Allogeneic hematopoietic stem cell transplantation
APS	Ammonium Persulfate
B-ALL	B cell acute lymphoblastic leukemia
BCP-ALL	B cell precursor acute lymphoblastic leukemia
BCR	B cell receptor
BM	Bone marrow
Bp	Base pair
Kb	Kilo base pair
BSA	Bovine serum albumin
CDKN2A	Cyclin-Dependent Kinase Inhibitor 2A
CDNA	Complementary DNA
CDS	Coding sequence
CEBP	CCAAT enhancer-binding protein
CLL	Chronic lymphoblastic leukemia
CLP	Common lymphoid progenitor
CML	Chronic myeloid leukemia
CMP	Common myeloid progenitor
CNS	Central nervous system
CNV	Copy number variation/variant
CRLF2	Cytokine receptor-like factor 2
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNTP	Deoxyribonucleotide triphosphate
DS	Down syndrome
DPBS	Dulbecco's phosphate buffered saline
ETP	Early T cell precursor
E.coli	*Escherichia coli*
FBS	Fetal bovine serum
FACS	Fluorescence-activated cell sorting
GEMMs	Genetically engineered mouse models

GWASs	Genome wide association studies
HEK293	Human embryonic kidney 293
HSC	Hematopoietic stem cell
HLA	Human leukocyte antigen
IAMP21	Intrachromosomal amplification of chromosome 21
IGH	Immunoglobulin heavy
IL-3	Interleukin-3
IL-7	Interleukin-7
IL-7R	IL-7-Receptor
Indel	Insertion and/or deletion
JAK	Janus kinase
LOH	Loss of heterozygosity
LYL1	Lymphoblastic leukemia-derived sequence 1
MPC	Magnetic particle concentrator
MTOR	Mammalian target of rapamycin
MAPK	Mitogen-activated protein kinase
MLL	Mixed lineage leukemia
MHV	Mouse hepatitis virus
MNV	Murine norovirus
NGS	Next generation sequencing
PAX5	Paired Box 5
PE	Paired-end
Pax5$^{+/-}$	*Pax5* heterozygous mice
PB	Peripheral blood
PBMCs	Peripheral blood mononuclear cells
Ph	Philadelphia
PI3K	Phosphatidylinositol-3'-kinase
PEG	Polyethylenglycol
PCR	Polymerase chain reaction
PMNCs	Polymorphonuclear cells
PI	Propidium iodide
RT-PCR	Real-time polymerase chain reaction
RNA	Ribonucleic acid
Rh	Rhesus
SSP	Sequence specific primer
SBS	Sequencing-by-synthesis
STAT	Signal transducers and activators of transcription

SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SPF	Specific pathogen free
TAL	T cell acute lymphocytic leukemia
TLX	T cell Leukemia Homeobox
TCR	T cell receptor
TEMED	Tetramethylethylendiamin
Taq	*Thermus aquaticus*
TKI	Tyrosine kinase inhibitor
UTR	Untranslated region
WES	Whole exome sequencing
WB	Western Blot
WT	Wildtype

10.2 Gene Nomenclature

Human gene   =   all uppercase, italic (e.g. *PAX5*)

Human protein   =   all uppercase (e.g. PAX5)

Murine gene   =   first letter uppercase, italic (e.g. *Pax5*)

Murine Protein   =   first letter uppercase (e.g. Pax5)

10.3 Supplementary Tables

Sample ID	Phenotype	Material	Preparation	cells/µl
70560	BCP-ALL, 9p deletion	BM	Buffy Coat	52100
80591	BCP-ALL, 9p deletion	BM	Buffy Coat	28700
90161	BCP-ALL, 9p deletion	BM	Buffy Coat	40600
90530	BCP-ALL, 9p deletion	BM	Buffy Coat	94000
100110	BCP-ALL, 9p deletion	BM	Buffy Coat	58200
100174	BCP-ALL, 9p deletion	PB	Buffy Coat	33600
100342	BCP-ALL, 9p deletion	BM	Buffy Coat	237600
100444	BCP-ALL, 9p deletion	BM	Buffy Coat	108000
100533	BCP-ALL, 9p deletion	BM	Buffy Coat	60400
110127	BCP-ALL, 9p deletion	BM	Buffy Coat	59100
110182	BCP-ALL, 9p deletion	PB	Buffy Coat	42400
110251	BCP-ALL, 9p deletion	PB	Buffy Coat	74000
110275	BCP-ALL, 9p deletion	BM	Buffy Coat	185000
110285	BCP-ALL, 9p deletion	PB	Buffy Coat	74800
110302	BCP-ALL, 9p deletion	BM	Buffy Coat	132800
120005	BCP-ALL, 9p deletion	BM	Buffy Coat	94800
120179	BCP-ALL, 9p deletion	BM	Buffy Coat	100000
120185	BCP-ALL, 9p deletion	BM	Buffy Coat	30200
120192	BCP-ALL, 9p deletion	BM	Buffy Coat	23100
120279	BCP-ALL, 9p deletion	BM	Buffy Coat	109600
120351	BCP-ALL, 9p deletion	BM	Buffy Coat	61200
120379	BCP-ALL, 9p deletion	PB	Buffy Coat	9300
120424	BCP-ALL, 9p deletion	BM	Buffy Coat	42400

Supplementary Table 1: BCP-ALL cohort harboring aberrations on chromosome 9p, which was screened for the presence of *PAX5* c.547G>A.

Timeline	CONVENTIONAL FACILITY HEALTH MONITORING REPORT				SPF FACILITY HEALTH MONITORING REPORT
	When mice are moved to the conventional facility	1 year later	2 years later		
VIRUSES					
MAD (Adenovirus type 1 and 2)	NEG	NEG	NEG		NEG
MCMV (Mouse cytomegalovirus)	NEG	NEG	NEG		NEG
LCMV (Lymph choriomeningitis)	NEG	NEG	NEG		NEG
Ectromelia	NEG	NEG	NEG		NEG
EDIM	NEG	NEG	NEG		NEG
Hantaan virus	NEG	NEG	NEG		NEG
MHV (Mouse hepatitis virus)	**POS**	**POS**	**POS**		NEG
MVM (Minute virus)	NEG	NEG	NEG		NEG
MPV (Parvovirus type 1 and 2)	NEG	NEG	NEG		NEG
PVM (Pneumonia virus)	NEG	NEG	NEG		NEG
Reovirus type 3	NEG	NEG	NEG		NEG
Sendai	NEG	NEG	NEG		NEG
TMEV (Theiler´s murine encephalomyel)	NEG	NEG	NEG		NEG
K virus	NEG	NEG	NEG		NEG
Polyoma virus	NEG	NEG	NEG		NEG
MNV (Norovirus)	**POS**	**POS**	**POS**		NEG
BACTERIA					
Bordetella bronchiseptica	NEG	NEG	NEG		NEG
Car-bacillus	NEG	NEG	NEG		NEG
Citrobacter rodentium	NEG	NEG	NEG		NEG
Clostridium piliforme (Tyzzer´ disease)	NEG	NEG	NEG		NEG
Corynebacterium kutscheri	NEG	NEG	NEG		NEG
Mycoplasma pulmonis	NEG	NEG	NEG		NEG
Pasteurella pneumotropica	NEG	NEG	NEG		NEG
Pasteurella multocida	NEG	NEG	NEG		NEG
Salmonella spp	NEG	NEG	NEG		NEG
Streptobacillus moniliformis	NEG	NEG	NEG		NEG
Streptococci beta hemolíticos (A,G)	NEG	NEG	NEG		NEG
Streptococcus pneumoniae	NEG	NEG	NEG		NEG
Helicobacter spp	**POS**	**POS**	**POS**		NEG
PARASITES					
Ectoparasites	NEG	NEG	NEG		NEG
Helminth (A.tetraptrera, Syphacia spp)	NEG	NEG	NEG		NEG
Cestods (Hymenolepis spp)	NEG	NEG	NEG		NEG
Eimeria spp	NEG	NEG	NEG		NEG
Entamoeba muris	NEG	NEG	NEG		NEG
Giardia spp	NEG	NEG	NEG		NEG
Trichomonas muris	**POS**	**POS**	NEG		NEG
Spironucleus spp	NEG	NEG	NEG		NEG
Encephalitozoon cuniculi	NEG	NEG	NEG		NEG
NECROPSY	No gross lesion.	No gross lesion.	No gross lesion.		No gross lesion.
NEG: Negative					
POS: Positive					

Supplementary Table 2: Health monitoring report from the conventional facility (left) and SPF facility (right) of the Instituto de Biologia Molecular y Celular del Cancer (IBMCC) in Salamanca (Spain). Positive tested pathogens are marked in red.

10.4 Figure Directory

10.5 Table Directory

10.6 Supplementary Table Directory

Curriculum Vitae

<u>Personal data</u>

Name:	Franziska Auer
Address:	Leichlingerstr. 19, 40591 Düsseldorf
Date of Birth:	31st of October 1986
Nationality:	German

<u>International Experiences</u>

10/2003 - 7/2004:	Northwestern High School, Connecticut, USA
5/2015 – 7/2015:	Cooperation work in Dr. Isidro Sánchez-García's laboratory in Salamanca, Spain

<u>Education</u>

10/2006 – 10/2009:	Technical University Munich, Molecular Biotechnology, Bachelor of Science
5/2010 – 5/2012:	Technical University Munich, Molecular Biotechnology, Master of Science
10/2012 – 11/2016	PhD thesis in the Department of Pediatric Oncology, Hematology and Clinical Immunology, Heinrich-Heine University Düsseldorf (Supervisor: Prof. Dr. Arndt Borkhardt)

<u>Internships</u>

10/2009 – 4/2010: (I.A.Z.).	Internship in the cell culture institute Munich
11/2011 – 06/2012:	Assistant student at the chair of Animal Hygiene (Technical University Munich)

<u>Languages</u>

German (native language)

English (very good knowledge)

Spanish (good knowledge)

French (basic knowledge)

Publications

Pre-Submission

Auer, F., et al., *Activation induced cytidine deaminase (Aid) acts as a gate keeper in pro-B cells and prevents pB-ALL*.

Martín-Lorenzo, A., Auer, F., Chan, L.N., et al., *Pax5 drives the metabolic shift in the development of pB-ALL as a result of BCR-ABLp190 susceptibility*.

Published

1. Auer, F., et al., *GEMMs addressing Pax5 loss-of-function in childhood pB-ALL*. Eur J Med Genet, 2015.

2. Martin-Lorenzo, A., et al., *Infection Exposure Is a Causal Factor in B-cell Precursor Acute Lymphoblastic Leukemia as a Result of Pax5-Inherited Susceptibility*. Cancer Discov, 2015. **5**(12): p. 1328-43.

3. Vicente-Duenas, C., et al., *Tumoral stem cell reprogramming as a driver of cancer: Theory, biological models, implications in cancer therapy*. Semin Cancer Biol, 2015. **32**: p. 3-9.

4. Auer, F., et al., *Inherited susceptibility to pre B-ALL caused by germline transmission of PAX5 c.547G>A*. Leukemia, 2014. **28**(5): p. 1136-8.

5. Kaspar, D., et al., *Temperature- and nitrogen source-dependent regulation of GlnR target genes in Listeria monocytogenes*. FEMS Microbiol Lett, 2014. **355**(2): p. 131-41.

Affirmation

Hereby, I declare on oath that I composed this dissertation independently by myself. I used only the references and resources indicated in this thesis. With the exception of such quotations, the work presented in this thesis is my own. I have accredited all the sources of help. This PhD thesis was never submitted or presented in a similar form to any other institution or examination board. I have not undertaken a doctoral examination without success so far.

Düsseldorf, 19th of October 2016

_____

Franziska Auer