

**Establishment of a Novel Technique Termed
GIPFEL to Determine the Frequency of
ETV6-RUNX1 Fusions in Healthy Newborns
and
Analysis of Cooperating Oncogenic Lesions
Leading to *ETV6-RUNX1* Positive Childhood
Leukemia**

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Daniel Schäfer
aus Hachenburg

Düsseldorf, Juni 2017

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

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Arndt Borkhardt

Korreferent: Prof. Dr. William F. Martin

Tag der mündlichen Prüfung: 07.08.2017

Table of Contents

Summary.....	7
Zusammenfassung	9
1 Introduction	11
1.1 Pediatric Leukemia	11
1.2 Acute Lymphoblastic Leukemia.....	13
1.2.1 Epidemiology	13
1.2.2 Genetics of ALL	14
1.2.3 Clinical Aspects.....	16
1.2.3.1 Symptoms and Diagnosis	16
1.2.3.2 Treatment.....	18
1.3 Translocations	21
1.3.1 <i>ETV6-RUNX1</i>	22
1.3.2 <i>TCF3-PBX1</i>	24
1.4 Methods for Detection of Translocations.....	27
2 Aim of the Thesis	31
3 Materials	32
3.1 Media / Cell Culture Reagents	32
3.2 Cell Lines.....	32
3.3 Patients.....	33
3.4 Newborns	35
3.5 <i>ETV6-RUNX1</i> ⁺ Mice	35
3.6 Enzymes	38
3.7 Chemicals.....	38
3.8 Kits, Size Markers, and Other Materials	39
3.9 Oligonucleotides.....	39
3.10 Software and Hardware.....	42
4 Methods	43
4.1 GIPFEL.....	43
4.1.1 Cell Culture	44
4.1.2 DNA Isolation	45
4.1.2.1 DNA Isolation from Cell Lines.....	45
4.1.2.2 DNA Isolation from Umbilical Cord Blood	45
4.1.3 Whole Genome Amplification.....	47
4.1.4 Determination of Nucleic Acid Concentration	47

4.1.5	Restriction Enzyme Digest	47
4.1.6	Purification of DNA after Enzymatic Digests	48
4.1.7	Ligation.....	49
4.1.8	Exonuclease Digest.....	49
4.1.9	Alcohol Precipitation.....	50
4.1.9.1	Ethanol Precipitation	50
4.1.9.2	Isopropanol Precipitation	50
4.1.10	PCR.....	51
4.1.10.1	Primer Design	51
4.1.10.2	Pre-amplification PCR.....	51
4.1.10.3	Real-Time PCR.....	52
4.1.11	Gel Electrophoresis.....	54
4.1.12	Gel Extraction	54
4.1.13	Sanger Sequencing	55
4.1.14	Reverse Transcription Real-Time PCR.....	55
4.2	Additional Methods	56
4.2.1	Cloning	56
4.2.1.1	Generation of Chemically Competent Cells.....	56
4.2.1.2	Transformation	57
4.2.1.3	Plasmid Preparation.....	58
4.2.2	Next Generation Sequencing	58
4.2.2.1	Library Preparation	59
4.2.2.2	Cluster Generation	59
4.2.2.3	Sequencing	60
4.2.2.4	Data Analysis.....	60
4.2.3	Genome-wide Methylation Analysis	61
5	Results	63
5.1	Establishment of the GIPFEL Method.....	63
5.1.1	Computational Groundwork	63
5.1.2	Proof of Principle.....	66
5.1.3	Optimization of the Method	69
5.1.3.1	Duration of Restriction Enzyme Digest.....	69
5.1.3.2	Clean-up of Restriction Enzyme Digest	70
5.1.3.3	Ligation Temperature.....	72
5.1.3.4	Duration of Ligation	73
5.1.3.5	Clean-up of Exonuclease Digest.....	74

5.1.3.6	Whole Genome Amplification	75
5.1.3.7	Multiplexing of Primers.....	76
5.1.3.8	Optimized GIPFEL Protocol	77
5.1.4	Final Results of the GIPFEL Establishment Phase	79
5.1.4.1	Sensitivity of GIPFEL	79
5.1.4.2	Accuracy and Specificity.....	81
5.2	Screening of Umbilical Cord Blood from Healthy Newborns	86
5.2.1	Adaption of the GIPFEL protocol.....	86
5.2.1.1	CD19 ⁺ Enrichment	86
5.2.1.2	Whole Genome Amplification	88
5.2.1.3	384 Well Real-Time PCR	89
5.2.2	Results from Cord Blood Screening for <i>ETV6-RUNX1</i>	90
5.2.3	Validation of Cord Blood Screening Results	95
5.3	<i>ETV6-RUNX1</i> ⁺ Mouse Model	96
5.3.1	DNA Methylation Analysis	97
5.3.2	Comparison of Murine and Human Mutations	99
6	Discussion.....	103
6.1	Establishment of the GIPFEL Method.....	103
6.2	Optimization of GIPFEL.....	104
6.3	Advantages and Limitations of GIPFEL	107
6.3.1	Advantages	107
6.3.2	Limitations	108
6.4	Frequency of <i>ETV6-RUNX1</i> in the Healthy Population.....	108
6.5	Characterization of <i>ETV6-RUNX1</i> Positive Leukemias	111
6.5.1	<i>ETV6-RUNX1</i> ⁺ Mouse Model	111
6.5.2	DNA Methylation.....	112
6.5.3	Characteristics of Human and Murine <i>ETV6-RUNX1</i> ⁺ ALL	113
7	Conclusion and Outlook	115
8	References	117
9	Acknowledgements	126
10	Appendix.....	127
10.1	Supplemental Figures	127
10.2	Abbreviations.....	130
10.3	Nomenclature.....	132
10.4	Figure Directory	133

10.5 Table Directory.....	135
10.6 Supplemental Figure Directory	136
Curriculum Vitae	137
Publications.....	138
Affirmation.....	139

Summary

Acute lymphoblastic leukemia (ALL) is often defined by genetic susceptibility factors and subsequently acquired somatic mutations. One of the most common genetic susceptibility factors is the presence of a chromosomal translocation. The most frequent one in pediatric ALL is the t(12;21)(p13;q22) translocation, resulting in the *ETV6-RUNX1* fusion gene. This translocation predominantly arises *in utero* and requires secondary mutations for leukemia development. Even though the rates of pediatric *ETV6-RUNX1* positive leukemia are known, the frequency of this fusion among healthy newborns is disputed.

Therefore, in this study a new screening method to identify chromosomal translocations was developed. This method, termed “Genomic Inverse PCR for Exploration of Ligated Breakpoints” (GIPFEL), was established with the help of the most common translocations in pediatric ALL. It could be shown that GIPFEL is sensitive and specific to the investigated translocations.

With the help of GIPFEL, it was possible to determine the frequency of the *ETV6-RUNX1* fusion in 1,000 umbilical cord bloods of healthy Danish newborns. It was found that 5% of the cord blood samples harbored the *ETV6-RUNX1* fusion. Not only did this support previous findings, the frequency in this study even exceeded the ones described by these works.

Additionally, *ETV6-RUNX1* positive ALL was further characterized using a mouse model and a human cohort. Histone related genes, especially members of the *KDM* histone demethylase family, were affected in many tumor samples. Additionally, the leukemic *ETV6-RUNX1* positive mice showed higher levels of DNA methylation, potentially leading to misregulation of transcription.

Taken together, a new powerful screening method for chromosomal translocations was established, and it could be shown that the frequency of the t(12;21)(p13;q22) translocation, leading to the *ETV6-RUNX1* fusion, is 500 times higher in the investigated cohort than the respective leukemia incidence. This finding has a major

impact on how to proceed after detection of the *ETV6-RUNX1* fusion in healthy children, because the better part of them is going to stay leukemia free.

Zusammenfassung

Akute lymphoblastische Leukämie (ALL) ist oftmals durch eine genetische Prädisposition und später erworbene somatische Mutationen gekennzeichnet. Eine der häufigsten genetischen Prädispositionen ist dabei das Vorhandensein einer chromosomalen Translokation, wobei die t(12;21)(p13;q22)-Translokation, welche zur Entstehung des *ETV6-RUNX1*-Fusiongens führt, in pädiatrischer ALL am häufigsten vorkommt. Sie entsteht überwiegend *in utero* und führt erst in Kombination mit sekundären Mutationen zu Leukämie. Während die Häufigkeit der *ETV6-RUNX1*-positiven Leukämie bekannt ist, ist die Frequenz dieser Fusion bei gesunden Neugeborenen umstritten.

Aus diesem Grund wurde im Rahmen dieser Studie eine neue Screening-Methode zur Identifizierung chromosomaler Translokationen entwickelt: „Genomic Inverse PCR for Exploration of Ligated Breakpoints“, kurz GIPFEL. Diese Methode wurde mit Hilfe der häufigsten Translokationen in pädiatrischer ALL etabliert und erwies sich als sensitiv und spezifisch für die untersuchten Translokationen.

Durch die Anwendung von GIPFEL war es möglich, die Frequenz der *ETV6-RUNX1*-Fusion im Nabelschnur-Blut von 1.000 gesunden dänischen Neugeborenen zu bestimmen. Dabei zeigte sich, dass 5% der Blutproben die Fusion trugen. Dieser Befund stützt nicht nur die bisher beschriebenen Häufigkeiten, vielmehr übertrifft er diese sogar.

Des Weiteren konnte die *ETV6-RUNX1*-positive ALL unter Verwendung eines Mausmodells sowie einer menschlichen Kohorte weitergehend charakterisiert werden. So erwiesen sich Histon-betreffende Gene, insbesondere solche der *KDM*-Histon-Demethylase-Familie, in zahlreichen Tumorproben als von Mutationen betroffen. Darüber hinaus zeigten die leukämischen *ETV6-RUNX1*-positiven Mäuse eine höhere DNA-Methylierungsrate, durch welche es möglicherweise zu einer transkriptionellen Fehlregulation kommt.

Zusammenfassend wurde mit GIPFEL eine neue leistungsstarke Screening-Methode für chromosomale Translokationen etabliert. Durch ihre Verwendung konnte gezeigt werden, dass die Häufigkeit der t(12;21)(p13;q22)-Translokation, welche zur *ETV6-RUNX1*-Fusion führt, in der untersuchten Kohorte 500-mal höher als die entsprechende Leukämie-Inzidenz ist. Dieses Erkenntnis hat einen großen Einfluss auf die Vorgehensweise nach Feststellung einer *ETV6-RUNX1*-Fusion bei gesunden Kindern, da der Großteil von ihnen im Laufe ihres Lebens nicht an Leukämie erkranken wird.

1 Introduction

1.1 Pediatric Leukemia

Cancer is the most common disease-related cause of death among children aged 0-14. It had an incidence rate of 15.5 per 100,000 children in the United States in 2013 [1]. The largest cancer groups in these children are leukemia and cancer of the central nervous system (CNS) with 31% and 23%, respectively [1]. In terms of cancer related deaths, leukemia is with 29% second to cancer of the CNS (Figure 1.1) [2].

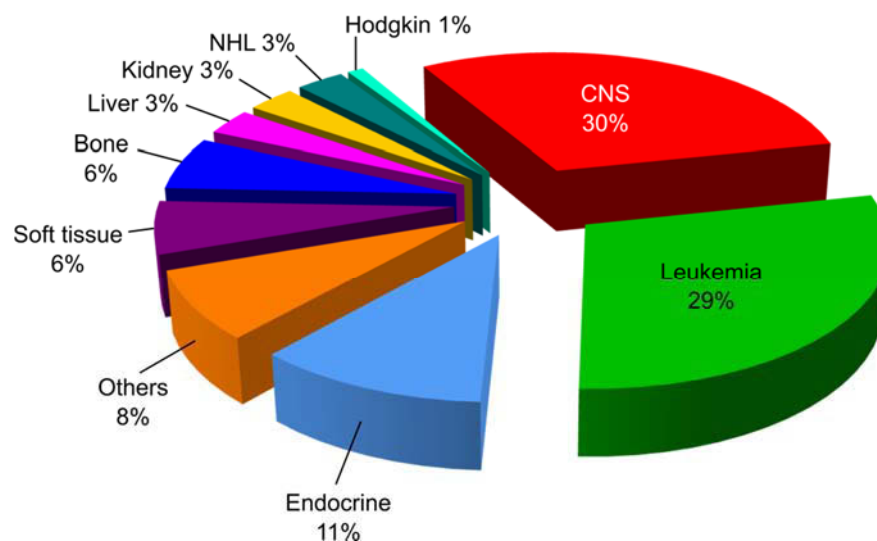


Figure 1.1: Cancer related deaths in children aged 0-14. With 29%, leukemia is the second most common cause of cancer related death. CNS = central nervous system, NHL = Non-Hodgkin lymphoma, adapted from [2].

Leukemia is a non-solid tumor of the white blood. The name derives from the Greek words leukós and haima which mean white and blood, respectively. It affects the white blood cells of the hematopoietic system, the leukocytes (Figure 1.2). Leukemia presents with high numbers of malignant leukocytes which are no longer able to meet their physiological obligations. Instead, they proliferate in an uncontrolled manner and infiltrate tissues and organs. These rapidly proliferating lymphocytes are called blasts. Depending on the cells in which the lesion occurs, leukemias are referred to as lymphoid or myeloid. Lymphoid leukemias arise from cells derived from the common

lymphoid progenitor, such as B cells and T cells, whereas myeloid leukemias arise from lymphocytes derived from the common myeloid progenitor [3] (Figure 1.2).

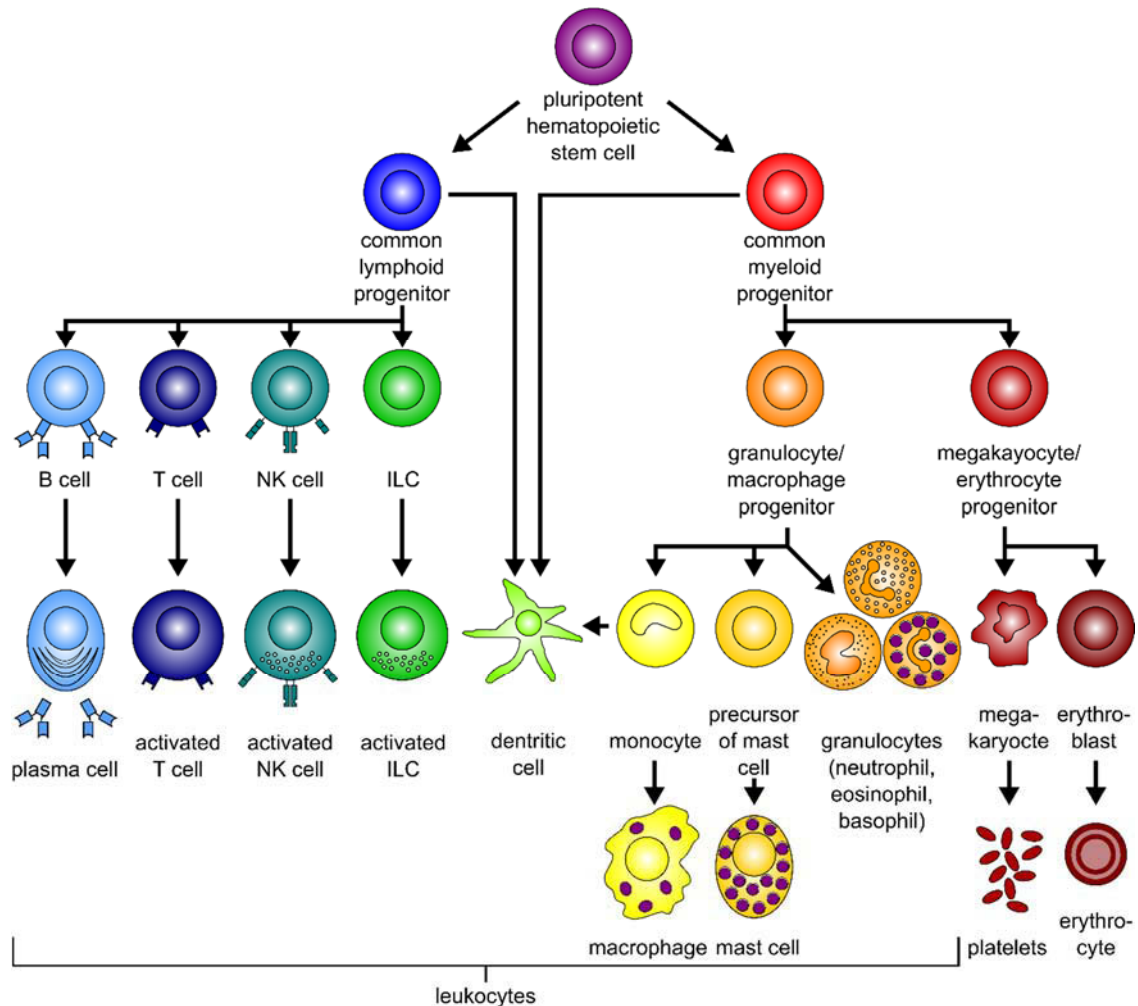


Figure 1.2: Scheme of hematopoiesis. All hematopoietic cells derive from a common pluripotent hematopoietic stem cell (purple). The common lymphoid progenitor (blue) gives rise to cells of the lymphoid compartment. These include B cells, T cells, natural killer (NK) cells, innate lymphoid cells (ILC), and dendritic cells. The common myeloid progenitor (red) differentiates into red and white blood cells. The red blood cells include platelets, erythrocytes, and their progenitors. The myeloid white blood cells derive from the granulocyte/macrophage progenitor and include dendritic cells, macrophages, mast cells, and granulocytes. Leukocytes are cells of the immune system and derive from both lineages. Leukemias are classified as lymphoid and myeloid, depending on the lineage that they arise in. Adapted from [4].

Leukemias can be further categorized by the speed of disease progression. Acute leukemias develop quickly and show rapid accumulation of blasts. This calls for immediate treatment. Acute leukemias mainly affect precursor cells and are more common in children [2]. Chronic leukemias on the other hand progress more slowly and

mainly affect cells in more mature developmental stages. Chronic leukemias are more common in adults [2].

Through these categorizations, leukemias can be subcategorized in four groups: acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphoid leukemia (CLL), and chronic myeloid leukemia (CML). This thesis focusses on ALL.

1.2 Acute Lymphoblastic Leukemia

1.2.1 Epidemiology

With over 80% of leukemia cases, ALL is the most common subtype in children [1] and with 25%, the most common childhood cancer. Unlike for AML, for ALL a peak incidence can be observed with highest incidences at ages 2 to 5 [5]. Furthermore, boys are affected more often than girls [6], and male sex used to be a markedly poor prognostic factor. Even though rates have improved, boys still have higher incidences and slightly poorer prognoses [6]. In the United States, ALL is most common among Hispanics (40.9 cases per million), followed by people of European descent and least common in those of African descent (35.6 and 14.8 cases per million, respectively) [7].








	stem cell	early pro-B cell	late pro-B cell	large pre-B cell	small pre-B cell	immature B cell	mature B cell
							
H-chain genes	germline	D-J rearrangement	V-DJ rearrangement	VDJ rearranged	VDJ rearranged	VDJ rearranged	VDJ rearranged
L-chain genes	germline	germline	germline	germline	V-J rearrangement	VJ rearranged	VJ rearranged
Ig status	none	none	none	μ heavy-chain, surrogate light-chain. Pre-B cell receptor on cell surface	μ heavy-chain in endoplasmic reticulum	μ heavy-chain, λ or κ light-chain. IgM on surface.	IgD and IgM on surface.

Figure 1.3: B cell development. The different stages of B cell development are shown. For each stage, the status of rearrangement of the immunoglobulin heavy-chain (H-chain) genes and light-chain (L-chain) genes is given as well as the status of the immunoglobulins (Ig). Pro-B and pre-B cells qualify as B cell progenitors. IgM is depicted in yellow, IgD in green, the pre-B cell receptor in yellow and red. Adapted from [8].

The vast majority of pediatric ALL cases (80-85%) express early B cell progenitor markers [2] and hence belong to the B cell precursor ALL (BCP-ALL) subtype (Figure 1.3). Approximately 15% carry markers of early T cell progenitors [2] and belong to the precursor T-ALL group.

1.2.2 Genetics of ALL

ALL is caused by a combination of genetic susceptibility factors and subsequent somatic mutations, often in genes that are critical for lymphoid development [7]. The genetic susceptibility factors include inherited mutations and also recurrent non-random mutations, for example translocations. Some translocations can arise prenatally; these include the t(12;21) translocation leading to *ETV6-RUNX1* gene fusion [9] and the t(4;11) translocation leading to *KMT2A-AFF1* gene fusion [10, 11]. Inherited mutations in genes important for lymphocyte development, like *PAX5*, lead to higher incidences within a family [12, 13]. Generally, siblings of ALL patients have a two- to four-fold higher risk of also developing an ALL [14]. Among further predisposing factors, there are gene defects associated with Fanconi-Anemia [15]. In these cases, AML is more common, but ALL also does occur. Furthermore, trisomy 21 is a risk factor for both, AML and ALL, with AML peaking in infancy [16-18].

Genetically, ALL can be divided into subgroups depending on characteristic lesions (Figure 1.4). One quarter of pediatric ALLs harbor the t(12;21) translocation [19], which will be explained in more detail in chapter 1.3.1. An equally large subgroup is hyperdiploid, meaning that more than 50 chromosomes are present, which occurs more frequently than the hypodiploid counterpart with less than 45 chromosomes [20-22]. Besides trisomy 21, which affects all body cells, trisomy of other chromosomes, like 5, 10, and 17, is common in ALL cells [22]. Translocations involving *KMT2A* are also frequent lesions (Figure 1.4). In the most common form, *KMT2A* on chromosome 11 is translocated to *AFF1* on chromosome 4, but more than 120 *KMT2A* rearrangements with different genes have been described [23]. These include *MLLT3*

on chromosome 9 and *MLLT1* on chromosome 19. Overall, almost half of all pediatric ALLs harbor translocations [19], including *TCF3-PBX1*, *BCR-ABL1*, and translocations involving *MYC* and *CRLF2* (Figure 1.4 A). Deletions of *ERG* on chromosome 21 and

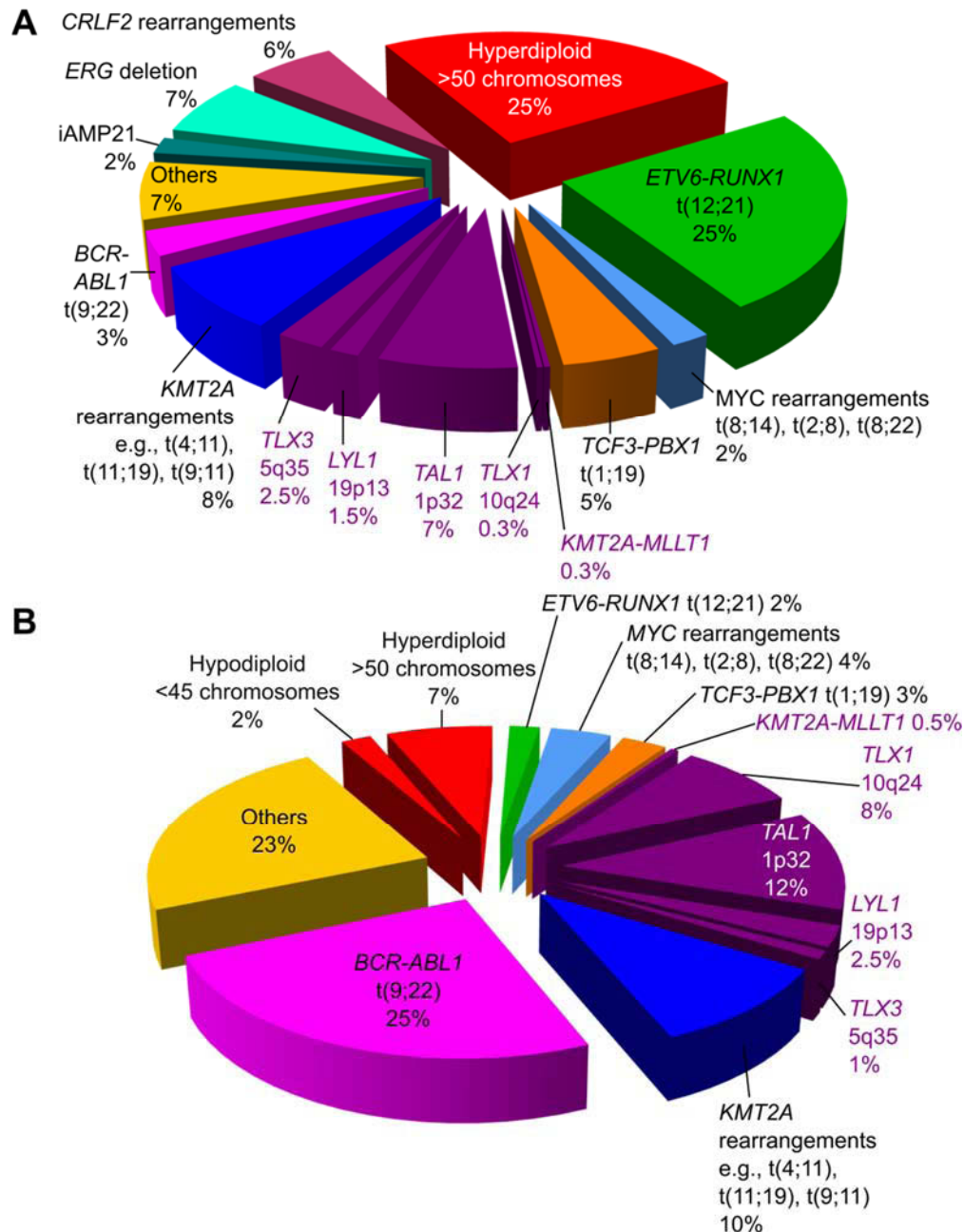


Figure 1.4: Subtypes of ALL in childhood and adulthood. **(A)** Frequency of ALL subtypes in childhood. Hyperdiploidy with more than 50 chromosomes and translocation t(12;21) with *ETV6-RUNX1* are the most common subtypes. T-ALL subtypes are indicated in purple. iAMP21 = intrachromosomal amplification of chromosome 21. **(B)** Frequency of subtypes in adulthood. Translocation t(9;22) with *BCR-ABL1* is the most common subgroup, whereas hyperdiploidy and *ETV6-RUNX1* are far less common than in children. T-ALL subtypes are indicated in purple. Corresponding subtypes in (A) and (B) are indicated in the same color. (A) adapted from [19], (B) adapted from [6].

intrachromosomal amplification of chromosome 21 (iAMP21) are further frequent aberrations (Figure 1.4 A).

In adulthood, the distribution of subtypes is different (Figure 1.4 B). Aneuploidy with aberrant chromosome numbers decreases as well as translocation t(12;21) with *ETV6-RUNX1*. The percentage of B-ALL cases with translocations, however, rises to almost 60%, mainly due to the increase in *BCR-ABL1* cases. The percentage of adulthood T-ALLs is more than twice that of childhood T-ALL [19] (Figure 1.4). Other lesions, like *KMT2A* translocations, *TCF3-PBX1*, or translocations including *MYC*, are roughly equally frequent in adults and children (Figure 1.4).

1.2.3 Clinical Aspects

1.2.3.1 Symptoms and Diagnosis

In leukemic patients the normal hematopoiesis is faulty. This leads to anemia, thrombocytopenia, and neutropenia which then in turn lead to clinical features, including hepatosplenomegaly, splenomegaly, fever, lymphadenopathy, bleeding, and bone pain [3]. In leukemia especially long bones are affected by bone pain, but joints can also hurt. This may lead to a false diagnosis as it resembles symptoms of other nonmalignant disorders, e.g. juvenile rheumatoid arthritis. Further common symptoms are pallor, fatigue, petechiae, and purpura [3].

For a correct diagnosis, ALL has to be distinguished from other malignancies affecting the bone marrow. The most prominent ones are neuroblastoma and non-Hodgkin lymphoma (NHL). Therefore, the percentage of blasts in the bone marrow is measured for diagnosis. If more than 25% of cells are blasts, it is categorized as ALL and not as lymphoma [3]. However, this differentiation is becoming more and more obsolete as lymphomas show excellent treatment response to ALL treatment [24].

Depending on several prognostic factors, leukemias can be categorized in risk groups. These risk groups have different outcome prognoses and require different treatments. For B-ALL, higher initial leukocyte counts are associated with poorer outcome, whereas

this correlation is not present in T-ALL [25]. The age at diagnosis is an additional factor: Children aged 1-10 at diagnosis have the best prognosis, which decreases with increased age. Infants under one year of age present with the poorest outcome [26], mainly because *KMT2A* translocations with dismal prognoses are very common in this age group [27]. Even though the difference used to be more pronounced, girls still have a slightly better prognosis than boys [3]. The same holds true for racial differences: patients with Caucasian and Asian ethnicity have the best outcome, patients with African ethnicity the worst [1].

A very important prognostic factor is the immunophenotype. Generally, B-ALL has better outcome than T-ALL [26]. The determination of the immunophenotype is carried out by flow cytometry with the help of antibodies against surface molecules that are specific for a subgroup (Table 1.1). Besides the immunophenotype, genetic alterations play a major role in determining the prognosis: when treated with current protocols, lesions like the t(12;21) and t(1;19) translocations, resulting in *ETV6-RUNX1* fusion and *TCF3-PBX1* fusion, respectively, confer a good prognosis, whereas others like the t(4;11) translocation, leading to *KMT2A-AFF1* fusion, have a dismal prognosis (Figure 1.5) [28].

Table 1.1: Immunophenotype of B- and T-ALL. Adapted from [29]. Classification of subtypes according to the European Group for the Immunological Characterization of Leukemias (EGIL) [30]. Characterization of early T-precursor ALL by [31]. c = cytoplasmic, s = surface, H-chain = heavy-chain, L-chain = light-chain.

Lineage	Subtype	Classification	Markers
B cell	pro-B-ALL	EGIL B-I	CD19 ⁺ , CD22 ⁺ , CD79a ⁺
	common B-ALL	EGIL B-II	CD19 ⁺ , CD22 ⁺ , CD79a ⁺ , CD10 ⁺
	pre-B-ALL	EGIL B-III	CD19 ⁺ , CD22 ⁺ , CD79a ⁺ , CD10 ⁺ , c-μ H-chain
	mature B-ALL	EGIL B-IV	CD19 ⁺ , CD22 ⁺ , CD79a ⁺ , CD10 ⁺ , s-L-chains
T cell	pro-T-ALL	EGIL T-I	cCD3 ⁺ , CD7 ⁺
	pre-T-ALL	EGIL T-II	cCD3 ⁺ , CD7 ⁺ , CD5/CD2 ⁺
	cortical-T-ALL	EGIL T-III	cCD3 ⁺ , sCD3 ⁺ , CD1a ⁺
	mature-T-ALL	EGIL T-IV	cCD3 ⁺ , sCD3 ⁺ , CD1a ⁻
	early T-precursor ALL	-	CD5 ^{low} , CD1a ⁻ , CD8 ⁻ , presence of stem cell or myeloid markers

ALL can affect extramedullary sites, along which CNS, testes, liver, kidneys, lymph nodes, and spleen are the most common [3]. In less than 5% of B-ALL cases, the CNS is involved. Affection of the CNS is associated to poorer outcome [32], but through

improved treatment protocols, only cases with high blast counts in the cerebrospinal fluid have a dismal prognosis [33].

Further adverse prognostic factors are poor response to treatment, known as induction failure, leading to high relapse rates as well as nutritional aspects, including malnutrition and obesity [3].

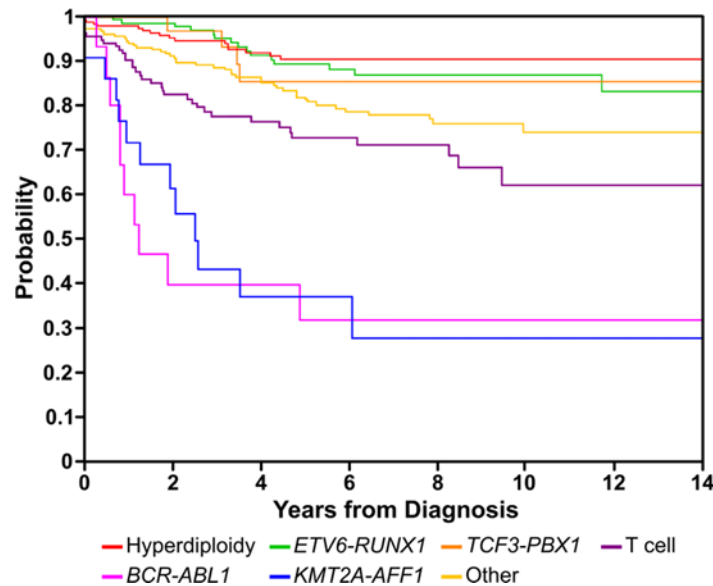


Figure 1.5: Event-free survival by ALL sybtype. Kaplan-Meier analysis for different ALL subtypes. The probability of event-free survival is given. Number of investigated cases: 205 hyperdiploidy, 163 *ETV6-RUNX1*, 40 *TCF3-PBX1*, 138 T cell ALL, 22 *BCR-ABL1*, 15 *KMT2A-AFF1*, and 261 other B lineage ALL (other). Adapted from [28].

1.2.3.2 Treatment

Treatment of ALL is divided into different steps. It starts with the induction of remission, followed by CNS preventive therapy, interim maintenance, delayed intensification, and maintenance which is sometimes called continuation [3]. The intensity of the chemotherapy depends on the risk stratification. Intense treatment is maintained for 6-12 month, depending on the risk. This is also the case for overall duration of treatment and for dosage of the chemotherapeutic agents [3].

The first step of treatment is the induction of remission which means that peripheral blood values must be within normal range and the cellularity of the bone marrow must be normal with less than 5% blasts [34]. A further requirement is the absence of

detectable or extramedullary disease. The leukemic cell burden of an overt ALL is approximately 10^{12} cells (Figure 1.6). Chemotherapy has to kill 99% of these cells in order to induce complete remission.

When chemotherapy is induced with vincristine [35] and a glucocorticoid, 85% of cases reach complete remission [3]. Prednisone, prednisolone, and dexamethasone are frequently used as glucocorticoids with dexamethasone being more effective in inducing bone marrow remission and treating the CNS [36]. Addition of L-asparaginase leads to better results, with 95% of ALL cases reaching complete remission [3]. Medication with L-asparaginase induces apoptosis in B- and T-ALL blasts because they are not capable of synthesizing asparagine. However, addition of L-asparaginase can also lead to side effects like pancreatitis and thrombosis. For high risk groups, an additional anthracycline can be added to the compounds. This leads to higher toxicity for blast cells but also bears the risk of higher mortality as the overall amount of drugs

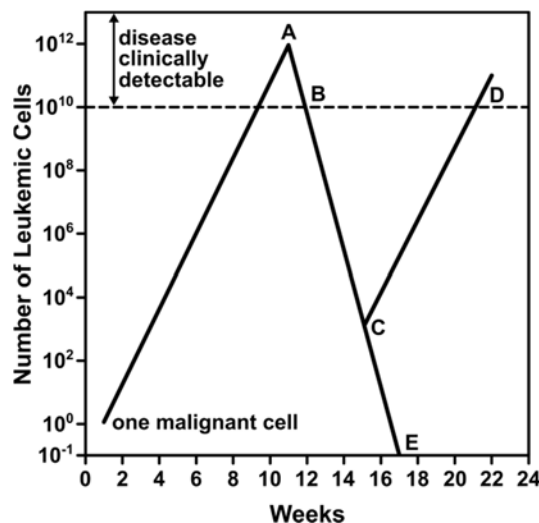


Figure 1.6: Schematic presentation of treatment results. The course of ALL development and treatment is shown. The ALL starts with a single cell that rapidly proliferates. The diagnosis follows at point **A**, leading to induction of therapy. At point **B**, complete remission is achieved. Therapy is stopped or becomes ineffective at point **C**. This can lead to proliferation of the blasts and a relapse (**D**). If treatment is successful and the patient is cured, more than 10^{12} cells were killed during treatment (**E**). ALL is clinically detectable when more than 10^{10} blasts are present. Adapted from [3, 37].

is increased [3]. In fewer than 3% of childhood ALL cases, more than 5% of blasts remain in the bone marrow after induction therapy. This so-called induction failure is a

dismal prognostic factor. An event free survival rate of 16% used to be achieved with older protocols, mainly through bone marrow transplants. With the help of newer treatment protocols, this rate could be doubled but remains low [38].

CNS preventive therapy is carried out in order to minimize the risk of a leukemic CNS relapse. The CNS is especially prone to relapses because at diagnosis undetected leukemic cells in the CNS are protected from chemotherapeutic compounds by the blood-brain barrier [3]. There are three ways of CNS relapse prophylaxis: irradiation, intrathecal chemotherapy, and application of high-dose systemic agents. Nowadays, less than 10% of ALL patients receive cranial irradiation. Still in the 1960s and 1970s, almost 100% of patients were irradiated [3]. Patients with good prognostic factors usually only receive intrathecal methotrexate, whereas patients with higher risks often receive intrathecal cytarabine and hydrocortisone in addition [3]. The overall risk of CNS relapse is under 10% and under 5% for cases with good prognoses. Cases with high initial leukocyte counts, T-ALL, infant leukemia, and translocation $t(1;19)$, however, have a higher risk of CNS involvement and CNS relapse [39, 40].

After complete remission is achieved, therapy has to be continued. Otherwise, a relapse within one or two months is very likely as there still remain approximately 10^{10} blasts in the body (Figure 1.6) [3]. This continued post induction therapy is not as intense as the induction therapy. It usually starts with a one to two month long consolidation phase in which lower doses are given and the interval between drug applications is longer. Following the consolidation, a two to six month long phase is started. In this so-called delayed intensification, the chemotherapy is intensified again. Afterwards, two to three years of maintenance with low dose chemotherapy follow. This stage is also known as continuation phase [3].

If a relapse occurs, a more intense chemotherapy becomes necessary as these cells have already survived the first chemotherapy. Therefore, drugs are often changed in order to be able to reach a second complete remission. Treatment of relapses, especially early relapses, often involves hematopoietic stem cell transplantation [3]. In

these cases, the entire hematopoietic system is destroyed and replaced by the one of a healthy donor with the goal of also destroying the leukemic cells. However, it is important to distinguish a relapse from a newly arising secondary leukemia. These new leukemias usually appear later than relapses [41] and arise from an independent leukemic event. Secondary leukemias often arise from treatment of the first leukemia and are mostly AMLs [42]. These cases have a poorer prognosis than those with secondary *de novo* leukemia [42].

1.3 Translocations

Translocations arise from illegitimate joining of at least two different chromosomes. Potential mechanisms underlying translocations are illegitimate V(D)J recombination, the recombination of homologous sequences such as Alu elements, topoisomerase II-mediated breakage, or sequences with alternating purines and pyrimidines [43-48]. After DNA breakage, the chromosomes are joined in an attempt to repair the damage. This can be accomplished by homologous recombination or non-homologous end joining (NHEJ) [49].

In contrast to other malignancies, leukemia frequently harbors translocations [50]. These non-random breaks occur in breakpoint cluster regions (BCR), which usually are represented by one or two introns of a gene [51-53]. Through the translocation, exons from two different genes are joined together and transcribed as a fusion transcript which then in turn is translated into a fusion protein. These fusion proteins have other properties than their respective wild-type counterparts [54-58]. The t(9;22) translocation, creating the so-called Philadelphia chromosome, fuses the genes *BCR* and *ABL1* and was the first to be described [59]. It leads to the constitutive activation of the *ABL1* kinase [60]. Besides the activation of a kinase, translocations can also lead to altered transcriptional regulation. This is most obvious through the frequent involvement of transcription factors like *RUNX1*, *RARA*, *KMT2A*, and *ETV6*, which are critical for stem cell development or lineage specification of hematopoiesis [61].

The most common translocations in childhood B-ALL are *ETV6-RUNX1*, *TCF3-PBX1*, and several translocations involving *KMT2A* [19]. All these translocations involve transcription factors and therefore lead to altered transcriptional regulation and not to constitutive activation of a kinase.

As the investigated lesions, the *ETV6-RUNX1* and *TCF3-PBX1* fusions, arise in BCP-ALL, this work focusses on this subgroup.

1.3.1 *ETV6-RUNX1*

The *ETV6-RUNX1* fusion is the result of a balanced t(12;21)(p13;q22) translocation, merging *ETV6* exons 1 to 5 to either *RUNX1* exons 2-9 or *RUNX1* exons 3-9 (Figure 1.7). The *RUNX1* gene, initially known as *AML1*, was first identified as a translocation partner in t(8;21)(q22;q22) positive acute myeloid leukemia where it was fused to *RUNX1T1* [62]. After that, *RUNX1* was identified in further translocations, among them the t(12;21) translocation with *ETV6*, resulting in the *ETV6-RUNX1* gene [63, 64]. *RUNX1* is a transcription factor with a Runt DNA binding domain (Figure 1.7 B), closely related to the *Drosophila melanogaster* gene *runt*. *RUNX1* deficiency leads to an early block in hematopoietic differentiation. *RUNX1* knock out animals completely lack mature hematopoiesis and die embryonically [65, 66]. *ETV6* also encodes for a transcription factor; it has an ETS DNA binding domain and a pointed domain (PNT) for oligomerization [67] (Figure 1.7 B). *ETV6*, formerly known as *TEL*, was initially identified as the fusion partner of *PDGFRB* in t(5;12) translocations in chronic myelomonocytic leukemia [68]. Besides the fusions to *RUNX1* and *PDGFRB*, *ETV6* is also involved in further rearrangements, including kinases like *ABL1* [69] and *JAK2* [70, 71].

While only 2% of adult ALL cases present with *ETV6-RUNX1* [6], it is the most common aberration in childhood ALL, where 25% of cases harbor the translocation [19, 72]. The t(12;21) translocation is linked to favorable outcome with a five-year event free

survival rate of up to 96.8% and a five-year overall survival rate of up to 98.9%, depending on the treatment protocol [73].

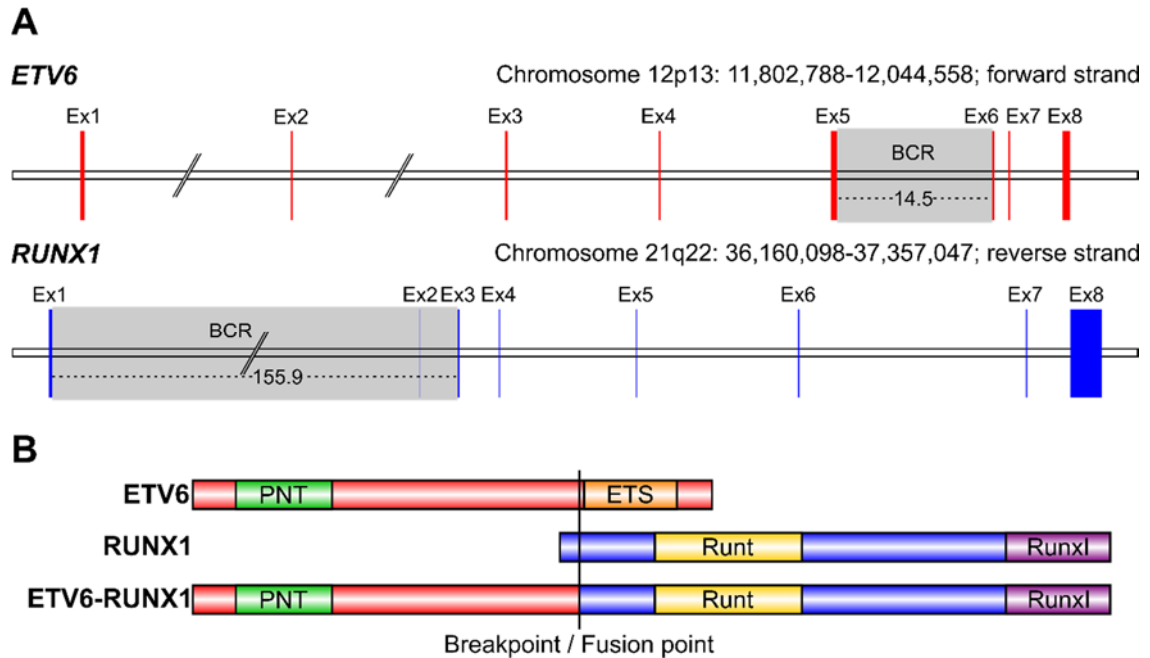


Figure 1.7: ETV6-RUNX1 fusion. **(A)** Genomic loci of *ETV6* (upper panel) and *RUNX1* (lower panel). The breakpoint cluster regions (BCR) are indicated in gray. For *ETV6* the BCR is identical to intron 5, the *RUNX1* BCR equals introns 1 and 2. Sizes of BCRs are given in kb. **(B)** The ETV6, RUNX1, and ETV6-RUNX1 proteins. The ETV6 pointed domain (PNT) serves as oligomerization domain [67], the ETS domain is for DNA binding. RUNX1 has a Runt DNA binding domain and a Runx inhibition domain (Runx1) at its C-terminus that interacts with cofactors [74]. In the ETV6-RUNX1 fusion protein, the ETV6 ETS domain is lost, but almost the entire RUNX1, including all functional domains, is maintained. For *RUNX1*, only the transcript (NM_001754.4) and protein (NP_001745.2) for isoform 1 are shown as isoforms 2 and 3 do not contain the BCR.

The *ETV6-RUNX1* fusion gene is an extreme example of genotype-phenotype association as it is only linked to BCP-ALL [75]. It predominantly arises prenatally in committed B cell progenitors [9]. However, the frequency of this event in the population is highly disputed, as contradicting results have been published [76-80]. Mori *et al.* [76] and Zuna *et al.* [80] state that 1% to 2% of all newborns harbor the t(12;21) translocation, while other studies could not validate this finding [77-79].

The fact that the translocation arises prenatally shows that the fusion of *ETV6* to *RUNX1* poses an initiating event [9]. It is, however, not sufficient for leukemia development: monozygotic twins who share the same *ETV6-RUNX1* clone do not necessarily both develop BCP-ALL [9]. Hence, a second lesion has to occur in order to

develop an overt leukemia. These secondary mutations are mostly deletions affecting genes of the B lymphocyte development and differentiation, such as *BTG1*, *CDKN2A*, *PAX5*, *RAG1*, *RAG2*, *TBL1XR1*, and the second *ETV6* allele [81]. Deletions of the wild-type *ETV6* have been reported in up to 70% of cases, deletions of *PAX5* and *CDKN2A* in 28% and 26%, respectively [81]. The deletions are predominantly caused by aberrant RAG1/2 activity [82]. These lesions are subclonal to *ETV6-RUNX1* [83] and differ in monozygotic twins [84].

RUNX1 functions as a transcriptional activator whose targets involve blood cell development associated genes [85-87]. B cell differentiation is dependent on RUNX1 activity, whereas hematopoietic stem cells (HSC) are not dependent on RUNX1 for their maintenance and self-renewal [65, 66]. Enforced *ETV6-RUNX1* expression leads to similar results like loss of RUNX1 activity: an increase in multipotent progenitors, namely early pro-B cells [88]. *ETV6-RUNX1* represses hematopoiesis-specific genes that are activated by RUNX1 and by that antagonizes the wild-type RUNX1 function [54].

1.3.2 *TCF3-PBX1*

The t(1;19)(q23;p13) translocation fuses the transcription factor *TCF3* on chromosome 19 to the transcription factor *PBX1* on chromosome 1. The *PBX1* BCR spans the 229.2 kb of intron 2 and the *TCF3* BCR comprises intron 16 (Figure 1.8 A). Furthermore, *TCF3* breakpoints cluster near transposable elements [52].

TCF3, previously known as E2A, is a basic helix-loop-helix (bHLH) transcription factor, first identified as such in 1989 [89]. The bHLH domain binds the DNA. There are two different products transcribed from the *TCF3* gene, named *E12* and *E47*, which differ in one exon [89]. The differences mainly regard the bHLH domain which is not included in the fusion protein (Figure 1.8 B). Additionally, *TCF3* has two activation domains, one at the N-terminal and one in the middle (Figure 1.8 B) [90, 91]. Like other related proteins, *TCF3* is an activating transcription factor which predominantly forms hetero-dimers.

TCF3 is widely expressed, its specificity is guaranteed by the dimerization partner [92]. Still, *TCF3* seems to play a major role in B cells as its expression in B cells, especially those of the germinal centers, is higher [93, 94]. Furthermore, enforced *TCF3* transcription leads to immunoglobulin rearrangements and transcription of B cell specific genes, e.g. *EBF1* and *PAX5*, in non-B cells [95-97]. *TCF3* also plays a role in Immunoglobulin class switch [93, 94, 98, 99], which explains the elevated expression in germinal center B cells. Additionally, the E47 form of the protein is able to form homo-dimers exclusively in B cells [100].

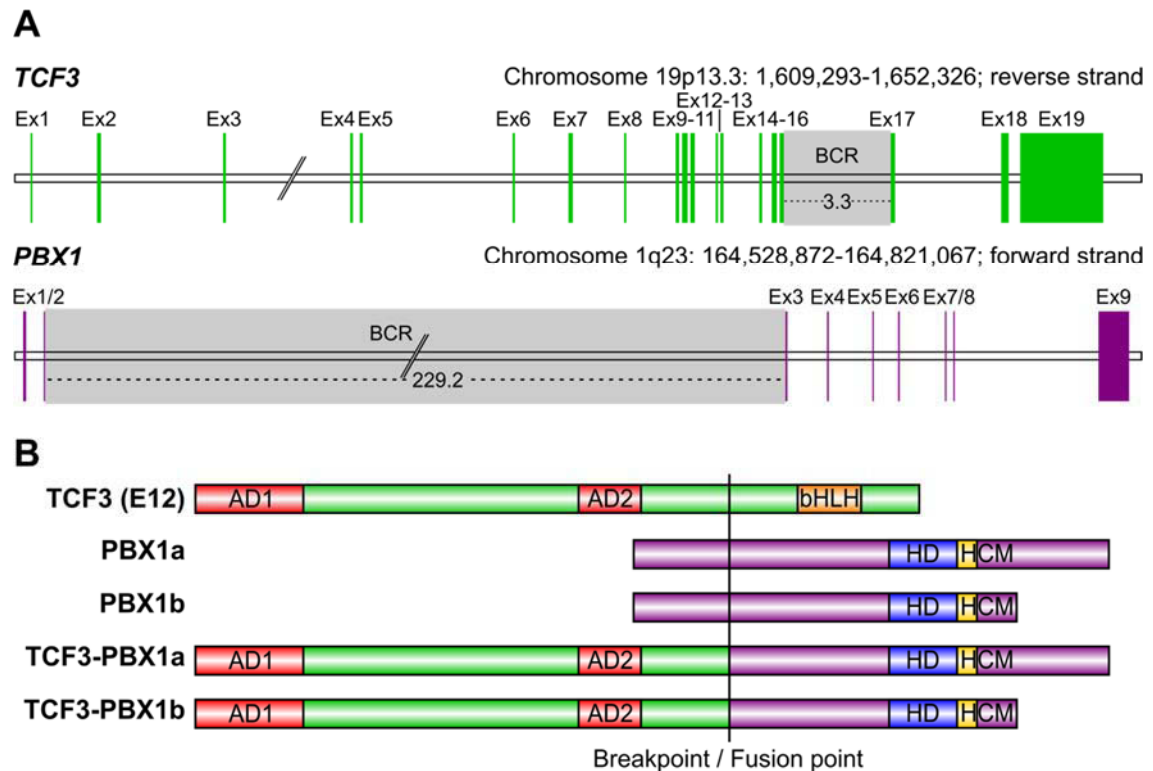


Figure 1.8: TCF3-PBX1 fusion. **(A)** Genomic loci of *TCF3* (upper panel) and *PBX1* (lower panel). The breakpoint cluster regions (BCR) are indicated in gray. For *TCF3*, the BCR is identical to intron 16 and the *PBX1* BCR equals intron 2. For *TCF3*, the isoform *E12* (NM_003200.3) is shown, for *PBX1* the isoform *PBX1a* (NM_002585.3) is shown. Sizes of BCRs are given in kb. **(B)** The *TCF3* (*E12*), *PBX1a/b*, and *TCF3-PBX1a/b* proteins. For *TCF3*, only the *E12* isoform (NP_003191.1) is shown as the *E47* isoform (NP_001129611.1) only differs by the use of an alternative exon C-terminal of the breakpoint [89]. Hence, the *TCF3* part in *TCF3-PBX1* is identical. The activation domains (AD1 and AD2) are retained in the fusion, whereas the basic helix-loop-helix (bHLH) DNA binding domain is lost. Two isoforms *PBX1a* (NP_002576.1) and *PBX1b* (NP_001191890.1) are shown for *PBX1*. *PBX1b* is shorter but has the same domains as *PBX1a*: a homeodomain (HD) for DNA binding and a HOX cooperativity domain (HCM) for the interaction with HOX proteins. Both domains are retained in the fusion proteins. (B) adapted from [101]. Note: As the isoform is not relevant for the detection of the breakpoint, only isoforms *E12* and *PBX1a* are used henceforward.

PBX1 was first identified through its fusion to *TCF3* [102]. It encodes for a transcription factor with a homeodomain for DNA binding (Figure 1.8 B). There are different forms of *PBX1* which differ in their respective C-termini, with *PBX1b* being shorter than *PBX1a* (Figure 1.8 B). *PBX1* plays a role in organogenesis and osteogenesis [103-105]. Although it is expressed in early stages of embryonic hematopoiesis [106], *PBX1* is not expressed in lymphoid cells [103]. *PBX1*, like other PBX proteins, builds complexes with HOX proteins which can act as transcriptional activators or repressors, depending on cofactors [107].

Unlike *PBX1*, the *TCF3-PBX1* fusion exclusively acts as a transcriptional activator, leading to misregulation of *PBX1* target genes [55-58]. Besides this misregulation of *PBX1* targets, *TCF3-PBX1* dimers bind *TCF3* cofactors and by that repress *TCF3* targets [108, 109]. The fusion protein remodels the oncogenic signaling in BCP-ALL by perturbing signaling pathways upstream of PLC γ 2 [110]. Furthermore, it interferes with signaling pathways relevant for apoptosis and cell cycle control as well as with WNT signaling [58]. *TCF3-PBX1* on its own does not seem to be sufficient for leukemia development but rather requires secondary mutations [111]. Mice with inducible *TCF3-PBX1* had secondary mutations in important B cell genes like *Pax5* and genes affecting the Jak/Stat pathway [112].

The translocation occurs in a balanced (-1, -19, +der(1), +der(19)) and in an unbalanced (-19, +der(19)) form, with three fourth of the cases being unbalanced [113]. Initially, the unbalanced form was reported to have a more favorable outcome [114, 115]; later studies including advanced treatment protocols could not find this difference [116]. This development mirrors the overall prognosis of patients with the t(1;19) translocation: initially linked to a dismal prognosis [114, 117], *TCF3-PBX1* is now linked to a good prognosis in childhood ALL [115, 118]. Moreover, prognosis for adults is not worse than for children [118]. In children, 5% of ALLs harbor this fusion, making it one of the most common aberrations in childhood ALL [19]. The incidence in adults is only a little lower with 3% of ALL cases [6].

1.4 Methods for Detection of Translocations

There are several methods that can be used to detect translocations. All of them have advantages and disadvantages. The oldest method for the detection of translocations is karyotyping [119] which is done by Giemsa staining [120]. This allows for discrimination of the chromosomes in order to identify aberrations (Figure 1.9 A). Karyotyping leads to unambiguous results when the translocation switches material of different size from two chromosomes. In contrast, when only small fragments are switched, it is hard to identify the translocation by karyotyping. Another downside of this technique is the need to culture the cells, as metaphase cells are required. This also lowers the detection sensitivity to approximately three to four in 10^2 cells [121]. As a variant of the Giemsa based karyotyping, translocations can also be identified by whole chromosome painting. This technique labels every chromosome fluorescently in a different color, allowing for discrimination of the chromosomes and identification of translocations [122]. Therefore, it is possible to also identify small translocations and those which are hard to detect by Giemsa stained karyotyping. However, culturing of cells and the need for metaphase cells are still a major downside of this method. Furthermore, karyotyping is usually done with bone marrow cells, as opposed to more easily accessible blood cells.

The need for metaphase cells can be eliminated by the use of interphase fluorescence *in situ* hybridization (FISH) [123]. For interphase FISH, fluorescently labeled probes that cover the investigated genes close to the breakpoint are used. If a translocation is present, the signals overlap and a different color appears (Figure 1.9 B). This can in turn be a problem when the signals lie in close proximity to each other, with no translocation present, because the chromosomes are close to each other. Generally, interpreting FISH results requires some experience. The sensitivity of interphase FISH is with the detection of two to five in 10^3 cells higher than that of karyotyping [121], as more cells can be investigated due to the use of interphase cells. Even though the cells can be obtained from blood, they still have to be cultured.

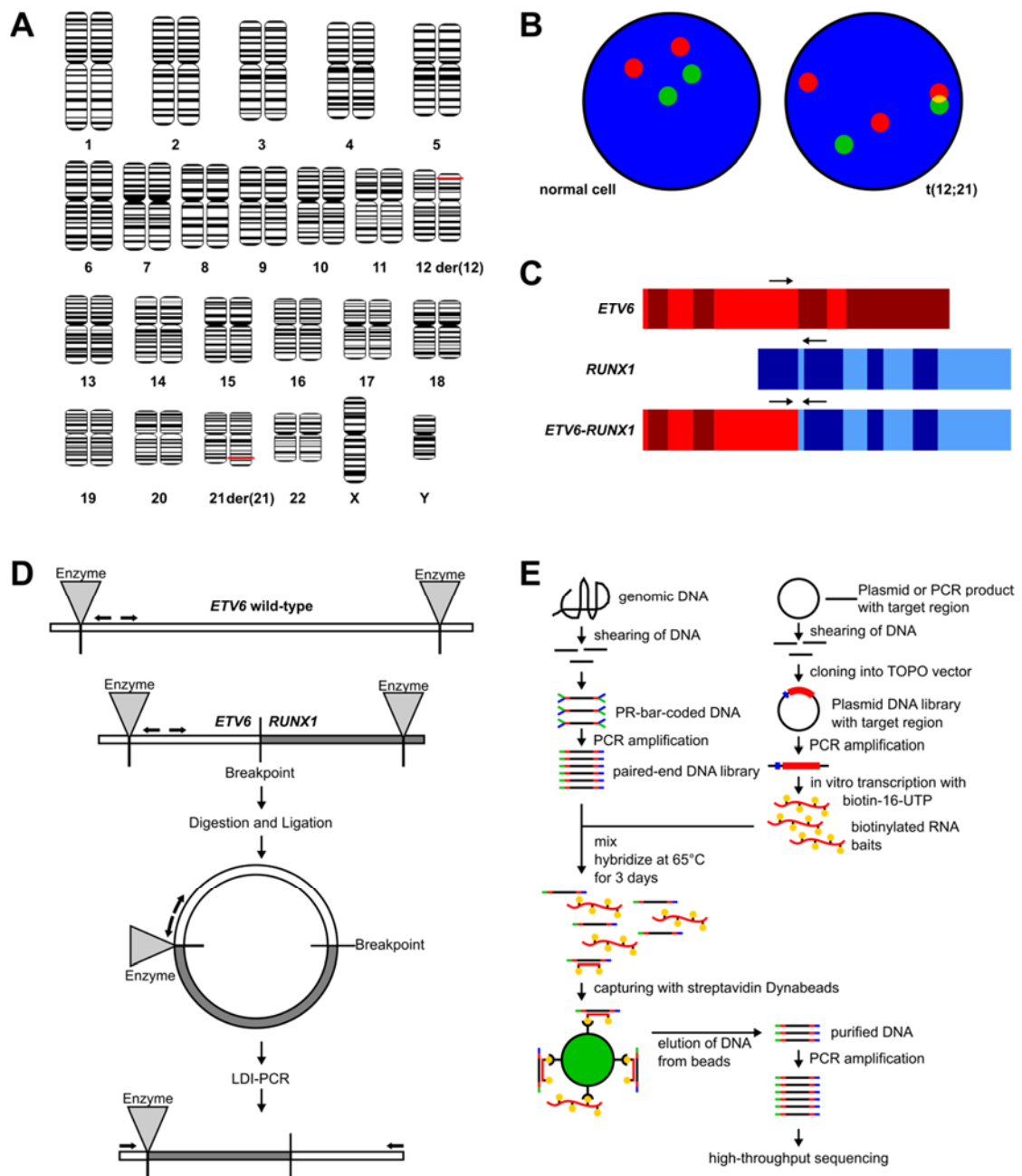


Figure 1.9: Overview of methods used for detection of translocations. **(A)** Schematic presentation of karyotyping with the use of Giemsa staining. A balanced translocation $t(12;21)$ in a male human is shown exemplarily. The red lines indicate the break- and fusion points. Chromosome numbers are shown below the chromosomes. The prefix *der* indicates the changed chromosomes. **(B)** Exemplary result of interphase FISH. The left panel shows a normal cell and the right panel a cell with $t(12;21)$. Probes for *ETV6* are green, those for *RUNX1* red. The left panel shows two signals for each gene and therefore no translocation. On the right panel, a yellow fusion signal is detectable, indicating the translocation. Depending on the probes that are used, a normal *RUNX1* signal can still be seen, as the probe also covers a non-translocated part. This is the case in this example. **(C)** Detection of the $t(12;21)$ translocation with RT-PCR. The exons are shown in different shades of red and blue for *ETV6* and *RUNX1*, respectively. Primers are shown as arrows and can only produce a product in case of translocation. **(D)** Overview of long distance inverse PCR. The DNA is digested and religated. The breakpoint is then detected by long distance PCR. The invers primers are shown as arrows. The restriction enzyme has to be chosen for every patient separately. Adapted from [124]. **(E)** Principle of anchored chromPET. Genomic DNA is sheared and then used to establish a paired-end library with the help of Y-shaped adapters. In parallel, a plasmid or PCR product with the target region is sheared and cloned into a TOPO vector. The insert is PCR amplified and used to generate biotinylated RNA baits through *in vitro* transcription from the T7 promoter (blue) and integration of biotin-16-UTP. Library and baits are then mixed and hybridized. Bait bound DNA is captured with streptavidin beads and subsequently released from those. Afterwards, the purified DNA is PCR amplified and subjected to high-throughput sequencing. Adapted from [121].

A highly sensitive and widely used method is the detection of translocations via reverse transcriptase polymerase chain reaction (RT-PCR). This PCR method uses complementary DNA (cDNA), reverse transcribed from RNA. Therefore, the cDNA has no introns and the translocation can be validated on exome level, sparing the often large intronic BCRs. The detection is a simple PCR reaction, which usually amplifies a small product that is independent of the exact breakpoint (Figure 1.9 C). With a detection sensitivity of one in 10^4 cells [121, 125], RT-PCR is more sensitive than karyotyping and FISH. The major downside of RT-PCR is, however, the use of RNA. Opposed to DNA, RNA is highly unstable and prone to degradation. This is important when handling preserved material. Detection of translocations was five-fold more sensitive on DNA than on RNA from formalin-fixed cells and 10,000 times more efficient when using DNA instead of RNA from extracellular medium [121]. These findings are of importance as cells may die during the process of preservation.

Long distance inverse PCR (LDI-PCR) [124, 126] is a DNA based method used to detect rearrangements, including translocations (Figure 1.9 D). When using this technique, DNA is digested with a fitting restriction enzyme that has to be chosen for every patient. The restriction fragment with the translocation has to be smaller than the wild-type fragment [124]. After the digest, the DNA is religated to form a circle. Inverse primers are designed in close proximity to one restriction site and used to amplify a PCR product including the breakpoint. This PCR product can then be sequenced. The use of DNA instead of RNA is advantageous because it is more stable. However, there are also limitations to this method: The translocation carrying fragment has to be rather small and smaller than the wild-type fragment. Furthermore, the restriction enzyme has to be chosen individually for every patient [124]. That makes this method more suitable for the exact localization of the breakpoint rather than for a large scale screening.

Another DNA based method for detection of translocations is the use of anchored chromosomal paired-end tags (chromPET) [121]. Anchored chromPET is a Next Generation Sequencing based technique (Figure 1.9 E). Genomic DNA is sheared and

a sequencing library is created with the use of Y-shaped adapters. In parallel, a DNA library containing one of the translocated genes is generated and used to produce biotinylated RNA baits. These are then hybridized with the DNA and captured with streptavidin beads. The captured DNA is subsequently submitted to high-throughput sequencing. With the detection of three in 10^4 cells, this method is almost as sensitive as RT-PCR [121]. Through the use of RNA baits, the cost of sequencing is much lower than for whole genome sequencing, which can also be used to detect translocations. Anchored chromPET can only predict breakpoints which then have to be validated by Sanger sequencing [121]. Furthermore, it is time consuming as the hybridization step on its own takes three days, which makes the technique less attractive for screenings. Anchored chromPET is a good method to establish a follow-up of diagnosed patients as it allows for the design of primers in close proximity of the breakpoint. Normal PCR can then be used for the follow-up.

2 Aim of the Thesis

Over the past decades, several studies investigated the childhood cancer incidence in close proximity to nuclear facilities in England, Wales, and Germany [127-134]. All but one study were able to find elevated levels of pediatric leukemia in the closest radius of the nuclear power plants [127-131, 134]. However, the radiation dose caused by nuclear facilities is with 0.01 mSv per year very low, especially compared to natural radiation with 2.1 mSv per year and the average medical radiation dose of 1.8 mSv per year [135]. In order to identify further risk factors, the German Federal Office for Radiation Protection called for new screening methods, including one DNA based method for the detection of chromosomal translocations. Accordingly, one aim of this work is to establish a new screening method called “Genomic Inverse PCR for Exploration of Ligated Breakpoints” (GIPFEL).

If the GIPFEL method will be established successfully, it should be used to screen 1,000 umbilical cord bloods for *ETV6-RUNX1*, because contradicting results regarding the frequency of this fusion have been published. Studies by Mori *et al.* [76] and Zuna *et al.* [80] found that 1% and 2% of children harbor this translocation at birth, respectively. However, a series of other publications could not validate these findings [77-79]. Hence, a further aim of this work is to determine the frequency of *ETV6-RUNX1* amongst healthy newborns.

An additional aim of this thesis is to analyze *ETV6-RUNX1* positive leukemia. To that end, the disease should be reproduced in an *in vivo* mouse model. These mice will then be whole genome and whole exome sequenced and the thereby identified mutations will be compared to those of a human cohort.

3 Materials

3.1 Media / Cell Culture Reagents

Table 3.1: Media / cell culture reagents

Name	Company	Order No.
autoMACS Running Buffer	Miltenyi Biotec, Bergisch Gladbach, Germany	130-091-221
DMEM GlutaMAX	Gibco, invitrogen, Carlsbad, USA	31966-021
RPMI 1640	Gibco, invitrogen, Carlsbad, USA	31870-025
L-Glutamine	Gibco, invitrogen, Carlsbad, USA	25030-024
Penicillin/Streptomycin	Sigma-Aldrich, St. Louis, USA	P4333-100ML
FBS (Fetal Bovine Serum), heat inactivated	PAN Biotech, Aidenbach, Germany	P30-1902
Dulbecco's PBS	Sigma-Aldrich, St. Louis, USA	D8537
TrypLE Select	Gibco, invitrogen, Carlsbad, USA	12563-029
DMSO	Sigma-Aldrich, St. Louis, USA	D2650
Human Serum Albumin	Octapharma, Langenfeld, Germany	B.664.003.D

3.2 Cell Lines

Three cell lines were cultured as source of DNA: REH, 697, and HEK-293. All cell lines were acquired from the German Collection of Microorganisms and Cell Cultures (DSMZ, Brunswick, Germany). Properties and culture conditions are given in Table 3.2 for REH, in Table 3.3 for 697, and in Table 3.4 for HEK-293.

Table 3.2: Properties and culturing conditions for cell line REH. Information on the cell line was taken from DSMZ (www.dsmz.de). Handling is described as it was carried out.

REH	
DMSZ accession no.	ACC-22
Species:	human (15-year-old female)
Cell type:	B cell precursor leukemia
Morphology:	small, round, single cells; suspension cells
Characteristic:	t(12;21)(p13;q22.3) with <i>ETV6-RUNX1</i> fusion
Medium:	RPMI 1640 (Gibco) supplemented with 10% heat inactivated FBS (20% after thawing) (PAN Biotech), 2 mM L-Glutamine (Gibco), 100 U/ml Penicillin (Sigma-Aldrich), and 100 µg/ml Streptomycin (Sigma-Aldrich)
Incubation:	at 37°C with 5% CO ₂
Subculture:	Cells were seeded out at 2 x 10 ⁶ cells/ml. Cells were cultured to be between 0.5 x 10 ⁶ and 5 x 10 ⁶ cells/ml.
Doubling time:	50-70 hours
Storage:	frozen in 90% human serum albumin (Octapharma) + 10% DMSO (Sigma-Aldrich)
References:	[136-138], DSMZ (www.dsmz.de)

Table 3.3: Properties and culturing conditions for cell line 697. Information on the cell line was taken from DSMZ (www.dsmz.de). Handling is described as it was carried out.

697	
DMSZ accession no.	ACC-42
Species:	human (12-year-old male)
Cell type:	B cell precursor leukemia
Morphology:	round, single cells and clusters of cells; suspension cells
Characteristic:	t(1;19)(q23;p13.3) with <i>TCF3-PBX1</i> fusion
Medium:	RPMI 1640 (Gibco) supplemented with 10% heat inactivated FBS (PAN Biotech), 2 mM L-Glutamine (Gibco), 100 U/ml Penicillin (Sigma-Aldrich), and 100 µg/ml Streptomycin (Sigma-Aldrich)
Incubation:	at 37°C with 5% CO ₂
Subculture:	Cells were seeded out at 1 x 10 ⁶ cells/ml. Cells were cultured to be between 0.5 x 10 ⁶ and 1.5 x 10 ⁶ cells/ml.
Doubling time:	30-40 hours
Storage:	frozen in 90% human serum albumin (Octapharma) + 10% DMSO (Sigma-Aldrich)
References:	[139], DSMZ (www.dsmz.de)

Table 3.4: Properties and culturing conditions for cell line HEK-293. Information on the cell line was taken from DSMZ (www.dsmz.de). Handling is described as it was carried out.

HEK-293	
DMSZ accession no.	ACC-305
Species:	human (female embryo)
Cell type:	human embryonic kidney
Morphology:	adherent fibroblastoid cells growing as monolayer
Characteristic:	human near-triploid karyotype
Medium:	DMEM (Gibco) supplemented with 10% heat inactivated FBS (PAN Biotech), 2 mM L-Glutamine (Gibco), 100 U/ml Penicillin (Sigma-Aldrich), and 100 µg/ml Streptomycin (Sigma-Aldrich)
Incubation:	at 37°C with 5% CO ₂
Subculture:	Cells were seeded out at 2-3 x 10 ⁶ cells/25 cm ² . Cells were split 1:10 every 2-3 days.
Doubling time:	24-30 hours
Storage:	frozen in 90% heat inactivated FBS (PAN Biotech) + 10% DMSO (Sigma-Aldrich)
References:	[140], DSMZ (www.dsmz.de)

3.3 Patients

A total of eleven pediatric patients diagnosed with t(12;21) *ETV6-RUNX1* positive ALL from Germany were selected. Patients were treated according to protocols of the cooperative study group for childhood acute lymphoblastic leukemia (CoALL) and the ALL-BFM 2000 (Berlin-Frankfurt-Münster) trial, respectively. Written informed consent was given by all parents, and the local ethics committee approved the research.

Patients 5 through 11 suffered a relapse and were part of the ALL-REZ BFM 96 and 2002 relapse trials. Bone marrow or peripheral blood samples were drawn at diagnosis, relapse, and, as a germline control, at remission. The percentage of blasts exceeded

70% in all leukemic samples. Chromosomal breakpoints were known for all but one case (Table 3.5).

Table 3.5: Breakpoints of *ETV6-RUNX1* positive patients. The sequences 5' and 3' of the junction are given, as well as the inserted nucleotides. Positions are taken from human chromosomal built GRCh37.p13. Patient 8 had two different *ETV6-RUNX1* fusions, and data is shown for both. The breakpoints for patient 11 could not be determined exactly. For patients 5, 6, and the second fusion of patient 8 breakpoints of the *RUNX1-ETV6* fusion are shown, all other breakpoints are from *ETV6-RUNX1* fusions.

Pat. No.	Breakpoint <i>ETV6</i>	Breakpoint <i>RUNX1</i>	5' Junction sequence	Insertion (N-Nucleotides)	3' Junction sequence
1	12:12,030,752	21:36,266,983	AATAGTAATA	–	AGAGAGAGCA
2	12:12,038,290	21:36,268,551	CCACTGACAG	CA CTAGCATAGA	
3	12:12,031,207	21:36,302,680	TCAGGACAAT	–	CCTCTAGTCA
4	12:12,029,375	21:36,295,743	CCTGCTTTAA	ACCCCCCAAACA	TCACAACTTA
5	12:12,033,312	21:36,396,634	CTTCACAATC	–	AGGGAAACGG
6	12:12,031,596	21:36,327,573	CCTCCAACCT	–	GTGCCACCGT
7	12:12,027,897	21:36,264,202	GTTAATGGGT	–	TACTGTTTTTC
8	12:12,032,589	21:36,265,122	AGAAAATCCA	–	AATTTAACCA
	12:12,033,064	21:36,268,577	AGAATGGTCT	T	ATAATGTTAT
9	12:12,033,079	21:36,304,719	TCTGAGCTTG	–	TAGGTTTGAA
10	12:12,025,708	21:36,265,112	GGATTACAGG	–	AATAATTAAG
11	<i>ETV6</i> intron 5	<i>RUNX1</i> intron 1	–	–	–

DNA from five patients with t(1;19) *TCF3-PBX1* positive ALL was obtained [141] (Table 3.6). Patients were recruited from different countries within the International BFM Study Group and enrolled in multicenter trials on treatment of pediatric ALL. The respective national institutional review boards approved treatment and experiments, informed consent was obtained.

Table 3.6: Breakpoints of *TCF3-PBX1* patients. The exact breakpoint is given for all five patients, as well as the sequences 5' (*TCF3*) and 3' (*PBX1*) of the junction. There are no inserted nucleotides. Positions are taken from human chromosomal built GRCh38.p7. Data was taken from [141].

Pat. No.	Breakpoint <i>TCF3</i>	Breakpoint <i>PBX1</i>	5' Junction sequence	3' Junction sequence
12	19:1,617,932	1:164,695,248	CATCTCACCG	TGCTGAGAAT
13	19:1,617,932	1:164,783,319	CATCTCACCG	TCCTTAGCCT
14	19:1,617,931	1:164,787,839	ATCTCACCGC	TTGGGTGGAT
15	19:1,617,929	1:164,786,577	CTCACCGCAG	AAAGAAATTA
16	19:1,617,930	1:164,787,678	TCTCACCGCA	CACAATGCTG

In order to determine the specificity of the GIPFEL method, a screening of another set of patients was carried out. Fifty *ETV6-RUNX1* positive and 25 *TCF3-PBX1* positive ALL patient samples were obtained. For every translocation five additional samples, negative for the respective translocation, were interspersed. All samples were blinded

prior to the screening by the supervisor of the study, so that the status of the samples was unknown to the experimenter.

3.4 Newborns

Mononuclear cells from umbilical cord blood were provided by the laboratory of Kjeld Schmiegelow (Department of Pediatrics and Adolescent Medicine, University Hospital Rigshospitalet, Copenhagen, Denmark). Cord blood storage for every newborn is a standard practice in Denmark and donor's informed consent was given, in agreement with local ethics guidelines. Kjeld Schmiegelow's group provided a total of 1,000 anonymized cord blood samples, which were withdrawn from the umbilical cords of healthy newborns. From every umbilical cord 0.5 to 6 ml of blood were collected in blood collection tubes with EDTA as anticoagulation additive and processed within 24 h. Mononuclear cells were separated by Ficoll density centrifugation and washed in RPMI 1640. The washed cells were then resuspended in 2 ml RPMI 1640 containing 10% BSA. The equal amount of cryopreservative solution was added and the cells were transferred to cryovials. After initial freezing at -80°C , the vials were stored in liquid nitrogen and sent to Düsseldorf on dry ice.

3.5 *ETV6-RUNX1*⁺ Mice

DNA from mice carrying the human *ETV6-RUNX1* fusion and from wild-type littermates was obtained from the laboratory of Isidro Sánchez-García (Institute of Molecular and Cellular Biology of Cancer, University of Salamanca, Salamanca, Spain). First, the human *ETV6-RUNX1* cDNA was inserted into the ClaI site of the pLy6 vector [142] and amplified. Following the amplification, the cDNA was cut from the vector using NotI and injected into fertilized eggs from a cross between CBA [143] and C57BL/6J mice (CBA x C57BL/6J). This NotI fragment was, with the exception of the inserted cDNA, homologous to the *Ly6a* locus on chromosome 15, which had previously been known

as *Sca1*. By homologous recombination, the human *ETV6-RUNX1* cDNA was entered into the murine *Ly6a* locus and put under the control of its promoter (Figure 3.1).

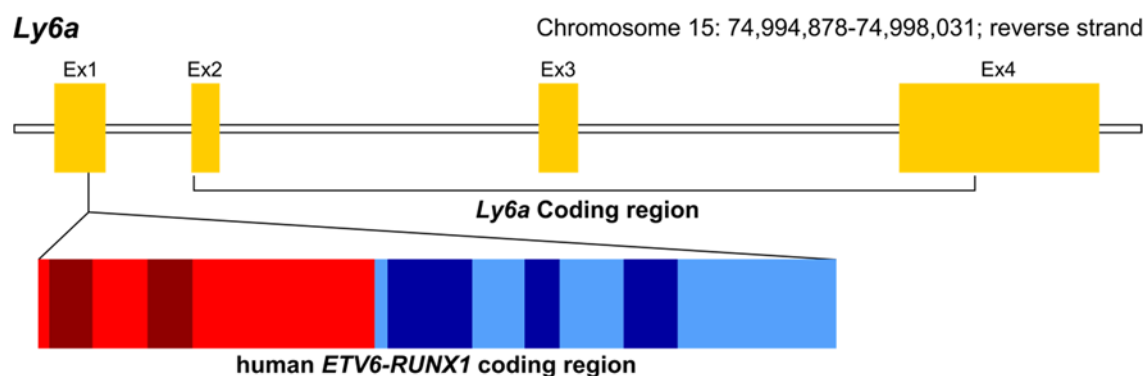


Figure 3.1: Insertion of the human *ETV6-RUNX1* cDNA into the *Ly6a* locus. The murine *Ly6a* locus is shown. The human *ETV6-RUNX1* coding region was inserted in exon 1, 5' of the *Ly6a* coding region. *ETV6-RUNX1* is displayed in the same size scale as *Ly6a* with the *ETV6* part in red and the *RUNX1* part in blue. *ETV6-RUNX1* exons are displayed with alternating color intensities.

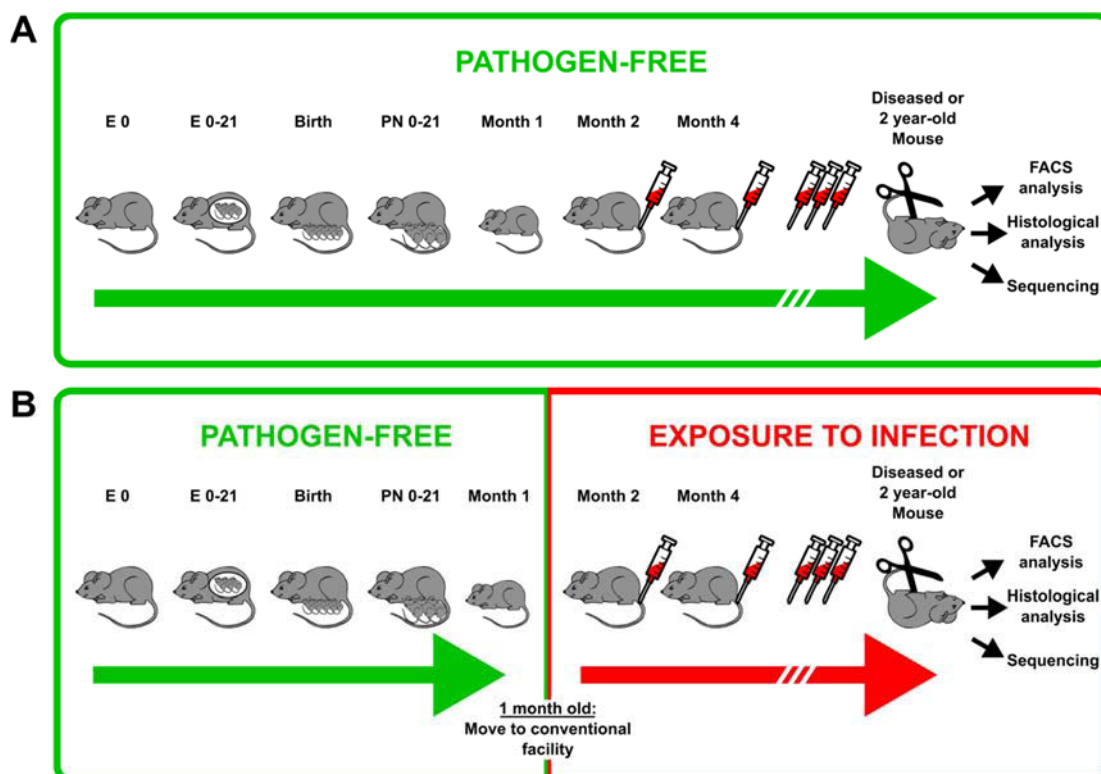


Figure 3.2: Keeping of the mice. **(A)** Mice were born and kept in a specific pathogen free facility (green). Bimonthly bleedings were performed to monitor the health status. Mice were sacrificed after two years of age or at signs of sickness. **(B)** Initially, mice were born and kept as in (A), but the mice were moved to a conventional facility (red) at one month of age, where they were exposed to common pathogens. E0 = 0 days old embryo, E 0-21 = 0-21 days old embryo, PN 0-21 = post natal, 0-21 days old. Adapted from [144].

The mice were born in a specific pathogen free facility (SPF) and either kept there for two years until they were sacrificed or moved to a conventional facility (CF) at one month of age (Figure 3.2). Bimonthly bleedings were performed in order to monitor their health status on a regular basis. At signs of sickness or after two years of age, the mice were sacrificed. Six leukemic mice were whole exome sequenced (Table 3.7), and DNA from 22 mice was used for DNA methylation analysis (Table 3.8).

Table 3.7: Information on mice that were whole exome sequenced. All mice suffered from leukemia. The DNA was obtained from the tissue with the highest blast count.

Mouse ID	Housing	Sex	Source of DNA	Blasts [%]
J406	conventional	female	peripheral blood	43.0
J408	conventional	female	peripheral blood	56.0
K213	conventional	female	spleen	50.0
K955	conventional	female	spleen	22.0
R222	conventional	female	spleen	33.0
S825	conventional	female	peripheral blood	84.4

Table 3.8: Information on mice that were subjected to DNA methylation analysis. Pro-/pre-B cells were identified by the surface markers IgM⁺ B220^{low}, pre-B cells by IgM⁺ B220^{low} CD25⁺. SPF = specific pathogen free facility, CF = conventional facility.

Mouse ID	Housing	ETV6-RUNX1 Status	Health Status	Sex	Source of DNA
T447	SPF	ETV6-RUNX1 ⁺	healthy	female	pro-/pre-B cells from bone marrow
T448	SPF	ETV6-RUNX1 ⁺	healthy	female	pro-/pre-B cells from bone marrow
T450	SPF	ETV6-RUNX1 ⁺	healthy	female	pro-/pre-B cells from bone marrow
R583	SPF	ETV6-RUNX1 ⁺	healthy	female	pro-/pre-B cells from bone marrow
R585	SPF	ETV6-RUNX1 ⁺	healthy	female	pro-/pre-B cells from bone marrow
R591	SPF	ETV6-RUNX1 ⁺	healthy	female	pro-/pre-B cells from bone marrow
T020	SPF	ETV6-RUNX1 ⁺	healthy	female	pro-/pre-B cells from bone marrow
R225	CF	ETV6-RUNX1 ⁺	healthy	female	pro-/pre-B cells from bone marrow
R256	CF	ETV6-RUNX1 ⁺	healthy	female	pro-/pre-B cells from bone marrow
K953	CF	ETV6-RUNX1 ⁺	healthy	male	pro-/pre-B cells from bone marrow
K954	CF	ETV6-RUNX1 ⁺	healthy	male	pro-/pre-B cells from bone marrow
J406	CF	ETV6-RUNX1 ⁺	leukemic	female	total bone marrow
J408	CF	ETV6-RUNX1 ⁺	leukemic	female	total bone marrow
S825	CF	ETV6-RUNX1 ⁺	leukemic	female	total bone marrow
WT1	SPF	wild-type	healthy	female	pre-B cells from bone marrow
WT2	SPF	wild-type	healthy	female	pre-B cells from bone marrow
WT3	CF	wild-type	healthy	female	pre-B cells from bone marrow
WT4	CF	wild-type	healthy	female	pre-B cells from bone marrow
WT5	SPF	wild-type	healthy	female	pre-B cells from bone marrow
WT6	SPF	wild-type	healthy	female	pre-B cells from bone marrow
WT7	CF	wild-type	healthy	female	pre-B cells from bone marrow
WT8	CF	wild-type	healthy	female	pre-B cells from bone marrow

All animal experiments have been approved by the Bioethics Committee and the Superior Council of Scientific Investigations of the University of Salamanca. They were conducted according to Spanish and international guidelines.

3.6 Enzymes

Table 3.9: Restriction enzymes

Name	Restriction site	Company	Order No.
MfeI-HF	C AATTG	New England Biolabs, Frankfurt/Main, Germany	R3589
SacI-HF	GAGCT C	New England Biolabs, Frankfurt/Main, Germany	R3156

Table 3.10: Further enzymes and enzyme containing kits

Name	Company	Order No.
BigDye Terminator v1.1 Cycle	Applied Biosystems, Waltham, USA	4337450
Brilliant II SYBR Green QPCR Master Mix	Agilent, Santa Clara, USA	600828
DNase I	QIAGEN, Hilden, Germany	79254
Exonuclease III	New England Biolabs, Frankfurt/Main, Germany	M0206
LongAmp Hot Start 2x Master Mix	New England Biolabs, Frankfurt/Main, Germany	M0533
Phusion HF PCR Maser Mix	New England Biolabs, Frankfurt/Main, Germany	M0531
ROX Reference Dye	Agilent, Santa Clara, USA	600828
RNase Out	invitrogen, Carlsbad, USA	10777019
SuperScript™ III	invitrogen, Carlsbad, USA	18080044
Reverse Transcriptase		
T4 DNA Ligase	New England Biolabs, Frankfurt/Main, Germany	M0202
T4 DNA Ligase	Promega, Madison, USA	M1801

3.7 Chemicals

Table 3.11: General chemicals

Name	Company	Order No.
Acetic Acid	Merck, Darmstadt, Germany	160305
Agarose	Biozym, Hessisch Oldendorf, Germany	840004
Ampicillin	Sigma-Aldrich, St. Louis, USA	A9393
Bromophenol Blue	Sigma-Aldrich, St. Louis, USA	B-8026
Calcium Chloride Dihydrate	Merck, Darmstadt, Germany	208290
Dynabeads CD19 Pan B	invitrogen, Carlsbad, USA	11143D
Dynabeads MyOne		
Streptavidin T1	invitrogen, Carlsbad, USA	65601
Ethanol	VWR Chemicals, Fontenay-sous-Bois, France	20821.330
Ethidium Bromide Solution	Sigma-Aldrich, St. Louis, USA	E1510-10ML
Glycerol (85%)	Merck, Darmstadt, Germany	356352
Isopropanol	VWR Chemicals, Fontenay-sous-Bois, France	20842.330
LB Agar	Roth, Karlsruhe, Germany	X965.2
LB Medium	Roth, Karlsruhe, Germany	X964.3
Manganese(II) Chloride Dihydrate	Merck, Darmstadt, Germany	105934
MOPS	Ambion, Thermo-Fisher, Waltham, USA	AM9570
Nuclease Free Water	Ambion, Thermo-Fisher, Waltham, USA	AM9937
Potassium Acetate	Merck, Darmstadt, Germany	529543
Rubidium Chloride	Sigma-Aldrich, St. Louis, USA	R2252
Sodium Acetate	Sigma-Aldrich, St. Louis, USA	S8750-250G
Sodium Hydroxide	Merck, Darmstadt, Germany	199062
Sucrose	Sigma-Aldrich, St. Louis, USA	S-1888
50x TAE (Tris/Acetic Acid/EDTA) Buffer	Bio-Rad, Munich, Germany	161-0743
Xylene Cyanol	Sigma-Aldrich, St. Louis, USA	X-2751

3.8 Kits, Size Markers, and Other Materials

Table 3.12: Composition of 6x loading dyes. Either bromophenol blue or xylene cyanol was used for the loading dyes.

Substance	Bromophenol Blue Loading Dye	Xylene Cyanol Loading Dye
Sucrose	20 g	20 g
Bromophenol Blue	125 mg	-
Xylene Cyanol	-	125 mg
Deionized H ₂ O	to 50 ml	to 50 ml

Table 3.13: Kits, size markers, and other materials

Name	Company	Order No.
6x Purple Loading Dye	New England Biolabs, Frankfurt/Main, Germany	B70245
DyeEx® 2.0 Spin Kit	QIAGEN, Hilden, Germany	63206
GelPilot 1 kb Plus Ladder	QIAGEN, Hilden, Germany	239095
GelPilot 100 bp Plus Ladder	QIAGEN, Hilden, Germany	239045
HiSpeed Plasmid Maxi Kit	QIAGEN, Hilden, Germany	12663
innuPREP PCRpure Kit	Analytik Jena, Jena, Germany	845-KS-5010050
ipsogen <i>ETV6-RUNX1</i> Kit	QIAGEN, Hilden, Germany	675113
Monarch Plasmid Miniprep Kit	New England Biolabs, Frankfurt/Main, Germany	T1010
One Shot™ Stbl3™	invitrogen, Carlsbad, USA	C737303
Chemically Competent Cells		
pGEM®-T Easy Vector	Promega, Madison, USA	A1360
QIAamp DNA Blood Mini Kit	QIAGEN, Hilden, Germany	51106
QIAquick Gel Extraction Kit	QIAGEN, Hilden, Germany	28706
REPLI-g UltraFast Mini Kit	QIAGEN, Hilden, Germany	150035
RNeasy Mini Kit	QIAGEN, Hilden, Germany	74104
SureSelect ^{XT} Kit	Applied Biosystems, Waltham, USA	G9611A
TruSeq SBS Kit v3	Illumina, Inc., San Diego, USA	FC-401-3001

3.9 Oligonucleotides

Table 3.14: Primers for validation of SNVs. The *KDM6B* primers were used for both aberrations. Primers with all uppercase gene names are for human samples, the others for murine samples. T_M indicates the melting temperature, G/C the content of guanine and cytosine nucleotides, and *nt* stands for nucleotides.

Name	Sequence [5'-3']	Length [nt]	T_M [°C]	G/C [%]
Ebf1-f	GTCGTGGTGTCTACACAG	19	59.5	58
Ebf1-r	ATGATTCGCCTACCATGTTCC	21	59.5	48
Kdm5c-Seq-f	AGGTTAGAGGGACTCTTCAG	20	58.4	50
Kdm5c-Seq-r	CCTCACATCAACATACCCAG	20	58.4	50
KDM5C-Seq-f	CTTCTCCCTCCCTACCCCTTAT	20	64.2	55
KDM5C-Seq-r	ATTTACCAGCCTCCAGAACTCC	20	62.1	50
KDM6A-Seq-f	TATTGTCTTCATCAGGCCTGCTGAGC	26	67.9	50
KDM6A-Seq-r	AGAGAAACCAACAGTGGAGAGGGAAG	26	67.9	50
KDM6B-Seq-f	CACCTCTTGTAACCCCTGACTCTTG	24	66.9	54
KDM6B-Seq-r	TGGTAGGGCTGGTGGTGGCTTC	22	67.9	64

Table 3.15: GIPFEL primers for *ETV6-RUNX1*. Primers are sorted by their names, and the respective (color coded) primer bundle is given. Note that there is no primer RUNX1-S9f as the *SacI* restriction sites for RUNX1-S9f and RUNX1-S10f were only 4 bp apart, which did not allow for a primer to be designed. All primers are named according to the *SacI* (S) restriction site they were designed for. The abbreviation *rev-n* stands for nested reverse primers, *rev* for reverse primers, and *ctrl* for *RUNX1* reaction control primers. T_M indicates the melting temperature, G/C the content of guanine and cytosine nucleotides, and *nt* stands for nucleotides.

Name	Sequence [5'-3']	Length [nt]	T_M [°C]	G/C [%]	Bundle
RUNX1-S1f	CAGAGGCAAGACGGGCTGATAACC	24	68.5	58	red
RUNX1-S2f	AGGGACTCATGGTGACGGGAGC	22	67.9	64	green
RUNX1-S3f	GACTCTATATTGGAACCTCGGAAACGC	27	68.2	48	yellow
RUNX1-S4f	TTATCTGGTGGGCTGTTAGGAGGCTC	26	69.5	54	red
RUNX1-S5f	GGTGTGTTTCATAGGGAAGTGGTTTTGC	28	68.5	46	yellow
RUNX1-S6f	CCCACACCCTAGTTTGCATCGGTTTG	26	69.5	54	green
RUNX1-S7f	GAGGTGGAAGTAGTCATTATGGGATAACC	29	69.1	45	green
RUNX1-S8f	TGGTGACAAGTTGCTTCAGGCTGATG	26	67.9	50	green
RUNX1-S10f	CCGGGATGACAACAGTTCAAGGAATAC	27	68.2	48	red
RUNX1-S11f	ACCAGGCACTTGACTCTTAGGATGTTTG	28	68.5	46	red
RUNX1-S12f	GTGTCATCTCAACCATGGAAAGGGTAC	27	68.2	48	blue
RUNX1-S13f	GGAGGACCTAGTGGGATGCAAGTG	24	68.5	58	yellow
RUNX1-S14f	CTGACTGGGCAGCTCCACTATGTC	24	68.5	58	red
RUNX1-S15f	CCTAGTGAGTTCAGTGTGGTTTTGTCAG	28	68.5	46	blue
RUNX1-S16f	AGTGAGCTGGGGAATCCATTCAAGTG	26	67.9	50	yellow
RUNX1-S17f	CGTTTCTAGAAGGAGTGCCGGCAG	24	68.5	58	blue
RUNX1-S18f	GCTACCAGTCAAGTTTCCTTTCGGGC	26	69.5	54	green
RUNX1-S19f	AGACACAAAAGGTCAGACGCATGACAC	27	68.2	48	yellow
RUNX1-S20f	TTGGGGAGAGAAGGATGATGGTCTTG	26	67.9	50	green
RUNX1-S21f	AGTGGAAAAGGAGGTGGCAAGTACAG	26	67.9	50	yellow
RUNX1-S22f	AAGGAAAGAAGCTAGTTGGGGTAGCG	26	67.9	50	blue
RUNX1-S23f	AACAGAGAAGTCGCAATAGTGCAGCAG	27	68.2	48	blue
RUNX1-S24f	TCTCATGTTTTCCAGTTGCTTAGGCGTG	28	68.5	46	red
RUNX1-S25f	TGTCTTGGGGATCATTCTCGCCTGC	25	69.1	56	yellow
RUNX1-S26f	CATCAGGCAGAAAGGAAGAAGGGAAG	26	67.9	50	blue
RUNX1-S27f	TGCAGTCACTTAGAAGCACCCATCTG	26	67.9	50	red
RUNX1-S28f	CAGAAAATCTTGCAGCAGTCAGCTTGC	27	68.2	48	blue
RUNX1-S29f	TCGGTTAGCTTTCACGGAGGCAGTG	25	69.1	56	green
ETV6-S1r-n	GGAGGACGCTGGGCAGTGATTATTC	25	69.1	56	rev-n
ETV6-S2r-n	TGGCAGCACCTTGATGGTCAGCTAG	25	69.1	56	rev-n
ETV6-S3r-n	TAGGACTGTTCTGGGGCCATCTGTC	24	68.5	58	rev-n
ETV6-S1r	GATGTGGTTCATGTAAGCCAGGTCTTC	27	68.2	48	rev
ETV6-S2r	AAAGGGACAGTACCTCAAGGCAGAAG	26	67.9	50	rev
ETV6-S3r	GGGACATTATGCACCTGCTTGGGAG	25	69.1	56	rev
RUNX1-S0f	CTTGGTTCAGAGTGTATCTCACCCCTTG	27	68.2	48	ctrl
RUNX1-S1r	GTGAAGCCAGGGACACACACTAAATG	26	67.9	50	ctrl

Table 3.16: GIPFEL primers for *TCF3-PBX1*. Primers are sorted by their names, and the respective (color coded) primer bundle is given. All primers are named according to the MfeI (M) restriction site they were designed for. The abbreviation *rev-n* stands for nested reverse primers, *rev* for reverse primers, and *ctrl* for *PBX1* reaction control primers. T_M indicates the melting temperature, *G/C* the content of guanine and cytosine nucleotides, and *nt* stands for nucleotides.

Name	Sequence [5'-3']	Length [nt]	T_M [°C]	G/C [%]	Bundle
PBX1-M1f	ACTTAAAACTTGGCCCTAGAGTCCCTC	27	68.2	48	blue
PBX1-M2f	GTGAAGCTGAGAAAACATGTGTGTGCG	29	69.1	45	green
PBX1-M3f	ATGGTGTAAAGGATGGGGTGAGTGCTG	26	69.5	54	yellow
PBX1-M4f	CAAGGATGTAACCTGATGGGGAATAGTG	28	68.5	46	red
PBX1-M5f	TTGGTCTGTGCCTACATGTATGTGCTC	27	68.2	48	yellow
PBX1-M6f	CCAGGTGTGAGAGGCAGTGTAAACATC	26	69.5	54	blue
PBX1-M7f	CCATCTGTAAAATTGGGTGGCAGTGTAG	28	68.5	46	purple
PBX1-M8f	TCAAGGTAAAGCTCTGAAATCCCACGC	27	68.2	48	red
PBX1-M9f	GATGGTGTCCCAGGAGCAAGCAAC	24	68.5	58	red
PBX1-M10f	GGATTGACACAGACCAAGGGGTCTTG	26	69.5	54	red
PBX1-M11f	AGAGAGGTCAGGAAGGGAAAGGGATG	26	69.5	54	green
PBX1-M12f	CGATCCCACCATTGGTCAACACAGAC	26	69.5	54	blue
PBX1-M13f	TAGAATGAGGCAGAGCTTCCAGGATAG	27	68.2	48	blue
PBX1-M14f	GAGAGAGACTCAGCTTCAGTAACCTG	26	67.9	50	yellow
PBX1-M15f	CCCTAGGCTGAACGAAACGAAACTC	26	67.9	50	purple
PBX1-M16f	TCAAAGGCAGGAGTGAGATGTCATCC	26	67.9	50	red
PBX1-M17f	TCTCTGACCTTCTGTCTCTGGGCAC	25	69.1	56	purple
PBX1-M18f	CTCTGAGACACGGAACACTAGTTGTG	26	67.9	50	yellow
PBX1-M19f	TCCCTCTAGTCATATGTCTGTGCTGC	26	67.9	50	yellow
PBX1-M20f	CAAAGTATGTTGAAGTGTGTTGGCGCC	27	68.2	48	green
PBX1-M21f	GTACATAGGCGTTATCACCTCATTGGAAG	29	69.1	45	red
PBX1-M22f	GACCCCTTCTCTCTTAACCTATAATGGC	28	68.5	46	purple
PBX1-M23f	CAGGAACAAGAACAAGAAGGCATGTAGG	28	68.5	46	red
PBX1-M24f	AGCATCATAGGTGACAAGGGGCCATG	26	69.5	54	yellow
PBX1-M25f	TGCCTGGTGCATGTTAAGCCTCACAG	26	69.5	54	green
PBX1-M26f	TAGAACATGCAGAAATGCCCACCGTGG	26	69.5	54	blue
PBX1-M27f	TGAGTGTGTTGGTACCGATGTGTGGC	26	69.5	54	purple
PBX1-M28f	GTGAATGCCTGTGTGTACACTTAACGTG	28	68.5	46	green
PBX1-M29f	CTGGCGTCATAACAGAAGTAGTCACAG	27	68.2	48	blue
PBX1-M30f	TGGCATCTGAAGCACCTGTCCTAATG	26	67.9	50	purple
PBX1-M31f	CTGAGCTTGACCTTCCAGTCTGCTTC	26	69.5	54	purple
PBX1-M32f	TTGGCATTGTGACCAGGAGATCTATTGC	28	68.5	46	yellow
PBX1-M33f	GATGCAAGGGAACAATTACTGGACTGTTC	29	69.1	45	blue
PBX1-M34f	ACATTCTGAGGAAGATACATGGTTGTTCC	29	67.4	41	green
PBX1-M35f	TGGTGGTAATGGGGTTGGTGGGATAG	26	69.5	54	green
PBX1-M36f	ATACACACATGCACGTAACACCCCAAAG	28	68.5	46	blue
TCF3-M1r-n	AGCGAGATGAGACCGCAGGAGTG	23	68.3	61	rev-n
TCF3-M1r	CTGTGCTGGAGCGGGAAGTATGC	23	68.3	61	rev
PBX1-M0f	GCCCTGTAACCTGGGAGGTCTATTAG	26	69.5	54	ctrl
PBX1-M1r	AACCATCTGTGGAGTGCCCGGATTAG	26	69.5	54	ctrl

Table 3.17: Primers used for sequencing of cloned PCR products. The primers match to the SP6 and T7 promoter sequences of the pGEM-T Easy vector.

Name	Sequence [5'-3']	Length [nt]	T_M [°C]	G/C [%]
SP6	ATTTAGGTGACACTATAGAA	20	50.2	30
T7	TAATACGACTCACTATAGGG	20	54.3	40

3.10 Software and Hardware

Table 3.18: Software

Software	Company/available at
ApE – A Plasmid Editor	http://biologylabs.utah.edu/jorgensen/wayned/apel/
Burrows-Wheeler Aligner (BWA)	http://bio-bwa.sourceforge.net
CorelDRAW Graphics Suite X7	Corel, Ottawa, Canada
GraphPad Prism 5	GraphPad Software, Inc., La Jolla, USA
Illumina Genome Studio v2011.1	https://support.illumina.com/downloads.html
Inkscape 0.91	Free Software Foundation, Inc., Boston, USA
Intas GDS	Intas, Göttingen, Germany
MuTect	http://archive.broadinstitute.org/cancer/cga/mutect
Oligo Calc	http://biotools.nubic.northwestern.edu/OligoCalc.html
OligoAnalyzer 3.1	IDT, Coralville, USA - https://eu.idtdna.com/calc/analyzer
Partek Genomics Suite	Partek, St. Louis, USA
Picard	https://broadinstitute.github.io/picard/
SAMtools	http://samtools.sourceforge.net/
Sequencher 4.8	Gene Codes, Ann Arbor, USA
VarScan	http://varscan.sourceforge.net/

Table 3.19: Hardware

Hardware	Product	Company
Fresco 21	Centrifuge	Heraeus, Hanau, Germany
Mikro 22R	Centrifuge	Hettich, Tuttlingen, Germany
Multifuge 3SR+	Centrifuge	Heraeus, Hanau, Germany
Mini-Sub Cell GT Cell	Electrophoresis system	Bio-Rad, Hercules, USA
Wide Mini-Sub Cell GT Cell	Electrophoresis system	Bio-Rad, Hercules, USA
Thermomixer Comfort	Heating Block	Eppendorf, Hamburg, Germany
B 6030	Incubator	Heraeus, Hanau, Germany
Certomat BS-T	Incubator Shaker	Braun Biotech, Melsungen, Germany
Dynal	Magnet	invitrogen, Carlsbad, USA
GeneAmp PCR System 2700	PCR Cyclor	Applied Biosystems, Waltham, USA
T-Gradient Thermoblock	PCR Cyclor	Biometra, Göttingen, Germany
BioPhotometer	Photometer	Eppendorf, Hamburg, Germany
P25 Powerpack	Power Supply	Biometra, Göttingen, Germany
7900 HT Fast Real-Time PCR System	Real-Time PCR Cyclor	Applied Biosystems, Waltham, USA
CFX Connect™ Real-Time PCR Detection System	Real-Time PCR Cyclor	Bio-Rad, Hercules, USA
CFX384 Touch™ Real-Time PCR Detection System	Real-Time PCR Cyclor	Bio-Rad, Hercules, USA
3130 Genetic Analyzer	Sanger Sequencer	Applied Biosystems, Waltham, USA
HiSeq2500	Sequencer	Illumina, Inc., San Diego, USA
NanoDrop ND-1000	Spectrophotometer	PeqLab, Erlangen, Germany
UVC/T-M-AR	UV Hood	Kisker Biotech, Steinfurt, Germany
MultiSync EA190M	UV Imaging System	Intas, Göttingen, Germany
UVT 2035	UV Table	Herolab, Wiesloch, Germany
SpeedVac Plus SC110A	Vacuum Centrifuge	Savant Instruments, Holbrook, USA
UVS400A	Vacuum System	Savant Instruments, Holbrook, USA

4 Methods

4.1 GIPFEL

In order to screen patient material for translocations, a new method called “Genomic Inverse PCR for Exploration of Ligated Breakpoints” (GIPFEL) [145] was established. This method relies on digestion of DNA and its subsequent religation into circular DNA. The resulting ligation joints can then be validated by PCR based methods without

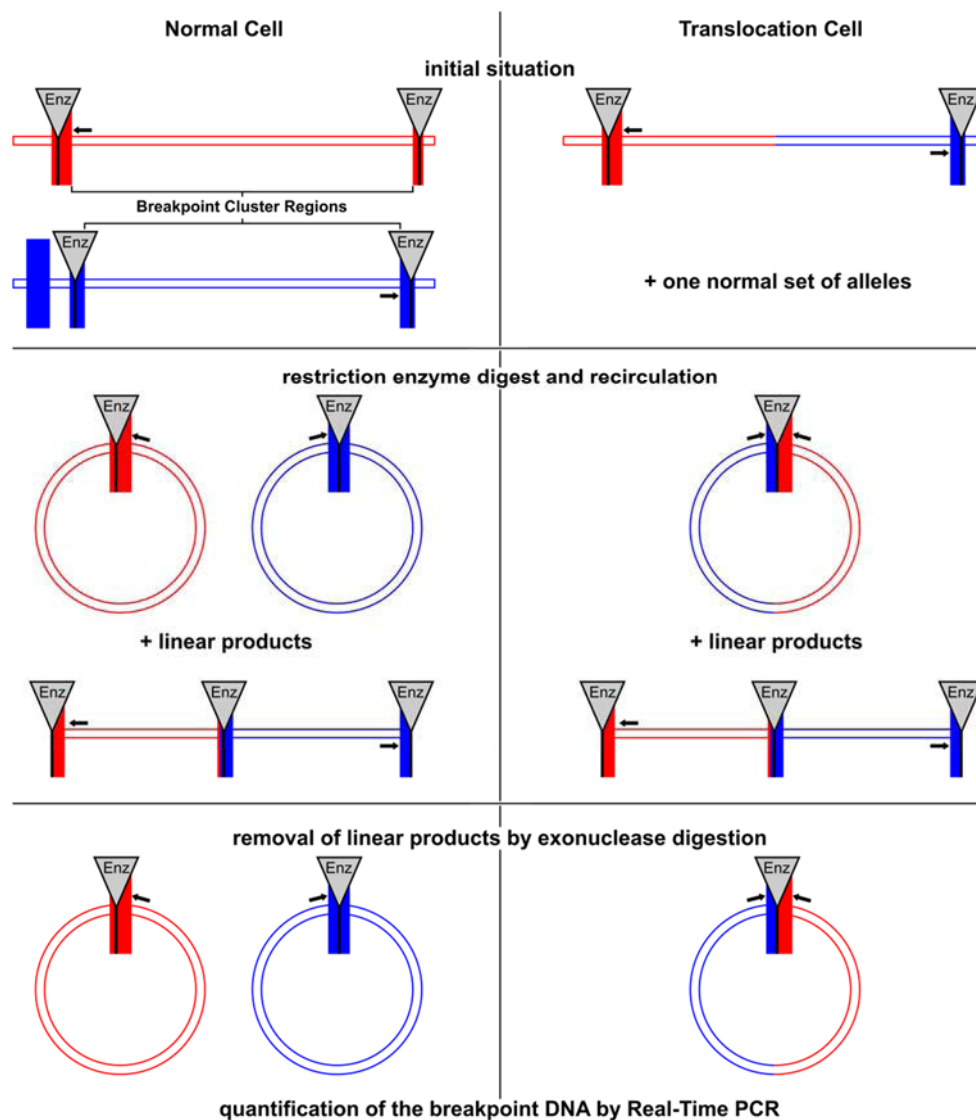


Figure 4.1: Principle of GIPFEL. The situation in a normal cell is shown on the left, the situation in a translocation bearing cell on the right. Unidirectional primers (arrows) were designed for every restriction site of the breakpoint cluster region. The DNA was digested with a restriction enzyme and ligated. This led to circular DNA and linear fragments. The circular DNA in a normal cell only consisted of material from one gene whereas the translocation bearing cell had one fragment containing the fusion point and hence material from both genes. It was only in this case that forward and reverse primers ended up on the same DNA fragment in the right orientation and were able to produce a PCR product.

knowing the break- and fusion points of the translocation. This was made possible by the design of unidirectional primers for each gene which can only amplify a product in case of a translocation (Figure 4.1).

With this principle in mind, an initial workflow was designed (Figure 4.2): At first, DNA was isolated and subsequently digested enzymatically. A clean-up step, ligation, and exonuclease digest followed before the products were used as template for a Real-Time PCR. This workflow was tested and optimized using the cell line REH for the t(12;21) translocation and the cell line 697 for the t(1;19) translocation.

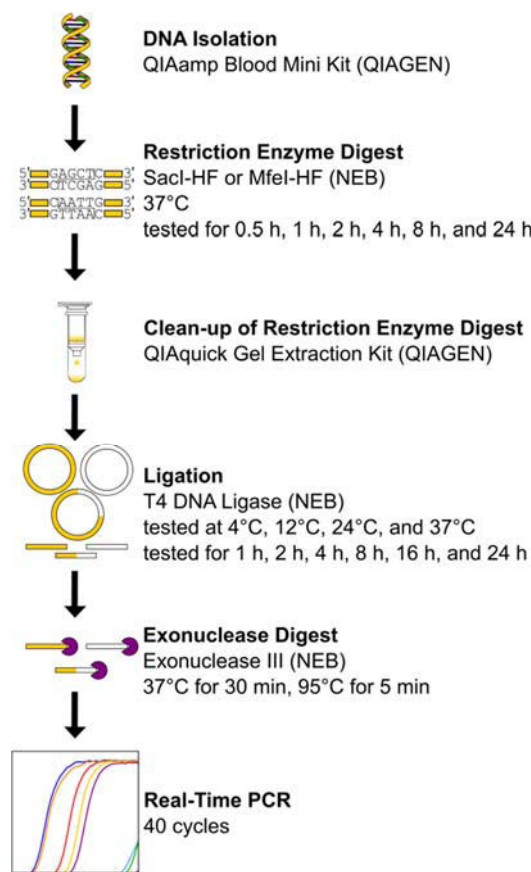


Figure 4.2: Initial workflow of GIPFEL. DNA was isolated and then enzymatically digested. The restriction enzyme digest was tested for six different durations. SacI-HF (NEB) was used for t(12;21) and MfeI-HF (NEB) for t(1;19). A clean-up step, religation, and an exonuclease digest followed. The ligation was tested at four different temperatures and for six durations. Eventually, a Real-Time PCR was carried out.

4.1.1 Cell Culture

The cell lines REH, 697, and HEK-293 were cultured as a source for DNA. The cell line REH [136-138] harbors the t(12;21)(p13;q22.3) translocation, the cell line 697 [139]

harbors the t(1;19)(q23;p13.3) translocation. Both are B cell precursor leukemia cell lines and were used as positive controls during the establishment of GIPFEL. Additionally, the human embryonic kidney cell line HEK-293 [140] was cultured as a source of DNA that is negative for both investigated translocations. All cell lines were acquired from DSMZ. They were cultured and cryopreserved as described in chapter 3.2.

4.1.2 DNA Isolation

4.1.2.1 DNA Isolation from Cell Lines

Since GIPFEL is a DNA based method, DNA was isolated from the cell lines REH and 697 in order to establish the GIPFEL method and as positive controls. In addition, DNA from the cell line HEK-293 was isolated and solely used as a translocation negative control.

Up to 5×10^6 cells were washed in 1 ml 1x PBS before DNA was isolated. The isolation was carried out using the QIAamp Blood Mini Kit (QIAGEN) according to manufacturer's instructions. The DNA was eluted in 200 μ l AE buffer (QIAGEN) after the columns were incubated for five minutes with AE buffer at room temperature.

4.1.2.2 DNA Isolation from Umbilical Cord Blood

To determine the frequency of the *ETV6-RUNX1* fusion in the healthy population, DNA from umbilical cord blood of 1,000 healthy Danish newborns (see chapter 3.4) was isolated. Per newborn, 0.5 to 6 ml of cord blood were drawn, equaling approximately 1.4 to 16.5×10^6 mononuclear cells. In cases with large volumes of blood withdrawal, the cells were divided on more than one tube.

Mononuclear cells were isolated and cryopreserved within 24 h after withdrawal [77]. These steps were carried out by the cooperation partner in Denmark, the group of Kjeld Schmiegelow. In order to improve sensitivity of GIPFEL, cells were CD19⁺ enriched as *ETV6-RUNX1* positive leukemias are CD19⁺ [75, 76, 146, 147].

The enrichment was carried out using Dynabeads CD19 pan B (Invitrogen). These magnetic beads are coated with CD19 antibodies which enable them to bind CD19⁺ cells. After binding, these cells can be isolated with the help of a magnet. The cord blood samples were thawed and washed by addition of the same volume of MACS running buffer (Miltenyi) and centrifugation for 10 min at 600 x g and 4°C. The supernatant was discarded, the cell pellets were resuspended in 1 ml MACS running buffer.

Prior to use, the beads had to be washed. To that end, for every sample 50 µl beads were transferred to a 2 ml tube and 1 ml MACS running buffer was added. This amount of beads was sufficient for 2.5×10^7 cells. The tubes were placed in a Dynal magnet (Invitrogen) for 1 min. While still in the magnetic field, the supernatant was discarded. The tubes were removed from the magnet, and the beads were resuspended in 50 µl of MACS running buffer.

Afterwards, the washed beads were added to the washed cells, mixed, and incubated at 4°C with gentle tilting for 20 min. Subsequently, the tubes were placed in the magnet for 2 min. The supernatant was discarded, and the bead bound cells were resuspended in 1 ml MACS buffer after removing the tube from the magnet. This washing step was repeated twice to achieve high purity. Finally, the cells were resuspended in 200 µl 1x PBS and subjected to DNA isolation.

The beads from the enrichment were not removed prior to the DNA isolation as the column held them back but was not clogged by the beads. The DNA isolation from CD19⁺ cells basically followed the same protocol as the one from cell lines. Only the following minor modifications were made: There was no need to wash the cells prior to the isolation as they already have been washed during CD19⁺ enrichment. Furthermore, the elution of the DNA was done using 81 µl AE buffer (QIAGEN) in order to account for the dead volume of the column and to receive a final eluate of approximately 80 µl which was then subjected to SmaI digest (see chapter 4.1.5).

4.1.3 Whole Genome Amplification

In order to be able to carry out multiple tests with patient material during the establishment phase of GIPFEL, DNA from patients with little amounts of DNA was whole genome amplified. The whole genome amplification (WGA) was carried out with the REPLI-g UltraFast Kit (QIAGEN) according to manufacturer's instructions in a T-Gradient Thermoblock PCR cycler (Biometra). After amplification, the DNA was diluted 1:4 with nuclease free water.

QIAGEN's REPLI-g kits use random hexamer primers that bind genomic DNA after denaturation. The elongation is carried out by a Phi 29 polymerase that does not dissociate from the DNA. This is supposed to prevent a bias in amplification. However, depending on the DNA integrity and hence on the fragment sizes, this bias can still occur.

4.1.4 Determination of Nucleic Acid Concentration

Nucleic acids have their absorption maximum at 260 nm, and the absorption is proportional to the amount of nucleic acid for commonly used DNA/RNA concentrations. Therefore it is possible to determine the amount of DNA/RNA in a sample by determining the absorption at 260 nm. This was done using the NanoDrop ND-1000 (PeqLab). The respective solvent in which the DNA/RNA was solved was used to normalize the measurement, and 1 µl of each sample was measured. The purity of DNA was determined by measuring of the absorption at 280 nm, the absorption maximum of proteins, in parallel.

4.1.5 Restriction Enzyme Digest

As part of the GIPFEL method, genomic DNA was digested enzymatically to achieve defined fragmentation. These fragments were later circularized (see chapter 4.1.7) to identify a translocation, if present.

In order to identify the t(12;21)(p13;q22.3) translocation, the genomic DNA was digested with *SacI*, whereas *MfeI* served as the restriction enzyme for the analysis of the t(1;19)(q23;p13.3) translocation. In both cases, high fidelity (HF) enzymes from NEB were used to reduce star activity. 2.5 µg of DNA were digested to establish the GIPFEL method. For the umbilical cord blood screening process, the entire vial of cells, a maximum of 1.8 ml, was subjected to the GIPFEL process. The exact amount of B cells was unknown but roughly estimated to be approximately 2.5×10^5 .

All digests were carried out at 37°C. After optimization, the DNA was digested for 2 h. The exact composition of the digests is shown in Table 4.1 and Table 4.2.

Table 4.1: Composition of DNA restriction enzyme reactions during the establishment phase of GIPFEL. *SacI*-HF was used for the detection of *ETV6-RUNX1* and *MfeI*-HF for the detection of *TCF3-PBX1*.

Substance	Volume
DNA	2.5 µg
<i>SacI</i> -HF/ <i>MfeI</i> -HF (NEB)	200 U
10x Cut Smart Buffer (NEB)	10 µl
H ₂ O	to 100 µl

Table 4.2: Composition of DNA restriction enzyme reactions during the umbilical cord blood screening for *ETV6-RUNX1*. 80 µl of DNA represent the entire DNA derived from CD19⁺ enriched B cells.

Substance	Volume
DNA	80 µl
<i>SacI</i> -HF (NEB)	200 U
10x Cut Smart Buffer (NEB)	10 µl

4.1.6 Purification of DNA after Enzymatic Digests

Nucleic acids can be purified with commercially available column based methods. These columns usually have a silica membrane at the bottom which binds the DNA at high salt concentrations and acidic pH. This allows for washing of the bound DNA with high salt buffers, removing impurities, such as agarose, enzymes, and primers that do not bind to the membrane. Elution then follows with low salt buffers at basic pH.

After the restriction enzyme digest, the DNA was purified to rule out interference with the ligation reaction that followed. The purification was conducted with the QIAquick Gel Extraction Kit (QIAGEN), following manufacturer's instructions for PCR clean-up. The only modifications to that protocol concerned the final DNA elution: 50 µl

preheated water were added to the columns, incubation at 60°C for 5 min followed prior to centrifugation. During an early phase of testing, an alternative column based kit for DNA purification (innuPREP PCRpure Kit from Analytik Jena) was also used according to manufacturer's protocol.

4.1.7 Ligation

After the DNA was digested with the respective restriction enzymes and column purified, it was religated to achieve circularization. The GIPFEL method is based on the detection of these ligation joints.

The entire DNA eluate from the clean-up of the enzymatic digest was used for the ligation. Table 4.3 shows the composition of the ligation reactions, all of which were carried out at 24°C for 2 h.

Table 4.3: Composition of the ligation reactions. The 48 µl of DNA represent the entire eluate from the purification (chapter 4.1.6). The missing 2 µl represent the dead volume.

Substance	Volume
DNA	48 µl
10x T4 Ligase Buffer (NEB)	10 µl
T4 DNA Ligase (NEB)	200 U
H ₂ O	40 µl

4.1.8 Exonuclease Digest

After the ligation, residual linear DNA fragments remained in the preparation. As they may interfere with downstream PCR reactions, they were removed by an exonuclease digest. Exonuclease III was chosen, as it is active in T4 Ligase Buffer (NEB) which was still present from the ligation. Therefore, another purification step with loss of DNA and thereby sensitivity could be avoided.

100 U of Exonuclease III (NEB) were added to the sample directly after ligation, and incubation at 37°C for 30 min followed. After that, the exonuclease and the ligase were heat inactivated at 95°C for 5 min.

4.1.9 Alcohol Precipitation

DNA can be precipitated with salt, such as sodium acetate (NaAc), and alcohol. The alcohol deprives the DNA of its hydration and allows it to react with cations, such as sodium cations. In the presence of high molar concentrations of NaAc, the DNA precipitates as a salt.

4.1.9.1 Ethanol Precipitation

In order to purify and concentrate DNA, 1/10 volume (v/v) of 3 M NaAc (pH 5.2) solution and three volumes (v/v) of absolute ethanol were added to the DNA and the preparation was mixed. The samples were then incubated at -20°C for at least 45 min. After the incubation, the DNA was pelleted by centrifugation in a table top centrifuge (Mikro 22R, Hettich) at 4°C and 23880 x g for 20 min, and the supernatant was discarded. The pellet was washed in 400 µl 70% (v/v) ethanol, a second centrifugation step followed. The supernatant was discarded as well and the pellet dried in a vacuum centrifuge (SC110A, Savant Instruments). Eventually, the dried pellet was resuspended in 10 µl nuclease free water. For the cord blood screening, the pellets were resuspended in 25 µl nuclease free water.

4.1.9.2 Isopropanol Precipitation

For the DNA precipitation, 1/10 volume (v/v) of 3 M NaAc (pH 5.2) solution and 3/5 volumes (v/v) of isopropanol were added, the preparation was mixed and incubated at room temperature for 10 min. A centrifugation step at 4°C and 23880 x g followed for 30 min (Mikro 22R, Hettich). The supernatant was discarded, and the pellet was washed in 200 µl 70% (v/v) ethanol. Subsequently, the sample was centrifuged again for further 15 min. This washing step was repeated once, the supernatant was discarded. Afterwards, the pellet was dried in a vacuum centrifuge (SC110A, Savant Instruments) and resuspended in 10 µl nuclease free water.

4.1.10 PCR

PCR [148] was used to detect the ligation joints produced by recircularization. To this end, the DNA was pre-amplified by 25 cycles of PCR followed by a Real-Time PCR.

4.1.10.1 Primer Design

All primers were designed to have a G/C content of approximately 50% but no less than 30% and no more than 70%. The melting temperature of the primers was set to be 2-4°C above the annealing temperature used in the PCR. These properties were checked with the online tool Oligo Calc [149]. The oligonucleotides were tested for potential adverse effects, such as hairpin formation, 3' complementarity, and homo-dimers with the same tool. Furthermore, potential hetero-dimers of forward and reverse primers as well as these of multiplexed primers were precluded as far as possible by the use of the OligoAnalyzer 3.1 tool (IDT). Sequence specificity was assured by the Primer Blast tool (NCBI), and primers were ordered from MWG Eurofins Genomics (Ebersberg, Germany). The primers for the detection of *ETV6-RUNX1* are presented in Table 3.15, those for the detection of *TCF3-PBX1* in Table 3.16.

4.1.10.2 Pre-amplification PCR

Ethanol precipitated DNA was subjected to a first round of PCR. During the establishment of GIPFEL, 2.5 µg of DNA were used at the start, and half of the DNA solution was used for each PCR reaction. A total of five and six reactions were needed for the detection of the t(12;21) and t(1;19) translocations, respectively. Therefore, it was necessary to use three samples. During the screening of umbilical cord blood, the DNA samples were split in five to make five PCR reactions possible.

Each PCR was run with multiplexed primers. To detect the *ETV6-RUNX1* fusion, a total of 36 primers were used. Three reverse primers covered the *ETV6* BCR, and 28 forward primers were spread over the *RUNX1* BCR. The forward primers were pooled in four color coded groups of seven primers each, and the reverse primers were also

pooled together. The remaining two primers amplified a product outside of the *RUNX1* BCR and served as a control for a successful reaction (Table 3.15).

For the conformation of *TCF3-PBX1*, 39 primers were needed. The *TCF3* BCR could be covered with a single reverse primer, whereas the *PBX1* BCR required 36 forward primers. These were divided into four groups of seven and one group of eight primers. In addition, two primers produced a control product outside of the *PBX1* BCR (Table 3.16)

DNA polymerase and buffer were provided by Brilliant II SYBR Green Master Mix (Agilent), because the products later served as template for a Real-Time PCR (see chapter 4.1.10.3) and the conditions of the two PCR steps were supposed to be as identical as possible. Composition of the PCR is shown in Table 4.4, and conditions are shown in Table 4.5. PCRs were performed on T-Gradient Thermoblock (Biometra) and GeneAmp PCR System 2700 cyclers (Applied Biosystems).

Table 4.4: Composition of the pre-amplification PCRs. Each primer of every bundle was used at a concentration of 1 μ M.

Substance	Volume
2x Brilliant II SYBR Green Master Mix (Agilent)	12.5 μ l
1 μ M each Forward Primers	2.5 μ l
1 μ M each Reverse Primers	2.5 μ l
DNA	5 μ l
H ₂ O	2.5 μ l

Table 4.5: Conditions of the pre-amplification PCRs.

Step	Temperature	Time
Initial Denaturation	95°C	10 min
Denaturation	95°C	15 s
Annealing	65°C	30 s
Elongation	72°C	30 s
Until Further Use	4°C	∞

4.1.10.3 Real-Time PCR

Choosing a Real-Time PCR over a normal PCR offered some important advantages. First, it was possible to compare different dilutions of target DNA during the establishment phase. Relative quantification of Real-Time PCR products from different dilutions could be carried out. This would not have been possible with normal PCR

reactions. Second, this also allowed for quantification of positive results with the help of the *RUNX1* and *PBX1* controls. These controls were present in every cell and therefore allowed for relative quantification of Real-Time PCR products. Furthermore, the melt curves are an additional source of information whether or not a PCR product is the desired amplification product. The melt curves helped to identify primer dimers and also multiple products from a single PCR reaction.

The PCR products from the first PCR were used as templates for the Real-Time PCR, the same forward primer bundles were used. However, the reverse primers were substituted by nested primers laying several base pairs closer to the forward primer, in order to improve template specificity. The *RUNX1* and *PBX1* control primers, on the other hand, remained unchanged.

The Real-Time PCRs were performed on a 7900 HT Fast Real-Time PCR System (Applied Biosystems), a CFX Connect™ Real-Time PCR Detection System (Bio-Rad), or a CFX384 Touch™ Real-Time PCR Detection System (Bio-Rad). Table 4.6 shows the composition of the PCR and Table 4.7 the PCR conditions. Duplets and triplets were used for every sample.

When a sample yielded a positive result, a new Real-Time PCR was done, this time with demultiplexed forward primers. If one of the forward primers still produced a positive result, the PCR was repeated with this forward primer and demultiplexed reverse primers. This PCR was run in a normal T-Gradient Thermoblock PCR cycler

Table 4.6: Composition of the Real-Time PCRs. Every primer was used at a concentration of 1 μ M. Nested primers were used as reverse primers. Depending on the cycler used, the preparation was adjusted. A "-" indicates that the respective substance was not used.

Substance	7900 HT	CFX Connect™	CFX384 Touch™
2x Brilliant II SYBR Green Master Mix (Agilent)	12.5 μ l	12.5 μ l	5 μ l
200 nM ROX Reference Dye (Agilent)	0,375 μ l	-	-
1 μ M each Forward Primers	2.5 μ l	2.5 μ l	2 μ l
1 μ M each Reverse Primers	2.5 μ l	2.5 μ l	2 μ l
Template DNA (PCR product)	1 μ l	1 μ l	1 μ l
H ₂ O	6.125 μ l	6.5 μ l	-

Table 4.7: Conditions of the Real-Time PCRs. "Ramp" indicates a melt curve analysis.

step	temperature	time	
Initial Denaturation	95°C	10 min	
Denaturation	95°C	15 s	} 40x
Annealing	65°C	30 s	
Elongation	72°C	30 s	
	95°C	15 s	
Melt Curve Analysis	55°C	ramp	+0.5°C
	95°C ↓		

(Biometra) with the same conditions as the Real-Time PCR. The products were then analyzed on an agarose gel and sequenced if a specific DNA band presented.

4.1.11 Gel Electrophoresis

Depending on the expected product size, 1% to 2% (w/v) agarose gels were used, with higher agarose amounts for products smaller than 100 bp. The agarose was solved in 1x TAE buffer (Bio-Rad) which was also used as running buffer. Wide Mini-Sub Cell GT Cell and Mini-Sub Cell GT Cell chambers (Bio-Rad) were used for the gel electrophoresis.

The samples were weighted and dyed with 1/6 loading dye (purple loading dye, NEB; xylene xyanol, bromophenyl blue) and loaded onto the gel. To detect the DNA, 2 µg/ml ethidium bromide (Sigma-Aldrich) was added to the gel solution before casting. Ethidium bromide is a fluorescent dye that intercalates sequence independent between the bases of the DNA. After excitation with UV light (312 nm), DNA was visualized. The results were captured with the MultiSync EA190M System (Intas).

4.1.12 Gel Extraction

To purify PCR products for subsequent Sanger sequencing, they were run on an agarose gel. If the band had the expected size, it was cut from the gel using a sterile scalpel. The DNA was purified from the gel using the QIAquick Gel Extraction Kit (QIAGEN), following manufacturer's instructions. Subsequently, the DNA concentration was determined by photometric measurement (see chapter 4.1.3).

4.1.13 Sanger Sequencing

In total, 20 ng of DNA were used for sequencing. The composition of the sequencing PCR was completed by addition of 0.5 µl of either a forward or a reverse 10 µM primer, 4 µl Big Dye Terminator (Applied Biosystems), and nuclease free water to 20 µl. A PCR of 26 cycles followed: each cycle consisted of 96°C for 30 s, 55°C for 30 s, and 60°C for 4 min.

The PCR product was purified with the DyeEx 2.0 Kit (QIAGEN), according to manufacturer's instructions, and 20 µl HiDi (Applied Biosystems) were added. The samples then were sequenced on a 3130 Genetic Analyzer (Applied Biosystems), and the results were analyzed with the software Sequencher 4.8 (Gene Codes) and ApE (M. Wayne Davis).

4.1.14 Reverse Transcription Real-Time PCR

In order to validate positive results from the cord blood screening, *ETV6-RUNX1* positive and negative samples were checked for the presence of the *ETV6-RUNX1* transcript. To that end, RNA was isolated from cell lines REH and HEK-293 as well as from *ETV6-RUNX1* positive cord blood samples N005, N260, and 50 *ETV6-RUNX1* negative cord blood samples. The isolation was done with the RNeasy Mini Kit (QIAGEN), according to manufacturer's instructions and including a DNase digest with 30 U DNase I (QIAGEN). The concentration of the RNA was subsequently determined with NanoDrop ND-1000 (PeqLab), and 1 µg of RNA was reverse transcribed to cDNA, using SuperScript™ III Reverse Transcriptase (Invitrogen), according to manufacturer's instructions.

The newly synthesized cDNA was then used as a template for a Real-Time PCR in order to determine the amount of *ETV6-RUNX1* transcripts (Table 4.8 and Table 4.9). To ensure that only the desired product was amplified, the ipsogen *ETV6-RUNX1* Kit (QIAGEN), including a FAM-TAMRA-labeled probe and primers specific for *ETV6-RUNX1*, was used. This kit also comes with a FAM-TAMRA-labeled probe and specific

primers for *ABL1* as well as standards for both targets, allowing for normalization and quantification of the results. The transcription was normalized to 10^4 *ABL1* transcripts and water was used as a non-template control.

Table 4.8: Composition of the reverse transcription Real-Time PCRs. Each sample was amplified with *ETV6-RUNX1* and *ABL1* specific primers and probes. Besides the investigated cDNA, *ETV6-RUNX1* and *ABL1* standards as well as non-template controls for both genes were included.

Substance	Volume
2x Brilliant II SYBR Green Master Mix (Agilent)	12.5 μ l
<i>ETV6-RUNX1</i> or <i>ABL1</i> Probe & Primers (QIAGEN)	1 μ l
cDNA / <i>ETV6-RUNX1</i> or <i>ABL1</i> Standard (QIAGEN) / Non-Template Control (H ₂ O)	5 μ l
H ₂ O	6.5 μ l

Table 4.9: Conditions of the reverse transcription Real-Time PCRs.

Step	Temperature	Time
Initial Denaturation	95°C	10 min
Denaturation	95°C	30 s
Annealing and Elongation	60°C	30 s

} 50x

4.2 Additional Methods

4.2.1 Cloning

4.2.1.1 Generation of Chemically Competent Cells

One vial of One Shot™ Stbl3™ Chemically Competent *E. coli* cells (Invitrogen) was added to 15 ml LB medium and incubated at 37°C over night. Afterwards, the optical density at 600 nm (OD₆₀₀) was measured with a BioPhotometer (Eppendorf). The equivalent of 0.1 OD₆₀₀ was filled up to 200 ml with LB medium and incubated at 37°C until an OD₆₀₀ of 0.5 to 0.6 was reached.

The cells were then centrifuged at 6,600 rpm and 4°C for 5 min in a J2-21 centrifuge with a JA-12 rotor (both Beckman Coulter). Subsequently, the supernatant was discarded, the cells were resuspended in 68 ml buffer RF1 (Table 4.10), and incubated on ice for approximately 5 h. After a second centrifugation step at the same conditions, the supernatant was discarded, the cells were resuspended in 16 ml buffer RF2 (Table 4.11), and incubated on ice for 15-30 min. Eventually, the cells were split into 75 μ l aliquots on ice, shock frozen in liquid nitrogen, and stored at -80°C.

Table 4.10: Composition of buffer RF1. The pH was adjusted to 5.8 with 0.2 M acetic acid and the buffer was sterile filtered with a 0.2 µm filter.

Substance	Chemical Formula	Molarity / Percentage	Mass per 500 ml
Rubidium Chloride	RbCl	100 mM	6.05 g
Manganese(II) Chloride	MnCl ₂ x 2 H ₂ O	50 mM	4.05 g
Potassium Acetate	C ₂ H ₃ KO ₂	30 mM	1.47 g
Calcium Chloride	CaCl ₂ x 2 H ₂ O	10 mM	0.735 g
Glycerol (85%)	C ₃ H ₈ O ₃	15%	88.2 ml

Table 4.11: Composition of buffer RF2. The pH was adjusted to 6.8 with 0.5 M sodium hydroxide and the buffer was sterile filtered with a 0.2 µm filter.

Substance	Chemical Formula	Molarity / Percentage	Mass per 500 ml
Rubidium Chloride	RbCl	10 mM	0.6 g
MOPS	C ₇ H ₁₅ NO ₄ S	10 mM	1.05 g
Calcium Chloride	CaCl ₂ x 2 H ₂ O	75 mM	5.51 g
Glycerol (85%)	C ₃ H ₈ O ₃	15%	88.2 ml

4.2.1.2 Transformation

In order to sequence them, the PCR products from the validation were cloned into the vector pGEM-T Easy (Promega) (Supplemental Figure 10.1). To that end, the vector and the PCR products were ligated with 3 U T4 DNA ligase (Promega) at 4°C over night. Insert and vector were used at a molecular ratio of 3:1. The pGEM-T Easy vector allows for direct ligation of PCR products generated by polymerases that add a 3' terminal adenine to their product, because the vector is already linearized and has 3' thymine at either end. These 3' thymine then pair with the adenines.

For every transformation, one vial of chemically competent Stbl3 cells was thawed on ice and 3 µl of the ligated vector were added. The cells were incubated on ice for 30 min and heat shocked at 42°C for 30 s, allowing the vector to enter the bacteria. After the heat shock, the cells were incubated on ice for at least 2 min, 250 µl of pre-warmed LB medium were added. The cells were then shaken horizontally at 450 rpm in a Certomat BS-T incubator shaker (Braun Biotech) at 37°C for 1 h to allow the ampicillin resistance to develop. Eventually, 25 µl of the transformation were plated out on a pre-warmed LB agar plate containing 100 µg/ml ampicillin and incubated at 37°C over night in a B 6030 incubator (Heraeus).

4.2.1.3 Plasmid Preparation

For every transformation, several colonies were picked from the LB agar plates, transferred to 5 ml LB medium containing 100 µg/ml ampicillin, and incubated at 37°C and 450 rpm in a Certomat BS-T incubator shaker (Braun Biotech) over night. The plasmids were isolated using the Monarch Plasmid Miniprep Kit (NEB), according to manufacturer's instructions.

When larger amounts of plasmid were needed, 500 µl of the culture used for the first preparation were transferred to 250 ml LB medium containing 100 µg/ml ampicillin and incubated under the same conditions as the 5 ml sample. The plasmids were then isolated using the HiSpeed Plasmid Maxi Kit (QIAGEN), following manufacturer's instructions.

4.2.2 Next Generation Sequencing

This sequencing technique relies on incorporation of fluorescently labeled nucleotides. With every addition of a nucleotide, light is emitted and can be captured by a camera. As every base is labeled differently, it is possible to determine which nucleotide is incorporated. Next Generation Sequencing (NGS) sequences short reads, but sequencing is massively parallelized. This results in an immense increase in simultaneously sequenced bases which makes it possible to sequence the entire genome or the entire exome.

The sequencing was performed on an Illumina HiSeq2500 which uses four dyes to differentiate between the bases. NGS can be divided into four basic steps: library preparation, cluster generation, sequencing, and data analysis (Figure 4.3).

DNA from the *ETV6-RUNX1* positive ALL patient samples 1 through 11 (Table 3.5) was whole exome and whole genome sequenced. Additionally, DNA from four wild-type mice from either facility, SPF and CF, were subjected to whole exome sequencing. Furthermore, the exomes of six leukemic *ETV6-RUNX1* positive mice were sequenced.

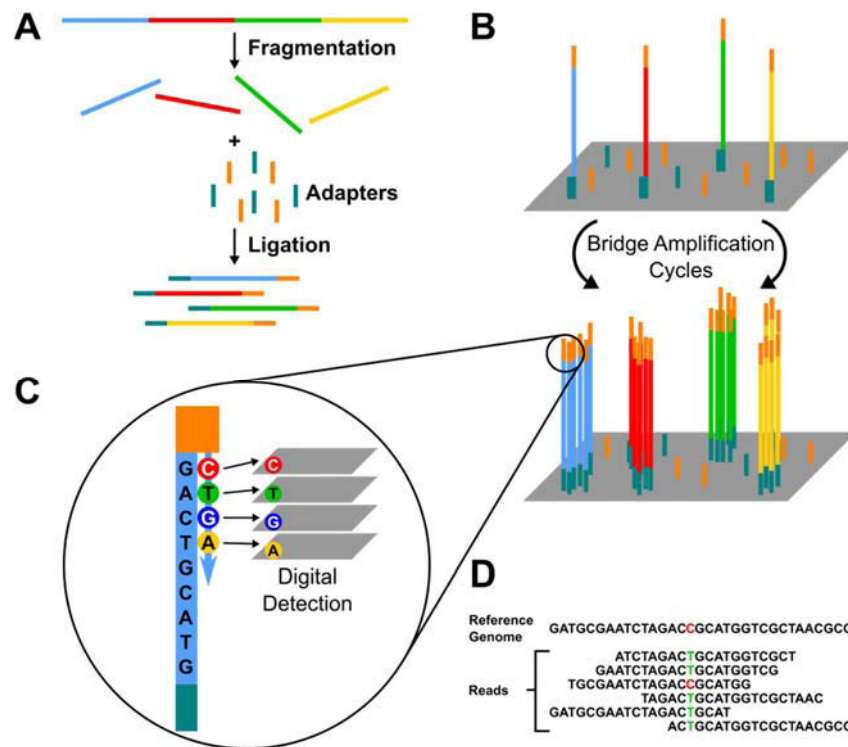


Figure 4.3: Workflow of Next Generation Sequencing. **(A)** Library preparation: the DNA is fragmented and 5' and 3' adapters are ligated to the fragments. **(B)** For the cluster generation, the fragments are loaded onto a flow cell (gray) and amplified into clusters by bridge amplification. **(C)** Sequencing by synthesis allows for the identification of the newly incorporated nucleotides. **(D)** The reads are aligned to a reference genome for data analysis. Adapted from [150, 151].

4.2.2.1 Library Preparation

In order to create sequencing libraries, the DNA was randomly fragmented by shearing (Figure 4.3 A). Subsequently, 5' and 3' adapters were ligated to the fragments (Figure 4.3 A). These adapters were important for the binding to the flow cell.

For whole exome sequencing only, an exon capture using the SureSelect^{XT} Human or Mouse Exon kits (Agilent) was performed prior to the adapter ligation. Adaptions were made according to Fisher *et al.* [152]. The exon capture allowed for the enrichment of exon sequences while intronic and intergenic sequences became underrepresented. The libraries were purified by binding to MyOne Streptavidin T1 Dynabeads (invitrogen).

4.2.2.2 Cluster Generation

The adapter bound fragments were PCR amplified and the original fragment was removed, leaving flow cell bound fragments. Following the binding, clusters were

generated by bridge amplification (Figure 4.3 B). This is a PCR technique in which the bound fragment bends until the other adapter can also bind to the flow cell. The fragment was then amplified; the two strands were separated again. Through repetition of this step, clusters were generated. Subsequently, the reverse strands were cleaved off, leaving only fragments with the same orientation.

4.2.2.3 Sequencing

Sequencing was carried out on a HiSeq2500 (Illumina) using a sequencing by synthesis approach and the TruSeq SBS Kit v3 (Illumina) according to manufacturer's instructions. Every base was detected as it was incorporated. The technique uses fluorescently labelled reversible terminators which were removed after every cycle. Prior to the removal, an image of the flow cell was taken, capturing the emitted light. As all reads in a cluster had the same base incorporated, the entire cluster emitted light of the same wavelength. After the sequencing of the forward strand, the synthesized strand was washed off, and the remaining one bent again, forming another bridge. By this, the reverse strands were produced with the help of a polymerase and the forward strands were washed off. Synthesis of a small part of the reverse adapter allowed for identification of the samples. Subsequently, the reverse strand was sequenced in the same manner as the forward strand. Read lengths of 2 x 100 bp were used.

4.2.2.4 Data Analysis

The fastq files were generated using the bcl2fastq conversion software (Illumina), and alignment to the reference genomes (GRCh37.p13 for human and GRCm38.p4 for mouse) was done with BWA version 0.7.4 [153]. The conversion steps were executed with SAMtools [154, 155] and followed by removal of duplicate reads by Picard (<https://broadinstitute.github.io/picard/>). Afterwards, the GATK 2.4.9 software was used for local alignment around indels, SNP calling, and recalibration. Resulting variation calls were annotated by the Variant Effect Predictor [156] using the Ensembl database

(v84). Subsequently, the data was imported to an in-house MySQL database, developed by Sebastian Ginzl (Hochschule Bonn-Rhein-Sieg, Bonn) and called Single Nucleotide Polymorphism (SNUPy). SNUPy was used for automatic and manual annotation, reconciliation, and data analysis. The prediction tools PolyPhen2 [157] and SIFT [158] are included in SNUPy and predicted the impact of the mutations. The tools MuTect [159] and VarScan [160] were used for the somatic calls. For further analysis, only entries with at least 9% difference in allele frequency between tumor and control were kept. Cancer related genes were taken from COSMIC's cancer gene consensus [161]. Analysis of the structural aberrations from whole genome sequencing was carried out in Münster by Martin Dugas' group (Institute for Medical Informatics, Westfälische Wilhelms-University Münster). Finally, conversion of the genomic positions from GRCh37.p13 to GRCh38.p7 was carried out with Ensembl's Assembly Converter (<http://www.ensembl.org/>).

4.2.3 Genome-wide Methylation Analysis

Generally, DNA methylation is very prominent in promoter regions, often expanding into the first exon and first intron [162]. Usually, promoter methylation represses gene expression, there are, however, exceptions where methylated promoters are active or unmethylated promoters are silent [162]. An altered methylation signature is a very common feature among tumor cells, interfering with gene regulation. For instance, the expression of tumor suppressors can be repressed by methylation. Therefore, the leukemic *ETV6-RUNX1*⁺ mice should be compared to their healthy counterparts and wild-type mice from SPF and CF regarding their DNA methylation. Mammals methylate cytosines of CpG dinucleotides in CpG islands [163] (Figure 4.4), which helps the DNA methyltransferases to scan DNA for hemimethylated CpGs during replication and methylate the newly synthesized strand.

The methylation of the murine DNA was measured with an Infinium

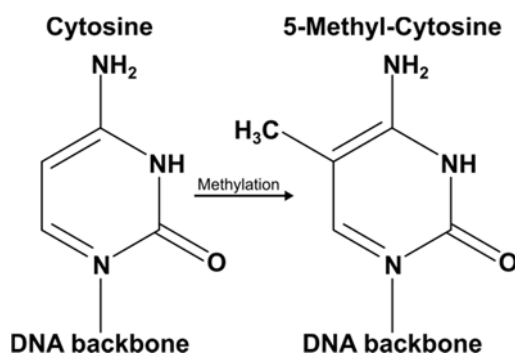


Figure 4.4: Methylation of cytosine. A methyl group is added to carbon 5 of cytosine, resulting in 5-methyl-cytosine.

HumanMethylation450 bead chip (Illumina). This chip analyzes the methylation status of 485,577 human CpG dinucleotides in parallel. Even though it was not designed for murine DNA, it can also be used for it as the DNA of both organisms shows similarities. However, only approximately 3% of probes bind uniquely to murine CpG dinucleotides [164]. For every sample, 1 μg of DNA was diluted in 40 μl H_2O , and the analysis was performed at the German Cancer Research Center (DKFZ) in Heidelberg.

Analysis of the raw data was performed with Partek Genomics Suite (Partek) in cooperation with Marc Remke's group (Junior Research Group "Pediatric Neuro-Oncogenomics", German Cancer Consortium and Department of Pediatric Oncology, Hematology, and Clinical Immunology, University Hospital, Düsseldorf). A principal component analysis (PCA) was carried out. This analysis introduces three new variables, termed principal components (PC) 1 to 3. These PCs are used to graphically display the differences between the investigated samples. This is achieved by keeping the differences that are present in the original variables and transforming them into only three PCs. The impact of each PC determines the sample's position in the three-dimensional graph. Here, the general methylation status was compared between leukemic *ETV6-RUNX1*⁺ mice, healthy *ETV6-RUNX1*⁺ mice, and wild-type mice, housed in SPF and CF.

5 Results

5.1 Establishment of the GIPFEL Method

The GIPFEL method, presented in this chapter, was established in cooperation with Elisa Füller and Robert K. Slany (Friedrich Alexander University, Erlangen). The most common translocations in childhood leukemia [19] were included in the establishment: *ETV6-RUNX1*, *TCF3-PBX1*, and three *KMT2A* translocations (*KMT2A-AFF1*, *KMT2A-MLLT3*, and *KMT2A-MLLT1*). The first two translocations were investigated by this work, whereas the *KMT2A* translocations were investigated in Erlangen. Eventually, the established method, including the used methods (chapter 4.1) and primers (Table 3.15 and Table 3.16), was published in 2014 [145].

5.1.1 Computational Groundwork

The GIPFEL method is based on enzymatic restriction of DNA and subsequent relegation. Consequently, it was necessary to identify restriction enzymes which best fit the demands of the GIPFEL method: the restriction sites should be distributed equally across the BCRs of the analyzed genes, and the number of restriction sites should be as little as possible without being separated by more than 30,000 bp. To this end, the genomic sequences of *ETV6* (NG_011443.1), *RUNX1* (NG_011402.2), *TCF3* (NG_029953.1), and *PBX1* (NG_028246.1) were acquired from NCBI (<http://www.ncbi.nlm.nih.gov>). The *ETV6* (NM_001987.4) BCR is located in intron 5 and the *RUNX1* (NM_001754.4) BCR spans introns 1 and 2 [63, 64] (Figure 5.1). For *TCF3* (NM_003200.3), the BCR is in intron 16, and the *PBX1* (NM_002585.3) BCR is in intron 2 [56, 165] (Figure 5.2).

The restriction enzyme with the best distribution across the BCRs of *ETV6* and *RUNX1* was *SacI* (Figure 5.1), the best fit for *TCF3-PBX1* was *MfeI* (Figure 5.2). After digestion with *SacI*, three DNA fragments spanning the *ETV6* BCR were left. For *RUNX1*, 29 fragments spanned the BCR after digestion with *SacI*. An additional fragment lay 3' of the BCR and was used as a control, as it had to be present in either cell, translocation

positive or negative. This control was placed in *RUNX1*, because the second *ETV6*

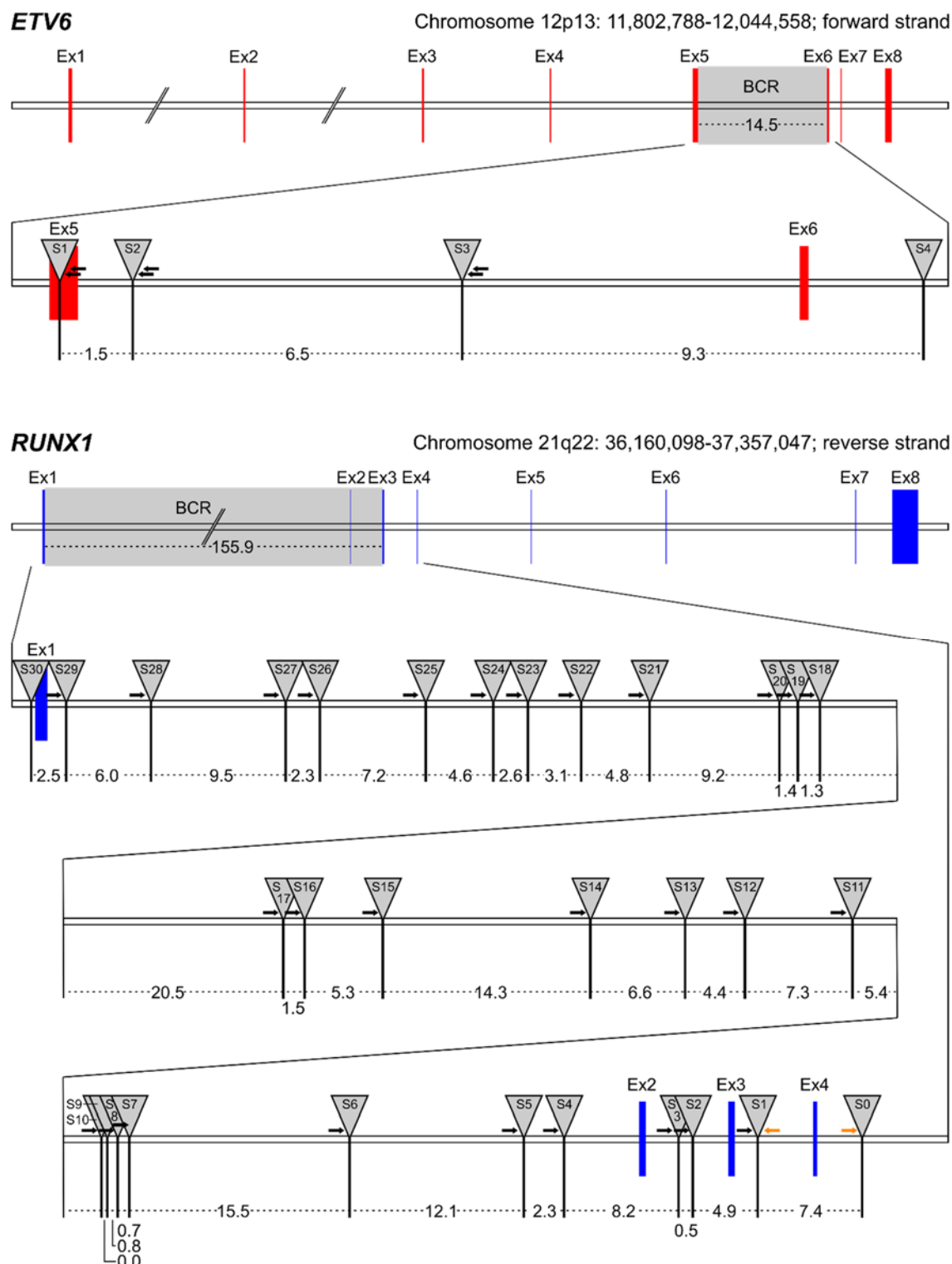


Figure 5.1: *Sacl* restriction sites in the BCRs of the *ETV6* and *RUNX1* genes. Red boxes mark *ETV6* exons, blue boxes mark *RUNX1* exons. The areas marked in gray are the breakpoint cluster regions (BCR), the segments in which the chromosomal breaks occur. This part is zoomed in below each gene. The gray triangles indicate *Sacl* restriction sites, every restriction site is numbered. Distances between restriction sites and length of the BCR are given in kb. Black arrows indicate primers, orange arrows the *RUNX1* control primers. All primers were named according to the restriction site they are closest to. For *ETV6* one reverse and one nested reverse primer was designed per *Sacl* site, for *RUNX1* a single forward primer was designed per site.

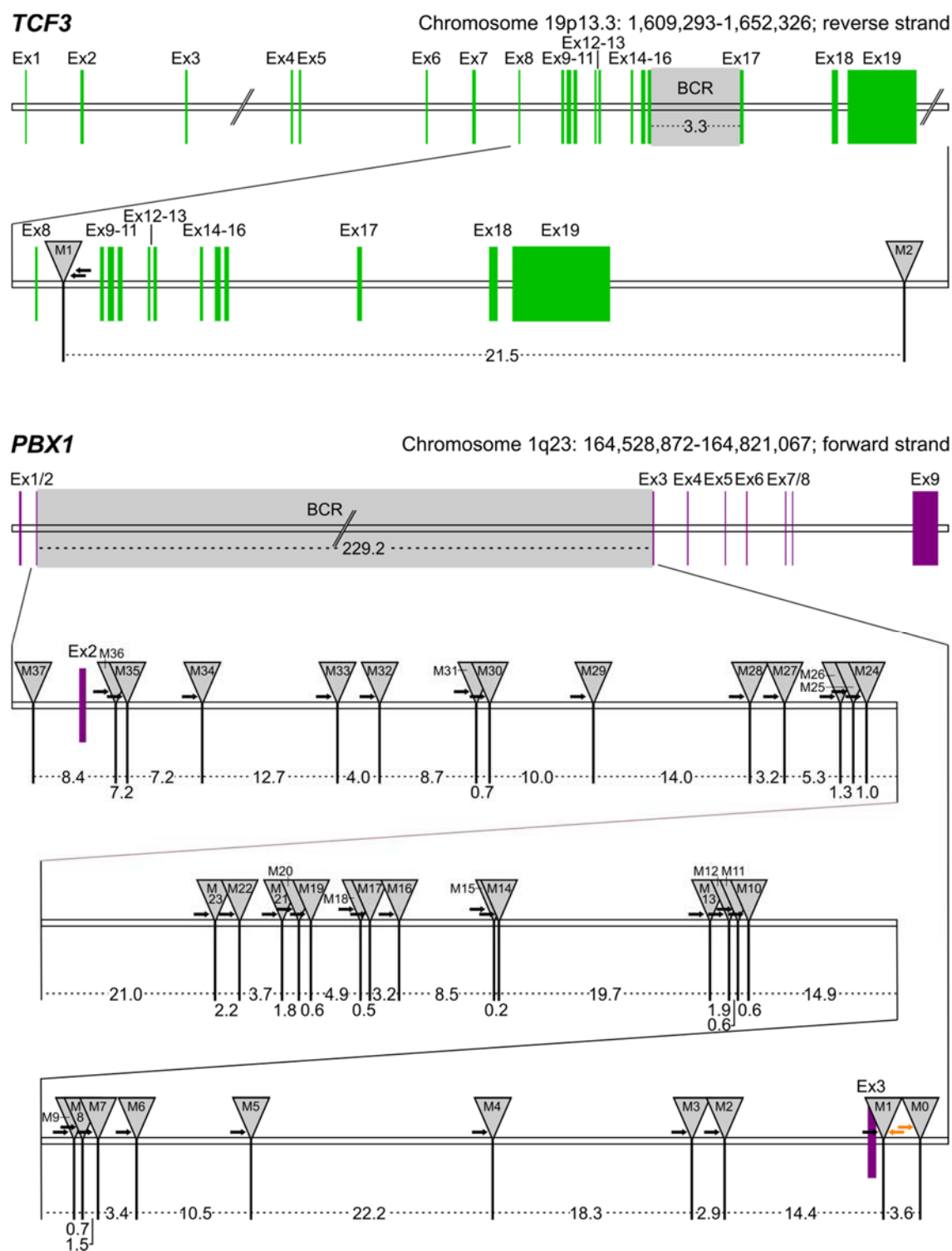


Figure 5.2: MfI restriction sites in the BCRs of the *TCF3* and *PBX1* genes. Green boxes mark *TCF3* exons, purple boxes mark *PBX1* exons. The areas marked in gray are the breakpoint cluster regions (BCR), the segments in which the chromosomal breaks occur. This part is zoomed in below each gene. The gray triangles indicate MfI restriction sites, every restriction site is numbered. Distances between restriction sites and length of the BCR are given in kb. Black arrows indicate primers, orange arrows the *PBX1* control primers. All primers were named according to the restriction site they are closest to. For *TCF3* one reverse and one nested reverse primer were designed, for *PBX1* a single forward primer was designed per MfI site.

allele is commonly deleted in *ETV6-RUNX1* positive leukemia patients [81]. In t(1;19) positive leukemia patients, the control was put 3' of the *PBX1* BCR and its 36 fragments after MfeI digest. For *TCF3*, only one fragment remained after MfeI digest. Unidirectional primers were designed for every restriction site, meeting the criteria stated in chapter 4.1.10.1. For *ETV6* and *TCF3*, only reverse primers were designed. All primers for *RUNX1* and *PBX1* were forward primers. In case of *ETV6* and *TCF3*, two primers per restriction site were present; one primer for the pre-amplification PCR and one nested primer for the semi-nested Real-Time PCR.

The nomenclature of the primers derived from their location: all primers were named after the restriction site for which they were designed. The *RUNX1* *SacI* sites 9 and 10 (Figure 5.1) were only 4 bp apart. Therefore, no primer was designed for site 9, because the fragment was too short for primer alignment. Numbering started at the 5' ends of the BCRs of *ETV6* and *TCF3* and at the 3' end of the BCRs of *RUNX1* and *PBX1*.

In all three translocations investigated in Erlangen, the 5' fusion partner was *KMT2A*, and the BCRs were small in comparison to those of the *ETV6-RUNX1* and *TCF3-PBX1* translocations. Therefore, it was possible to find a common restriction enzyme, *Bam*HI. However, *Bam*HI was not suited for the translocations investigated by this work: there was no *Bam*HI restriction site near the 5' end or inside the *TCF3* BCR and only one far 5' of the *ETV6* BCR.

5.1.2 Proof of Principle

To prove that the GIPFEL method works, the procedure was carried out with DNA from the cell lines REH and 697 which are known to carry the t(12;21) and t(1;19) translocations, respectively. At these early stages GIPFEL differed slightly from the optimized protocol: The DNA was digested, followed by a clean-up with the QIAquick Gel Extraction Kit (QIAGEN). The elution was done with water at room temperature as opposed to 60°C and with an elution step of 1 min instead of 5 min. A ligation step and

an exonuclease digest followed. Subsequently, the DNA was subjected to a single Real-Time PCR. As the fusion points of the cell lines were known, only the respective primers as well as the *RUNX1* and *PBX1* control primers were used. The procedure was done in parallel with a non-template control (NTC) and a negative control (NC). The cell line which does not carry the respective translocation served as NC, water as NTC. A preparation without the ligation step was used as an additional negative control. In this case, no amplification should take place as the DNA was not ligated, and amplification with the used primers should not be possible. Furthermore, the exonuclease digest should remove all linear DNA. However, in some cases unspecific background amplification took place (Figure 5.4).

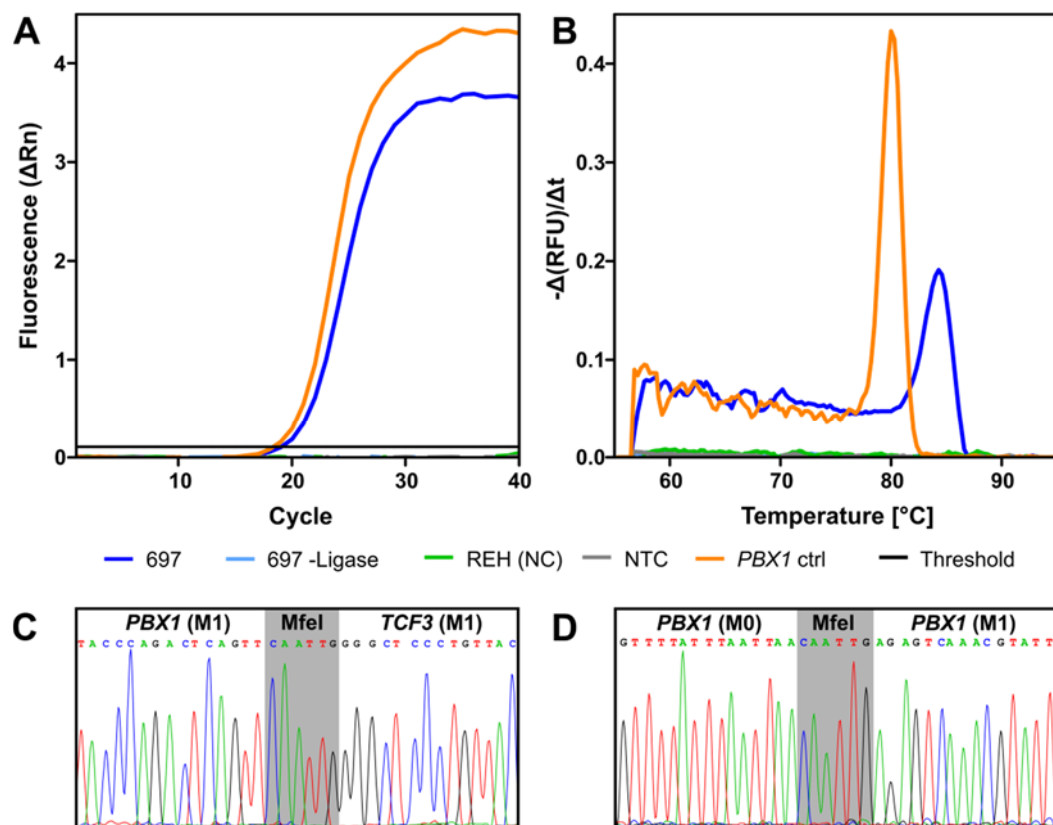


Figure 5.3: Proof of principle of GIPFEL for *TCF3-PBX1*. DNA of the cell line 697 was amplified with the primers PBX1-M1f and TCF3-M1r-n. These primers were also used for the negative controls: one sample without the ligation step (697 -Ligase), t(1;19) negative DNA from the cell line REH (NC), and a non-template control (NTC). A *PBX1* control (ctrl) product was amplified as positive control. **(A)** Amplification plot showing the amplification of the desired product (697) in dark blue and the *PBX1* control in orange. None of the negative control samples was amplified. **(B)** The dissociation plot shows a single main peak for both amplified products, indicating the amplification of one product in both cases. **(C)** Chromatogram of the sequencing for the detection of the translocation. The sequence of the ligation joint is shown in gray. **(D)** As in (C) but for the *PBX1* control. Rn = normalized reporter signal, RFU = relative fluorescent units.

The GIPFEL method was proven to work with DNA from both cell lines [145]. Amplification with PBX1-M1f and TCF3-M1r-n was only successful on 697 DNA, not on the NC, the NTC, or the sample without the ligation step (Figure 5.3 A). The dissociation curve (Figure 5.3 B) showed a single product for the amplification. Amplification with the *PBX1* control primers PBX1-M0f and PBX1-M1r also led to a single product (Figure 5.3 A-B). Both products could be validated by Sanger sequencing (Figure 5.3 C-D).

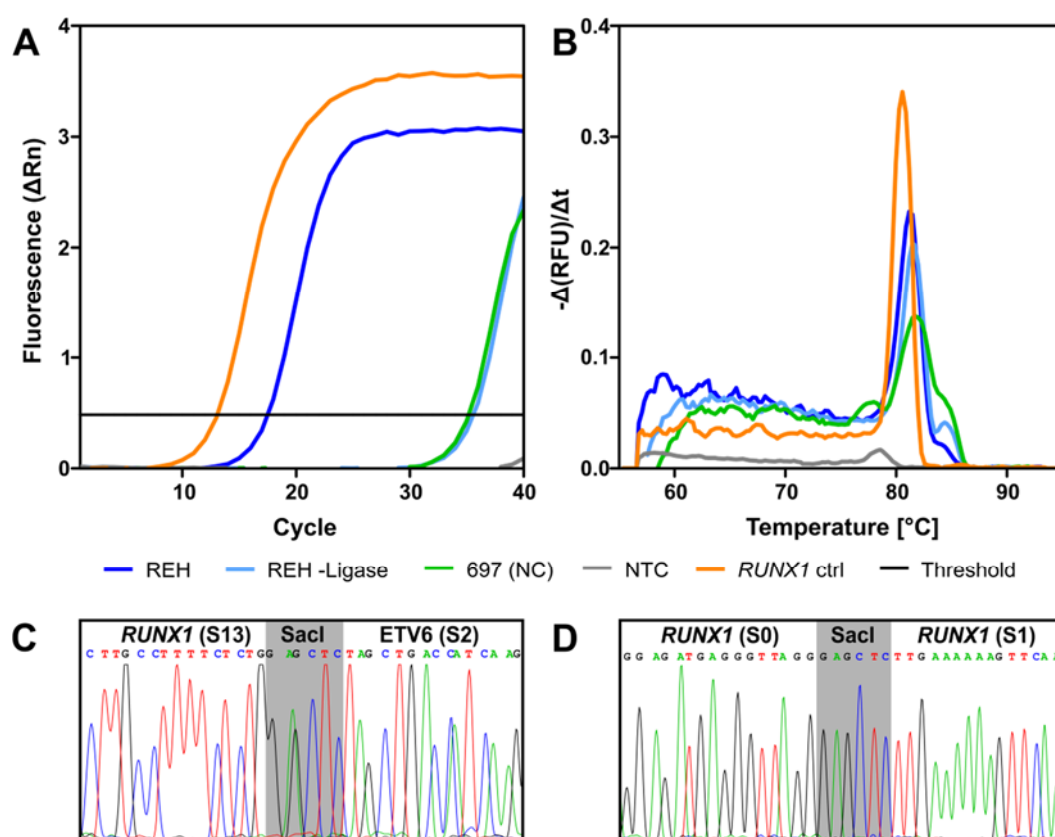


Figure 5.4: Proof of principle of GIPFEL for *ETV6-RUNX1*. DNA of the cell line REH was used with the primers RUNX1-S13f and ETV6-S2r-n. These primers were also used for the negative controls: one sample without the ligation step (REH -Ligase), t(12;21) negative DNA from the cell line 697 (NC), and a non-template control (NTC). A *RUNX1* control (ctrl) product was amplified as positive control for a successful reaction. **(A)** Amplification plot showing the amplification of the desired product (REH) in dark blue and the *RUNX1* control in orange. The negative control samples are amplified late. **(B)** The dissociation plot shows a single peak for the two desired amplified products, indicating the amplification of one product in both cases. The products seen for the negative controls cannot be validated by Sanger sequencing. **(C)** Sequencing result for the detection of the translocation. The ligation joint is shown in gray. **(D)** As in (C) but for the *RUNX1* control. Rn = normalized reporter signal, RFU = relative fluorescent units.

As for the DNA from 697, carrying the t(1;19) translocation, the DNA from REH, carrying the t(12;21) translocation, produced the desired result when amplified with

RUNX1-S13f and ETV6-S2r-n (Figure 5.4 A-C). However, in contrast to 697, the amplification plot (Figure 5.4 A) showed very late product amplifications for the NC and the sample without ligase. Both products seemed to have a comparable size with the target product, as indicated by the dissociation curve (Figure 5.4 B), but when subjected to Sanger sequencing, both samples could not be verified. This is also the case for the products which were amplified in some of the later experiments. When the *RUNX1* control primers RUNX1-S0f and RUNX1-s1r were used, the *RUNX1* control was amplified (Figure 5.4 A-D).

Employing two cell lines carrying the investigated translocations, this experimental setup repeatedly gave proof of principle evidence that the GIPFEL method works in principle with any given translocation.

5.1.3 Optimization of the Method

New steps were introduced into the protocol in order to optimize specificity and sensitivity of GIPFEL. The exonuclease digest was purified to reduce the sample volume and to remove buffers and enzymes, possibly interfering with the PCR reaction. In addition, a pre-amplification PCR was placed before Real-Time amplification to improve sensitivity. The Real-Time PCR itself became a semi-nested PCR to serve the same purpose. Further attempts at optimization involved the restriction duration, the ligation duration and temperature, clean-up steps, primer multiplexing, and the use of WGA.

5.1.3.1 Duration of Restriction Enzyme Digest

All restriction enzyme digests were carried out at 37°C. The duration of the restriction digests, however, was subject to optimization. DNA from the cell lines REH and 697 was digested with SacI-HF (NEB) and MfeI-HF (NEB), respectively. For each enzyme, six incubation times were tested: 0.5 h, 1 h, 2 h, 4 h, 8 h, and 24 h. Nine biological replicates per enzyme were used for every time point. The samples were divided prior

to PCR amplification in order to also amplify a *RUNX1* or *PBX1* control, respectively. This control was used for normalization.

The restriction with *SacI*-HF was mostly independent of time (Figure 5.5). Samples digested for 1 h, 2 h, and 4 h were slightly, but not significantly, better amplified than the one digested for 0.5 h (Figure 5.5). Starting with 8 h, the product amplification got weaker, probably due to off-target activity of the enzyme, resulting in less DNA template for the PCR.

The restriction with *MfeI*-HF, however, was highly dependent on the restriction time (Figure 5.5). Prolonged restriction led to better results, peaking at 2 h. After that, results got worse, as more off-targets were cut by the enzyme.

As a restriction time of 2 h turned out to be the best for *MfeI* digestion and also to be one of the best for *SacI* digestion, all further digests were carried out for 2 h.

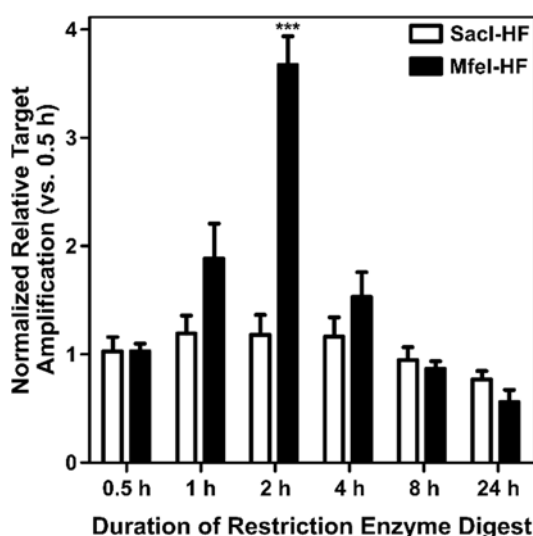


Figure 5.5: Results of the Real-Time PCR for the optimization of restriction enzyme digests. DNA from REH was digested with *SacI*-HF (NEB); DNA from 697 was digested with *MfeI*-HF (NEB). The digests were carried out for 0.5 h, 1 h, 2 h, 4 h, 8 h, and 24 h. Afterwards, the samples were subjected to the GIPFEL procedure, normalized to the respective *RUNX1* or *PBX1* control, and the quantification was compared to 0.5 h. Mean values and standard deviations of nine measurements are shown. *MfeI* performed best when used for 2 h (Dunnett's Multiple Comparison Test, *** $P < 0.001$).

5.1.3.2 Clean-up of Restriction Enzyme Digest

Two different kits were compared for the clean-up of the restriction digest: the QIAquick Gel Extraction Kit (QIAGEN) and the innuPrep PCRpure Kit (Analytik Jena). This clean-

up was important to rule out any interference from buffers or enzymes in the following ligation reaction.

DNA from the cell line REH, isolated from nine biological replicates at different time points, was used to compare the kits. The DNA was digested for 2 h at 37°C and then purified with both kits. The samples then underwent the same further steps of GIPFEL. For the PCR steps, each sample was split in half and amplified with *RUNX1* control primers and REH fusion point specific primers, respectively. The *RUNX1* control results were used to normalize the fusion point results. Each sample was normalized against its internal *RUNX1* control to rule out differences due to uneven amounts of input DNA. The QIAquick Gel Extraction Kit showed significantly better results than the innuPREP PCR pure Kit. The product was 14.7-fold better amplified after clean-up with the QIAquick Kit (Figure 5.6 A). The innuPREP Kit is optimized for a fast protocol and not for high quantitative recovery of DNA, which probably is the reason for the lower yield.

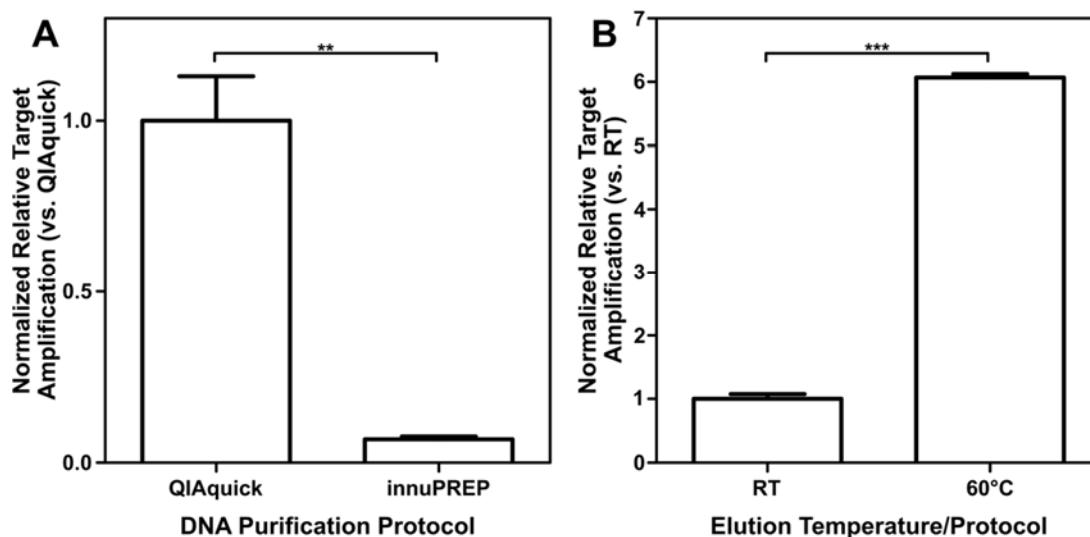


Figure 5.6: Results of the Real-Time PCR for the optimization of the DNA clean-up after restriction enzyme digest. **(A)** Comparison of clean-up using QIAquick Gel Extraction Kit (QIAGEN) and innuPREP PCRpure Kit (Analytik Jena) as part of GIPFEL. Results were normalized and compared to the QIAquick Kit (t test, ** P = 0.0021). **(B)** Different elution methods are compared. The QIAquick Kit was used. Elution at room temperature (RT) was carried out according to manufacturer's protocol, with incubation at RT for 1 min; for elution at 60°C, the samples were incubated at 60°C for 5 min (t test, *** P < 0.001). Mean values and standard deviations of nine measurements are shown.

In another optimization step, different elution methods were tested with the QIAquick Gel Extraction Kit. Again, nine biological replicates from the cell line REH were used

per method, and all samples underwent the same procedure, except for the elution step. For one half of the samples the elution step was carried out at room temperature (RT) with a one minute incubation, and for the other half it was carried out at 60°C with pre-warmed water and a five minute incubation at 60°C.

The prolonged incubation in combination with the raised temperature had an enormous influence on the amount of DNA that was eluted. Hence, it had also a vast influence on the amplification. The warm and prolonged elution led to an approximately six-fold increase in DNA amplification (Figure 5.6 B). Further clean-ups were therefore carried out with the QIAquick Gel Extraction Kit and with incubation at 60°C for 5 min.

5.1.3.3 Ligation Temperature

Four different temperatures were tested for the ligation. Nine biological replicates of REH DNA were used for every temperature. The ligation was carried out at 4°C, 12°C, 24°C, and 37°C for 2 h. Besides the ligation temperature, no other parameters were

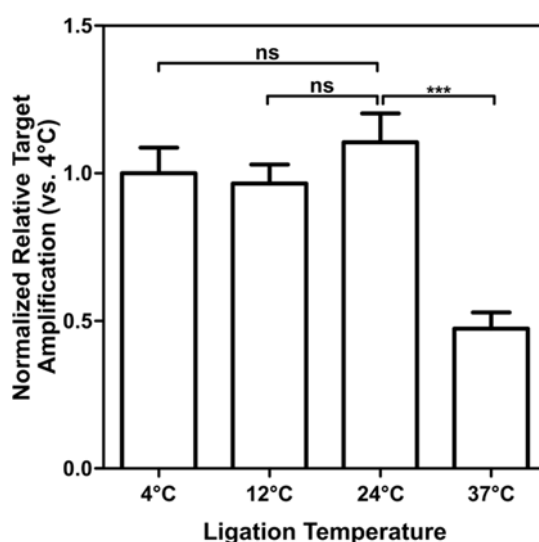


Figure 5.7: Results of the Real-Time PCR for the optimization of the ligation temperature. As part of GIPFEL, four different ligation temperatures were tested, normalized and compared to ligation at 4°C. Ligation at 24°C yielded the best result and is significantly better than ligation at 37°C (Dunnett's Multiple Comparison Test, *** $P < 0.001$, ns = not significant). Mean values and standard deviations of nine measurements are shown.

changed, and the conditions were the same for all samples. *RUNX1* control results were used for normalization.

GIPFEL worked best with ligation at 24°C (Figure 5.7). The difference between 24°C and 37°C was highly significant (Figure 5.7). Even though the results for 4°C, 12°C, and 24°C were not significantly different, ligation at 24°C yielded the best results and was consequently used for further ligations.

5.1.3.4 Duration of Ligation

The experimental setup to determine the best ligation time resembled that for the optimization of the ligation temperature. Six different ligation times were tested: 1 h, 2 h, 4 h, 8 h, 16 h, and 24 h. Each ligation was carried out at 24°C with nine biological replicates per duration. The REH cell line served as the source of DNA, *RUNX1* controls were used for normalization.

The results for ligation at 2 h and 4 h were best and almost identical, with no significant difference. The other ligation durations, including 1 h, were less effective. Starting with 8 h, the performance got weaker over time (Figure 5.8). This was probably due to more

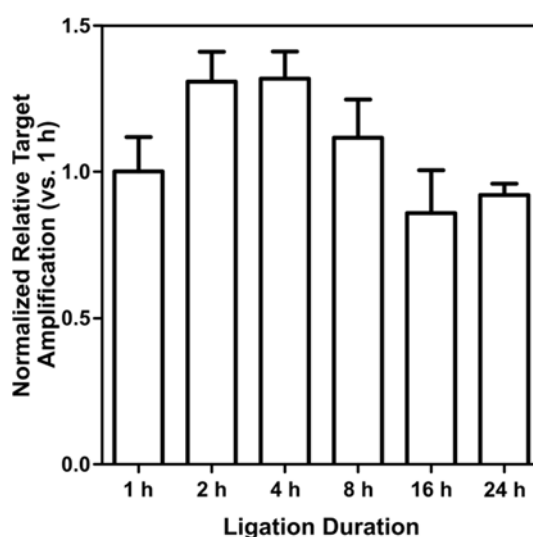


Figure 5.8: Real-Time PCR results for the optimization of ligation duration. Different ligation durations were tested for GIPFEL. The data was normalized and compared to 1 h. The results do not differ significantly from the 2 h result when compared with Dunnett's Multiple Comparison Test. Mean values and standard deviations of nine measurements are shown.

unspecific ligated DNA in the Real-Time PCR. For further ligations, a ligation time of 2 h appeared the best. It ranked among the best performing durations, but still is a rather short time.

5.1.3.5 Clean-up of Exonuclease Digest

After the exonuclease digest, another clean-up step was introduced. This was done to remove buffers and enzymes which could possibly interfere with the PCR reaction and to reduce the sample volume in order to increase sensitivity. To that end, three different methods were tested. The first method was a precipitation protocol using ethanol, the second a precipitation protocol using isopropanol. This protocol was later used by the cooperation partner in Erlangen. The last method tested was a column purification using the innuPREP PCRpure Kit (Analytik Jena). The protocols were carried out as described in chapter 4.1.9 and according to manufacturer's instructions, respectively. REH served as the source of DNA, and nine biological replicates per protocol were tested.

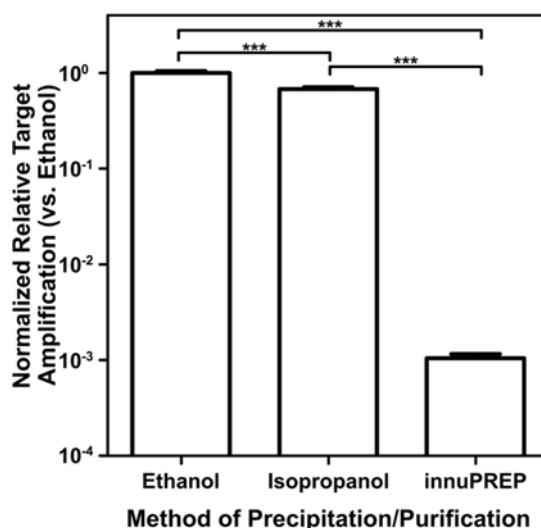


Figure 5.9: Real-Time PCR results after the introduction of a second clean-up step, following the exonuclease digest. Two DNA precipitation protocols, one using ethanol, one using isopropanol, and one column based clean-up (innuPREP PCRpure Kit, Analytik Jena) were tested for GIPFEL. The results were normalized and compared to the ethanol protocol. The Real-Time PCR results after application of the protocols differ significantly from each other (Dunnett's Multiple Comparison Test, *** $P < 0.001$). Mean values and standard deviations of nine measurements are shown.

The precipitation protocols were better suited for this step than the column purification (Figure 5.9): the ethanol precipitation led to an almost 1000-fold higher yield compared to the column purification. The difference between the two alcohol precipitations was not as pronounced, but still significant. The ethanol protocol led to a 1.5-fold higher yield when compared to the isopropanol protocol (Figure 5.9). Hence, the ethanol precipitation protocol was used for further experiments.

5.1.3.6 Whole Genome Amplification

Since the amounts of patient derived DNA, which was used to establish GIPFEL, were very limited, this DNA was whole genome amplified. The WGA was performed with the REPLI-g UltraFast Mini Kit (QIAGEN). In order to test the effects of the WGA on the GIPFEL results, REH DNA from nine different time points was amplified. In parallel, the respective non-amplified REH samples were also subjected to the GIPFEL procedure. The samples were split in half before the PCR, with one part being used to validate the translocation and the other being used to amplify the *RUNX1* control for normalization. The whole genome amplified as well as the non-amplified DNA produced correct results. Both, the *RUNX1* control and the translocation, could be validated. Figure 5.10 A shows a representative example of an amplification plot. It is apparent that WGA led to an increase in the quantification cycle (C_q). The products themselves, however, were correct in either case, as shown by the dissociation curves (Figure 5.10 B). The difference in target amplification between non-amplified and whole genome amplified DNA is highly significant, though (Figure 5.10 C). There is a 140-fold change between the two of *RUNX1* positive controls and a 42-fold change between the two sets of samples used to detect the translocation. This showed that the sensitivity decreases when WGA is used.

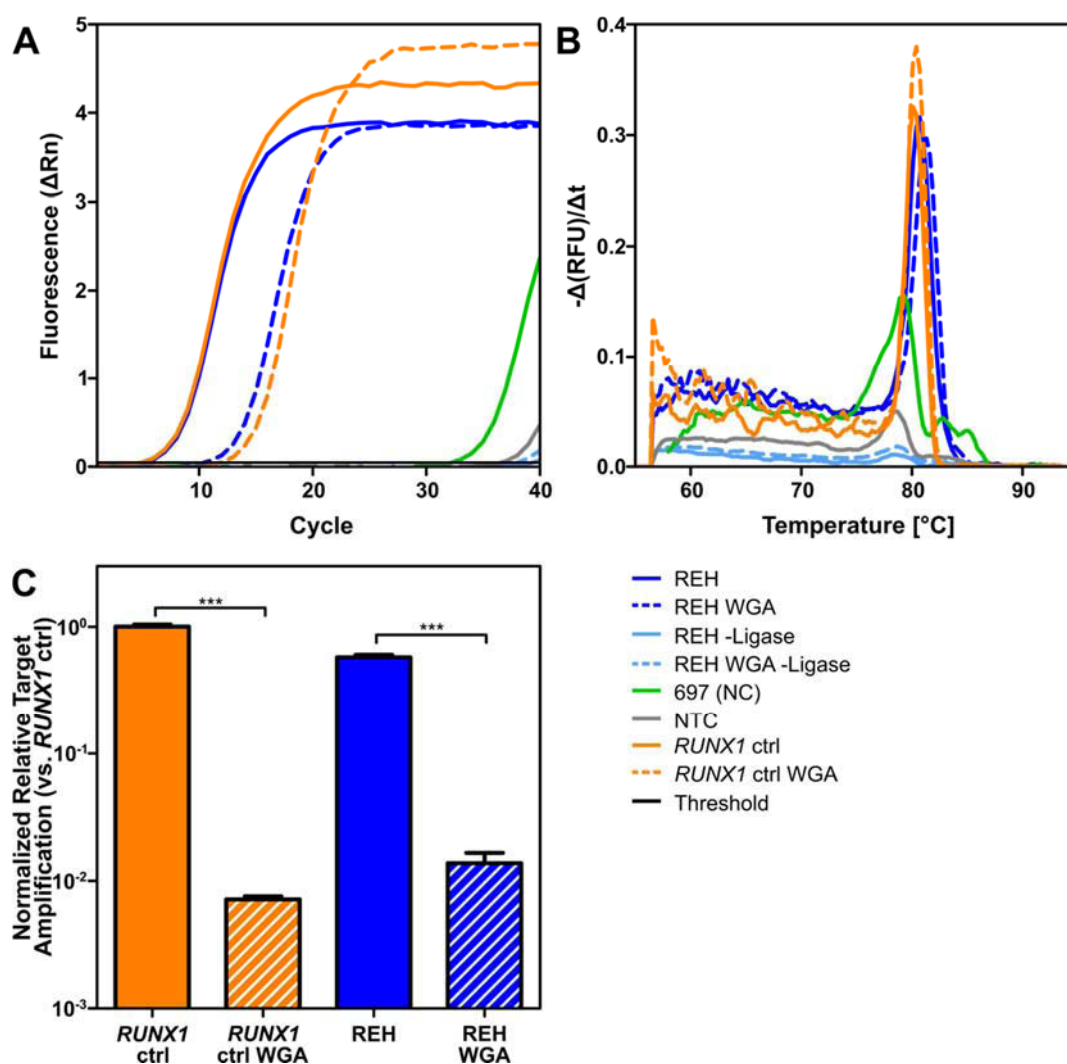


Figure 5.10: Impact of whole genome amplification on GIPFEL. **(A)** Amplification plot showing the amplification of the REH specific joining sequence (REH), the REH specific joining sequence but without the ligation step (REH -Ligase), and the *RUNX1* control (ctrl) with and without whole genome amplification (WGA). DNA from 697 cells serves as negative control (NC), water as non-template control (NTC). **(B)** Dissociation plot showing that the amplified and non-amplified samples produce the same products. **(C)** The quantification results were normalized and compared to the *RUNX1* control without WGA. The whole genome amplified samples score significantly worse than the non-amplified samples (t test, *** $P < 0.001$). Rn = normalized reporter signal, RFU = relative fluorescent units.

5.1.3.7 Multiplexing of Primers

A total of 28 forward primers for *ETV6-RUNX1* and 36 forward primers for *TCF3-PBX1* would have made the method time, labor, cost, and material intensive. To reduce all four, primers were multiplexed. The 28 *RUNX1* forward primers were multiplexed in four bundles with seven primers each. These bundles were color coded. All three *ETV6* reverse and nested reverse primers, respectively, were multiplexed together. For the *PBX1* forward primers, there were five bundles, also color coded. In each bundle the primers with the lowest possible hetero-dimerization were put together. This was

checked with OligoAnalyzer 3.1 (IDT). The final bundles are shown in Table 5.1 and Table 5.2. After a bundle amplified a product, the primers were demultiplexed. The REH cell line, for instance, used the primers RUNX1-S13f and ETV6-S2r-n. Hence, the yellow bundle scored positive and the primers were used separately in another Real-Time PCR. In this Real-Time PCR RUNX1-S13f was the only primer to score positive. Without multiplexing, this would have meant 87 PCRs and 87 different Real-Time PCRs. Through multiplexing, however, this number could be reduced to five PCRs and 12 different Real-Time PCRs.

Table 5.1: Multiplexing of *ETV6-RUNX1* primers.

Blue Bundle	Red Bundle	Green Bundle	Yellow Bundle	Reverse Bundle
RUNX1-S12f	RUNX1-S1f	RUNX1-S2f	RUNX1-S3f	ETV6-S1r
RUNX1-S15f	RUNX1-S4f	RUNX1-S6f	RUNX1-S5f	ETV6-S2r
RUNX1-S17f	RUNX1-S10f	RUNX1-S7f	RUNX1-S13f	ETV6-S3r
RUNX1-S22f	RUNX1-S11f	RUNX1-S8f	RUNX1-S16f	Nested Reverse Bundle
RUNX1-S23f	RUNX1-S14f	RUNX1-S18f	RUNX1-S19f	ETV6-S1r-n
RUNX1-S26f	RUNX1-S24f	RUNX1-S20f	RUNX1-S21f	ETV6-S2r-n
RUNX1-S28f	RUNX1-S27f	RUNX1-S29f	RUNX1-S25f	ETV6-S3r-n

Table 5.2: Multiplexing of *TCF3-PBX1* primers.

Blue Bundle	Red Bundle	Green Bundle	Yellow Bundle	Purple Bundle
PBX1-M1f	PBX1-M4f	PBX1-M2f	PBX1-M3f	PBX1-M7f
PBX1-M6f	PBX1-M8f	PBX1-M11f	PBX1-M5f	PBX1-M15f
PBX1-M12f	PBX1-M9f	PBX1-M20f	PBX1-M14f	PBX1-M17f
PBX1-M13f	PBX1-M10f	PBX1-M25f	PBX1-M18f	PBX1-M22f
PBX1-M26f	PBX1-M16f	PBX1-M28f	PBX1-M19f	PBX1-M27f
PBX1-M29f	PBX1-M21f	PBX1-M34f	PBX1-M24f	PBX1-M30f
PBX1-M33f	PBX1-M23f	PBX1-M35f	PBX1-M32f	PBX1-M31f
PBX1-M36f				

5.1.3.8 Optimized GIPFEL Protocol

A new standard protocol resulted from the optimization process (Figure 5.11). All further screenings followed the new protocol as described in Methods (chapter 4.1) and depicted in Figure 5.11. WGA (see chapter 4.1.3) was done in cases where only minute amounts of DNA were available. Ethanol precipitation was chosen over isopropanol precipitation.

In cases that led to positive Real-Time PCR results, further steps were carried out (Figure 5.11 B): a Real-Time PCR and a normal PCR were conducted to identify the

forward and the reverse primer that amplified a product, respectively. The PCR product was then separated on an agarose gel and subjected to Sanger sequencing to validate the findings. If one of these steps did not lead to putative positive results, no further steps were carried out.

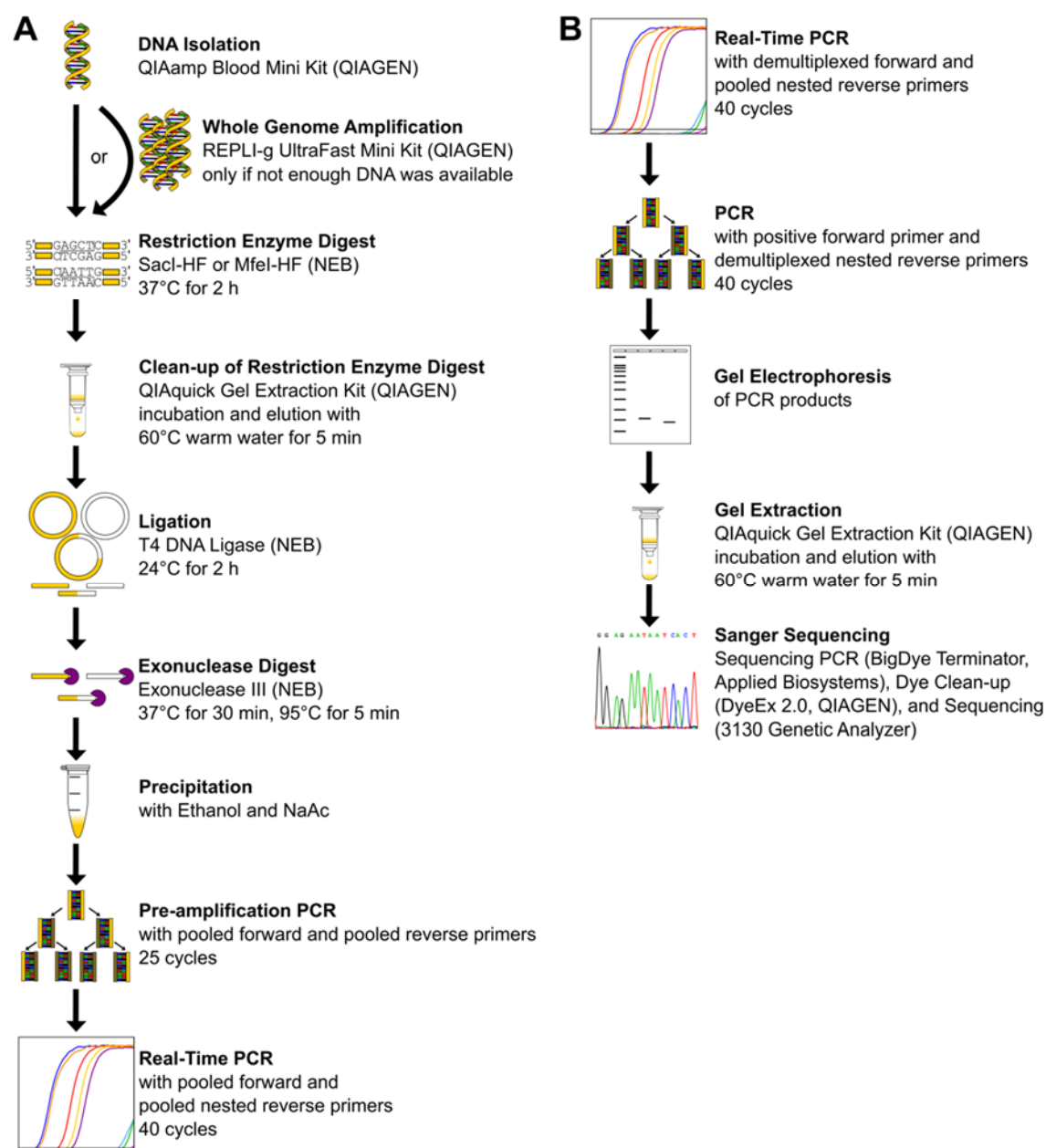


Figure 5.11: Optimized GIPFEL protocol. **(A)** Optimized protocol that was carried out with each sample. **(B)** Optional continuative protocol that was carried out if the Real-Time PCR scored a putative positive result. In case that a step did not produce a positive result, no further validation steps were done.

5.1.4 Final Results of the GIPFEL Establishment Phase

5.1.4.1 Sensitivity of GIPFEL

Cells from cell lines carrying either translocation were diluted down to 10^{-6} in cells not carrying the respective translocations. These samples then underwent the optimized GIPFEL procedure with the intention to find the detection limit of the method.

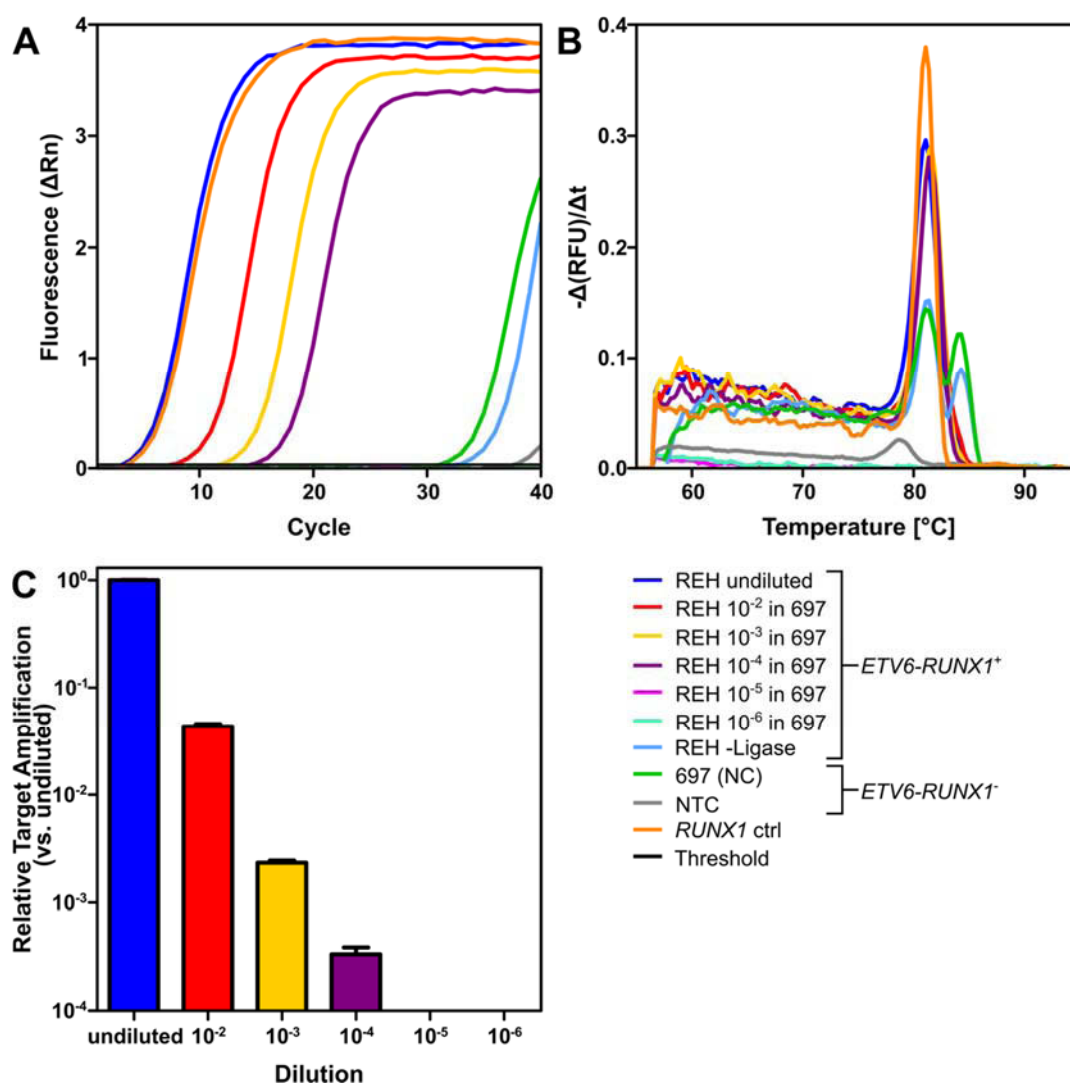


Figure 5.12: Sensitivity of GIPFEL for *ETV6-RUNX1*. **(A)** Amplification plot showing the amplification of REH products down to a dilution of one in 10^4 cells. **(B)** The dissociation plot indicates that all amplified dilutions produced the same amplification product. **(C)** The quantification was compared to undiluted REH cells. Dilutions down to 10^{-4} were amplified. Further dilutions could not be amplified by GIPFEL. Rn = normalized reporter signal, RFU = relative fluorescent units.

For the *ETV6-RUNX1* fusion, the detection was possible down to one in 10^4 cells (Figure 5.12 A). All products of dilutions down to 10^{-4} were the right product (Figure

5.12 B) and in the expected gradation range (Figure 5.12 C). Amplification of further dilutions was not repeatable (Figure 5.12).

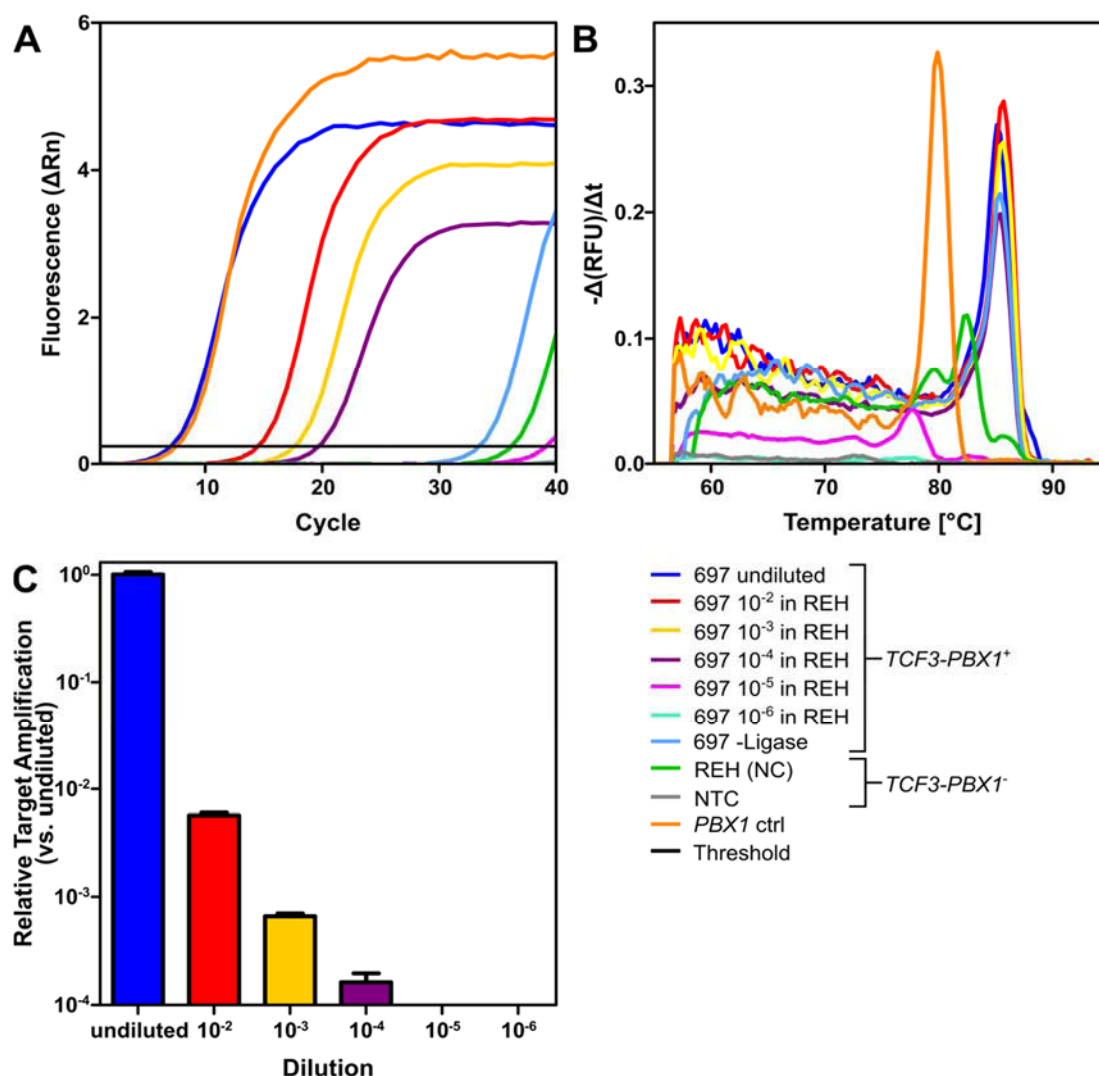


Figure 5.13: Sensitivity of GIPFEL for *TCF3-PBX1*. **(A)** Amplification plot showing the amplification of 697 products down to a dilution of one in 10^4 cells. **(B)** The dissociation plot indicates that all amplified dilutions produced the same amplification product. The product of 697 without ligase (-Ligase) could not be validated by Sanger sequencing and therefore was an artefact. **(C)** The quantification was compared to undiluted 697 cells. Dilutions down to 10^{-4} were amplified as expected. Further dilutions could not be amplified by GIPFEL. Rn = normalized reporter signal, RFU = relative fluorescent units.

The detection of the t(1;19) translocation had the same limitations as the one for translocation t(12;21): a dilution down to 10^{-4} was detectable and provided the correct products, whereas further dilutions showed no amplification of the expected product (Figure 5.13). Hence, the detection limit of the GIPFEL method is 10^{-4} . This finding was

also supported by the results of the cooperation partner in Erlangen regarding the three *KMT2A* translocations.

5.1.4.2 Accuracy and Specificity

Two screenings using DNA from patients with translocations t(12;21) and t(1;19) were carried out. In the first one, a set of ten t(12;21) and five t(1;19) translocation positive patients was screened. Patient 11 was not used because the breakpoint was unknown, for the 15 used patients the breakpoint was known. In a second screening, 50 t(12;21) and 25 t(1;19) translocation positive patients were screened. The breakpoints for these leukemic patients were unknown. For each translocation five healthy samples were interspersed, and the samples were blinded. This was done in order to analyze specificity of the method. These screenings allowed for determination of the accuracy of GIPFEL for either translocation.

Nine out of the ten leukemic patients, carrying a t(12;21) translocation with known breakpoints, could be validated. Patient 8 had two chromosomal breakpoints. Both were validated with the same primer pair as they were in close proximity (Table 5.3). Of the 50 blinded samples with unknown breakpoints, 33 could be identified as *ETV6-RUNX1* positive by GIPFEL. Patient P83 also had two breakpoints (Table 5.3).

An example of the validation process in patients is given in Figure 5.14. Patient P43 was screened with the four primer bundles. The amplification plot showed a higher amplification with the yellow bundle (Figure 5.14 A). The respective dissociation curve implicated that the product was of expected size (Figure 5.14 B). Therefore, the primers of the yellow bundle were demultiplexed and used separately. The amplification with primer RUNX1-S25f led to product amplification whereas the other primers produced no specific signal (Figure 5.14 C-D). This interpretation was strengthened by the agarose gel electrophoresis (Figure 5.14 E) where RUNX1-S25f was the sole primer to produce a product. Sequencing of this product eventually proved the presence of the translocation (Figure 5.14 F).

Table 5.3: *ETV6-RUNX1* positive samples that were validated by GIPFEL. The primers used for validation are shown. *Patients 8 and P83 had two chromosomal breakpoints each. Primers are given for both breakpoints.

Sample ID	Sample type	forward primer	reverse primer
REH	cell line	RUNX1-S13f	ETV6-S2r-n
1	patient with known breakpoint	RUNX1-S3f	ETV6-S3r-n
2	patient with known breakpoint	RUNX1-S3f	ETV6-S3r-n
3	patient with known breakpoint	RUNX1-S10f	ETV6-S3r-n
5	patient with known breakpoint	RUNX1-S25f	ETV6-S3r-n
6	patient with known breakpoint	RUNX1-S14f	ETV6-S3r-n
7	patient with known breakpoint	RUNX1-S3f	ETV6-S2r-n
8*	patient with known breakpoint	RUNX1-S3f	ETV6-S3r-n
		RUNX1-S3f	ETV6-S3r-n
9	patient with known breakpoint	RUNX1-S10f	ETV6-S3r-n
10	patient with known breakpoint	RUNX1-S3f	ETV6-S2r-n
P31	patient from blinded screening	RUNX1-S4f	ETV6-S3r-n
P32	patient from blinded screening	RUNX1-S15f	ETV6-S3r-n
P33	patient from blinded screening	RUNX1-S28f	ETV6-S3r-n
P35	patient from blinded screening	RUNX1-S6f	ETV6-S3r-n
P37	patient from blinded screening	RUNX1-S2f	ETV6-S1r-n
P38	patient from blinded screening	RUNX1-S5f	ETV6-S3r-n
P39	patient from blinded screening	RUNX1-S14f	ETV6-S3r-n
P40	patient from blinded screening	RUNX1-S14f	ETV6-S3r-n
P41	patient from blinded screening	RUNX1-S10f	ETV6-S1r-n
P43	patient from blinded screening	RUNX1-S25f	ETV6-S3r-n
P44	patient from blinded screening	RUNX1-S13f	ETV6-S1r-n
P45	patient from blinded screening	RUNX1-S4f	ETV6-S2r-n
P47	patient from blinded screening	RUNX1-S11f	ETV6-S2r-n
P49	patient from blinded screening	RUNX1-S11f	ETV6-S3r-n
P51	patient from blinded screening	RUNX1-S13f	ETV6-S3r-n
P52	patient from blinded screening	RUNX1-S28f	ETV6-S2r-n
P54	patient from blinded screening	RUNX1-S8f	ETV6-S3r-n
P57	patient from blinded screening	RUNX1-S5f	ETV6-S3r-n
P58	patient from blinded screening	RUNX1-S2f	ETV6-S2r-n
P59	patient from blinded screening	RUNX1-S11f	ETV6-S3r-n
P61	patient from blinded screening	RUNX1-S28f	ETV6-S3r-n
P62	patient from blinded screening	RUNX1-S6f	ETV6-S2r-n
P63	patient from blinded screening	RUNX1-S12f	ETV6-S2r-n
P64	patient from blinded screening	RUNX1-S14f	ETV6-S3r-n
P65	patient from blinded screening	RUNX1-S28f	ETV6-S3r-n
P70	patient from blinded screening	RUNX1-S29f	ETV6-S2r-n
P75	patient from blinded screening	RUNX1-S14f	ETV6-S2r-n
P77	patient from blinded screening	RUNX1-S10f	ETV6-S1r-n
P79	patient from blinded screening	RUNX1-S5f	ETV6-S3r-n
P80	patient from blinded screening	RUNX1-S8f	ETV6-S3r-n
P82	patient from blinded screening	RUNX1-S28f	ETV6-S1r-n
P83*	patient from blinded screening	RUNX1-S4f	ETV6-S1r-n
		RUNX1-S4f	ETV6-S2r-n
P84	patient from blinded screening	RUNX1-S28f	ETV6-S2r-n

All five t(1;19) translocation positive patients with known breakpoints were validated by GIPFEL. Six of the 25 blinded samples were identified as being *TCF3-PBX1* positive (Table 5.4), and the validation of patient P26 is shown exemplarily (Figure 5.15). Identification of the right primer bundle was not as clear as in the *ETV6-RUNX1*

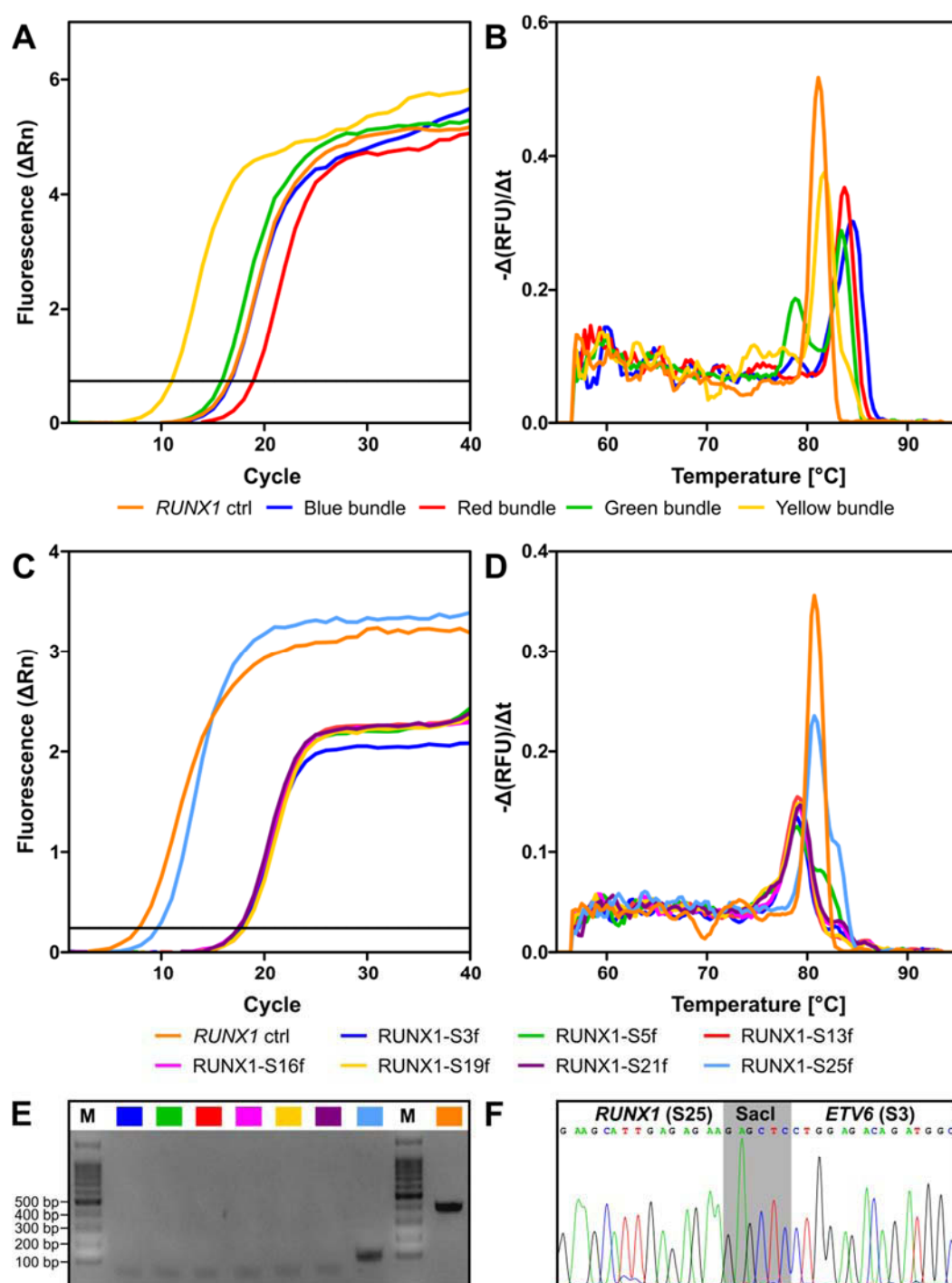


Figure 5.14: Example for the validation of an *ETV6*-*RUNX1* positive patient (P43). **(A)** Amplification plot of the Real-Time PCR conducted with multiplexed primers. Each primer bundle is represented by its respective color, the *RUNX1* control (ctrl) is depicted in orange. **(B)** Dissociation plot of the Real-Time PCR shown in (A). **(C)** Amplification plot of the Real-Time PCR with demultiplexed forward primers of the yellow bundle. **(D)** Dissociation plot of the demultiplexed Real-Time shown in (C). The plot shows that all primers but *RUNX1*-S25f solely produced primer dimers. **(E)** Gel electrophoresis of the products from the Real-Time PCR shown in (C). The colors shown above the lanes represent the same primers as in (C) and (D). Besides the *RUNX1* control, *RUNX1*-S25f was the only primer with a product. The product had the expected size of 122 bp. QIAGEN's 100 bp plus marker served as marker (M). **(F)** Result of the Sanger sequencing of the product with *RUNX1*-S25f and *ETV6*-S3r-n. The sequencing unveiled the expected sequence. Rn = normalized reporter signal, RFU = relative fluorescent units.

Table 5.4: *TCF3-PBX1* positive samples that were validated by GIPFEL. The primers used for validation are shown.

Sample ID	Sample type	forward primer	reverse primer
697	cell line	PBX1-M1f	TCF3-M1r-n
12	patient with known breakpoint	PBX1-M13f	TCF3-M1r-n
13	patient with known breakpoint	PBX1-M1f	TCF3-M1r-n
14	patient with known breakpoint	PBX1-M1f	TCF3-M1r-n
15	patient with known breakpoint	PBX1-M1f	TCF3-M1r-n
16	patient with known breakpoint	PBX1-M1f	TCF3-M1r-n
P5	patient from blinded screening	PBX1-M1f	TCF3-M1r-n
P22	patient from blinded screening	PBX1-M9f	TCF3-M1r-n
P25	patient from blinded screening	PBX1-M9f	TCF3-M1r-n
P26	patient from blinded screening	PBX1-M13f	TCF3-M1r-n
P27	patient from blinded screening	PBX1-M13f	TCF3-M1r-n
P30	patient from blinded screening	PBX1-M13f	TCF3-M1r-n

example. Besides the *PBX1* control product, several bundles produced putative positive results (Figure 5.15 A). The dissociation curve (Figure 5.15 B) showed that the green bundle produced an unspecific product and the yellow bundle amplified three products. The red and purple bundles produced one product only and therefore were candidates. Primers from the blue bundle produced two products which, however, lay in the expected range. The blue, red, and purple bundles were demultiplexed, the primers were used separately. Primer PBX1-M13f of the blue bundle was the only one to produce a product (Figure 5.15 C-D) which could be visualized on an agarose gel (Figure 5.15 E). Sequencing of this product proved the patient to be translocation t(1;19) positive (Figure 5.15 F).

None of the interspersed translocation negative samples was called positive. On the contrary, these samples were all identified as translocation negative. This result shows that the GIPFEL method has a specificity of 100%, regardless of the investigated translocation. This is also the case for the *KMT2A* translocations investigated in Erlangen [145]. For *ETV6-RUNX1*, 61 translocation positive samples were tested, and 43 of them were identified as translocation positive. A total of 31 translocation t(1;19) positive samples was tested, and 12 of them could be identified as translocation positive. The accuracy was dependent on the translocation: for *ETV6-RUNX1*, the accuracy is 70%, whereas the accuracy of *TCF3-PBX1* is lower and lies at 39%. The *KMT2A* translocations investigated in Erlangen also have different accuracies: for

KMT2A-AFF1, *KMT2A-MLLT3*, and *KMT2A-MLLT1*, 83%, 65%, and 24% of positive samples are called correctly [145].

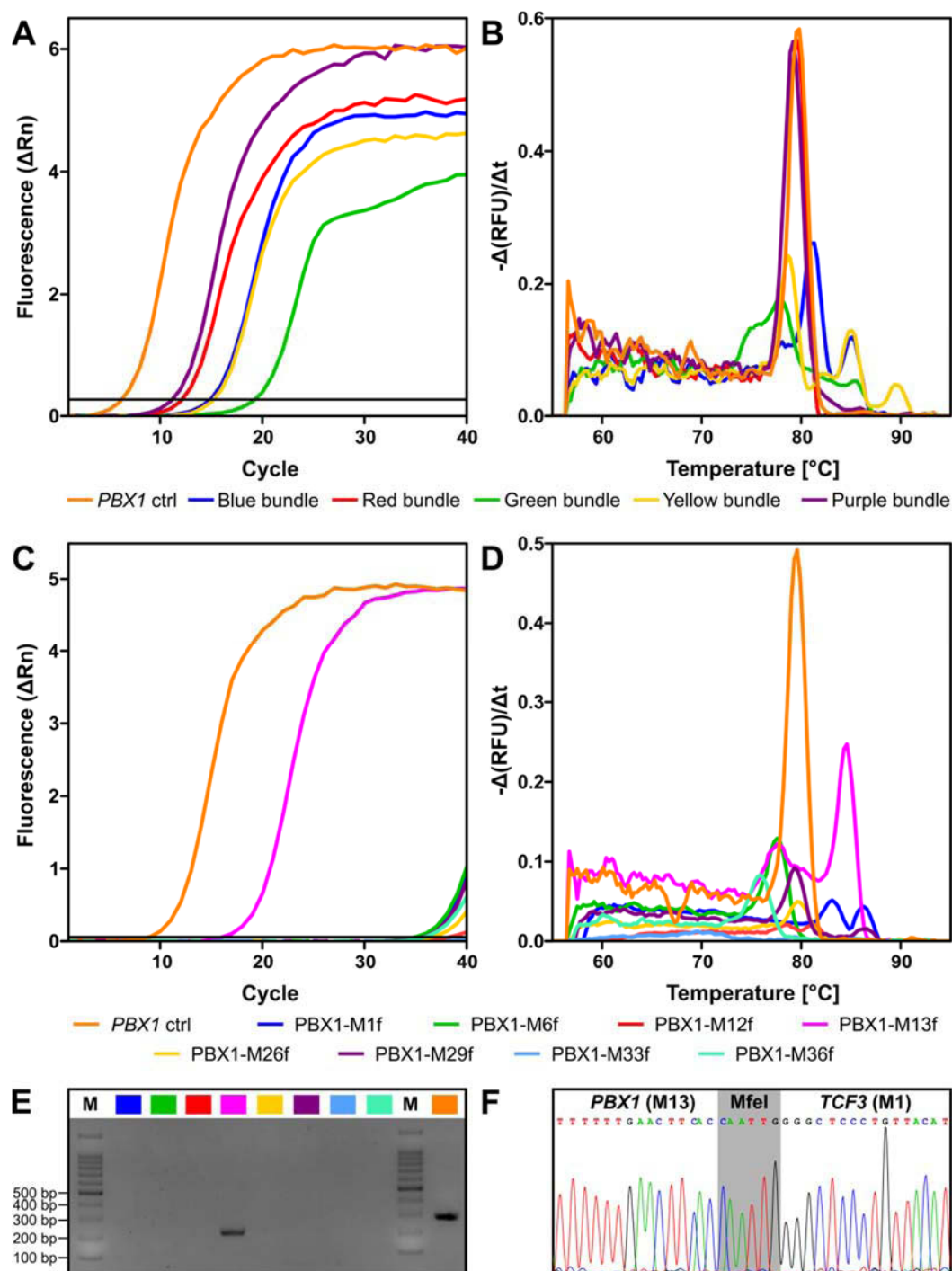


Figure 5.15: Example for the validation of a *TCF3-PBX1* positive patient (P26). **(A)** Amplification plot of the Real-Time PCR conducted with multiplexed primers. Each primer bundle is represented by its respective color, the *PBX1* control (ctrl) is depicted in orange. **(B)** Dissociation plot of the Real-Time PCR shown in (A). **(C)** Amplification plot of the Real-Time PCR with demultiplexed forward primers of the blue bundle. **(D)** Dissociation plot of the demultiplexed Real-Time shown in (C). The plot shows that none but primer *PBX1-M13f* amplified a product. **(E)** Gel electrophoresis of the products from the Real-Time PCR shown in (C). The colors shown above the lanes represent the same primers as in (C) and (D). Besides the *PBX1* control, *PBX1-M13f* was the only primer with a product. The product had the expected size of 224 bp. QIAGEN's 100 bp plus marker served as marker (M). **(F)** Result of the Sanger sequencing of the product with *PBX1-M13f* and *TCF3-M1r-n*. The sequencing unveiled the expected sequence. Rn = normalized reporter signal, RFU = relative fluorescent units.

5.2 Screening of Umbilical Cord Blood from Healthy Newborns

5.2.1 Adaption of the GIPFEL protocol

5.2.1.1 CD19⁺ Enrichment

ETV6-RUNX1 positive leukemia usually derives from CD19⁺ cells [75, 76, 146, 147]. Therefore, the mononuclear cells from umbilical cord blood were CD19⁺ enriched with Dynabeads CD19 pan B (invitrogen). The major goals of this enrichment were to reduce unspecific target amplification and to increase sensitivity by reducing the amount of translocation negative input DNA.

In order to test this procedure, t(12;21) positive REH cells were diluted in t(12;21) negative HEK-293 cells (Figure 5.16 D). The highest possible amount of 2.5×10^7 cells was used for the enrichment. As a further control, the test amount of REH DNA, used during establishment, was subjected to the procedure. This equals the amount of DNA of approximately 2.0×10^5 cells.

The sensitivity of GIPFEL was not increased by CD19⁺ enrichment. *ETV6-RUNX1* positive REH cells diluted 10^{-5} and 10^{-6} in HEK-293 cells could not be detected (Figure 5.16). The test showed that an input of 2.5×10^7 cells was too high, as it led to template inhibition. The translocation could still be detected, but the template inhibition led to drastically worse amplification. The 10^{-2} dilution and the test amount led to comparable results, which was expected, as both samples had roughly the same amount of translocation positive cells.

The CD19⁺ enrichment had no negative effect on the GIPFEL result and was kept in the protocol in order to remove dead cells and unspecific DNA.

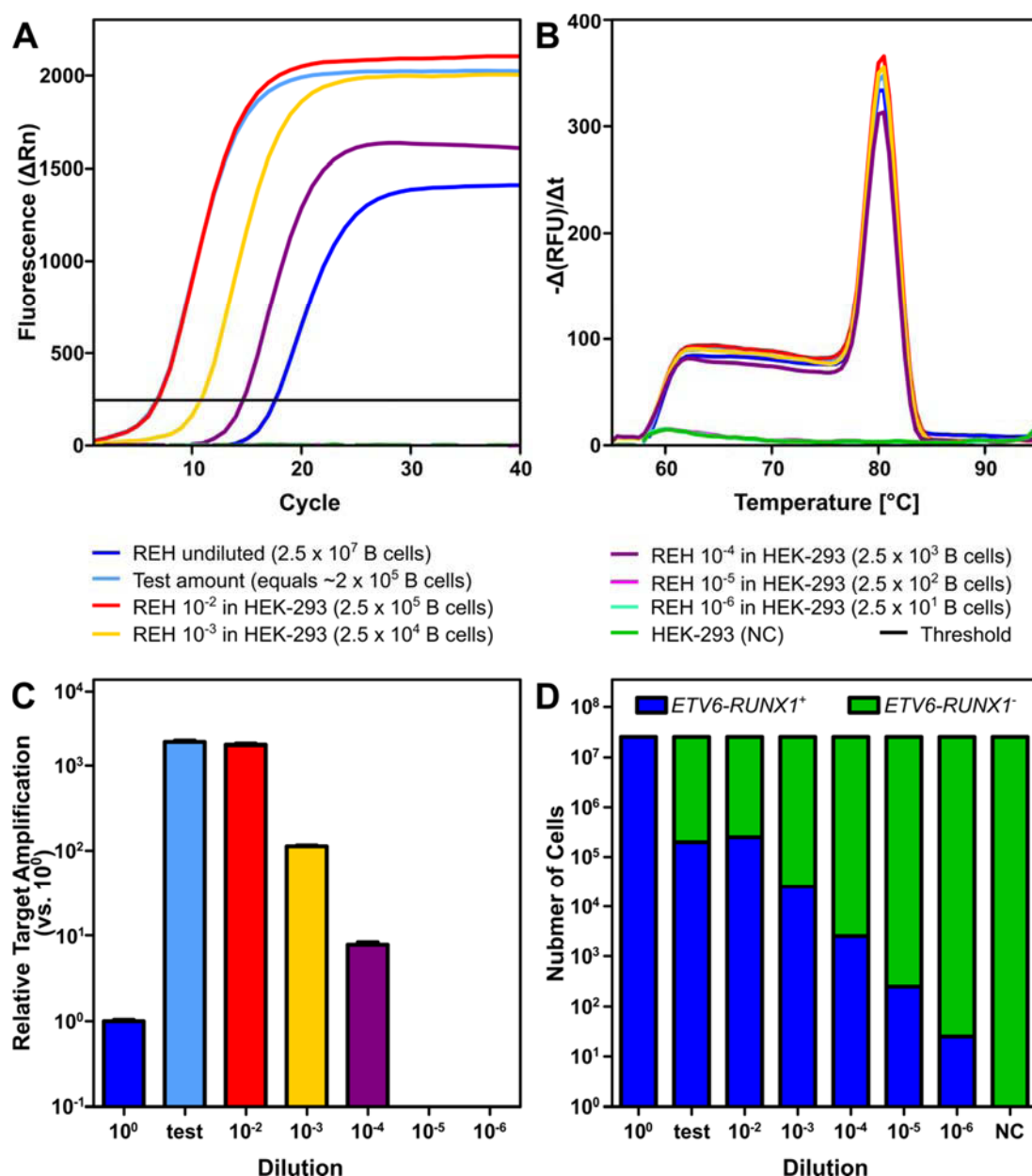


Figure 5.16: GIPFEL results including CD19⁺ enrichment. **(A)** Amplification Plot: The CD19⁺ enrichment did not increase sensitivity of GIPFEL. The undiluted REH cell DNA led to target inhibition during the PCRs. **(B)** Dissociation curves of the Real-Time PCRs. All amplified products were of the expected size. **(C)** The amplification was compared to the undiluted REH DNA (10^0). The undiluted sample ranged lower due to target inhibition. Dilutions of 10^{-5} and 10^{-6} showed no amplification. Note: For the test amount, measured DNA was used and not REH cells diluted in HEK-293 cells. This might be the reason why test amount led to slightly better amplification than the 10^{-2} dilution, which should have had 50,000 cells more. **(D)** Overview of the amount of cells in each dilution. The *ETV6-RUNX1* positive cells were REH cells (blue), the negative ones HEK-293 cells (green). NC = negative control, Rn = normalized reporter signal, RFU = relative fluorescent units.

5.2.1.2 Whole Genome Amplification

With the objective of receiving more template DNA and increasing the sensitivity by that, WGA was tested after CD19⁺ enrichment. To that end, the DNA isolated from the CD19⁺ REH cells diluted in HEK-293 cells was subjected to the amplification. As a control, the test amount of REH DNA was used in parallel.

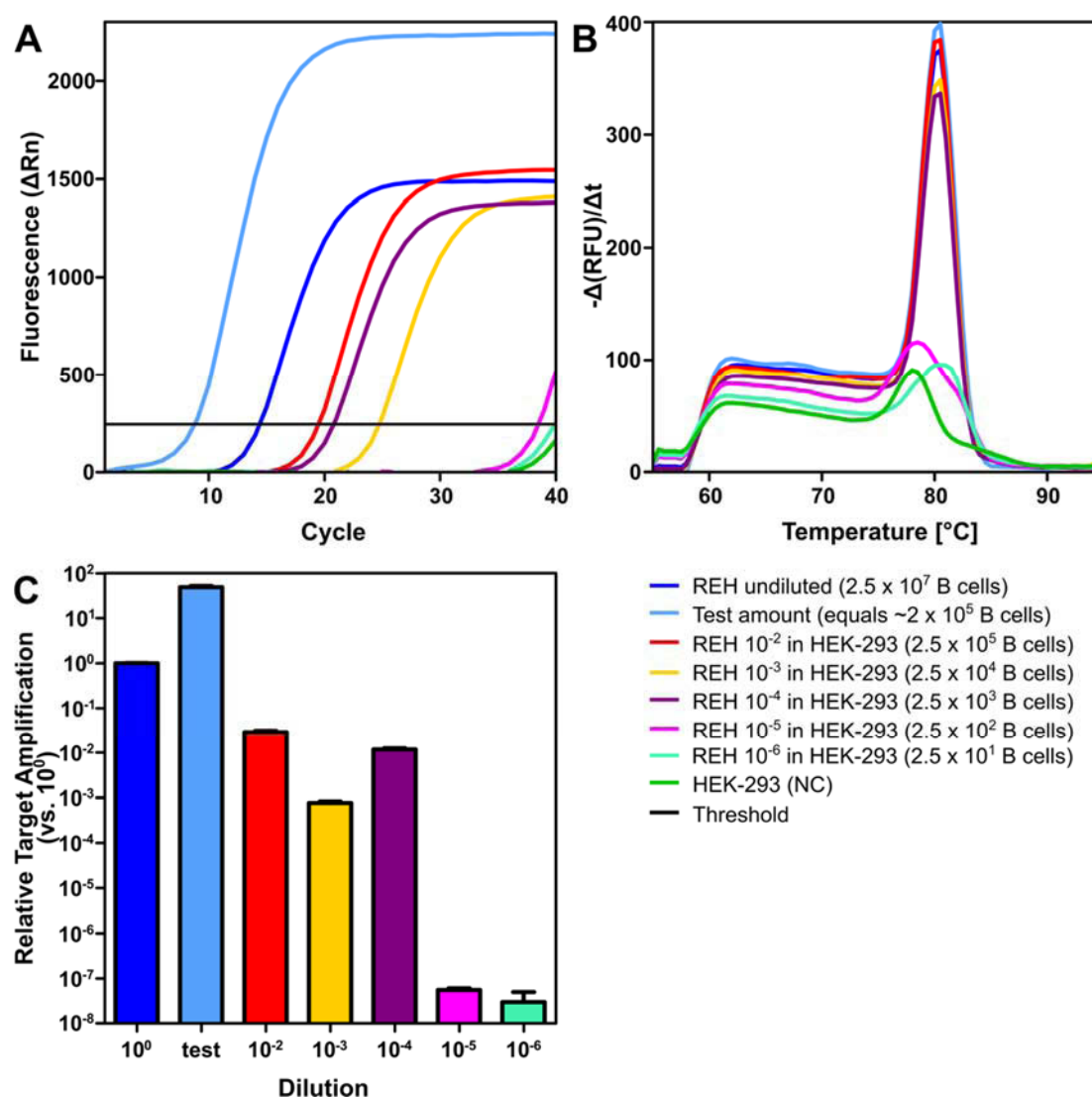


Figure 5.17: GIPFEL results including WGA after CD19⁺ enrichment. **(A)** Amplification plot: Sensitivity was not increased through WGA. Instead, samples were amplified unevenly. **(B)** Dissociation plot: the amplified samples produced the right amplification product. **(C)** The amplification was compared to the undiluted REH DNA (10⁰). Due to the uneven amplification, the dilutions were no longer distributed as expected. The 10⁻⁴ dilution led to better results than the 10⁻³ dilution. NC = negative control, Rn = normalized reporter signal, RFU = relative fluorescent units.

WGA did not increase sensitivity after CD19⁺ enrichment. When the test amount of DNA was used, GIPFEL worked as before, but when the CD19⁺ enriched DNA was

used, less product was amplified, resulting in higher C_q values (Figure 5.17). Furthermore, it is apparent that WGA influenced the results. The 10⁻⁴ dilution was better amplified than the 10⁻³ dilution. The products, however, were all correct (Figure 5.17 B). As WGA lowers the sensitivity of GIPFEL, it was not used for the screening of the umbilical cord blood samples.

5.2.1.3 384 Well Real-Time PCR

Due to the large number of samples to be screened, the Real-Time PCRs for the screening of 1,000 cord blood was carried out on a CFX384 Touch™ Real-Time PCR Detection System (Bio-Rad). This machine can read 384 well plates, but the wells themselves are smaller than those of 96 well plates. Hence, the sample volume had to be reduced from 25 µl to 10 µl. Three different compositions were tested (Table 5.5). Nine biological replicates of REH DNA were tested per composition and normalized to their respective *RUNX1* controls.

Table 5.5: Tested compositions of Real-Time PCRs on a 384 well cycler. Three different compositions were tested: A, B, and C.

Reagent	A	B	C
Template DNA	1 µl	1 µl	0.5 µl
Forward Primers (1 µM each)	2 µl	1 µl	1 µl
Reverse Primers (1 µM each)	2 µl	1 µl	1 µl
H ₂ O	-	2 µl	2.5 µl
2x Brilliant II SYBR Green Master Mix (Agilent)	5 µl	5 µl	5 µl

Overall, composition “A” showed the best results and was therefore used for the screening. When the amplification was compared to compositions “B” and “C”, “A” led to the best amplification results (Figure 5.18 A). The difference was not significant to composition “B”, but still 1.4x fold higher. The difference to composition “C” was significant. However, in “C”, the amount of template DNA was only half of that in “A”. Hence, a value of 0.5 would have been expected but was not reached (Figure 5.18 A). When the target DNA was diluted, all three compositions showed good results down to a dilution factor of 10⁻³. After that, composition “B” performed worse than the other two

compositions (Figure 5.18 B). Overall, composition “C” had the best results in this serial dilution. The results of “A” also lay in the expected range (Figure 5.18 B).

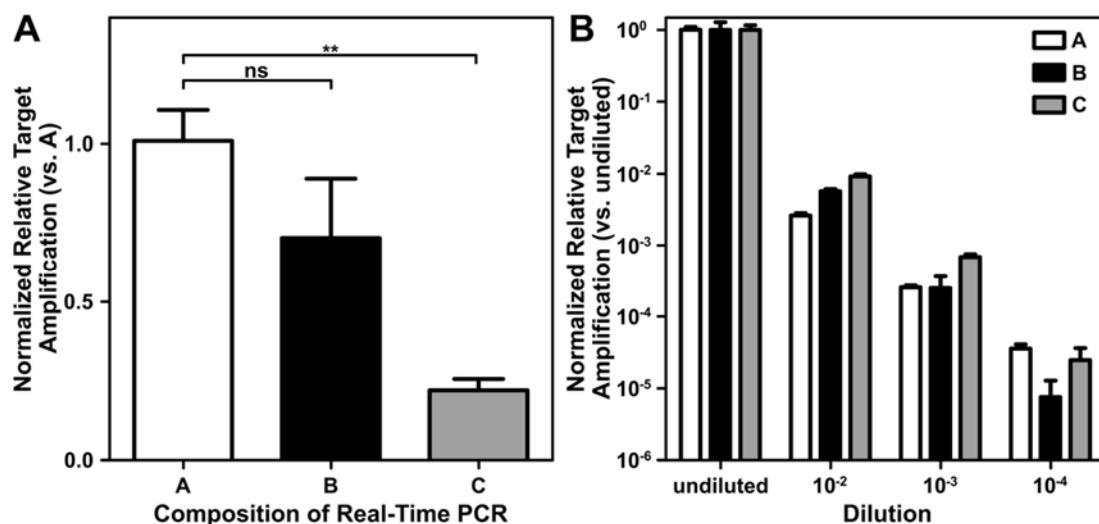


Figure 5.18: Tested compositions of 384 well Real-Time PCR. **(A)** The undiluted samples for each composition were normalized and compared to composition “A”. Composition “C” had a significantly worse amplification result compared to “A” (Dunnett’s Multiple Comparison Test, ** $P < 0.01$). However, the amount of template in “C” was half of that in “A” and “B”. **(B)** Serial dilution of the three compositions. The 10⁻⁴ dilution of composition “B” was only amplified at levels of 10⁻⁵.

5.2.2 Results from Cord Blood Screening for *ETV6-RUNX1*

Mononuclear cells from umbilical cord blood of 1,000 healthy Danish newborns were screened for the *ETV6-RUNX1* fusion, caused by the t(12;21) translocation. This was done by using the optimized GIPFEL protocol (chapter 5.1.3.8) and the adjustments made for this screening, introducing CD19⁺ enrichment and 384 well Real-Time PCRs. The identification of *ETV6-RUNX1* positive samples followed the same procedure as the one for the patients. When a sample appeared to carry the translocation in the first Real-Time PCR, a second Real-Time PCR with demultiplexed forward primers was done. This procedure is shown exemplarily for umbilical cord blood N926 in Figure 5.19 and Figure 5.20.

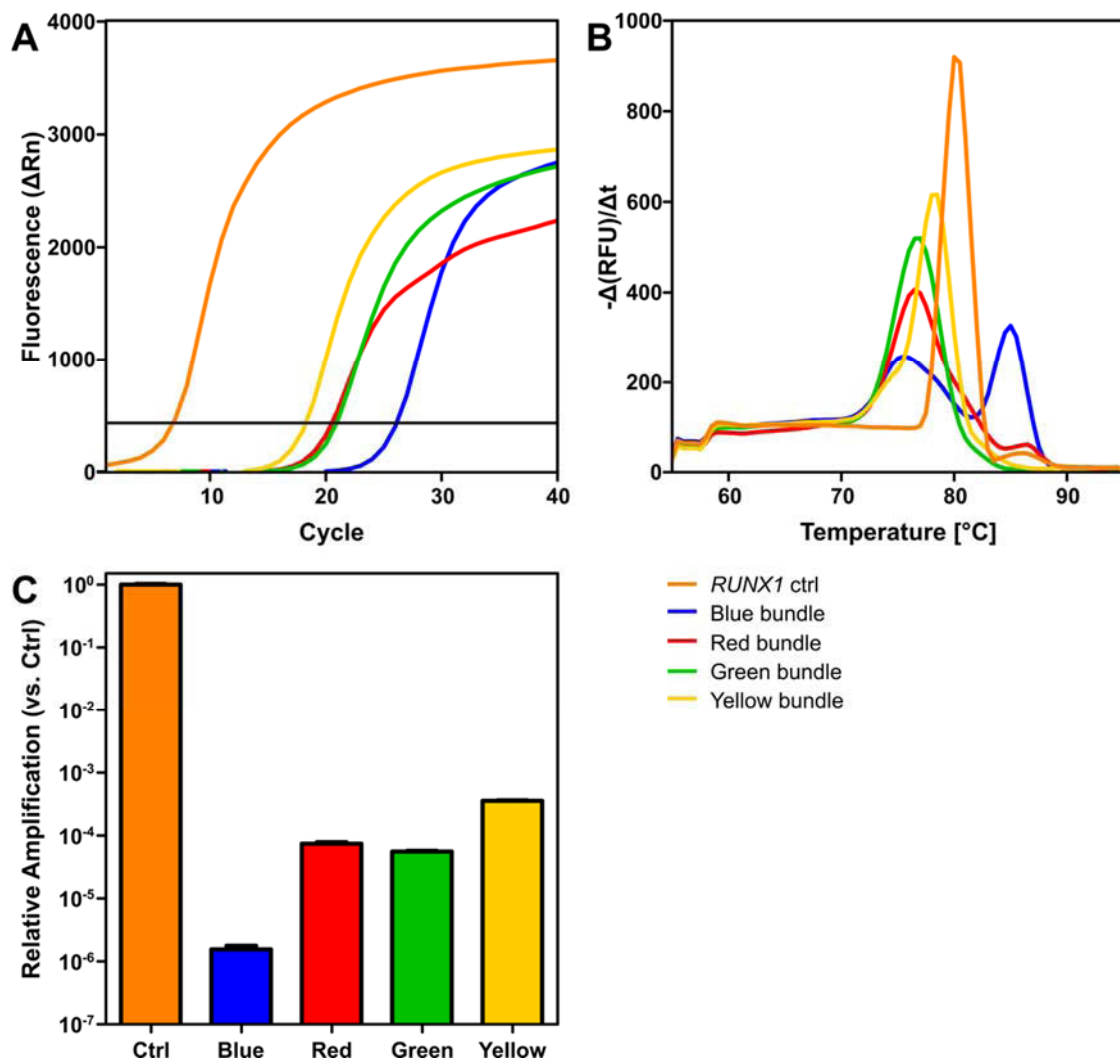


Figure 5.19: Example for the first step of the validation of an *ETV6-RUNX1* positive cord blood (N926). **(A)** Amplification plot of the Real-Time PCR conducted with multiplexed primers. Each primer bundle is represented by its respective color, the *RUNX1* control (ctrl) is depicted in orange. The yellow bundle amplified best. **(B)** Dissociation plot of the Real-Time PCR shown in (A). The yellow bundle is the only one with a peak in the expected temperature range. **(C)** The amplification was compared to the *RUNX1* ctrl. The results of the yellow bundle hint at a frequency of approximately 3×10^{-4} . Rn = normalized reporter signal, RFU = relative fluorescent units.

Within a cohort of 1,000, 50 umbilical cord blood samples could be identified as *ETV6-RUNX1* positive. Two samples, N505 and N531, harbored two coexistent different *ETV6-RUNX1* breaks (Table 5.6, Figure 5.21) [166]. For every positive sample the frequency of the *ETV6-RUNX1* carrying cells was estimated from the Real-Time PCR results (Table 5.6). These are not measured but estimated values and represent more a trend than an exact number. Most *ETV6-RUNX1* carrying cells had estimated frequencies of 1×10^{-4} to 5×10^{-3} , few had a higher frequency of up to 4×10^{-2} , and

others had a lower frequency of down to 1.5×10^{-5} . Latter showed the estimating character of these numbers, as a dilution of one in 10^5 cells usually could not be validated with GIPFEL (chapter 5.1.4.1).

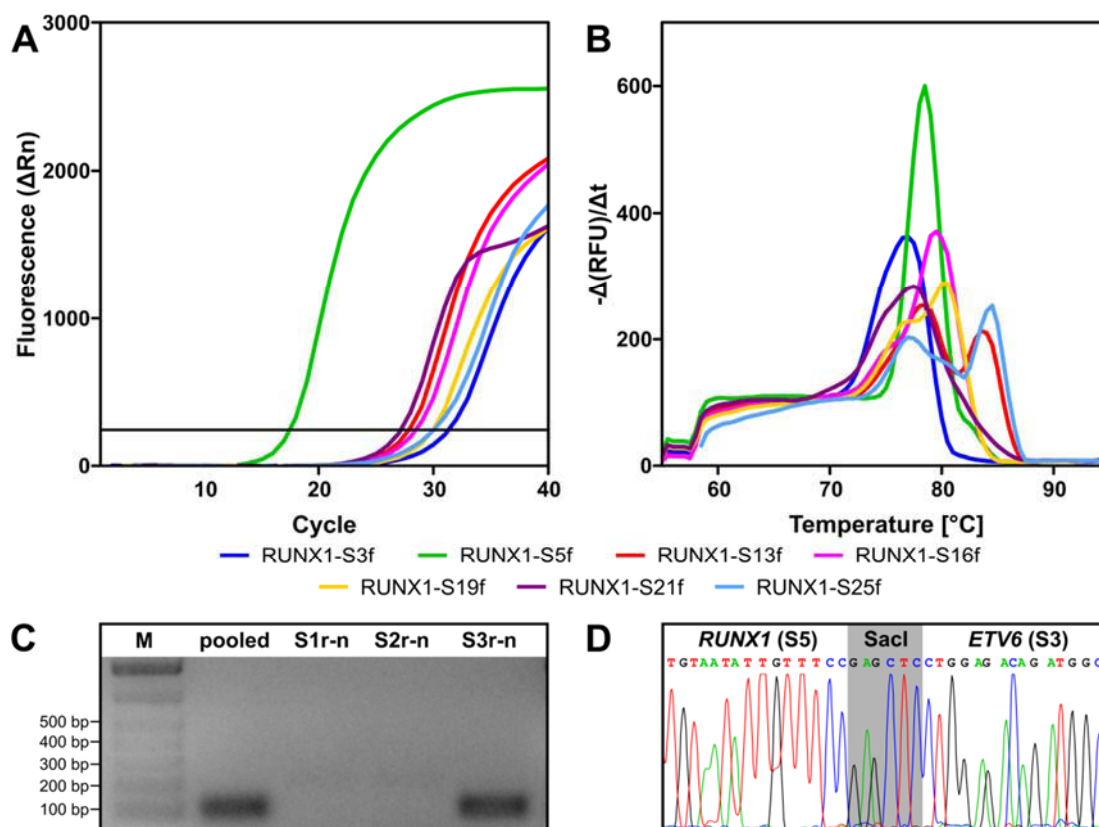


Figure 5.20: Example for the further steps of the validation of an *ETV6-RUNX1* positive cord blood (N926). **(A)** Amplification plot of the Real-Time PCR with demultiplexed forward primers from the yellow bundle. RUNX1-S5f showed the best amplification. **(B)** Dissociation plot of the demultiplexed Real-Time shown in (A). The plot shows that primer RUNX1-S5f amplified a product of expected size. **(C)** Gel electrophoresis of the products from the Real-Time PCR shown in (A). Used reverse primers are given. Besides the pooled reverse primers, *ETV6-S3r-n* was the only primer with a specific product of the expected size of 106 bp. QIAGEN's 1 kb plus marker served as marker (M). **(D)** Result of the Sanger sequencing of the product with RUNX1-S5f and *ETV6-S3r-n*. The sequencing unveiled the expected sequence. The *SacI* restriction site is depicted in gray. Rn = normalized reporter signal, RFU = relative fluorescent units.

Table 5.6: *ETV6-RUNX1* positive newborns from umbilical cord blood screening. The forward primer bundle as well as forward and reverse primers used for detection are given. Newborns N505 and N531 had two different breakpoints, both primer combinations are shown. The frequencies of the *ETV6-RUNX1* carrying cells were estimated on basis of the Real-Time PCR results [166].

Newborn ID	Forward Primer Bundle	Forward Primer	Reverse Primer	Estimated Frequency
N005	green	RUNX1-S6f	ETV6-S1r-n	1×10^{-4}
N059	yellow	RUNX1-S13f	ETV6-S2r-n	6×10^{-3}
N099	green	RUNX1-S2f	ETV6-S2r-n	2×10^{-4}
N260	red	RUNX1-S14f	ETV6-S2r-n	1.5×10^{-4}
N285	red	RUNX1-S11f	ETV6-S1r-n	8×10^{-4}
N286	green	RUNX1-S8f	ETV6-S1r-n	1×10^{-4}
N382	red	RUNX1-S11f	ETV6-S3r-n	1×10^{-3}
N424	blue	RUNX1-S23f	ETV6-S1r-n	5×10^{-4}
N439	blue	RUNX1-S23f	ETV6-S1r-n	4×10^{-4}
N440	green	RUNX1-S18f	ETV6-S2r-n	1×10^{-3}
N441	blue	RUNX1-S22f	ETV6-S2r-n	1×10^{-4}
N447	blue	RUNX1-S28f	ETV6-S3r-n	2×10^{-3}
N463	red	RUNX1-S10f	ETV6-S2r-n	3×10^{-4}
N472	yellow	RUNX1-S21f	ETV6-S3r-n	1×10^{-4}
N479	yellow	RUNX1-S13f	ETV6-S2r-n	1×10^{-3}
N493	green	RUNX1-S18f	ETV6-S2r-n	1×10^{-2}
N496	yellow	RUNX1-S5f	ETV6-S1r-n	1.5×10^{-3}
N505	red	RUNX1-S11f	ETV6-S3r-n	4×10^{-3}
	yellow	RUNX1-S13f	ETV6-S1r-n	7×10^{-4}
N506	blue	RUNX1-S28f	ETV6-S1r-n	5×10^{-3}
N521	red	RUNX1-S4f	ETV6-S2r-n	4×10^{-2}
N522	red	RUNX1-S10f	ETV6-S2r-n	1×10^{-2}
N527	yellow	RUNX1-S21f	ETV6-S1r-n	4×10^{-3}
N531	green	RUNX1-S29f	ETV6-S2r-n	5×10^{-3}
	yellow	RUNX1-S13f	ETV6-S2r-n	1.5×10^{-3}
N548	yellow	RUNX1-S21f	ETV6-S2r-n	9×10^{-5}
N563	red	RUNX1-S11f	ETV6-S2r-n	8×10^{-4}
N578	red	RUNX1-S24f	ETV6-S3r-n	8×10^{-5}
N590	blue	RUNX1-S26f	ETV6-S1r-n	4×10^{-4}
N599	red	RUNX1-S10f	ETV6-S3r-n	5×10^{-3}
N619	yellow	RUNX1-S5f	ETV6-S3r-n	6×10^{-3}
N622	blue	RUNX1-S26f	ETV6-S2r-n	3×10^{-4}
N630	blue	RUNX1-S12f	ETV6-S2r-n	1×10^{-2}
N651	red	RUNX1-S4f	ETV6-S2r-n	1×10^{-4}
N670	blue	RUNX1-S12f	ETV6-S3r-n	1×10^{-2}
N673	green	RUNX1-S8f	ETV6-S2r-n	1×10^{-2}
N674	blue	RUNX1-S28f	ETV6-S3r-n	9×10^{-3}
N726	blue	RUNX1-S22f	ETV6-S1r-n	4×10^{-5}
N729	blue	RUNX1-S23f	ETV6-S3r-n	5×10^{-5}
N731	red	RUNX1-S4f	ETV6-S3r-n	1.5×10^{-5}
N732	red	RUNX1-S10f	ETV6-S2r-n	1×10^{-4}
N770	yellow	RUNX1-S13f	ETV6-S3r-n	4×10^{-4}
N775	red	RUNX1-S14f	ETV6-S2r-n	1×10^{-3}
N784	blue	RUNX1-S28f	ETV6-S2r-n	3×10^{-4}
N791	green	RUNX1-S6f	ETV6-S2r-n	1×10^{-4}
N795	red	RUNX1-S24f	ETV6-S3r-n	1×10^{-4}
N817	yellow	RUNX1-S25f	ETV6-S1r-n	2×10^{-3}
N823	yellow	RUNX1-S13f	ETV6-S1r-n	1×10^{-3}
N890	blue	RUNX1-S23f	ETV6-S3r-n	1×10^{-3}
N908	blue	RUNX1-S15f	ETV6-S3r-n	6×10^{-4}
N912	red	RUNX1-S24f	ETV6-S2r-n	3×10^{-3}
N926	yellow	RUNX1-S5f	ETV6-S3r-n	3×10^{-4}

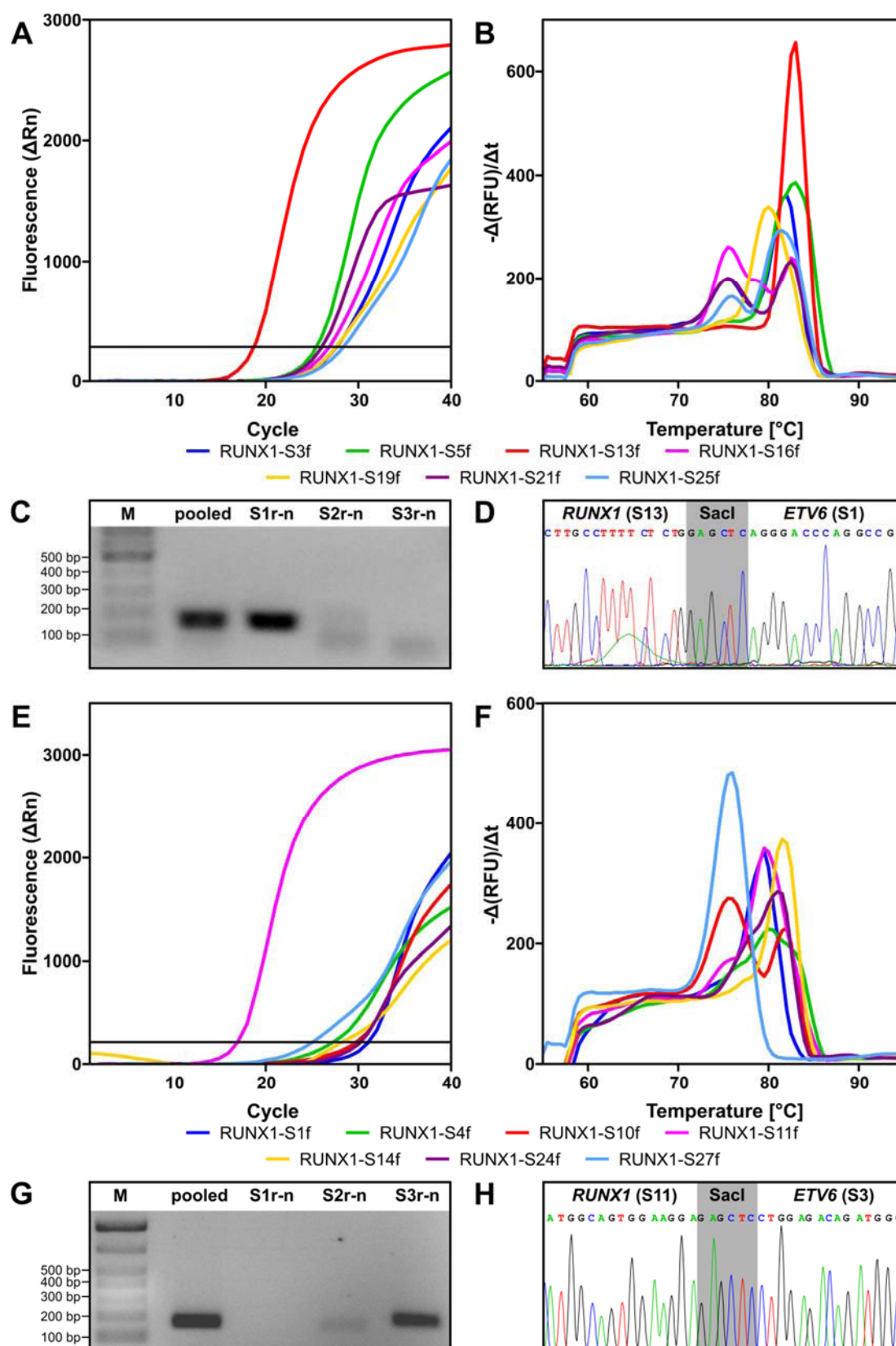


Figure 5.21: Example for a cord blood sample with two *ETV6-RUNX1* fusions (N505). **(A)** Amplification plot with demultiplexed primers from the yellow bundle. RUNX1-S13f shows amplification. **(B)** Dissociation plot of the Real-Time PCR shown in (A). **(C)** The gel electrophoresis unveils *ETV6-S1r-n* to be the reverse primer with an expected product size of 159 bp. **(D)** Sequencing with RUNX1-S13f and *ETV6-S1r-n* leads to the expected sequence. The *SacI* restriction site is depicted in gray. **(E)-(H)** As (A)-(D) but with primers from the red bundle, RUNX1-S11f and *ETV6-S3r-n* as positive primers, and a product size of 166 bp. Rn = normalized reporter signal, RFU = relative fluorescent units, M = marker, 100 bp plus (QIAGEN) for (C) and 1 kb plus (QIAGEN) for (G).

5.2.3 Validation of Cord Blood Screening Results

The results obtained in the cord blood screening should be validated. Therefore, RNA from 52 cord bloods and the cell lines REH and HEK-293 was isolated and reverse transcribed to cDNA. As the number of leftover cord blood samples was limited, only two *ETV6-RUNX1* positive samples (N005 and N260) and 50 *ETV6-RUNX1* negative samples were included in this experiment. Additionally, the cell lines REH and HEK-293 served as positive and negative control, respectively. The cDNA was subjected to a reverse transcription Real-Time PCR using primers and a FAM-TAMRA-labeled probe supplied by the ipsogen *ETV6-RUNX1* Kit (QIAGEN).

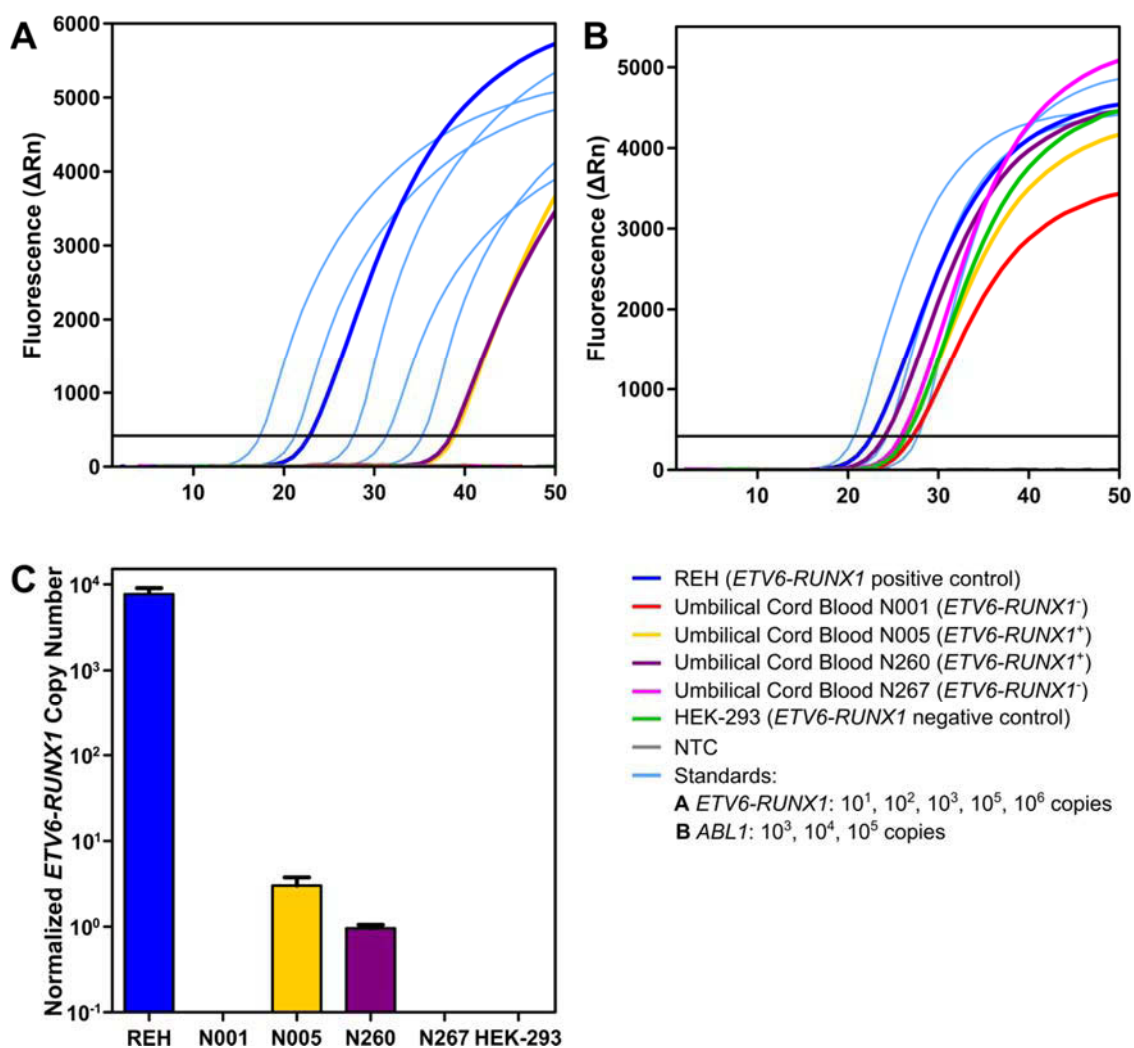


Figure 5.22: Results from reverse transcription Real-Time PCR regarding *ETV6-RUNX1* transcription in cord blood samples and cell line controls. **(A)** Amplification plot for *ETV6-RUNX1*. The thin light blue lines indicate the *ETV6-RUNX1* standards. **(B)** Amplification plot for the internal control gene *ABL1*. The thin light blue lines indicate the *ABL1* standards. **(C)** Normalized *ETV6-RUNX1* copy number. Each sample was normalized with its *ABL1* control to *ETV6-RUNX1* transcripts per 10^4 *ABL1* transcripts. N001 and N267 are shown exemplarily for the 50 *ETV6-RUNX1* negative samples. NTC = non-template control.

The transcription of *ETV6-RUNX1* (Figure 5.22 A) was normalized to that of *ABL1* (Figure 5.22 B), and the transcript could only be detected in the samples that scored *ETV6-RUNX1* positive in the GIPFEL screening (Figure 5.22). The REH cell line, which served as a positive control, showed approximately 8×10^3 *ETV6-RUNX1* transcripts per 10^4 *ABL1* transcripts (Figure 5.22 C). The *ETV6-RUNX1* negative samples (N001 and N267 are shown as examples) as well as the negative control HEK-293 showed no transcription of *ETV6-RUNX1* at all (Figure 5.22 C), neither did the non-template control. The *ETV6-RUNX1* positive cord bloods N005 and N260 had approximately one to three *ETV6-RUNX1* transcripts per 10^4 *ABL1* transcripts (Figure 5.22 C) [166].

The PCR products of REH, N005, and N260 were purified using the PCR clean-up protocol of the QIAquick Gel Extraction Kit (QIAGEN) and subsequently cloned into the vector pGEM-T Easy (Promega). After amplification and plasmid isolation, the PCR products were Sanger sequenced using the primers T7 and SP6 (Table 3.17). All three PCR products turned out to be the expected fusion of *ETV6* exon 1 to *RUNX1* exon 2 (Supplemental Figure 10.2).

5.3 *ETV6-RUNX1*⁺ Mouse Model

A mouse model with human *ETV6-RUNX1* was developed by Isidro Sánchez-García's group in Salamanca. These mice expressed *ETV6-RUNX1* under the control of the *Ly6a* promoter. They did not only function as a leukemia model but were also used to compare healthy *ETV6-RUNX1* carrying mice to wild-type mice.

A total of 93 *ETV6-RUNX1* positive mice were included in this study. Of these 93 mice, 10 (10.75%) developed ALL as revealed by flow cytometry and histological experiments carried out by Isidro Sánchez-García's group. Interestingly, the mice only developed leukemia after being moved from SPF to CF [167]. This correlated with an earlier finding which showed that *Pax5*^{+/-} heterozygous mice developed leukemia only after being exposed to infection [144].

5.3.1 DNA Methylation Analysis

DNA from three leukemic and four healthy *ETV6-RUNX1*⁺ mice housed in CF, seven healthy *ETV6-RUNX1*⁺ mice housed in SPF, and four wild-type mice from each facility was submitted to DNA methylation analysis and subsequent PCA (Table 3.8).

When looking at the 150 most differentially methylated genes, the mice clustered in four groups (Figure 5.23 A). One cluster was generated by the wild-type mice, regardless of the housing. The healthy *ETV6-RUNX1*⁺ mice formed two clusters. However, mice from both housings were found in either group. The fourth cluster consisted of the leukemic mice and was more diverse. When the 500 most differentially methylated genes were used for the analysis, one group of healthy *ETV6-RUNX1*⁺ mice and the wild-type mice clustered together (Figure 5.23 B).

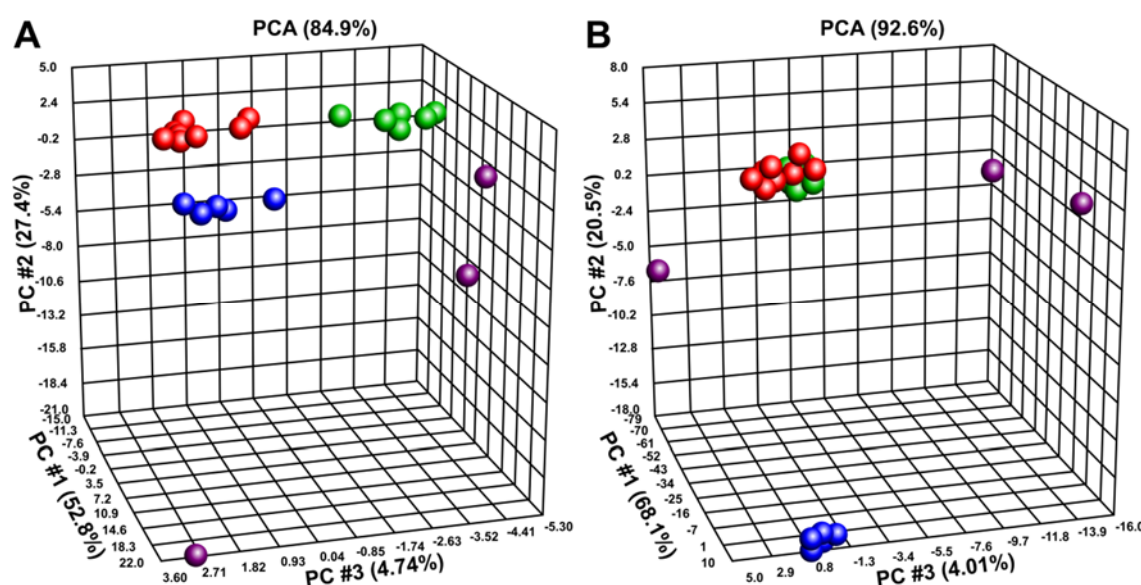


Figure 5.23: Principal component analysis of DNA methylation. DNA methylation of leukemic *ETV6-RUNX1*⁺ mice, healthy *ETV6-RUNX1*⁺ mice from SPF and CF, and wild-type mice from SPF and CF was investigated. **(A)** The 150 most differentially methylated genes were taken into account. Four groups could be identified: wild-type mice (red), two clusters of healthy *ETV6-RUNX1*⁺ mice (blue and green), and leukemic *ETV6-RUNX1*⁺ mice (purple). Except for the cluster with the leukemic mice, mice from SPF and CF are present in all clusters. **(B)** As (A), but with the 500 most differentially methylated genes. The wild-type cluster and the green cluster form a new cluster. Percentages at the axes represent the degree by which the newly introduced principal component (PC) represented the diversity seen in the original data. The percentage above the graph does the same for all three PCs combined.

The four subgroups were further characterized by the methylation status of the genes and the time spent in their respective housing facility (Figure 5.24). This demonstrated

the differences between the four groups. While group 1 (wild-type) and group 3 (healthy *ETV6-RUNX1*⁺) hardly showed any differences (Figure 5.23 B and Figure 5.24), the other group of healthy *ETV6-RUNX1*⁺ mice (group 2) showed substantially less methylation (Figure 5.24). In general, the leukemic mice (group 4) had higher degrees of DNA methylation, with mouse S825 having the highest (Figure 5.24, bottom row). Except for group four, there was no apparent difference between the housings. The leukemic mice had a different methylation pattern and were all housed in the CF. Furthermore, these mice were older than the healthy mice. The difference between the two healthy *ETV6-RUNX1*⁺ subgroups could neither be linked to the housing facility nor to the age. As all but two mice were female, the difference between sexes could not be investigated.

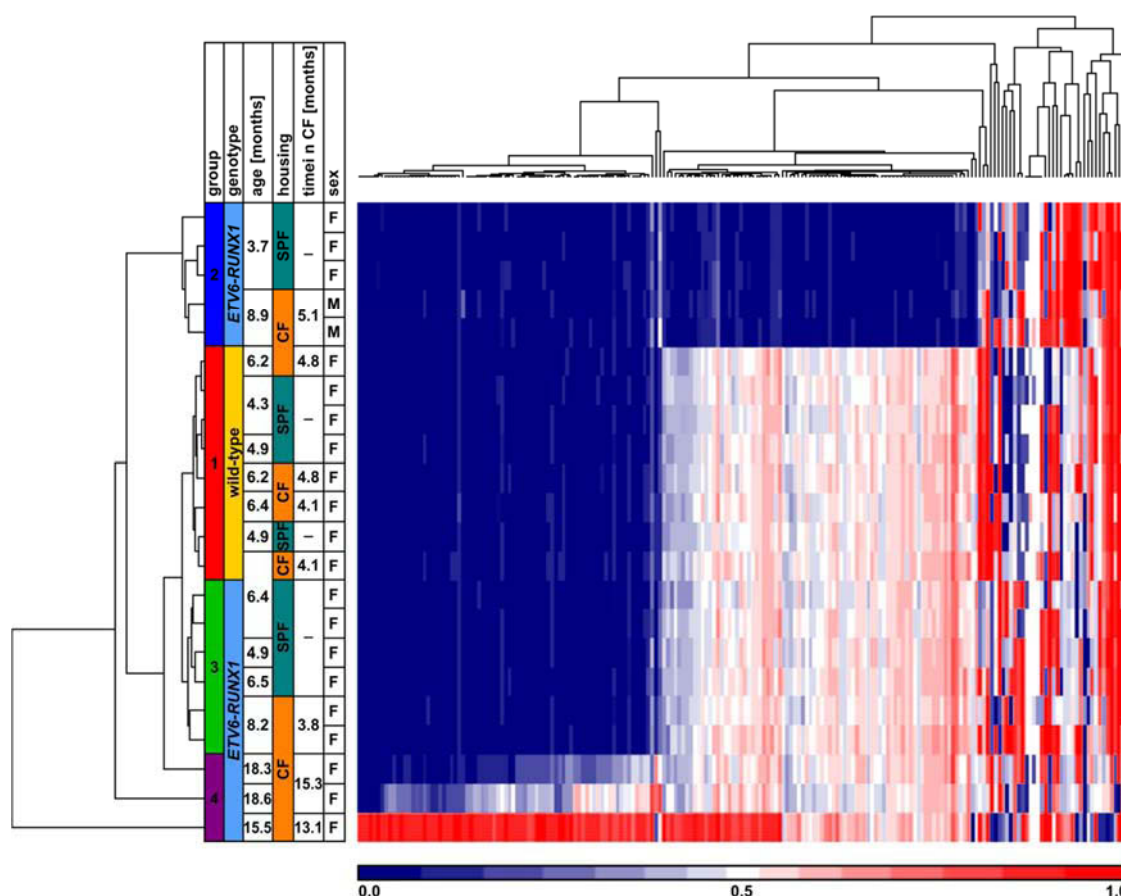


Figure 5.24: Heat map of unsupervised clustering of differentially methylated genes. On the left, the four groups, the genotype, and the housing are given. For every mouse the age, the time spent in conventional facility (CF), and the sex are shown (F = female, M = male). Blue shows no or little methylation, whereas red shows higher methylation.

5.3.2 Comparison of Murine and Human Mutations

Blasts of six of the ten mice which developed leukemia were whole exome sequenced (Table 3.7). The other four had too little blast cells. The mice had an average of 30 somatic mutations (range 12-56) (Figure 5.25). Four showed no mutations in known

sample ID	J406		J408		S825		K213		R222		K955	
	tumor	control	tumor	control	tumor	control	tumor	control	tumor	control	tumor	control
overall	29522	30818	27240	26919	29720	34001	15572	15897	16570	16690	16294	16297
somatic	28		56		53		15		12		15	
somatic cancer	2		3		0		0		0		0	

Figure 5.25: Results from mouse whole exome sequencing. *Overall* shows the number of every called mutation in every single sample. Underneath, the numbers of somatic mutations and at the bottom the numbers of somatic mutations affecting known cancer related genes are shown. Cancer related genes were taken from COSMIC's cancer gene consensus [161].

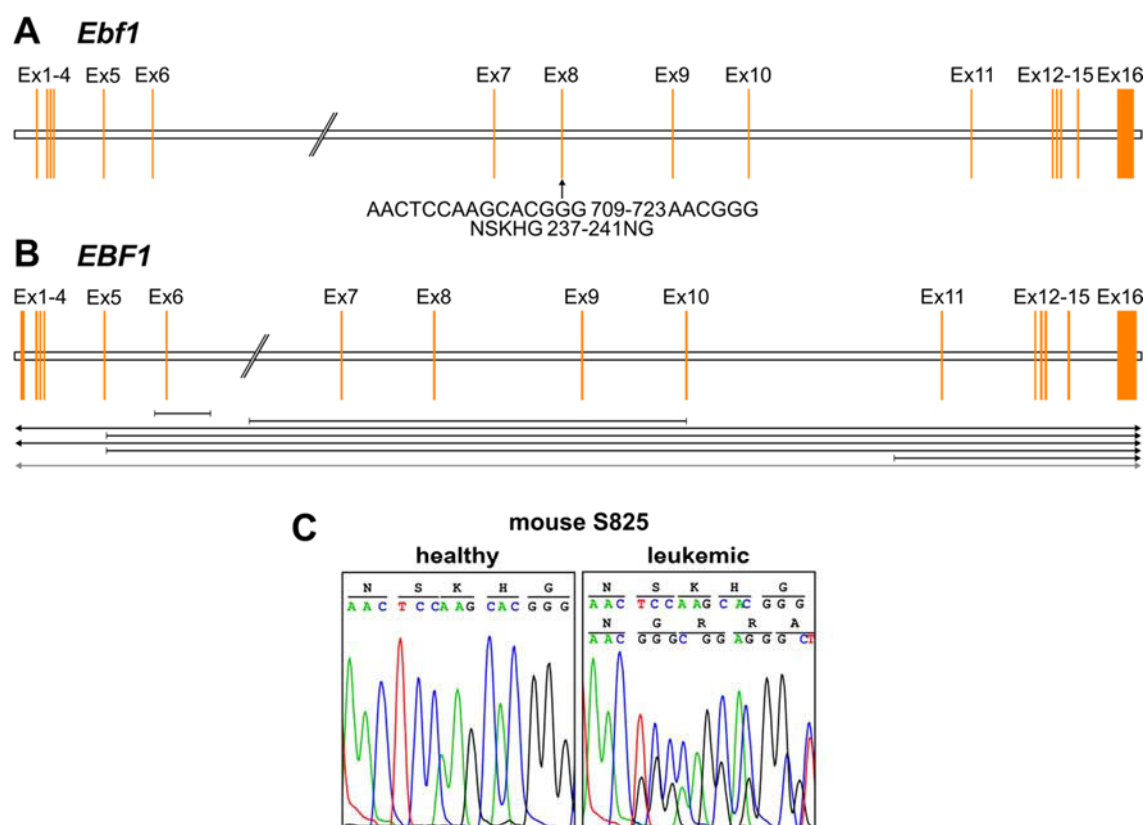


Figure 5.26: Mutations of murine *Ebf1* and human *EBF1*. **(A)** 9 bp deletion in exon 8 of murine *Ebf1*, revealed by whole exome sequencing of mouse S825, leading to the loss of a DNA binding region. **(B)** Deletions in human *EBF1* described in the literature. Deleted regions are shown under the gene locus. Arrows indicate that the deletion extends beyond the gene. Deletions shown in black are taken from [81] and the deletion in gray from [168]. **(C)** Sanger sequencing of the 9 bp deletion in mouse S825. The healthy control is shown on the left; the sequencing of the leukemic sample proves the deletion to be heterozygous with a frequency of 0.33. Both resulting products are indicated above the chromatogram.

cancer related genes [161], whereas two had two and three mutations, respectively (Figure 5.25) [167].

The somatic mutations were investigated further and compared to mutations from the human cohort. Whole exome sequencing revealed a heterozygous deletion of 9 bp in the *early B cell factor 1* gene (*Ebf1*) of mouse S825, resulting in the loss of three amino acids (Figure 5.26 A, C). This mutation deletes a DNA binding site. Even though the investigated human cohort showed no mutation in *EBF1*, deletions in *EBF1* are well known from literature (Figure 5.26 B) [81, 168].

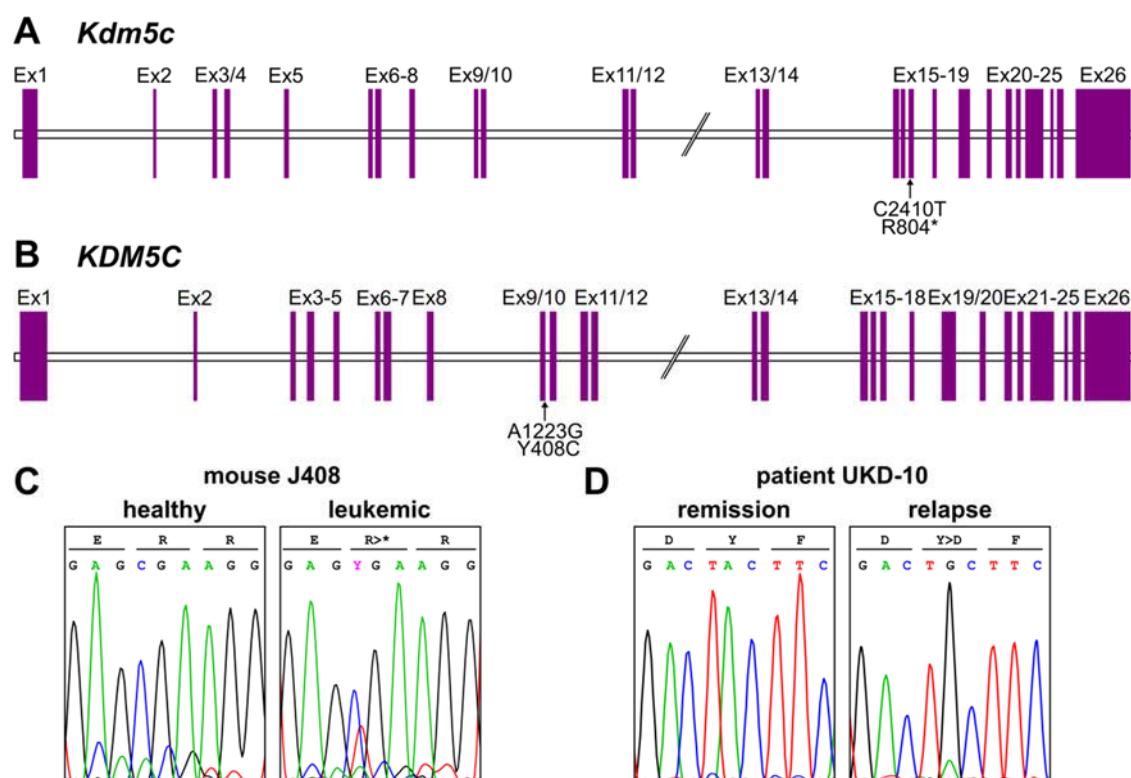


Figure 5.27: Mutations of murine *Kdm5c* and human *KDM5C*. **(A)** A single base pair substitution leading to a premature stop in exon 17 of murine *Kdm5c* in mouse J408. **(B)** Missense mutation in exon 9 of human *KDM5C*, revealed by whole exome sequencing of patient UKD-10. **(C)** Sanger sequencing results of mouse J408. The left panel shows the healthy control and the right panel the mutation in the leukemic mouse. **(D)** Chromatograms of the Sanger sequencing of UKD-10. The left panel shows the remission and the right panel the relapse with the mutation.

Furthermore, mouse J408 had a mutation in the *lysine demethylase 5c* gene (*Kdm5c*), leading to a premature stop (Figure 5.27 A, C). The same gene was affected in patient UKD-10, where a somatic missense mutation appeared (Figure 5.27 B, D). *KDM5C* demethylates di- or trimethylated lysine 4 on histone 3 (H3K4me2/3). Loss of function

therefore leads to elevated levels of H3K4me3 which is a known binding site for RAG2 [169] and hence for the entire RAG1/2 complex.

As the identified mutations in *Kdm5c/KDM5C* affected a histone regulating gene and featured a link to RAG1/2, the mice and patients were screened for further mutations in histone and histone regulating genes. Additional mutations were found in genes of the *KDM* family by whole exome sequencing, more frequently in patients at relapse (Table 5.7 and Supplemental Figure 10.3). Analysis of structural aberrations revealed further mutations in histone related genes, especially deletions. Once again, genes of the *KDM* family were affected in particular (Table 5.8). However, both, initial and relapsed samples, were affected equally [167].

In addition, a cohort published by Papaemmanuil *et al.* [82], consisting of 51 *ETV6-RUNX1* positive patients, was screened for single nucleotide variants (SNV), small indels (Table 5.7), and structural aberrations (Table 5.8) in histone related genes. Again, members of the *KDM* family were affected, especially by structural aberrations. In the own cohort, *KDM2B* was affected in two patients, one with a missense mutation and one with a duplication. Like *KDM5C*, *KDM2B* demethylates H3K4me2/3 [167].

Table 5.7: SNVs and small indels in histone related genes. Mice and patients from the own [167] cohort are shown above the dashed line, patients from Papaemmanuil *et al.* [82] below. D = diagnosis, R = relapse.

Gene	Patient / Mouse ID	Genomic Position	Nucleotide Exchange	Amino Acid Exchange
<i>KDM2B</i>	6 (R)	12:121,443,757	C2488T	H830Y
<i>KDM5C</i>	10 (R)	X:53,211,806	A1223G	Y408C
<i>Kdm5c</i>	J408	X:152,267,124	C2410T	R804*
<i>KDM6A</i>	5 (D/R)	X:45,107,510	C3979G	P1327A
<i>KDM6B</i>	5 (D)	17:7,848,213	C4000G	P1334A
<i>KDM6B</i>	5 (R)	17:7,848,541-7,848,543	C1925T	P642L
			CAC2253-2255>del	VT751-752V
<i>ATF2</i>	PD4031a	2:175,114,896	C476G	P159R
<i>ATF7IP</i>	PD3972a	12:14,436,172	G1712A	R571Q
<i>ATF7IP</i>	PD4023a	12:14,425,002	C1087T	R363*
<i>HDAC1</i>	PD4028a	1:32,330,868	AT939-940>del	Y314fs>nonsense
<i>HIST1H2BG</i>	PD3972a	6:26,217,028	G288C	Q96H
<i>KDM6A</i>	PD3959a	X:45,085,959	3528_3529insAGGT	L1177fs*4
<i>SETD5</i>	PD4024a	3:9,447,906	C2003T	S668L

Table 5.8: Structural aberrations in histone related genes. Patients from the own [167] cohort are shown above the dashed line, patients from Papaemmanuil *et al.* below [82]. D = diagnosis, R = relapse.

Patient ID	Genomic Position	Mutation	Genes in Region
7 (R)	10:70,075,392-70,076,000	duplication	<i>H2AFY2</i>
9 (D)	6:27,889,980-27,891,013	deletion	<i>HIST1H3J</i>
9 (D)	5:135,342,515-135,343,322	deletion	<i>H2AFY</i>
9 (D)	16:31,130,802-31,131,913	deletion	<i>KAT8</i>
9 (D)	2:86,449,898-86,450,186	deletion	<i>KDM3A</i>
9 (D)	12:121,549,465-121,549,760	duplication	<i>KDM2B</i>
9 (R)	1:202,758,035-202,759,189	deletion	<i>KDM5B</i>
9 (R)	11:67,206,526-67,207,576	deletion	<i>KDM2A</i>
10 (D)	1:43,694,420-43,695,739	deletion	<i>KDM4A</i>
10 (D)	1:202,807,993-202,809,056	deletion	<i>KDM5B</i>
10 (R)	12:14,367,484-14,499,521	deletion	<i>ATF7IP</i>
16 (D)	5:135,369,987-135,371,097	deletion	<i>H2AFY</i>
16 (D)	3:129,549,050-129,549,326	deletion	<i>HF100</i>
16 (D)	12:14,367,482-14,436,408	deletion	<i>ATF7IP</i>
16 (R)	5:141,634,616-141,635,995	deletion	<i>HDAC3</i>
PD3954a	12:233,827-19,257,786	deletion	<i>KDM5A, HIST4H4, H2AFY, ATF7IP</i>
PD3956a	5:103,321,431-141,722,866	tandem duplication	<i>KDM3B, HDAC3, H2AFY</i>
PD3958a	17:2,456,103-30,130,363	intrachromosomal inversion	<i>KDM6B</i>
PD3964a	11:101,335,608-122,475,886	deletion	<i>H2AFX</i>
PD3971a	12:11,416,000-18,252,434	deletion	<i>HIST4H4, H2AFJ, ATF7IP</i>
PD3971a	12:11,441,739-14,738,198	deletion	<i>ATF7IP</i>
PD4008a	3:46,980,792-50,018,051	deletion	<i>SETD2</i>
PD4009a	X:12,779,767-52,792,606	tandem duplication	<i>KDM6A, HDAC6</i>
PD4013a	12:9,587,821-117,607,518	deletion	<i>HDAC7, H1FNT, H3F3C, H2AFJ, ATF7IP</i>
PD4014a	12:10,613,519-14,366,444	deletion	<i>ATF7IP</i>
PD4021a	X:6,751,594-108,107,171	tandem duplication	<i>KDM6A, KDM5C, HDAC6, HDAC8</i>
PD4021a	X:51,389,390-75,013,398	deletion	<i>KDM5C, HDAC8</i>
PD4021a	X:63,491,761-105,543,681	deletion	<i>HDAC8</i>
PD4022a	12:8,637,963-15,102,210	deletion	<i>HIST4H4, H2AFJ, ATF7IP</i>
PD4022a	12:8,913,190-16,018,835	deletion	<i>HIST4H4, H2AFJ, ATF7IP</i>
PD4025a	11:93,784,108-99,697,061	deletion	<i>KDM4D, KDM4E</i>
PD4026a	9:2,829,313-27,538,401	deletion	<i>KDM4C</i>
PD4028a	6:113,953,122-159,160,389	deletion	<i>HDAC2</i>
PD4028a	12:14,367,286-14,644,542	deletion	<i>ATF7IP</i>
PD4031a	12:1,045,031-27,905,152	deletion	<i>HIST4H4, H2AFJ, ATF7IP</i>
PD4034a	4:92,738,002-154,319,641	deletion	<i>H2AFZ, SETD7</i>
PD4036a	12:31,983,253-125,873,960	tandem duplication	<i>HDAC7, H1FNT, KDM2B</i>
PD4037a	12:9,724,748-19,742,174	deletion	<i>HIST4H4, H2AFJ, ATF7IP</i>

6 Discussion

6.1 Establishment of the GIPFEL Method

The presence of a translocation is usually detected by either RT-PCR [125] or by FISH [170]. While both methods work, they also have downsides. To begin with, the RT-PCR relies on fragile and unstable RNA which can lead to false negative results. FISH, on the other hand, requires a high level of experience when interpreting the results. Through overlay of signals, it can come to false positive results. Another possible method to detect translocations is LDI-PCR [124, 126] which uses restriction enzyme digest and religation to detect translocation breakpoints via long distance PCR. For this method, however, several conditions have to be fulfilled. First, the restriction fragment containing the breakpoint has to be smaller than the wild-type fragment [124]. Second, the used restriction enzyme cutting close to the breakpoint has to be chosen for every individual breakpoint [124]. Therefore, this method is better suited for the determination of the exact genomic localization of a few breakpoints than for population-based high throughput screenings.

This called for the establishment of a new technique that avoids these pitfalls. This method has to be robust and use stable DNA as a template. Ideally, it does not require previous knowledge of the breakpoint and can be adapted to all translocations. The GIPFEL method fulfills all these requirements. It was established for five translocations including the most common ones in childhood leukemia: *ETV6-RUNX1*, *TCF3-PBX1* and *KMT2A* translocations involving *AFF1*, *MLLT3*, and *MLLT1* [19, 145]. GIPFEL was developed in cooperation with the group of Robert K. Slany at the Friedrich Alexander University Erlangen where Elisa Füller established the method for the *KMT2A* translocations [145].

Unlike LDI-PCR, which also relies on restriction enzyme digest, GIPFEL does not need different restriction enzymes depending on the breakpoint. Instead, one restriction enzyme was selected for each translocation type. *SacI*, *MfeI*, and *BamHI* were chosen for *ETV6-RUNX1*, *TCF3-PBX1*, and the *KMT2A* translocations, respectively. For every

restriction site in *RUNX1*, *PBX1*, *AFF1*, *MLLT3*, and *MLLT1* BCRs, a forward primer was designed, and for every restriction site in *ETV6*, *TCF3*, and *KMT2A* BCRs, a reverse and a nested reverse primer were designed. Afterwards, cell line DNA carrying the respective translocation underwent the GIPFEL procedure. In proof of principle experiments, it was shown that the method works. These experiments did not only proof the functionality of the method but also that GIPFEL is specific to DNA with the investigated translocation. Furthermore, the results from the cooperation partner demonstrated the same for the *KMT2A* translocations [145].

6.2 Optimization of GIPFEL

Having established the functionality of the GIPFEL method in principle, the method had to be optimized in order to achieve high sensitivity. To that end, optimizations were carried out concerning the duration of the restriction enzyme digest, the purification of this digest, the ligation duration and temperature, the purification of the exonuclease digest, the introduction of a pre-amplification PCR, the multiplexing of primers, and WGA.

Six different durations were tested for the restriction enzyme digest. While the time had only a minor impact on the performance of *SacI*-HF (NEB), *MfeI*-HF (NEB) was heavily influenced by the duration of the digest. Best results were achieved with digestion for 2 h. On one hand, longer restriction durations enable the enzyme to digest more DNA. On the other hand, if the duration is too long, off-target activity as well as decay of the enzyme is possible. High fidelity enzymes were designed to reduce these problems [171]. In the case of *SacI*, these effects were clearly reduced, whereas *MfeI* was still rather dependent on time. Both enzymes were subsequently used for 2 h.

Like the duration of the restriction enzyme digest, its following purification was also to be optimized. Two different kits were tested for this purpose: the QIAquick Gel Extraction Kit (QIAGEN) and the innuPREP PCRpure Kit (Analytik Jena). Use of the QIAGEN kit led to a more than 14-fold better product amplification in the Real-Time

PCR. This was probably due to the different protocols. The Analytik Jena protocol is optimized for speed, which probably led to the lower yield.

The elution step of the purification was also optimized. Eluting for 5 min instead of 1 min and at 60°C instead of room temperature increased the yield approximately six-fold. The warmth sped up chemical reactions and the prolonged time allowed for more DNA to be eluted.

Moreover, four different temperatures and six different durations were tested for the ligation. Although not significant when compared to 4°C and 12°C, best performance was achieved using 24°C, which was therefore used for further ligations. The used T4 DNA ligase is instable when heated. That is the reason why the ligation at 37°C was significantly less effective. As for the restriction enzymes, the ligase has an optimum time frame for the duration of the ligation. Initially, more time allows the enzyme to ligate more DNA, but with time the performance gets weaker for the same reasons as for the restriction enzymes: off-target activity and decay of the enzyme. Consequently, the ligations were carried out at 24°C for 2 h.

A purification step was introduced after the exonuclease digest. Three different protocols were tested: an ethanol precipitation, an isopropanol precipitation, and a column based purification with the innuPrep PCRpure Kit (Analytik Jena). The alcohol precipitation protocols led to a significantly higher yield than the column based protocol. This was probably due to the speed oriented nature of the Analytik Jena protocol and the fact that the DNA cycles were quite large. Larger fragments do not pass the column as good as small DNA fragments. The ethanol precipitation led to a 1.5-fold higher yield than the isopropanol precipitation and was therefore used for further experiments. Whole genome amplification was tested as numerous samples had only minute amounts of DNA. The WGA had a negative effect on GIPFEL. The amplified products were correct, but the amount of target DNA was reduced when compared to the same amount of non-amplified DNA. During the WGA, not every part of the DNA is amplified at the same rate. This leads to an amplification of the amount of the entire DNA, with

some sections of the DNA being over- or underrepresented. This effect led to the underrepresentation of the target DNA and hence to a worse amplification.

Furthermore, a pre-amplification PCR was introduced to increase the sensitivity of GIPFEL by increasing the total number of PCR cycles. For the same reason, the Real-Time PCR was turned into a semi-nested PCR. Primer multiplexing was introduced for both PCR steps in order to drastically reduce the number of PCR reactions from 174 to 17 for *ETV6-RUNX1* and from 74 to 14 for *TCF3-PBX1*. Moreover, the multiplexing had an additional positive effect as it allowed entering more DNA in every PCR reaction. This was possible due to the fact that patient samples only had to be split in five instead of 87 parts for *ETV6-RUNX1* and only in six instead of 37 parts for *TCF3-PBX1*, which drastically increased sensitivity.

The final protocol allowed GIPFEL to detect one translocation positive cell in 10^4 cells. The method was 100% specific to the investigated translocations; the accuracy depended on the translocation. For *ETV6-RUNX1*, an accuracy of 70% was reached, whereas only 39% accuracy was reached for *TCF3-PBX1*. There are multiple reasons for that. First of all, an accuracy of 100% was highly unlikely, as cryptic translocations involving several chromosomes were unlikely to be identified with this method. Furthermore, the number of positive controls for the establishment of the method was limited, for *TCF3-PBX1* more than for *ETV6-RUNX1*. Additionally, even though all primers fulfilled the quality criteria *in silico*, several primers could not be tested and optimized on a positive sample. Hence, it is very reasonable to assume that some of them were not optimal and therefore reduced the likelihood of detecting the translocation. This effect was more pronounced for *TCF3-PBX1* where only six samples with known breakpoints, including the cell line 697, could be used to establish GIPFEL. Finally, some of the samples were whole genome amplified, which probably had a negative effect on the sensitivity and by that also on accuracy. With the help of the cell line REH, it could be shown that GIPFEL is approximately 100 times less sensitive when whole genome amplified material is used. When WGA was used for the

first 300 samples of the cord blood screening, all samples were found to be negative. The same samples were then analyzed without WGA and six of them turned out to be translocation $t(12;21)$ positive. This emphasizes the negative effect of WGA and might therefore be an important factor for the measured accuracies.

6.3 Advantages and Limitations of GIPFEL

6.3.1 Advantages

The establishment of GIPFEL showed that the method has a specificity of 100%, meaning that so far no sample being negative for the investigated translocation was identified as positive. This was a major requirement for GIPFEL, as it made the method reliable.

The use of GIPFEL offers several advantages when compared to other methods. First of all, unlike RT-PCR, it uses DNA instead of RNA. DNA is easier to handle as it is much more stable and not as easily degraded. This makes GIPFEL very robust. Furthermore, the use of DNA allows GIPFEL to narrow down the breakpoint region to several kilobases through the knowledge of the used primers. There is, however, no need to know where the exact break occurred. FISH, another commonly used technique, demands an experienced investigator to identify translocation positive patients. GIPFEL on the other hand is PCR based and therefore easier to handle.

Although the method shows some similarities to the LDI-PCR [124, 126], it differs crucially from it: GIPFEL does not need an individual setup for every breakpoint. Instead it uses the same restriction enzyme and all primers for every possible breakpoint within the BCR. Therefore, GIPFEL is suited for large scale screenings. The method was initially established for five translocations [145] but can easily be adapted for every translocation of interest and even for large deletions. Hence, GIPFEL theoretically allows for the screening of a sample for several different translocations.

6.3.2 Limitations

Sensitivity tests showed that the detection limit of GIPFEL for all investigated translocations is one translocation positive cell in 10^4 cells [145]. This equals the detection limit of an RT-PCR approach [125]. Furthermore, the accuracy varied between 24% and 83%, depending on the translocation [145]. As mentioned earlier (chapter 6.2), this may be the result of complex translocations involving numerous chromosomes, the use of WGA, and the fact that not all primers could be optimized on a positive control. Whereas latter ones surely can be fixed in the future with the help of more patient material, the former one poses a real limitation to the method. Additionally, for cases where more than one putative positive result is obtained, the final decision whether or not the sample is positive requires testing of multiple primer combinations.

GIPFEL was established with costs of 36 € and 51 € for translocations t(12;21) and t(1;19), respectively, which makes it so far a rather costly method. These prices only cover chemicals and neither machines nor labor. They vary mostly because of the different prices for the restriction enzymes. In the future, prices could be lowered by testing low cost reagents. One main reagent that could be exchanged by a cheaper one is the Brilliant II Real-Time PCR Master Mix (Agilent), provided that the substitute performs equally well. As it is used in large volumes, it would have a major impact on the total price.

For every single sample, a total of at least 27 PCR reactions and 13 further reaction steps, including digests and purifications, had to be conducted. This makes GIPFEL a labor intense method, especially for larger scale screenings.

6.4 Frequency of *ETV6-RUNX1* in the Healthy Population

In the past, contradictory research results have been published regarding the frequency of the t(12;21) translocation in the population [76-80]. While the original study [76] and one following study [80] were able to find the translocation in 1% to 2%

of healthy cord bloods, using reverse transcriptase Real-Time PCR and FISH, later studies could not confirm this finding [77-79].

In order to resolve this problem, the GIPFEL method was used to screen a cohort of 1,000 umbilical cord blood samples from Danish newborns for *ETV6-RUNX1*. Of these 1,000 samples, 50 were positive for the t(12;21) translocation [166]. That means that 5% of the tested newborns carried the *ETV6-RUNX1* fusion. This number might even be higher, as GIPFEL is not 100% accurate [145]. Up to 7% of cord bloods could be positive when the known accuracy of GIPFEL is taken as a basis. A frequency of 5% also means a five-fold higher incidence than the one proposed by Mori *et al.* [76].

The studies of Lausten-Thomsen *et al.* [77] and Mori *et al.* [76] used RNA from cord blood in order to detect the *ETV6-RUNX1* fusion. Latter study used FISH to validate these results. Initially, both studies found 0.6% and 1% translocation positive samples, respectively. However, the Danish study could not validate their RT-PCR findings by FISH [77]. The considerably higher frequency in this work might be owing to the investigated material. While this work used DNA, the other studies investigated RNA and hence the transcription of the fusion gene. This might be one reason for the elevated frequency found by this work. The fact that a fusion gene is present due to a translocation does not necessarily mean that it is transcribed by all cells bearing it or that the transcription level is high enough for detection. Furthermore, RNA is much more instable than DNA, especially when released from dead cells [121]. However, the two *ETV6-RUNX1* positive cord blood samples that were investigated for *ETV6-RUNX1* transcription showed transcripts at a level of 10^{-4} , the detection limit of RT-PCR. The transcriptional levels roughly correlated with the frequencies of *ETV6-RUNX1* positive cells, estimated from the GIPFEL results of these cord bloods.

The present study, the study of Lausten-Thomsen *et al.* [77], and the study of Olsen *et al.* [78] all investigated umbilical cord blood samples from Denmark. Hence, it is highly unlikely that the different results are the product of differences in the population or the environment.

In the past, it could be shown that apoptotic signals can induce double-strand breaks in *ETV6* and *RUNX1* and that this can lead to the *ETV6-RUNX1* fusion [172]. Thus, it is possible that the cells in the stored cord blood underwent this process, leading to elevated levels of the *ETV6-RUNX1* fusion. However, these cells should not proliferate when frozen and therefore it is unlikely that estimated frequencies of up to one *ETV6-RUNX1* positive cell in 100 cells arose solely from double-strand breaks following apoptotic signals.

Two newborns (N505 and N531) had two different *ETV6-RUNX1* fusions [166]. By applying the GIPFEL method, it is not possible to determine whether these breaks occurred in the same or in different cells. If both occurred in the same cell, the newborns should have both developed B cell leukemia, as both *ETV6* alleles would have been lost and this would mark a critical step towards leukemia development [173]. In up to 70% of *ETV6-RUNX1* positive leukemias the second *ETV6* allele is deleted [81]. For reasons of data protection, a follow up of the newborns was not possible. Hence, the later health status could not be investigated. The different estimated frequencies of the fusions, however, indicated that two different clones were involved. If the fusions occurred in the same cell, both frequencies should have been identical. For each newborn, the frequency of one fusion was remarkably higher: for N505 the clone with *RUNX1-S11f* and *ETV6-S3r-n* had a more than five-fold higher frequency than the one with *RUNX1-S13f* and *ETV6-S1r-n*, and for N531 the clone with *RUNX1-S29f* and *ETV6-S2r-n* had a more than three-fold higher frequency than the one with *RUNX1-S13f* and *ETV6-S2r-n*. Whether one or two clones were involved could be tested by FISH using high numbers of cells. Then it would be possible to see, if the translocation positive cells have one or two fusions signals, indicating two and one clone, respectively. However, shortage of material prevented this experiment.

It is known that the presence of the t(12;21) translocation and the resulting *ETV6-RUNX1* fusion alone is not sufficient for leukemia development [9]. Further secondary

lesions are necessary for that. Very common secondary mutations are the deletion of the second *ETV6* allele [81] and further RAG1/2 mediated mutations [82].

The incidence of *ETV6-RUNX1* positive childhood leukemia is approximately 1 in 10,000 [76]. So, if 5% of newborns harbor the translocation, only 0.2% of those actually develop leukemia. This would be even less than the postulated 0.5% to 1% [76, 80]. This emphasizes the rareness of secondary mutations and confirms that the t(12;21) translocation bears a very low risk of leading to leukemia. This is also emphasized by the fact that 0.5% of adults carry the *ETV6-RUNX1* fusion [174] but only 2% of adult leukemias do so [6]. The frequency of *ETV6-RUNX1* carrying cells is, with 10^{-5} to 10^{-6} , lower in adults than in children [174].

6.5 Characterization of *ETV6-RUNX1* Positive Leukemias

With the help of a mouse model expressing the human *ETV6-RUNX1* and a human cohort of *ETV6-RUNX1* positive patients, the *ETV6-RUNX1* positive leukemias were characterized in more detail. The mouse model, for which the coding region was put under the regulation of the *Ly6a* promoter, was provided by Isidro Sánchez-García's group from Salamanca. The human group consisted of 11 patients [167].

6.5.1 *ETV6-RUNX1*⁺ Mouse Model

Of the 93 mice, only ten developed leukemia which equals 10.75%. This resembled the situation in humans, where also only a fraction of people with translocation t(12;21) develop leukemia (see chapter 6.4 and [76, 80]). When the mice were born and kept in the SPF, none developed leukemia, but ten mice became sick after they were moved to the CF [167]. This discrepancy in leukemia incidence between the facilities was probably caused by the exposure to infection in the CF. Infection was first proposed as a causal factor for acute leukemia in 1917 [175] and was proven for *Pax5*^{+/-} heterozygous mice in a similar experimental setup [144]. The mice in this *Pax5*^{+/-} experiment were kept in the same animal facility as the *ETV6-RUNX1*⁺ mice and had a

similar genetic background. Therefore, it is feasible to compare these experiments and to assume that infection also played a key role in leukemia development in *ETV6-RUNX1*⁺ mice. A recent study found space-time clustering for pediatric *ETV6-RUNX1*⁺ ALL, indicating that these clusters are the result of an infection [176]. These findings support the role of infection in leukemogenesis.

6.5.2 DNA Methylation

The DNA methylation of three leukemic mice, seven and four healthy *ETV6-RUNX1*⁺ positive mice from SPF and CF, respectively, and four wild-type mice from each facility was investigated. The mice clustered in four groups: group 1 consisted of all wild-type mice, regardless of the housing facility, groups 2 and 3 of healthy *ETV6-RUNX1*⁺ mice, and group 4 of the leukemic mice. In both groups with healthy *ETV6-RUNX1*⁺ mice were animals from both facilities. Therefore, it is reasonable to assume that the housing conditions had no direct influence on the DNA methylation. Otherwise the healthy *ETV6-RUNX1*⁺ mice and the wild-type mice would have clustered by their housing facility. Furthermore, the age does not seem to play a role for clustering as there were mice of different age in either group. Group 3, consisting of healthy *ETV6-RUNX1*⁺ mice, had a methylation pattern that was very similar to that of the wild-type mice and these mice also clustered together when the 500 most diversely methylated sites were taken into account. This suggests that the healthy *ETV6-RUNX1*⁺ mice from group 3 were very similar to the wild-type mice. None of these mice developed leukemia, and their similarity to wild-type mice might be the reason for that.

The mice of group 2 formed a distinct cluster with considerably less methylation. These mice were also healthy but differed from wild-type and group 3 mice. While these mice were hypomethylated, the leukemic mice were hypermethylated, especially mouse S825. Generally, DNA methylation inhibits transcription [177]. The higher DNA methylation in the leukemic mice is in line with the fact that *ETV6-RUNX1* recruits a corepressor complex including histone deacetylases (HDAC) [178]. A link between

DNA methylation and histone deacetylation has been well established [179-181]. If the mice in group 2 had been hypermethylated instead of hypomethylated, it would have been a clear indication for a tumor prone group. This was, however, not the case, and the different methylation of group 2 must have another reason. As the mice used in this experiment had a mixed background (CBA x C57BL/6J), it is possible that a different mouse strain is dominant for group 2 and therefore the DNA methylation differs.

6.5.3 Characteristics of Human and Murine *ETV6-RUNX1*⁺ ALL

In neither the human nor the murine cohort, there was a common secondary mutation [167]. Four out of six mice had no mutation in known cancer related genes [161], detectable by whole exome sequencing, but deletions exceeding a few base pairs could not be detected with this approach and can therefore not be ruled out. The absence of unifying point mutations is a known characteristic of *ETV6-RUNX1*⁺ ALL [81]. However, genes involved in B cell development are often targeted by deletions [81]. In the murine cohort, *Ebf1* had a small deletion affecting the DNA binding in blasts of one mouse and is known to be deleted in humans [81, 168]. The loss of the DNA binding region probably interferes with the protein function. *Ebf1* is an important gene for B cell development [182]: it promotes chromatin accessibility and DNA demethylation, leading to activation of B cell specific transcription [183]. Hence, loss of *Ebf1* function has an immense impact on B cells and can lead to developmental arrest at pro-B cell stage.

There was a considerable number of mutations in members of the *KDM* family [167]. These proteins demethylate lysines at histones. Methylation/demethylation of lysines at histones can have gene activating and gene silencing effects. Genes of this family showed point mutations and structural aberrations in the analyzed murine, human, and in a published human cohort [82]. A main focus lay on *KDM5C* and *KDM2B*, both of which demethylate H3K4me3 [184, 185]. Overexpression of *KDM5C* leads to drastically reduced H3K4me3 levels [169], loss of function of either of these proteins

leads to elevated levels of H3K4me3, which is a binding site for RAG2 [169, 186]. Subsequently, RAG1 can bind to RAG2 and form the RAG1/2 complex. This mechanism is also active in normal V(D)J recombination [186]. As off-target RAG-mediated recombination is the major oncogenic driver in *ETV6-RUNX1*⁺ ALL [82], this link is of special interest. The loss of function mutations might therefore explain why RAG off-target activity is so frequent in *ETV6-RUNX1*⁺ ALL.

On one hand, loss of function of *KDM5C* or *KDM2B* leads to higher levels of trimethylated H3K4 which in turn leads to transcriptional activity [187-189]. *ETV6-RUNX1*, on the other hand, is a transcriptional repressor [54] that recruits a corepressor complex including HDACs [178]. Histone deacetylation results in transcriptional repression and therefore antagonizes H3K4 trimethylation. These findings are contradicting as *ETV6-RUNX1* leads to gene repression while the mutations of *KDM5C* and *KDM2B* result in gene activation. As it is impossible to infer the order of the mutations from the sequencing, it is not known whether the *KDM* mutations happened early on or are just an effect that appeared later in leukemic cells. In this spirit, it would be possible to think of the *KDM* mutations and the consequential gene activations as a reaction to the gene repression by *ETV6-RUNX1*. This assumption is supported by the finding that SNVs and small indels in *KDM* family genes appeared mainly in relapsed patients. Nevertheless, it cannot be precluded that the mutations appeared early in leukemogenesis. This interpretation is backed by its link to RAG activity via H3K4me3 and by the fact that structural aberrations, especially deletions, of *KDM* genes also appear in initial samples.

However, the KDMs and HDACs do not necessarily act on the same histones and thereby affect the regulation of the same genes. Rather, it is possible that the regulation of different genes is affected and the *ETV6-RUNX1* induced repression and the KDM induced activation cooperatively lead to general misregulation of transcription.

7 Conclusion and Outlook

In the course of this work, a new DNA based screening method for the detection of chromosomal translocations, termed GIPFEL, was developed [145]. This method is robust and sensitive, detecting one translocation positive cell in 10,000 cells. Thereby, it is as sensitive as RT-PCR, but uses more stable DNA instead of RNA. GIPFEL is easily adaptable to all translocations and also suitable for large scale screenings.

The latter point is of special interest insofar as conflicting reports about the frequency of the *ETV6-RUNX1* fusion in newborns have been published in the past [76-80]. With the help of the GIPFEL method, it could be shown on DNA level that the frequency of *ETV6-RUNX1* in a cohort of 1,000 umbilical cord bloods is as high as 5%. These findings suggest that the translocation t(12;21)(p13;q22), which results in the *ETV6-RUNX1* gene fusion, is a rather common lesion among newborns. The leukemia incidence in children is, however, with approximately 0.01% much lower [76]. This implies that the part of *ETV6-RUNX1* positive newborns that develop leukemia is very low, leading to the conclusion that although the *ETV6-RUNX1* gene fusion is a risk factor for childhood leukemia, it is a rather weak one.

It is already known that the *ETV6-RUNX1* fusion alone is not sufficient for leukemia development [9]. In combination with the low penetrance identified by this work, the question which factors lead to leukemia development in *ETV6-RUNX1* positive children had to be raised. With the help of a mouse model and a human *ETV6-RUNX1* positive leukemia cohort it could be shown that a significant number of histone related genes was mutated, especially members of the *KDM* gene family. Whether these mutations are disease driving or acquired later on in the course of the disease could not be unequivocally clarified so far. Exposure to infection seems to be mandatory for disease development, as the mice only developed leukemia after being exposed to common infectious agents. This is in line with earlier findings in which *Pax5*^{+/-} mice also only developed leukemia after infection exposure [144].

The data acquired by this work should be backed up by further experiments. A larger cohort of umbilical cord bloods should be screened for the *ETV6-RUNX1* fusion. Ideally, it should be possible to follow up the donors in order to find out how many of the *ETV6-RUNX1* carriers actually develop *ETV6-RUNX1* positive leukemia in the course of their life. Additionally, the role of *KDM* mutations in *ETV6-RUNX1* positive leukemia has to be investigated further. Larger cohorts may be of help for that, and inducible mouse models could clarify the roles of distinct *KDM* family member mutations in leukemogenesis in *ETV6-RUNX1* positive mice.

8 References

1. Howlader, N., et al., *SEER Cancer Statistics Review, 1975-2013* , 2015, National Cancer Institute: Bethesda, MD.
2. Pizzo, P.A. and D.G. Poplack, *Biological Basis of Childhood Cancer*, in *Principles and Practice of Pediatric Oncology*. 2015, Wolters Kluwer: Philadelphia, PA. p. 1-100.
3. Pizzo, P.A. and D.G. Poplack, *Management of Common Cancers of Childhood*, in *Principles and Practice of Pediatric Oncology* . 2015, Wolters Kluwer: Philadelphia, PA. p. 463-965.
4. Murphy, K., et al., *Janeway's Immunobiology* . 9th ed. 2016, New York, NY: Garland Science.
5. Brown, P. and S.P. Hunger, *Acute leukemia in children* , in *Conn's current therapy*. 2013, Saunders: Philadelphia, PA. p. 765-768.
6. Pui, C.H., M.V. Relling, and J.R. Downing, *Acute lymphoblastic leukemia*. *N Engl J Med*, 2004. 350(15): p. 1535-48.
7. Lim, J.Y., et al., *Genomics of racial and ethnic disparities in childhood acute lymphoblastic leukemia*. *Cancer*, 2014. 120(7): p. 955-62.
8. Parham, P., *The Immune System* . 2nd ed. 2005, New York, NY: Garland Science.
9. Ford, A.M., et al., *Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia*. *Proc Natl Acad Sci U S A*, 1998. 95(8): p. 4584-8.
10. Gale, K.B., et al., *Backtracking leukemia to birth: identification of clonal gene fusion sequences in neonatal blood spots*. *Proc Natl Acad Sci U S A*, 1997. 94(25): p. 13950-4.
11. Ford, A.M., et al., *In utero rearrangements in the trithorax-related oncogene in infant leukaemias*. *Nature*, 1993. 363(6427): p. 358-60.
12. Shah, S., et al., *A recurrent germ line PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia*. *Nat Genet*, 2013. 45(10): p. 1226-31.
13. 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.
14. Infante-Rivard, C. and M. Guiguet, *Family history of hematopoietic and other cancers in children with acute lymphoblastic leukemia*. *Cancer Detect Prev*, 2004. 28(2): p. 83-7.
15. Kutler, D.I., et al., *A 20-year perspective on the International Fanconi Anemia Registry (IFAR)*. *Blood*, 2003. 101(4): p. 1249-56.
16. Krivit, W. and R.A. Good, *The simultaneous occurrence of leukemia and mongolism; report of four cases*. *AMA J Dis Child*, 1956. 91(3): p. 218-22.
17. Whitlock, J.A., et al., *Clinical characteristics and outcome of children with Down syndrome and acute lymphoblastic leukemia: a Children's Cancer Group study*. *Blood*, 2005. 106(13): p. 4043-9.
18. Bassal, M., et al., *Lymphoblast biology and outcome among children with Down syndrome and ALL treated on CCG-1952*. *Pediatr Blood Cancer*, 2005. 44(1): p. 21-8.
19. Pui, C.H., et al., *Biology, risk stratification, and therapy of pediatric acute leukemias: an update*. *J Clin Oncol*, 2011. 29(5): p. 551-65.
20. 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.
21. 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.
22. 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.

23. Meyer, C., et al., *The MLL recombinome of acute leukemias in 2013*. Leukemia, 2013. 27(11): p. 2165-76.
24. Reiter, A., et al., *Intensive ALL-type therapy without local radiotherapy provides a 90% event-free survival for children with T-cell lymphoblastic lymphoma: a BFM group report*. Blood, 2000. 95(2): p. 416-21.
25. Hastings, C., et al., *Increased post-induction intensification improves outcome in children and adolescents with a markedly elevated white blood cell count ($>10 \times 10^9/l$) with T cell acute lymphoblastic leukaemia but not B cell disease: a report from the Children's Oncology Group*. Br J Haematol, 2015. 168(4): p. 533-46.
26. 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.
27. Taki, T., et al., *Frequency and clinical significance of the MLL gene rearrangements in infant acute leukemia*. Leukemia, 1996. 10(8): p. 1303-7.
28. Pui, C.H., L.L. Robison, and A.T. Look, *Acute lymphoblastic leukaemia*. Lancet, 2008. 371(9617): p. 1030-43.
29. Chiaretti, S., G. Zini, and R. Bassan, *Diagnosis and subclassification of acute lymphoblastic leukemia*. Mediterr J Hematol Infect Dis, 2014. 6(1): p. e2014073.
30. 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.
31. Coustan-Smith, E., et al., *Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia*. Lancet Oncol, 2009. 10(2): p. 147-56.
32. Schultz, K.R., et al., *Risk- and response-based classification of childhood B-precursor acute lymphoblastic leukemia: a combined analysis of prognostic markers from the Pediatric Oncology Group (POG) and Children's Cancer Group (CCG)*. Blood, 2007. 109(3): p. 926-35.
33. Jeha, S. and C.H. Pui, *Risk-adapted treatment of pediatric acute lymphoblastic leukemia*. Hematol Oncol Clin North Am, 2009. 23(5): p. 973-90, v.
34. Pui, C.H. and D. Campana, *New definition of remission in childhood acute lymphoblastic leukemia*. Leukemia, 2000. 14(5): p. 783-5.
35. Svoboda, G.H., *Erratum: Alkaloids of vinca rosea (Catharanthus roseus)*. IX. Extraction and characterization of leurosidine and leurocristine. Lloydia, 1961. 24(4): p. 173.
36. Teuffel, O., et al., *Dexamethasone versus prednisone for induction therapy in childhood acute lymphoblastic leukemia: a systematic review and meta-analysis*. Leukemia, 2011. 25(8): p. 1232-8.
37. Valeriote, F., T.J. Vietti, and D.J. Fernbach, *Clinical pediatric oncology*. 1977, St. Louis, MO: CV Mosby.
38. Schrappe, M., et al., *Outcomes after induction failure in childhood acute lymphoblastic leukemia*. N Engl J Med, 2012. 366(15): p. 1371-81.
39. Jeha, S., et al., *Increased risk for CNS relapse in pre-B cell leukemia with the t(1;19)/TCF3-PBX1*. Leukemia, 2009. 23(8): p. 1406-9.
40. Pui, C.H. and S.C. Howard, *Current management and challenges of malignant disease in the CNS in paediatric leukaemia*. Lancet Oncol, 2008. 9(3): p. 257-68.
41. Hijjiya, N., et al., *Acute leukemia as a secondary malignancy in children and adolescents: current findings and issues*. Cancer, 2009. 115(1): p. 23-35.
42. Barnard, D.R., et al., *Acute myeloid leukemia and myelodysplastic syndrome in children treated for cancer: comparison with primary presentation*. Blood, 2002. 100(2): p. 427-34.
43. Gu, Y., et al., *The (4;11)(q21;q23) chromosome translocations in acute leukemias involve the VDJ recombinase*. Proc Natl Acad Sci U S A, 1992. 89(21): p. 10464-8.

44. Domer, P.H., et al., *Molecular analysis of 13 cases of MLL/11q23 secondary acute leukemia and identification of topoisomerase II consensus-binding sequences near the chromosomal breakpoint of a secondary leukemia with the t(4;11)*. *Leukemia*, 1995. 9(8): p. 1305-12.
45. Felix, C.A., et al., *Chromosome band 11q23 translocation breakpoints are DNA topoisomerase II cleavage sites*. *Cancer Res*, 1995. 55(19): p. 4287-92.
46. Negrini, M., et al., *Potential topoisomerase II DNA-binding sites at the breakpoints of a t(9;11) chromosome translocation in acute myeloid leukemia*. *Cancer Res*, 1993. 53(19): p. 4489-92.
47. So, C.W., et al., *MLL self fusion mediated by Alu repeat homologous recombination and prognosis of AML-M4/M5 subtypes*. *Cancer Res*, 1997. 57(1): p. 117-22.
48. Boehm, T., et al., *Alternating purine-pyrimidine tracts may promote chromosomal translocations seen in a variety of human lymphoid tumors*. *EMBO J*, 1989. 8(9): p. 2621-31.
49. Takata, M., et al., *Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells*. *EMBO J*, 1998. 17(18): p. 5497-508.
50. Rowley, J.D., *The critical role of chromosome translocations in human leukemias*. *Annu Rev Genet*, 1998. 32: p. 495-519.
51. Wiemels, J.L., et al., *Microclustering of TEL-AML1 translocation breakpoints in childhood acute lymphoblastic leukemia*. *Genes Chromosomes Cancer*, 2000. 29(3): p. 219-28.
52. Rodic, N., et al., *Translocation junctions in TCF3-PBX1 acute lymphoblastic leukemia/lymphoma cluster near transposable elements*. *Mob DNA*, 2013. 4(1): p. 22.
53. Yeung, J., et al., *Characterization of the t(17;19) translocation by gene-specific fluorescent in situ hybridization-based cytogenetics and detection of the E2A-HLF fusion transcript and protein in patients' cells*. *Haematologica*, 2006. 91(3): p. 422-4.
54. Hiebert, S.W., et al., *The t(12;21) translocation converts AML-1 B from an activator to a repressor of transcription*. *Mol Cell Biol*, 1996. 16(4): p. 1349-55.
55. LeBrun, D.P. and M.L. Cleary, *Fusion with E2A alters the transcriptional properties of the homeodomain protein PBX1 in t(1;19) leukemias*. *Oncogene*, 1994. 9(6): p. 1641-7.
56. Lu, Q., D.D. Wright, and M.P. Kamps, *Fusion with E2A converts the Pbx1 homeodomain protein into a constitutive transcriptional activator in human leukemias carrying the t(1;19) translocation*. *Mol Cell Biol*, 1994. 14(6): p. 3938-48.
57. Van Dijk, M.A., P.M. Voorhoeve, and C. Murre, *Pbx1 is converted into a transcriptional activator upon acquiring the N-terminal region of E2A in pre-B-cell acute lymphoblastoid leukemia*. *Proc Natl Acad Sci U S A*, 1993. 90(13): p. 6061-5.
58. Diakos, C., et al., *Direct and indirect targets of the E2A-PBX1 leukemia-specific fusion protein*. *PLoS One*, 2014. 9(2): p. e87602.
59. Nowell, P.C. and H.D. A., *A minute chromosome in human chronic granulocytic leukemia*. *Science*, 1960. 132: p. 1497-1501.
60. Kurzrock, R., et al., *A novel c-abl protein product in Philadelphia-positive acute lymphoblastic leukaemia*. *Nature*, 1987. 325(6105): p. 631-5.
61. Tenen, D.G., et al., *Transcription factors, normal myeloid development, and leukemia*. *Blood*, 1997. 90(2): p. 489-519.
62. Miyoshi, H., et al., *t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1*. *Proc Natl Acad Sci U S A*, 1991. 88(23): p. 10431-4.

63. Golub, T.R., et al., *Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia*. Proc Natl Acad Sci U S A, 1995. 92(11): p. 4917-21.
64. Romana, S.P., et al., *The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion*. Blood, 1995. 85(12): p. 3662-70.
65. Okuda, T., et al., *AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis*. Cell, 1996. 84(2): p. 321-30.
66. Wang, Q., et al., *Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis*. Proc Natl Acad Sci U S A, 1996. 93(8): p. 3444-9.
67. Lopez, R.G., et al., *TEL is a sequence-specific transcriptional repressor*. J Biol Chem, 1999. 274(42): p. 30132-8.
68. Golub, T.R., et al., *Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation*. Cell, 1994. 77(2): p. 307-16.
69. Papadopoulos, P., et al., *The novel activation of ABL by fusion to an ets-related gene, TEL*. Cancer Res, 1995. 55(1): p. 34-8.
70. Lacronique, V., et al., *A TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia*. Science, 1997. 278(5341): p. 1309-12.
71. Peeters, P., et al., *Fusion of TEL, the ETS-variant gene 6 (ETV6), to the receptor-associated kinase JAK2 as a result of t(9;12) in a lymphoid and t(9;15;12) in a myeloid leukemia*. Blood, 1997. 90(7): p. 2535-40.
72. 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*. Blood, 1997. 90(2): p. 571-7.
73. Bhojwani, D., et al., *ETV6-RUNX1-positive childhood acute lymphoblastic leukemia: improved outcome with contemporary therapy*. Leukemia, 2012. 26(2): p. 265-70.
74. Fukushima-Nakase, Y., et al., *Shared and distinct roles mediated through C-terminal subdomains of acute myeloid leukemia/Runt-related transcription factor molecules in murine development*. Blood, 2005. 105(11): p. 4298-307.
75. Hotfilder, M., et al., *Immature CD34+CD19- progenitor/stem cells in TEL/AML1-positive acute lymphoblastic leukemia are genetically and functionally normal*. Blood, 2002. 100(2): p. 640-6.
76. 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.
77. Lausten-Thomsen, U., et al., *Prevalence of t(12;21)[ETV6-RUNX1]-positive cells in healthy neonates*. Blood, 2011. 117(1): p. 186-9.
78. Olsen, M., et al., *RT-PCR screening for ETV6-RUNX1-positive clones in cord blood from newborns in the Danish National Birth Cohort*. J Pediatr Hematol Oncol, 2012. 34(4): p. 301-3.
79. Barbany, G., et al., *The ETV6/RUNX1 fusion transcript is not detected in RNA isolated from neonatal dried blood spots from children later diagnosed with the corresponding leukemia*. Leuk Lymphoma, 2013. 54(12): p. 2742-4.
80. Zuna, J., et al., *ETV6/RUNX1 (TEL/AML1) is a frequent prenatal first hit in childhood leukemia*. Blood, 2011. 117(1): p. 368-9; author reply 370-1.
81. Mullighan, C.G., et al., *Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia*. Nature, 2007. 446(7137): p. 758-64.
82. Papaemmanuil, E., et al., *RAG-mediated recombination is the pre dominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia*. Nat Genet, 2014. 46(2): p. 116-25.
83. Romana, S.P., et al., *Deletion of the short arm of chromosome 12 is a secondary event in acute lymphoblastic leukemia with t(12;21)*. Leukemia, 1996. 10(1): p. 167-70.

84. Maia, A.T., et al., *Molecular tracking of leukemogenesis in a triplet pregnancy*. Blood, 2001. 98(2): p. 478-82.
85. Hernandez-Munain, C. and M.S. Krangel, *Regulation of the T-cell receptor delta enhancer by functional cooperation between c-Myb and core-binding factors*. Mol Cell Biol, 1994. 14(1): p. 473-83.
86. Uchida, H., J. Zhang, and S.D. Nimer, *AML1A and AML1B can transactivate the human IL-3 promoter*. J Immunol, 1997. 158(5): p. 2251-8.
87. Okada, H., et al., *AML1(-/-) embryos do not express certain hematopoiesis-related genes including those of the PU.1 gene*. Oncogene, 1998. 17(18): p. 2287-93.
88. Tsuzuki, S., et al., *Modeling first-hit functions of the t(12;21) TEL-AML 1 translocation in mice*. Proc Natl Acad Sci U S A, 2004. 101(22): p. 8443-8.
89. Murre, C., P.S. McCaw, and D. Baltimore, *A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins*. Cell, 1989. 56(5): p. 777-83.
90. Quong, M.W., et al., *A new transcriptional-activation motif restricted to a class of helix-loop-helix proteins is functionally conserved in both yeast and mammalian cells*. Mol Cell Biol, 1993. 13(2): p. 792-800.
91. Aronheim, A., et al., *The E2A gene product contains two separable and functionally distinct transcription activation domains*. Proc Natl Acad Sci U S A, 1993. 90(17): p. 8063-7.
92. Dias, P., M. Dilling, and P. Houghton, *The molecular basis of skeletal muscle differentiation*. Semin Diagn Pathol, 1994. 11(1): p. 3-14.
93. Goldfarb, A.N., J.P. Flores, and K. Lewandowska, *Involvement of the E2A basic helix-loop-helix protein in immunoglobulin heavy chain class switching*. Mol Immunol, 1996. 33(11-12): p. 947-56.
94. Rutherford, M.N. and D.P. LeBrun, *Restricted expression of E2A protein in primary human tissues correlates with proliferation and differentiation*. Am J Pathol, 1998. 153(1): p. 165-73.
95. Kee, B.L. and C. Murre, *Induction of early B cell factor (EBF) and multiple B lineage genes by the basic helix-loop-helix transcription factor E12*. J Exp Med, 1998. 188(4): p. 699-713.
96. Schlissel, M., A. Voronova, and D. Baltimore, *Helix-loop-helix transcription factor E47 activates germ-line immunoglobulin heavy-chain gene transcription and rearrangement in a pre-T-cell line*. Genes Dev, 1991. 5(8): p. 1367-76.
97. Choi, J.K., et al., *E47 activates the Ig-heavy chain and TdT loci in non-B cells*. EMBO J, 1996. 15(18): p. 5014-21.
98. Quong, M.W., et al., *E2A activity is induced during B-cell activation to promote immunoglobulin class switch recombination*. EMBO J, 1999. 18(22): p. 6307-18.
99. Kee, B.L., M.W. Quong, and C. Murre, *E2A proteins: essential regulators at multiple stages of B-cell development*. Immunol Rev, 2000. 175: p. 138-49.
100. Shen, C.P. and T. Kadesch, *B-cell-specific DNA binding by an E47 homodimer*. Mol Cell Biol, 1995. 15(8): p. 4518-24.
101. LeBrun, D.P., *E2A basic helix-loop-helix transcription factors in human leukemia*. Front Biosci, 2003. 8: p. s206-22.
102. Nourse, J., et al., *Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor*. Cell, 1990. 60(4): p. 535-45.
103. Monica, K., et al., *PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1*. Mol Cell Biol, 1991. 11(12): p. 6149-57.
104. Roberts, V.J., M.A. van Dijk, and C. Murre, *Localization of Pbx1 transcripts in developing rat embryos*. Mech Dev, 1995. 51(2-3): p. 193-8.
105. Schnabel, C.A., et al., *Expression of Pbx1b during mammalian organogenesis*. Mech Dev, 2001. 100(1): p. 131-5.

106. DiMartino, J.F., et al., *The Hox cofactor and proto-oncogene Pbx1 is required for maintenance of definitive hematopoiesis in the fetal liver*. Blood, 2001. 98(3): p. 618-26.
107. Saleh, M., et al., *Cell signaling switches HOX-PBX complexes from repressors to activators of transcription mediated by histone deacetylases and histone acetyltransferases*. Mol Cell Biol, 2000. 20(22): p. 8623-33.
108. Massari, M.E., et al., *A conserved motif present in a class of helix-loop-helix proteins activates transcription by direct recruitment of the SAGA complex*. Mol Cell, 1999. 4(1): p. 63-73.
109. Qiu, Y., A. Sharma, and R. Stein, *p300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene*. Mol Cell Biol, 1998. 18(5): p. 2957-64.
110. Duque-Afonso, J., et al., *E2A-PBX1 Remodels Oncogenic Signaling Networks in B-cell Precursor Acute Lymphoid Leukemia*. Cancer Res, 2016. 76(23): p. 6937-6949.
111. Sera, Y., et al., *Identification of cooperative genes for E2A-PBX1 to develop acute lymphoblastic leukemia*. Cancer Sci, 2016. 107(7): p. 890-8.
112. Duque-Afonso, J., et al., *Comparative genomics reveals multistep pathogenesis of E2A-PBX1 acute lymphoblastic leukemia*. J Clin Invest, 2015. 125(9): p. 3667-80.
113. Alonso, C.N., *t(1;19)(q23;p13) TCF3-PBX1*. Atlas Genet Cytogenet Oncol Haematol, 2013. 17(1): p. 45-47.
114. Secker-Walker, L.M., et al., *Prognostic significance of the balanced t(1;19) and unbalanced der(19)t(1;19) translocations in acute lymphoblastic leukemia*. Leukemia, 1992. 6(5): p. 363-9.
115. Uckun, F.M., et al., *Clinical significance of translocation t(1;19) in childhood acute lymphoblastic leukemia in the context of contemporary therapies: a report from the Children's Cancer Group*. J Clin Oncol, 1998. 16(2): p. 527-35.
116. Felice, M.S., et al., *Prognostic impact of t(1;19)/TCF3-PBX1 in childhood acute lymphoblastic leukemia in the context of Berlin-Frankfurt-Munster-based protocols*. Leuk Lymphoma, 2011. 52(7): p. 1215-21.
117. Crist, W.M., et al., *Poor prognosis of children with pre-B acute lymphoblastic leukemia is associated with the t(1;19)(q23;p13): a Pediatric Oncology Group study*. Blood, 1990. 76(1): p. 117-22.
118. Burmeister, T., et al., *Clinical features and prognostic implications of TCF3-PBX1 and ETV6-RUNX1 in adult acute lymphoblastic leukemia*. Haematologica, 2010. 95(2): p. 241-6.
119. Sumner, A.T., H.J. Evans, and R.A. Buckland, *New technique for distinguishing between human chromosomes*. Nat New Biol, 1971. 232(27): p. 31-2.
120. Giemsa, G., *Eine Vereinfachung und Vervollkommnung meiner Methylenblau-Eosin-Färbemethode zur Erzielung der Romanowsky-Nocht'schen Chromatinfärbung*. Centralblatt für Bakteriologie, 1904. 37(2): p. 307-313.
121. Shibata, Y., A. Malhotra, and A. Dutta, *Detection of DNA fusion junctions for BCR-ABL translocations by Anchored ChromPET*. Genome Med, 2010. 2(9): p. 70.
122. Speicher, M.R., S. Gwyn Ballard, and D.C. Ward, *Karyotyping human chromosomes by combinatorial multi-fluor FISH*. Nat Genet, 1996. 12(4): p. 368-75.
123. Langer-Safer, P.R., M. Levine, and D.C. Ward, *Immunological method for mapping genes on Drosophila polytene chromosomes*. Proc Natl Acad Sci U S A, 1982. 79(14): p. 4381-5.
124. Wiemels, J.L. and M. Greaves, *Structure and possible mechanisms of TEL-AML1 gene fusions in childhood acute lymphoblastic leukemia*. Cancer Res, 1999. 59(16): p. 4075-82.
125. van Dongen, J.J., et al., *Standardized RT-PCR analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of*

- minimal residual disease. Report of the BIOMED-1 Concerted Action: investigation of minimal residual disease in acute leukemia.* Leukemia, 1999. 13(12): p. 1901-28.
126. Willis, T.G., et al., *Rapid molecular cloning of rearrangements of the IGHJ locus using long-distance inverse polymerase chain reaction.* Blood, 1997. 90(6): p. 2456-64.
 127. Gardner, M.J. and P.D. Winter, *Mortality in Cumberland during 1959-78 with reference to cancer in young people around Windscale.* Lancet, 1984. 1(8370): p. 216-7.
 128. Forman, D., et al., *Cancer near nuclear installations.* Nature, 1987. 329(6139): p. 499-505.
 129. Cook-Mozaffari, P.J., et al., *Geographical variation in mortality from leukaemia and other cancers in England and Wales in relation to proximity to nuclear installations, 1969-78.* Br J Cancer, 1989. 59(3): p. 476-85.
 130. Michaelis, J., et al., *Incidence of childhood malignancies in the vicinity of west German nuclear power plants.* Cancer Causes Control, 1992. 3(3): p. 255-63.
 131. Michaelis, J., et al., *Krebserkrankungen im Kindesalter in der Umgebung westdeutscher kerntechnischer Anlagen.* Dtsch Arztebl, 1992. 89: p. 2538-2544.
 132. Kaatsch, P., et al., *An extended study on childhood malignancies in the vicinity of German nuclear power plants.* Cancer Causes Control, 1998. 9(5): p. 529-33.
 133. Michaelis, J., P. Kaatsch, and U. Kaletsch, *Leukämien im Kindesalter – Epidemiologische Untersuchungen des Deutschen Kinderkrebsregisters.* Dtsch Arztebl, 1999. 96(14): p. 918-24.
 134. Kaatsch, P., et al., *Epidemiologische Studie zu Kinderkrebs in der Umgebung von Kernkraftwerken (KiKK-Studie) – Vorhaben 3602S04334 Bundesamt für Strahlenschutz, 2007.* urn:nbn:de:0221-20100317939.
 135. Reifenhäuser, C., et al., *Radioaktivität und Strahlung*, 2011, Bayerisches Landesamt für Umwelt: Augsburg.
 136. Rosenfeld, C., et al., *An effect human leukaemic cell line: Reh.* Eur J Cancer, 1977. 13(4-5): p. 377-9.
 137. Rosenfeld, C., et al., *Phenotypic characterisation of a unique non-T, non-B acute lymphoblastic leukaemia cell line.* Nature, 1977. 267(5614): p. 841-3.
 138. Matsuo, Y. and H.G. Drexler, *Establishment and characterization of human B cell precursor-leukemia cell lines.* Leuk Res, 1998. 22(7): p. 567-79.
 139. Findley, H.W., Jr., et al., *Two new acute lymphoblastic leukemia cell lines with early B-cell phenotypes.* Blood, 1982. 60(6): p. 1305-9.
 140. Graham, F.L., et al., *Characteristics of a human cell line transformed by DNA from human adenovirus type 5.* J Gen Virol, 1977. 36(1): p. 59-74.
 141. Fischer, U., et al., *Genomics and drug profiling of fatal TCF3-HLF-positive acute lymphoblastic leukemia identifies recurrent mutation patterns and therapeutic options.* Nat Genet, 2015. 47(9): p. 1020-9.
 142. Miles, C., et al., *Expression of the Ly-6E.1 (Sca-1) transgene in adult hematopoietic stem cells and the developing mouse embryo.* Development, 1997. 124(2): p. 537-47.
 143. Strong, L.C., *Production of the CBA Strain of Inbred Mice: Long Life Associated with Low Tumour Incidence.* British Journal of Experimental Pathology, 1936. 17(1): p. 60-63.
 144. 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.
 145. Füller, E., et al., *Genomic Inverse PCR for Exploration of Ligated Breakpoints (GIPFEL), a New Method to Detect Translocations in Leukemia.* PLoS One, 2014. 9(8): p. e104419.
 146. McLean, T.W., et al., *TEL/AML-1 dimerizes and is associated with a favorable outcome in childhood acute lymphoblastic leukemia.* Blood, 1996. 88(11): p. 4252-8.

147. Raynaud, S., et al., *The 12;21 translocation in involving TEL and deletion of the other TEL allele: two frequently associated alterations found in childhood acute lymphoblastic leukemia*. Blood, 1996. 87(7): p. 2891-9.
148. Mullis, K., et al., *Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction*. Cold Spring Harb Symp Quant Biol, 1986. 51 Pt 1: p. 263-73.
149. Kibbe, W.A., *OligoCalc: an online oligonucleotide properties calculator*. Nucleic Acids Res, 2007. 35(Web Server issue): p. W43-6.
150. Illumina, *An Introduction to Next-Generation Sequencing Technology*, 2016, Illumina, Inc.
151. Illumina, *Illumina Two-Channel SBS Sequencing Technology*, 2016, Illumina, Inc.
152. 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.
153. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform*. Bioinformatics, 2010. 26(5): p. 589-95.
154. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. Bioinformatics, 2009. 25(14): p. 1754-60.
155. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. 25(16): p. 2078-9.
156. 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.
157. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations*. Nat Methods, 2010. 7(4): p. 248-9.
158. Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm*. Nat Protoc, 2009. 4(7): p. 1073-81.
159. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples*. Nat Biotechnol, 2013. 31(3): p. 213-9.
160. Koboldt, D.C., et al., *VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing*. Genome Res, 2012. 22(3): p. 568-76.
161. 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.
162. Weber, M., et al., *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome*. Nat Genet, 2007. 39(4): p. 457-66.
163. Antequera, F. and A. Bird, *CpG Islands*, in *DNA Methylation: Molecular Biology and Biological Significance*, J.P. Jost and H.P. Saluz, Editors. 1993, Birkhäuser Verlag: Basel. p. 169-85.
164. Wong, N.C., et al., *Exploring the utility of human DNA methylation arrays for profiling mouse genomic DNA*. Genomics, 2013. 102(1): p. 38-46.
165. Mellentin, J.D., et al., *Molecular analysis of the t(1;19) breakpoint cluster region in pre-B cell acute lymphoblastic leukemias*. Genes Chromosomes Cancer, 1990. 2(3): p. 239-47.
166. Schäfer, D., et al., *5% of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening*. Blood, 2018.
167. Rodríguez-Hernández, G., et al., *Infection Exposure Promotes ETV6-RUNX1 Precursor B-cell Leukemia via Impaired H3K4 Demethylases*. Cancer Research, 2017. 77(16): p. 4365-4377.
168. Bateman, C.M., et al., *Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia*. Blood, 2010. 115(17): p. 3553-8.

169. Matthews, A.G., et al., *RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination*. *Nature*, 2007. 450(7172): p. 1106-10.
170. Pinkel, D., et al., *Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4*. *Proc Natl Acad Sci U S A*, 1988. 85(23): p. 9138-42.
171. Kamps-Hughes, N., et al., *Massively parallel characterization of restriction endonucleases*. *Nucleic Acids Res*, 2013. 41(11): p. e119.
172. Eguchi-Ishimae, M., et al., *Breakage and fusion of the TEL (ETV6) gene in immature B lymphocytes induced by apoptogenic signals*. *Blood*, 2001. 97(3): p. 737-43.
173. Greaves, M.F., et al., *Leukemia in twins: lessons in natural history*. *Blood*, 2003. 102(7): p. 2321-33.
174. Olsen, M., et al., *Preleukemic TEL-AML1-positive clones at cell level of 10⁻³ to 10⁻⁴ do not persist into adulthood*. *J Pediatr Hematol Oncol*, 2006. 28(11): p. 734-40.
175. Ward, G., *The infective theory of acute leukaemia*. *Br J Child Dis*, 1917. 14: p. 10-20.
176. Kreis, C., et al., *Space-Time Clustering of Childhood Leukemia: Evidence of an Association with ETV6-RUNX1 (TEL-AML1) Fusion*. *PLoS One*, 2017. 12(1): p. e0170020.
177. Busslinger, M., J. Hurst, and R.A. Flavell, *DNA methylation and the regulation of globin gene expression*. *Cell*, 1983. 34(1): p. 197-206.
178. Starkova, J., et al., *The identification of (ETV6)/RUNX1-regulated genes in lymphopoiesis using histone deacetylase inhibitors in ETV6/RUNX1-positive lymphoid leukemic cells*. *Clin Cancer Res*, 2007. 13(6): p. 1726-35.
179. Nan, X., et al., *Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex*. *Nature*, 1998. 393(6683): p. 386-9.
180. Ng, H.H., et al., *MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex*. *Nat Genet*, 1999. 23(1): p. 58-61.
181. Wade, P.A., et al., *Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation*. *Nat Genet*, 1999. 23(1): p. 62-6.
182. Kong, N.R., et al., *MEF2C and EBF1 Co-regulate B Cell-Specific Transcription*. *PLoS Genet*, 2016. 12(2): p. e1005845.
183. Boller, S., et al., *Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming*. *Immunity*, 2016. 44(3): p. 527-41.
184. Iwase, S., et al., *The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases*. *Cell*, 2007. 128(6): p. 1077-88.
185. Frescas, D., et al., *JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes*. *Nature*, 2007. 450(7167): p. 309-13.
186. Liu, Y., et al., *A plant homodomain in RAG-2 that binds Hypermethyated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement*. *Immunity*, 2007. 27(4): p. 561-71.
187. Santos-Rosa, H., et al., *Active genes are tri-methylated at K4 of histone H3*. *Nature*, 2002. 419(6905): p. 407-11.
188. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome*. *Nat Genet*, 2007. 39(3): p. 311-8.
189. Bernstein, B.E., et al., *Genomic maps and comparative analysis of histone modifications in human and mouse*. *Cell*, 2005. 120(2): p. 169-81.
190. Promega, *pGEM®-T and pGEM®-T Easy Vector Systems*, 2015, Promega Corporation.

9 Acknowledgements

First of all, I would like to thank Prof. Dr. Arndt Borkhardt for giving me the opportunity to complete my PhD thesis in the KMT lab and for this interesting project. I would like to thank you for offering me the chance to work on projects with outstanding international collaborators. Furthermore, I would like to thank Prof. Dr. William Martin evaluating my thesis as a second referee.

I want to thank Ute for all the support and her guidance during my PhD thesis. I very much appreciate the fact that we could always discuss the projects and that you always seemed to have an idea when something did not work immediately. Thank you for your help and for proofreading this thesis.

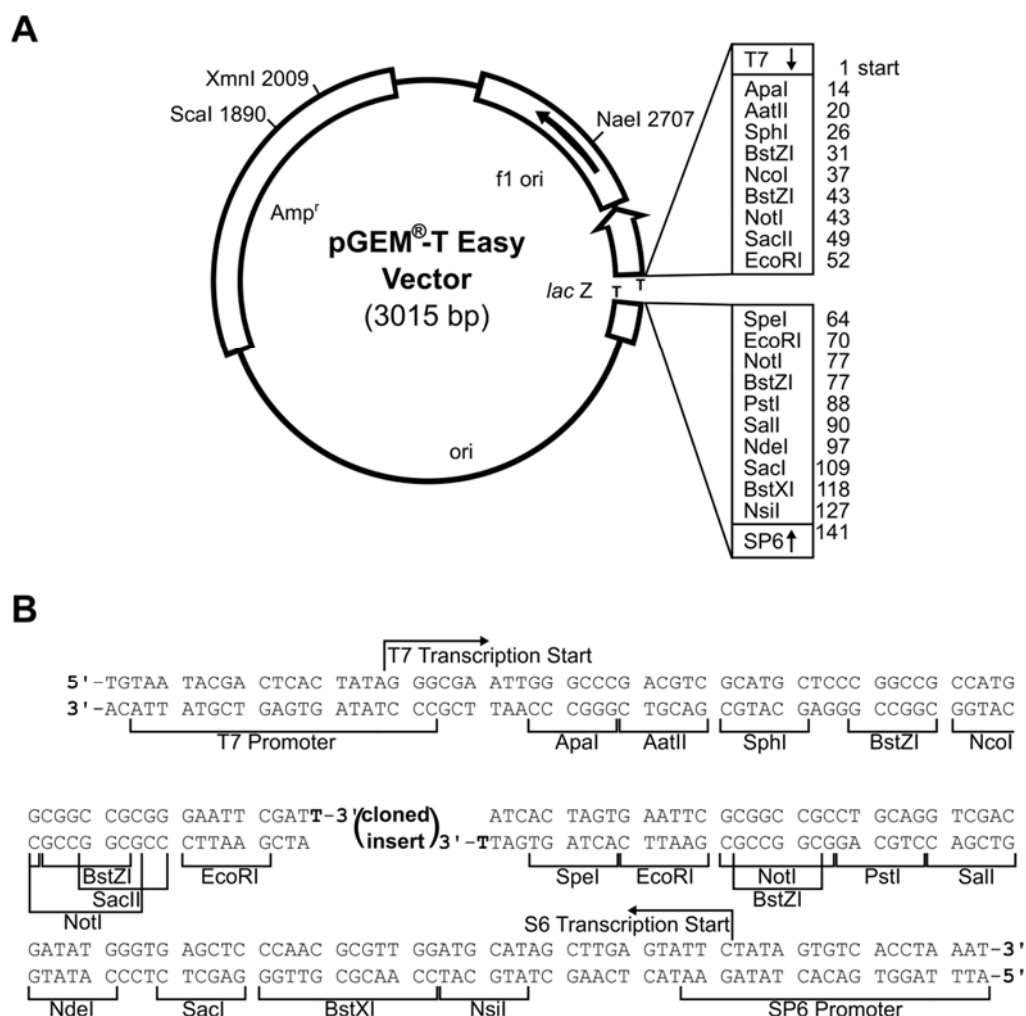
Moreover, I would like to thank the numerous collaborators I had: Prof. Dr Robert Slany and Elisa Füller for the fruitful and pleasant joint work during the establishment of GIPFEL and beyond; Prof. Dr. Julia Hauer as well as Prof. Dr. Isidro Sánchez-García and his group for a great and long lasting collaboration in a very friendly environment; Prof. Dr. Kjeld Schmiegelow and his group who provided the umbilical cord blood samples.

I also want to thank all present and past members of the KMT lab. You made this a very friendly place to work at. Thank you to Bianca and especially to Daniel for all your pipetting help. Without you I would probably still be pipetting. Thanks to Daniel for your help with the DNA methylation analysis. I would like to thank Deborah, Franzi and Julia for generating a nice atmosphere in the office, for fruitful discussions, and for just being friends. Special thanks to Franzi for all your help with experiments and all the discussions about our projects.

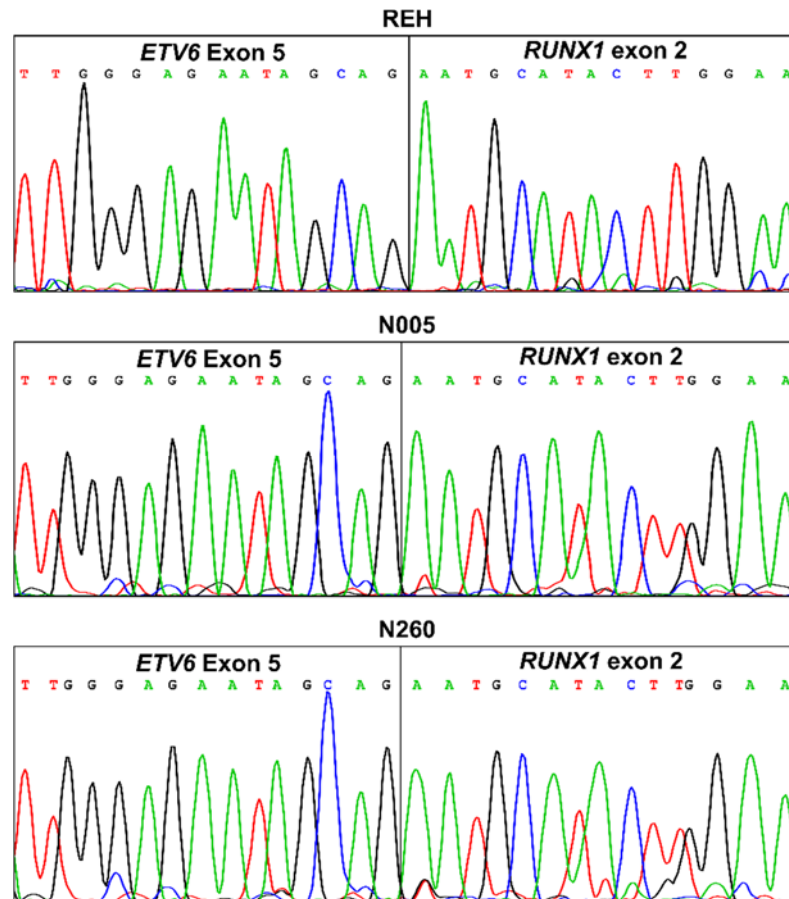
Last but not least, I would like to thank my family: My parents, who have always supported me and without whom all this would have never been possible and Sebastian for correcting my thesis. Finally, thank you Rebecca! Thank you for all your support, your patience, your help, not only with this thesis. Thank you for everything!

10 Appendix

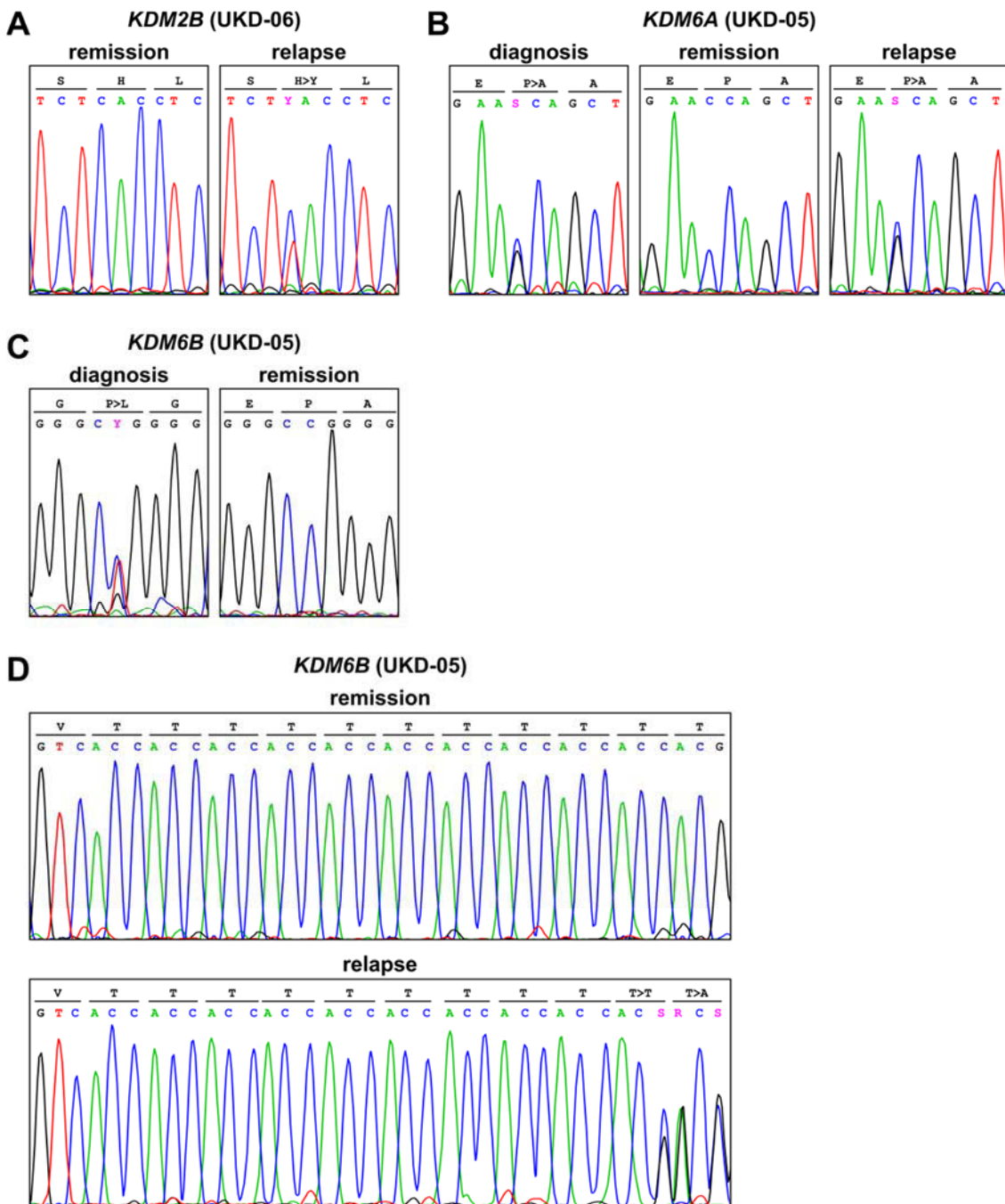
10.1 Supplemental Figures



Supplemental Figure 10.1: The vector pGEM-T Easy. **(A)** Map of the pGEM-T Easy vector (Promega). All enzymes of the multiple cloning site are given and the 3'-T overhangs are indicated. **(B)** Sequence of the multiple cloning site with restriction sites of enzymes and the T7 and SP6 promoter regions. Adapted from [190].



Supplemental Figure 10.2: Sanger sequencing results of the reverse transcriptase Real-Time PCR products. The PCR products from the reverse transcriptase Real-Time PCR were cloned into the pGEM-T Easy vector (Promega) and amplified. The sequencing was carried out with the primers T7 and SP6. The vertical line indicates the fusion of *ETV6* to *RUNX1*. As expected, *ETV6* exon 1 is fused to *RUNX1* exon 2 in all three cases.



Supplemental Figure 10.3: Validation of mutations in *KDM* family member genes. **(A)** *KDM2B* of patient UKD-06 was affected in the relapse. **(B)** *KDM6A* of patient UKD-05 was heterozygously mutated in diagnosis and relapse. **(C-D)** *KDM6B* of UKD-05 had a heterozygous base pair substitution in the diagnosis and a three base pair deletion in the relapse. Triplets with the corresponding amino acid and amino acid changes are given. Ambiguity codes: Y = pyrimidine (C or T), R = purine (A or G), S = strong (G or C).

10.2 Abbreviations

<i>ABL1</i> (<i>ABL</i>)	Abelson murine leukemia viral oncogene homolog 1 (previously known as <i>ABL</i>)
AD1	activation domain 1
AD2	activation domain 2
<i>AFF1</i> (<i>AF4</i>)	AF4/FMR2 family member 1 (previously known as <i>AF4</i>)
ALL	acute lymphoblastic leukemia
AML	acute myeloid leukemia
B-ALL	B cell acute lymphoblastic leukemia
BCP-ALL	B cell precursor acute lymphoblastic leukemia
<i>BCR</i>	BCR, RhoGEF and GTPase activating protein (previously known as breakpoint cluster region)
BCR	breakpoint cluster region
BFM	Berlin-Frankfurt-Münster
bHLH	basic helix-loop-helix
bp	base pair
BSA	bovine serum albumin
<i>BTG1</i>	B-cell translocation gene 1, anti-proliferative
CBA x C57BL/6J	cross of mouse strands CBA and C57BL/6J
CD	cluster of differentiation
<i>CDKN2A</i>	cyclin dependent kinase inhibitor 2A
cDNA	complementary DNA
CF	conventional facility
chromPET	chromosomal paired-end tags
CLL	chronic lymphoblastic leukemia
CML	chronic myeloid leukemia
CNS	central nervous system
CoALL	cooperative study group for childhood acute lymphoblastic leukemia
CpG	cytosine-guanine dinucleotide
Cq	quantification cycle
<i>CRLF2</i>	cytokine receptor-like factor 2
ctrl	control
DKFZ	German Cancer Research Center
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DSMZ	German Collection of Microorganisms and Cell Cultures
E12	transcription factor 3 isoform E12
E47	transcription factor 3 isoform E47
<i>EBF1</i>	early B cell factor 1
EGIL	European Group for the Immunological Characterization of Leukemias
<i>ERG</i>	v-ets erythroblastosis virus E26 oncogene like
ETS	E26 transformation-specific domain
<i>ETV6</i> (<i>TEL</i>)	ETS variant 6 (previously known as <i>TEL</i>)
FBS	fetal bovine serum
FISH	fluorescence in situ hybridization
GIPFEL	genomic inverse PCR for exploration of ligated breakpoints

H3K4	histone 3 lysine 4
H3K4me2/3	di-/trimethylated histone 3 lysine 4
HCM	HOX cooperativity domain
HD	homeodomain
HDAC	histone deacetylase
HSC	hematopoietic stem cell
iAMP21	intrachromosomal amplification of chromosome 21
Ig	immunoglobulin
ILC	innate lymphoid cell
indel	insertion and/or deletion
JAK2	Janus kinase 2
kb	kilobase pair
KDM	lysine demethylase
KDM2B	lysine demethylase 2B
KDM5C	lysine demethylase 5C
KDM6A	lysine demethylase 6A
KDM6B	lysine demethylase 6B
KMT2A (<i>MLL</i>)	lysine methyltransferase 2A (previously known as <i>MLL</i>)
LDI-PCR	long distance inverse polymerase chain reaction
<i>Ly6a</i> (<i>Sca1</i>)	lymphocyte antigen 6 complex, locus A (previously known as <i>Sca1</i>)
<i>LYL1</i>	lymphoblastic leukemia associated hematopoiesis regulator 1
MACS	magnetic activated cell sorting
<i>MLLT1</i> (<i>ENL</i>)	myeloid/lymphoid or mixed-lineage leukemia; translocated to, 1 (previously known as <i>ENL</i>)
<i>MLLT3</i> (<i>AF9</i>)	myeloid/lymphoid or mixed-lineage leukemia; translocated to, 3 (previously known as <i>AF9</i>)
mSv	millisievert
<i>MYC</i>	v-myc avian myelocytomatosis viral oncogene homolog
NaAc	sodium acetate
NaCl	sodium chloride
NC	negative control
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NHEJ	non-homologous end joining
NHL	non-Hodgkin lymphoma
NK cell	natural killer cell
ns	not significant
NTC	non-template control
<i>PAX5</i>	paired box 5
<i>Pax5</i> ^{+/-}	paired box 5 heterozygous
PBS	phosphate buffered saline
<i>PBX1</i>	pre-B-cell leukemia homeobox 1
PBX1a	pre-B-cell leukemia homeobox 1 isoform a
PBX1b	pre-B-cell leukemia homeobox 1 isoform b
PC	principal component
PCA	principal component analysis
PCR	polymerase chain reaction
PDGFRB	platelet derived growth factor receptor beta
PEG	polyethylene glycol

PLCy2	phospholipase C gamma 2
PNT	pointed domain
<i>RAG1</i>	recombination activating 1
<i>RAG2</i>	recombination activating 2
<i>RARA</i>	retinoic acid receptor alpha
RFU	relative fluorescent units
Rn	normalized reporter signal
RNA	ribonucleic acid
RPMI 1640	Roswell Park Memorial Institute medium 1640
RT	room temperature
RT-PCR	reverse transcription polymerase chain reaction
<i>RUNX1 (AML1)</i>	runt related transcription factor 1 (previously known as <i>AML1</i>)
<i>RUNX1T1 (ETO)</i>	RUNX1 translocation partner 1 (formerly known as <i>ETO</i>)
Runxl	Runx inhibition domain
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SPF	special pathogen free facility
SPRI	solid phase reversible immobilization
Stat	signal transducers and activators of transcription
TAE	Tris/Acetic Acid/EDTA
<i>TAL1</i>	T-cell acute lymphocytic leukemia 1
T-ALL	T cell acute lymphoblastic leukemia
<i>TBL1XR1</i>	transducin beta like 1 X-linked receptor 1
<i>TCF3 (E2A)</i>	transcription factor 3 (previously known as <i>E2A</i>)
<i>TLX1</i>	T-cell leukemia homeobox 1
<i>TLX3</i>	T-cell leukemia homeobox 3
U	unit
UTP	uridine triphosphate
v/v	volume per volume
w/v	weight per volume
WGA	whole genome amplification
WNT	wingless-related integration site

10.3 Nomenclature

human gene	all uppercase, italic (e.g. <i>RUNX1</i>)
human fusion gene	all uppercase, italic, separated by a dash (e.g. <i>ETV6-RUNX1</i>)
human protein	all uppercase (e.g. RUNX1)
human fusion protein	all uppercase, separated by a dash (e.g. ETV6-RUNX1)
murine gene	first letter uppercase, italic (e.g. <i>Kdm5c</i>)
murine protein	first letter uppercase (e.g. Kdm5c)
Drosophila gene	all lowercase, italic (e.g. <i>runt</i>)

10.4 Figure Directory

Figure 1.1: Cancer related deaths in children aged 0-14.....	11
Figure 1.2: Scheme of hematopoiesis.....	12
Figure 1.3: B cell development.....	13
Figure 1.4: Subtypes of ALL in childhood and adulthood.....	15
Figure 1.5: Event-free survival by ALL sybtype.....	18
Figure 1.6: Schematic presentation of treatment results.....	19
Figure 1.7: ETV6-RUNX1 fusion.....	23
Figure 1.8: TCF3-PBX1 fusion.....	25
Figure 1.9: Overview of methods used for detection of translocations.	28
Figure 3.1: Insertion of the human <i>ETV6-RUNX1</i> cDNA into the <i>Ly6a</i> locus.	36
Figure 3.2: Keeping of the mice.	36
Figure 4.1: Principle of GIPFEL.	43
Figure 4.2: Initial workflow of GIPFEL.....	44
Figure 4.3: Workflow of Next Generation Sequencing.....	59
Figure 4.4: Methylation of cytosine.	62
Figure 5.1: SacI restriction sites in the BCRs of the <i>ETV6</i> and <i>RUNX1</i> genes.....	64
Figure 5.2: MfeI restriction sites in the BCRs of the <i>TCF3</i> and <i>PBX1</i> genes.....	65
Figure 5.3: Proof of principle of GIPFEL for <i>TCF3-PBX1</i>	67
Figure 5.4: Proof of principle of GIPFEL for <i>ETV6-RUNX1</i>	68
Figure 5.5: Results of the Real-Time PCR for the optimization of restriction enzyme digests.....	70
Figure 5.6: Results of the Real-Time PCR for the optimization of the DNA clean-up after restriction enzyme digest.	71
Figure 5.7: Results of the Real-Time PCR for the optimization of the ligation temperature.....	72
Figure 5.8: Real-Time PCR results for the optimization of ligation duration.	73
Figure 5.9: Real-Time PCR results after the introduction of a second clean-up step, following the exonuclease digest.	74
Figure 5.10: Impact of whole genome amplification on GIPFEL.	76
Figure 5.11: Optimized GIPFEL protocol.	78
Figure 5.12: Sensitivity of GIPFEL for <i>ETV6-RUNX1</i>	79
Figure 5.13: Sensitivity of GIPFEL for <i>TCF3-PBX1</i>	80
Figure 5.14: Example for the validation of an <i>ETV6-RUNX1</i> positive patient (P43).	83
Figure 5.15: Example for the validation of a <i>TCF3-PBX1</i> positive patient (P26).	85
Figure 5.16: GIPFEL results including CD19 ⁺ enrichment.....	87
Figure 5.17: GIPFEL results including WGA after CD19 ⁺ enrichment.	88

Figure 5.18: Tested compositions of 384 well Real-Time PCR.....	90
Figure 5.19: Example for the first step of the validation of an <i>ETV6-RUNX1</i> positive cord blood (N926).	91
Figure 5.20: Example for the further steps of the validation of an <i>ETV6-RUNX1</i> positive cord blood (N926).	92
Figure 5.21: Example for a cord blood sample with two <i>ETV6-RUNX1</i> fusions (N505).	94
Figure 5.22: Results from reverse transcription Real-Time PCR regarding <i>ETV6-RUNX1</i> transcription in cord blood samples and cell line controls.....	95
Figure 5.23: Principal component analysis of DNA methylation.	97
Figure 5.24: Heat map of unsupervised clustering of differentially methylated genes. .	98
Figure 5.25: Results from mouse whole exome sequencing.	99
Figure 5.26: Mutations of murine <i>Ebf1</i> and human <i>EBF1</i>	99
Figure 5.27: Mutations of murine <i>Kdm5c</i> and human <i>KDM5C</i>	100

10.5 Table Directory

Table 1.1: Immunophenotype of B- and T-ALL	17
Table 3.1: Media / cell culture reagents	32
Table 3.2: Properties and culturing conditions for cell line REH.	32
Table 3.3: Properties and culturing conditions for cell line 697.....	33
Table 3.4: Properties and culturing conditions for cell line HEK-293.	33
Table 3.5: Breakpoints of <i>ETV6-RUNX1</i> positive patients.....	34
Table 3.6: Breakpoints of <i>TCF3-PBX1</i> patients.	34
Table 3.7: Information on mice that were whole exome sequenced.....	37
Table 3.8: Information on mice that were subjected to DNA methylation analysis.	37
Table 3.9: Restriction enzymes.....	38
Table 3.10: Further enzymes and enzyme containing kits.....	38
Table 3.11: General chemicals	38
Table 3.12: Composition of 6x loading dyes.	39
Table 3.13: Kits, size markers, and other materials	39
Table 3.14: Primers for validation of SNVs.	39
Table 3.15: GIPFEL primers for <i>ETV6-RUNX1</i>	40
Table 3.16: GIPFEL primers for <i>TCF3-PBX1</i>	41
Table 3.17: Primers used for sequencing of cloned PCR products.	41
Table 3.18: Software	42
Table 3.19: Hardware.....	42
Table 4.1: Composition of DNA restriction enzyme reactions during the establishment phase of GIPFEL.....	48
Table 4.2: Composition of DNA restriction enzyme reactions during the umbilical cord blood screening for <i>ETV6-RUNX1</i>	48
Table 4.3: Composition of the ligation reactions.	49
Table 4.4: Composition of the pre-amplification PCRs.	52
Table 4.5: Conditions of the pre-amplification PCRs.	52
Table 4.6: Composition of the Real-Time PCRs.	53
Table 4.7: Conditions of the Real-Time PCRs.	54
Table 4.8: Composition of the reverse transcription Real-Time PCRs.....	56
Table 4.9: Conditions of the reverse transcription Real-Time PCRs.....	56
Table 4.10: Composition of buffer RF1.	57
Table 4.11: Composition of buffer RF2.	57
Table 5.1: Multiplexing of <i>ETV6-RUNX1</i> primers.....	77
Table 5.2: Multiplexing of <i>TCF3-PBX1</i> primers.....	77
Table 5.3: <i>ETV6-RUNX1</i> positive samples that were validated by GIPFEL.	82

Table 5.4: <i>TCF3-PBX1</i> positive samples that were validated by GIPFEL.	84
Table 5.5: Tested compositions of Real-Time PCRs on a 384 well cyclor.....	89
Table 5.6: <i>ETV6-RUNX1</i> positive newborns from umbilical cord blood screening.	93
Table 5.7: SNVs and small indels in histone related genes.	101
Table 5.8: Structural aberrations in histone related genes.....	102

10.6 Supplemental Figure Directory

Supplemental Figure 10.1: The vector pGEM-T Easy.	127
Supplemental Figure 10.2: Sanger sequencing results of the reverse transcriptase Real-Time PCR products.	128
Supplemental Figure 10.3: Validation of mutations in <i>KDM</i> family member genes. ...	129

Curriculum Vitae

Personal Data

Name: Daniel Schäfer
 Address: Stresemannstraße 36, 40210 Düsseldorf
 Date and Place of Birth: September 10th, 1985 in Hachenburg
 Nationality: German

Education

10/2012 – Present	<p>PhD Thesis in the Department of Pediatric Oncology, Hematology and Clinical Immunology, Heinrich-Heine University Düsseldorf</p> <p>Title: “Establishment of a Novel Technique Termed GIPFEL to Determine the Frequency of <i>ETV6-RUNX1</i> Fusions in Healthy Newborns and Analysis of Cooperating Oncogenic Lesions Leading to <i>ETV6-RUNX1</i> Positive Childhood Leukemia”</p> <p>(Supervisor: Prof. Dr. Arndt Borkhardt)</p>
08/2011 – 05/2012	<p>Diploma Thesis at the Institute for Molecular Genetics, Johannes Gutenberg University Mainz</p> <p>Title: “<i>STK33</i> – Evolutionäre Entwicklung in Vertebraten“</p> <p>(Supervisor: Prof. Dr. Erwin R. Schmidt)</p>
10/2006 – 05/2012	Johannes Gutenberg University Mainz: Diploma in Biology
08/2005 – 04/2006	Civilian Service
08/1996 – 03/2005	Abitur at the Konrad Adenauer Gymnasium, Westerburg

Publications

1. Füller, E.*, Schäfer, D.*, Fischer, U., Krell, P.F.I., Stanulla, M., Borkhardt, A., Slany, R.K., *Genomic Inverse PCR for Exploration of Ligated Breakpoints (GIPFEL), a New Method to Detect Translocations in Leukemia*. PLoS One, 2014. **9**(8): p. e104419.
2. Lautwein, T., Lerch, S., Schäfer, D., Schmidt, E.R., *The serine/threonine kinase 33 is present and expressed in palaeognath birds but has become a unitary pseudogene in neognaths about 100 million years ago*. BMC Genomics, 2015. **16**: p. 543.
3. Fischer, U., Forster, M., Rinaldi, A., Risch, T., Sungalee, S., Warnatz, H.-J., Bornhauser, B., Gombert, M., Kratsch, C., Stütz, A.M., Sultan, M., Tchinda, J., Worth, C.L., Amstislavskiy, V., Badarinarayan, N., Baruchel, A., Bartram, T., Basso, G., Canpolat, C., Cario, G., Cavé, H., Dakaj, D., Delorenzi, M., Dobay, M.P., Eckert, C., Ellinghaus, E., Eugster, S., Frismantas, V., Ginzl, S., Haas, O.A., Heidenreich, O., Hemmrich-Stanisak, G., Hezaveh, K., Höll, J.I., Hornhardt, S., Husemann, P., Kachroo, P., Kratz, C.P., Le Kronnie, G., Marovca, B., Niggli, F., McHardy, A.C., Moorman, A.V., Panzer-Grümayer, R., Petersen, B.S., Raeder, B., Ralser, M., Rosenstiel, P., Schäfer, D., Schrappe, M., Schreiber, S., Schütte, M., Stade, B., Thiele, R., von der Weid, N., Vora, A., Zaliouva, M., Zhang, L., Zichner, T., Zimmermann, M., Lehrach, H., Borkhardt, A., Bourquin, J.-P., Franke, A., Korbel, J.O., Stanulla, M., Yaspo, M.-L., *Genomics and drug profiling of fatal TCF3-HLF-positive acute lymphoblastic leukemia identifies recurrent mutation patterns and therapeutic options*. Nat Genet, 2015. **47**(9): p. 1020-9.
4. Martín-Lorenzo, A., Hauer, J., Vicente-Duenas, C., Auer, F., Gonzalez-Herrero, I., Garcia-Ramirez, I., Ginzl, S., Thiele, R., Constantinescu, S.N., Bartenhagen, C., Dugas, M., Gombert, M., Schäfer, D., Blanco, O., Mayado, A., Orfao, A., Alonso-Lopez, D., De Las Rivas, J., Cobaleda, C., Garcia-Cenador, M.B., Garcia-Criado, F.J., Sanchez-Garcia, I., Borkhardt, A., *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.
5. Schipp, C., Nabhani, S., Bienemann, K., Simanovsky, N., Kfir-Erenfeld, S., Assayag-Asherie, N., Oommen, P.T., Revel-Vilk, S., Hönscheid, A., Gombert, M., Ginzl, S., Schäfer, D., Laws, H.J., Yefenof, E., Fleckenstein, B., Borkhardt, A., Stepensky, P., Fischer, U., *Specific antibody deficiency and autoinflammatory disease extend the clinical and immunological spectrum of heterozygous NFKB1 loss-of-function mutations in humans*. Haematologica, 2016. **101**(10): p. e392-e396.
6. Rodríguez-Hernández, G., Hauer, J., Martín-Lorenzo, A., Schäfer, D., Bartenhagen, C., García-Ramírez, I., Auer, F., González-Herrero, I., Ruiz-Roca, L., Gombert, M., Okpanyi, V., Fischer, U., Chen, C., Dugas, M., Bhatia, S., Linka, R.M., Garcia-Suquia, M., Rascón-Trincado, M.V., Garcia-Sanchez, A., Blanco, O., García-Cenador, M.B., García-Criado, F.J., Cobaleda, C., Alonso-López, D., De Las Rivas, J., Müschen, M., Vicente-Dueñas, C., Sánchez-García, I., Borkhardt, A., *Infection exposure promotes ETV6-RUNX1 precursor B cell leukemia via impaired H3K4 demethylases*. Cancer Res, 2017. **77**(16): p. 4365-4377.
7. Schäfer, D., Olsen, M., Lähnemann, D., Stanulla, M., Slany, R.K., Schmiegelow, K., Borkhardt, A., Fischer, U., *5% of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening*. Blood, 2018. Online first, doi: <https://doi.org/10.1182/blood-2017-09-808402>.

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, June 7th, 2017

Daniel Schäfer