Supplement

Material and Methods

Patients, healthy controls and cell lines

Probands were enrolled in this study after obtaining written informed consent. The samples were anonymized and all experiments were approved by the Danish Data Protection Agency, the Danish Scientific Ethics Committee, and the Ethical Review Board of the University Hospital, Heinrich-Heine-University, Düsseldorf.

All cell lines were acquired from the German Collection of Microorganisms and Cell Cultures (DSMZ, Brunswick, Germany) and cultured as recommended. The B cell precursor leukemia cell line REH harbors a t(12;21)(p13;q22.3) translocation coding for *ETV6-RUNX1*. Other cell lines used were the *ETV6-RUNX1*-negative B cell precursor leukemia cell lines 697 and HAL-01, the human embryonic kidney cell line HEK-293, the acute myeloid leukemia cell line HL-60, the acute monocytic leukemia cell lines THP-1 and MV4-11, the rhabdomyosarcoma cell line RH-30, the erythroleukemia cell line HEL, and the T cell leukemia cell line JURKAT.

Processing of umbilical cord blood samples

Umbilical cord blood samples from 1,000 healthy neonates were collected in blood collection tubes with EDTA as anticoagulation additive. All samples were processed within 24 h (median 12 h) from birth. Mononuclear cells (MNCs) were separated by Ficoll density centrifugation, washed with RPMI 1640 medium and resuspended in 2 ml RPMI 1640 medium containing 10% BSA. MNCs were transferred to cryovials and a cryopreservative solution was added 1+1. The vials were frozen at -80°C in a 5100 Cryo 1C Freezing Container and subsequently transferred to liquid nitrogen.

Enrichment of CD19-positive cells

CD19-positive B cells were enriched from at least 3 x 10^6 cells using magnetic Dynabeads CD19 pan B (Invitrogen, Carlsbad, CA), as recommended by the manufacturer. Briefly, MNCs derived from cord blood were thawed, diluted 1:1 with MACS buffer (Miltenyi, Bergisch Gladbach, Germany) and centrifuged at 600 x g and 4°C for 10 min. Cell pellets were resuspended in 1 ml MACS buffer. 50 µl magnetic beads coated with CD19 antibodies sufficient for binding of 2.5 x 10^7 $CD19^+$ cells were washed once with MACS buffer and added to each sample. Cells and beads were incubated at 4°C with gentle tilting for 20 min. Subsequently, cells were washed with MACS buffer and separated using a magnet twice. Cells were resuspended in 200 µl PBS. Purity of the $CD19^+$ cell fraction was tested by flow cytometry to be above 95%.

DNA isolation

Genomic DNA was isolated form $CD19^+$ cells and cell lines employing the QIAamp Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNA derived from $CD19^+$ cells was finally eluted in 80 µl AE buffer.

GIPFEL screening

The GIPFEL procedure was carried out basically as described.[1] Briefly, genomic DNA was digested enzymatically with *SacI* (NEB, Ipswich, MA) to achieve defined fragmentation. 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 number of B cells was estimated to range between 2.5 x 10^5 and 3 x 10^6.

After the restriction enzyme digest, the DNA was purified using the QIAquick Gel Extraction Kit (QIAGEN) with a final DNA incubation and elution in 50 µl preheated water at 60°C for 5 min. The purified DNA was ligated with T4 DNA ligase (NEB) at 24°C for 2 h to achieve circularization. To remove residual linear DNA 100 U of Exonuclease III (NEB) were added and incubated at 37°C for 30 min. Then exonuclease and ligase were heat inactivated at 95°C for 5 min. To purify and concentrate the DNA, an ethanol precipitation was carried out and the dried DNA pellet was resuspended in 25 µl nuclease free water. PCR was used to detect the ligation joints produced by circularization. To this end, the DNA was pre-amplified by 25 cycles of PCR with multiplexed primers using the Brilliant II SYBR Green Master Mix (Agilent, Santa Clara, CA) and a T-Gradient Thermoblock PCR cycler (Biometra, Göttingen, Germany). Three reverse primers covered the breakpoint cluster region (BCR) of *ETV6*, while 28 forward primers covered the BCR of *RUNX1*. Forward primers were pooled in four groups of seven primers each. Reverse primers were also pooled. Two additional primers amplified a product outside of the *RUNX1* BCR, which served as a positive control.

The PCR products were used as templates for a Real-Time PCR performed on a CFX Real-Time PCR Detection System (BioRad, Hercules, CA) using the same master mix with the same forward primer bundles, but nested primers were substituted for reverse primers (Figure 1B).

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. The products were then analyzed on an agarose gel and Sanger sequenced if a specific DNA band was still present (Figure 1C).

To screen the breakpoint cluster regions of *ETV6* and *RUNX1* a high number of primers is needed and the available patient material is usually restricted. Therefore primer multiplexing is a useful first step and allows screening of a high number of samples using only small amounts in a short time. However, multiplexing of PCR primers is usually associated with a risk of primer dimer formation and generation of false PCR products as a cause of mispriming and pairing of unrelated primers. This is not a frequent process and can usually easily be distinguished from real positive signals by the higher number of amplification cycles that are necessary to pass the threshold in real-time PCR (as can be seen in supplemental Figure S1 and S5). To identify the breakpoint region and to ensure that false positive signals are excluded, GIPFEL has implemented two demultiplexing steps and controls for the expected PCR product size and sequence as quality control steps.

Validation of GIPFEL results by RT Real-Time PCR

RNA was isolated from 2.9-7.2 x 10^6 cryopreserved mononuclear cord blood cells using the RNeasy Mini Kit (Qiagen) according to manufacturer's instructions. Subsequently, 1 µg of total RNA was reverse transcribed to cDNA in a final volume of 20 µl using the SuperScript III reverse transcriptase (Invitrogen) and oligo(dT) primers.

Reverse transcriptase (RT) Real-Time PCR was carried out using primers, probes, and standards contained in the Ipsogen *ETV6-RUNX1* Kit (Qiagen). For each sample, amplifications of the *ETV6-RUNX1* fusion transcript and the *ABL1* transcript were done in duplicates using 5 µl of cDNA as a template. The *ETV6-RUNX1* expression was then normalized to 10^4 *ABL1* transcripts (Figure 1D).

Validation of GIPFEL results by identification of patient-specific genomic ETV6-RUNX1 breakpoints

SacI-digested and circularized DNA was used as a template for targeted amplification carried out basically as described by Dean *et al.*[2] Briefly, primers flanking the identified ligation joint but facing towards the breakpoint were used in combination with *Phi29* polymerase, which enables the efficient amplification of longer DNA fragments of unknown length in an isothermal reaction (i.e. with continuous polymerase initiation and without the necessity of defined cycle lengths). Sulfur-modified nucleotides were added to the 3' ends of the primers to prevent 3'-5' degradation by nuclease activity of the *Phi29* polymerase. Template DNA and primers (0.5 µl each of 10 µM forward and reverse primers, supplemental Table S4) were mixed, the tube was placed in boiling water, and then slowly cooled to room temperature. 0.5 µl of 10 mM each dNTPs, 2 µl 10x reaction buffer, 3 U *Phi29* DNA polymerase, 0.02 U pyrophosphatase (all from Thermo Fisher Scientific, Waltham, MA) and H_2O_{dd} were added (20 µl final volume) and the amplification was carried out at 30°C for 24 h. A calf intestine alkaline phosphatase (NEB) digest at 37°C for 30 min followed to remove remaining dNTPs. The amplified DNA was ethanol precipitated and resuspended in nuclease-free water. Subsequently, the samples were Sanger sequenced (Figure 1E). In brief, DNA was mixed with 4 µl Big Dye Terminator (Applied Biosystems, Waltham, MA), 0.5 µl 10 µM sequencing primer (supplemental Table S4), and water (20 µl final volume). The samples were then PCR amplified with 26 cycles of 96°C for 30 s, 55°C for 30 s, and 60°C for 4 min. Subsequently, the sequencing was carried out on an ABI 3130 sequencer according to the recommendations of the manufacturer (Applied Biosystems).

Supplemental Tables

Supplemental Table S1: Overview of studies that investigated the *ETV6-RUNX1* frequency in healthy probands.

Study	Schäfer et al.	Eguchi-Ishimae et al.[3]	Eguchi-Ishimae et al.[3]	Mori et al.[4]	Olsen et al.[5]	Lausten-Thomsen et. al.[6]	Lausten-Thomsen et al.[7]	Lausten-Thomsen et al.[8]	Zuna et al.[9]	Olsen et al.[10]	Skorvaga et al.[11]	Ornelles et al.[12]	Kosik et al.[13]
Journal		Blood	Blood	PNAS	J Pediatr Hematol Oncol.	Eur J Haematol	Blood Cells Mol Dis	Blood	Blood	J Pediatr Hematol Oncol.	PLoS One	PLoS One	Oncotarget
Year	present study	2001	2001	2002	2006	2008	2010	2011	2011	2012	2014	2015	2017
Probands [n]	1000	67	147	567	2005	27	256	1417	253	1258	200	210	300
Positive [n]	50	1	13	6	10	0	0	0	5	3		5	
Carriers [%]	5%	1,49%	8,84%	1.06%	0,5%	0%	0%	0%	1,98%	0,24%	4%	2,4%	2.4%
Preleukemic Cells	10^{-3} to 10^{-6}			10^{-3} to 10^{-4}	10^{-5} to 10^{-6}			$<10^{-5}$		$<10^{-4}$	$\leq 10^{-5}$		$\leq 10^{-5}$
Population	Danish	Japanese	Japanese	British	Danish	Danish	Danish	Danish	Czech	Danish	Slovak	US-American	Slovak
Specimen	Cord Blood Frozen	Cord blood	Peripheral blood	Cord Blood Frozen	Peripheral Blood	Embryonic liver Fresh	Cord Blood Fresh	Cord Blood Fresh	Cord Blood	Cord Blood Fresh	Cord Blood Frozen	Cord Blood Fresh	Cord blood
Age Group	Newborns	Newborns	Children and Adults	Newborns	Adults	6th to 10th Week	Newborns	Newborns	Newborns	Newborns	Newborns	Newborns	Newborns
Screen	DNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
Detection	GIPFEL	nRT-PCR	nRT-PCR	RT-PCR	RT-qPCR	RT-PCR	RT-qPCR	RT-qPCR	RT-PCR	RT-qPCR	RT-qPCR	nRT-PCR	RT-qPCR
Sensitivity	10^{-4}			10^{-5}	10^{-4} to 10^{-5}	10^{-4} to 10^{-5}	10^{-4} to 10^{-5}	10^{-4} to 10^{-5}		10^{-4} to 10^{-5}	10^{-5}		10^{-5}

Supplemental Table S2: Validation of GIPFEL. Screen of 52 umbilical cord blood samples by RT-qPCR.

ID of Newborn	GIPFEL	*ETV6-RUNX1* Frequency (GIPFEL)	RT-qPCR	*ETV6-RUNX1* Frequency (RT-qPCR)
N001	negative	-	negative	-
N005	positive	1×10^{-4}	positive	3×10^{-4}
N009	negative	-	negative	-
N017	negative	-	negative	-
N027	negative	-	negative	-
N031	negative	-	negative	-
N036	negative	-	negative	-
N046	negative	-	negative	-
N055	negative	-	negative	-
N064	negative	-	negative	-
N072	negative	-	negative	-
N079	negative	-	negative	-
N087	negative	-	negative	-
N111	negative	-	negative	-
N119	negative	-	negative	-
N125	negative	-	negative	-
N130	negative	-	negative	-
N135	negative	-	negative	-
N139	negative	-	negative	-
N142	negative	-	negative	-
N145	negative	-	negative	-
N148	negative	-	negative	-
N150	negative	-	negative	-
N165	negative	-	negative	-
N173	negative	-	negative	-
N178	negative	-	negative	-
N180	negative	-	negative	-
N184	negative	-	negative	-
N186	negative	-	negative	-
N193	negative	-	negative	-
N198	negative	-	negative	-
N202	negative	-	negative	-
N205	negative	-	negative	-
N211	negative	-	negative	-
N218	negative	-	negative	-
N229	negative	-	negative	-
N230	negative	-	negative	-
N231	negative	-	negative	-
N233	negative	-	negative	-
N242	negative	-	negative	-
N245	negative	-	negative	-
N255	negative	-	negative	-
N257	negative	-	negative	-

N260	positive	1.5×10^{-4}	positive	1×10^{-4}
N267	negative	-	negative	-
N269	negative	-	negative	-
N271	negative	-	negative	-
N273	negative	-	negative	-
N276	negative	-	negative	-
N280	negative	-	negative	-
N288	negative	-	negative	-
N300	negative	-	negative	-

Table S3: Validation of GIPFEL screening of 10 human cell lines by RT-qPCR.

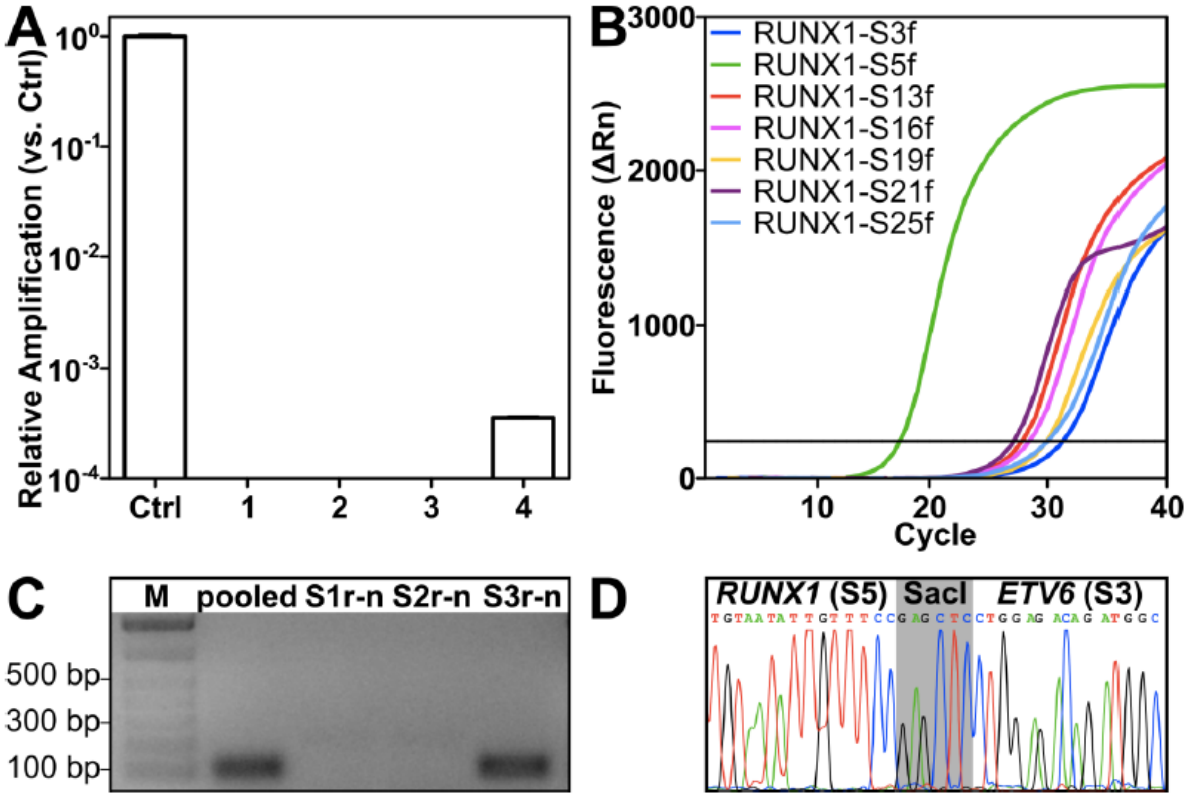
Cell line	Origin	Translocation t(12;21)	GIPFEL	RT-qPCR
REH	B-cell precursor leukemia	positive	positive	positive
697	B-cell precursor leukemia	negative	negative	negative
HAL-01	B-cell precursor leukemia	negative	negative	negative
HL60	Acute myeloid leukemia	negative	negative	negative
MV4-11	Acute monocytic leukemia	negative	negative	negative
THP-1	Acute monocytic leukemia	negative	negative	negative
JURKAT	T-cell leukemia	negative	negative	negative
HEL	Erythroleukemia	negative	negative	negative
RH-30	Rhabdomyosarcoma	negative	negative	negative
HEK293	Embryonal kidney cells	negative	negative	negative

Table S4: Primers for targeted amplification of DNA circles and for Sanger sequencing of chromosomal breakpoints.

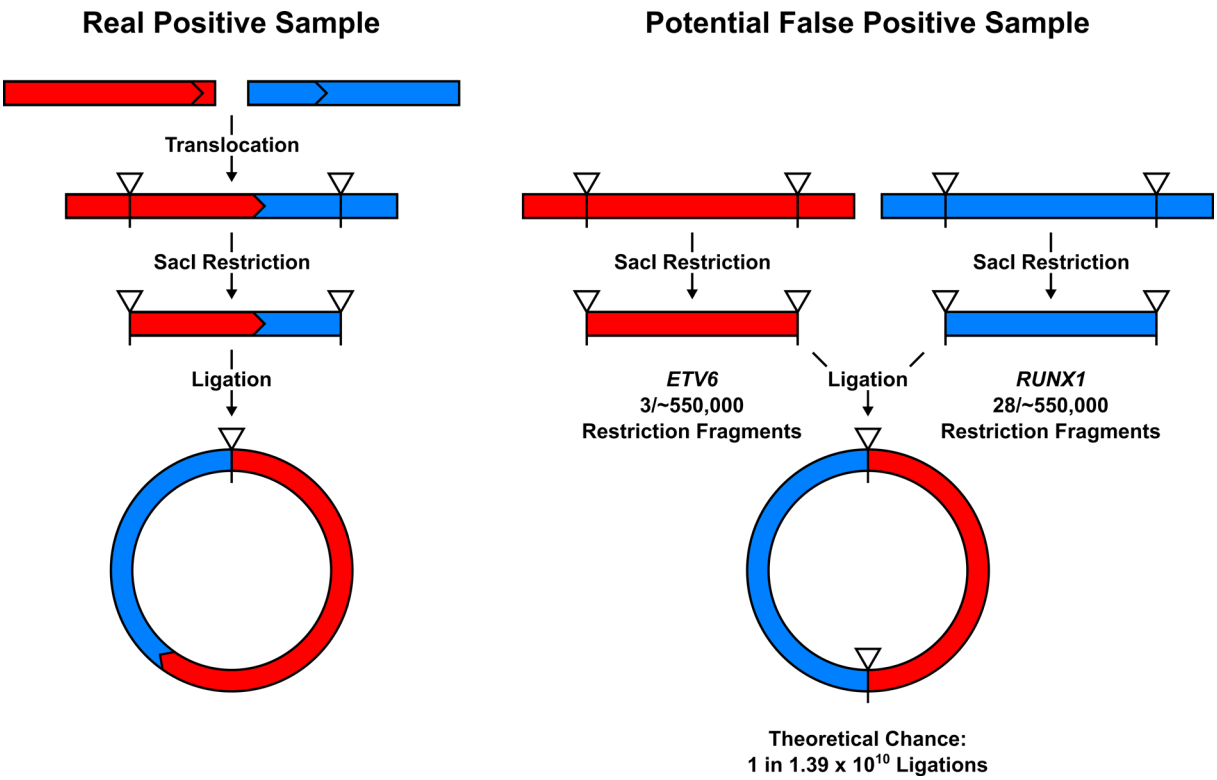
Name	Purpose	Sequence [5'-3']
ETV6-S1f-n	Targeted amplification	GAAGACCTGGCTTACATGAACCACA*T*C
ETV6-S2f-n	Targeted amplification	AGCCTACTTTACTTCTGCCTTGAGGT*A*C
ETV6-S3f-n	Targeted amplification	CTCCCAAGCAGGTGCATAATGTC*C*C
RUNX1-S13r	Targeted amplification	CACTTGCATCCCACTAGGTCCT*C*C
RUNX1-S22r	Targeted amplification	CTACCCCAACTAGCTTCTTTCCTTG*A*C
RUNX1-S23r	Targeted amplification	TGCTGCACTATTGCGACTTCTCTGT*T*C
ETV6-S1-Seq1-f	Sanger sequencing	TGCACCCTCTGATCCTGAAC
ETV6-S1-Seq2-f	Sanger sequencing	CCCTGTCCTGAAGTTTCCAG
ETV6-S1-Seq3-f	Sanger sequencing	ATATTTGTTGAGCACCTGCTGTG
ETV6-S3-Seq1-f	Sanger sequencing	CTCCCAAGCAGGTGCATAATGTCCC

* Indicates sulfur-modification.

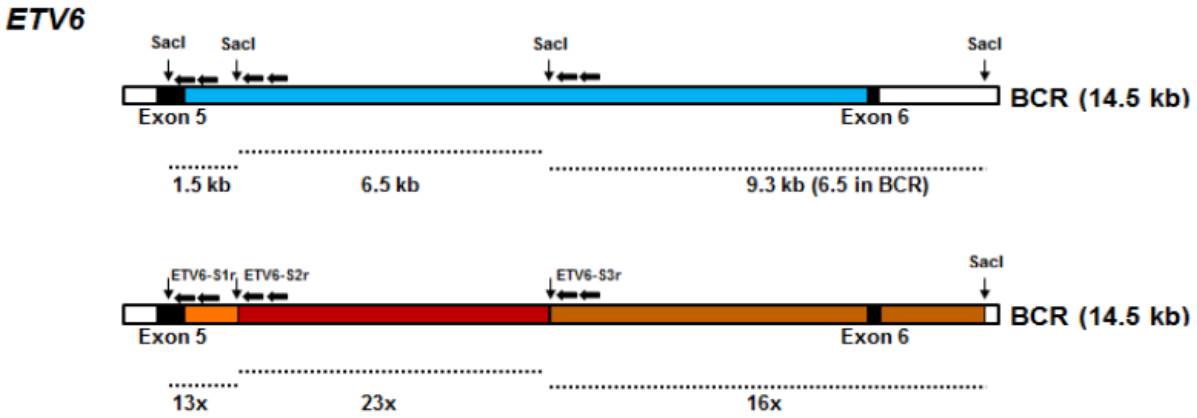
Supplemental Figures



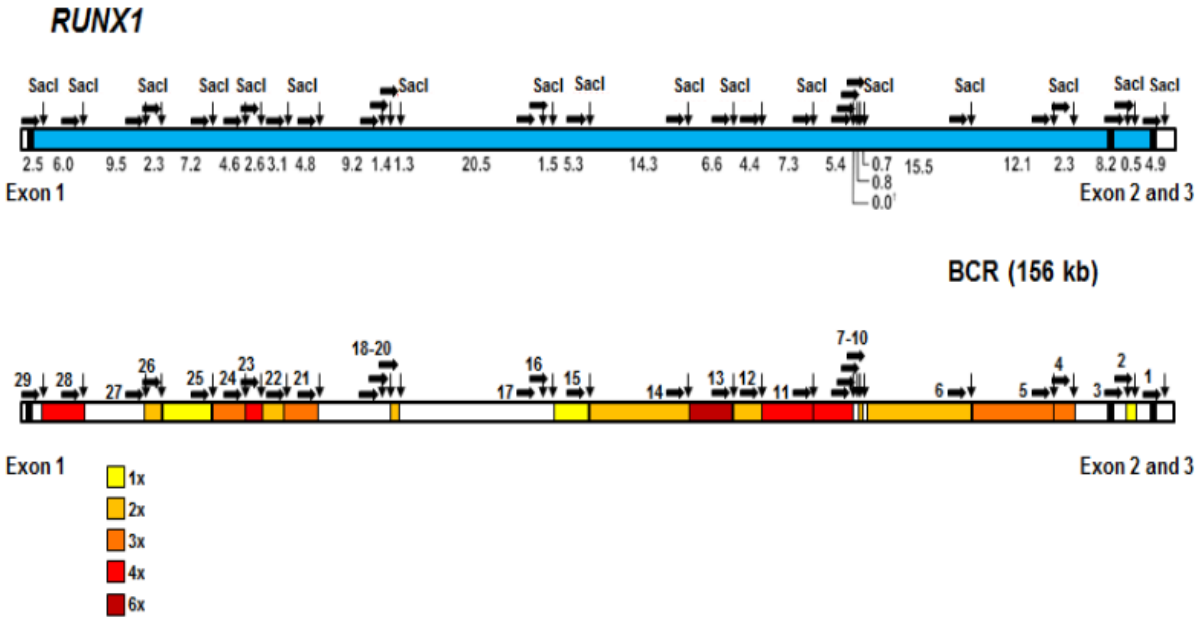
Supplemental Figure S1: Cord blood sample N926 harbors an *ETV6-RUNX1*-fusion as revealed by DNA-based GIPFEL screening. (A) Result of the first Real-Time PCR. The amplification of the four primer bundles (1 through 4) was compared to the amplification of the internal *RUNX1* wild-type control (Ctrl). Primer bundle 4 exceeds the detection limit of 10^{-4}. **(B)** Amplification plot of the second Real-Time PCR. The forward primers of bundle 4 were demultiplexed and primer RUNX1-S5f was identified as the one responsible for the amplification. **(C)** Agarose gel after a PCR with RUNX1-S5f and reverse primers, both pooled and demultiplexed. The pooled reverse primers and ETV6-S3r-n show the expected product of 106 bp. Hence, reverse primer ETV6-S3r-n and forward primer RUNX1-S5f led to target amplification. **(D)** Sanger sequencing result of cord blood N926. The expected sequences flanking the *SacI* ligation joint were identified, indicating the fusion of *RUNX1* segment S5 to *ETV6* segment S3.



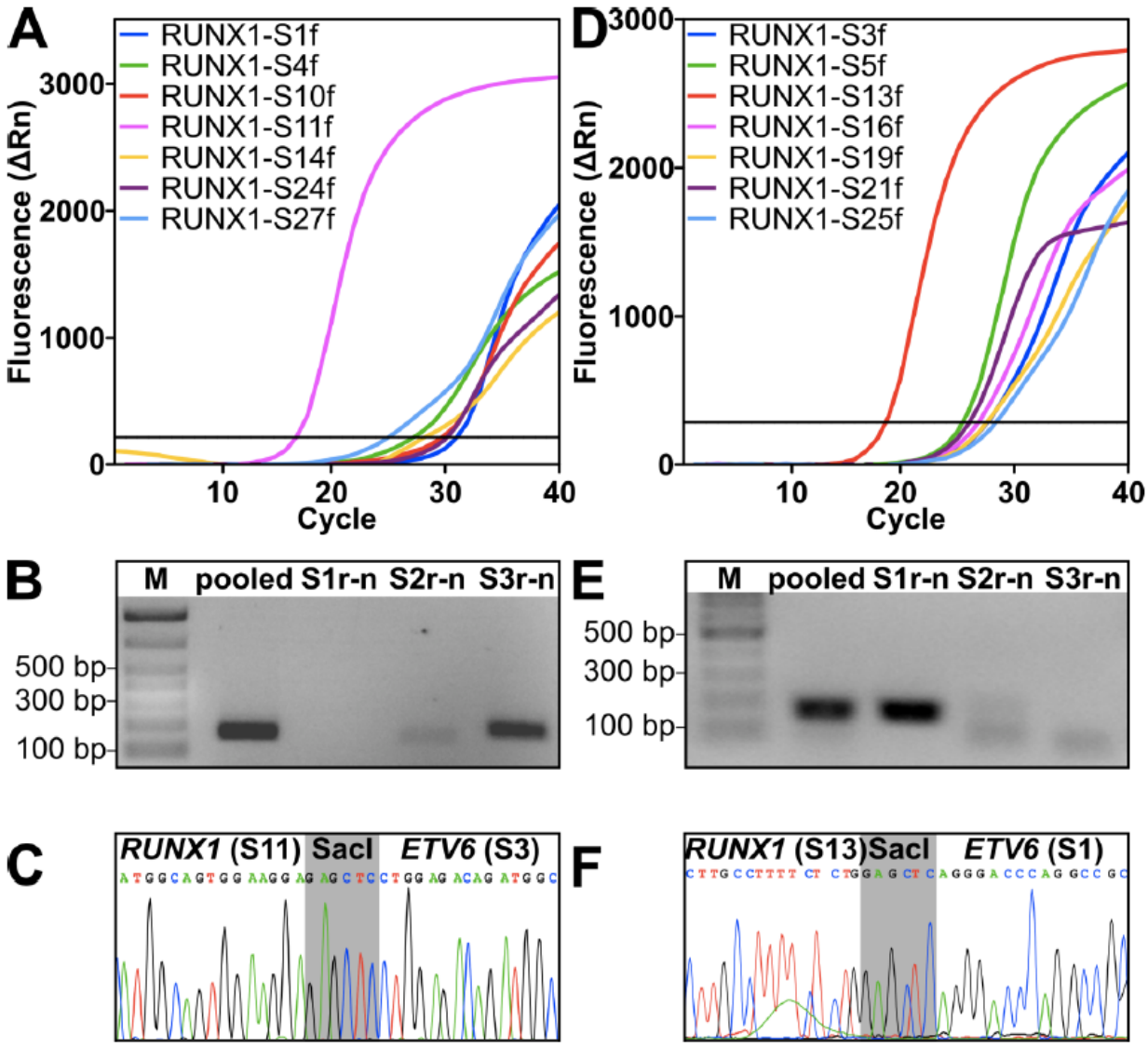
Supplemental Figure S2: Model explaining the high specificity of GIPFEL. *Left panel:* A true positive sample. The translocation generates a *SacI* restriction fragment that contains material from chromosome 12 and 21. After restriction and ligation, circular DNA allows the detection of the ligation joint through the GIPFEL procedure. *Right panel:* An assumed false positive signal. DNA without the investigated translocation is *SacI* restricted and the restriction fragments are ligated to form circular DNA without a true breakpoint. When the human genome is digested with *SacI*, ≈550,000 fragments are generated. Only three of these cover the *ETV6* breakpoint cluster region and can be detected by GIPFEL. For *RUNX1*, 28 fragments can be detected. The probability that these fragments are ligated together by chance is therefore very low. As these fragments have to be ligated in the correct orientation for the primers to generate a PCR product, the probability has to be divided by two. Thus, the theoretical probability that a false positive signal occurs is ≈1.39 x 10^{-10}. A false positive result in addition requires two independent ligation events.



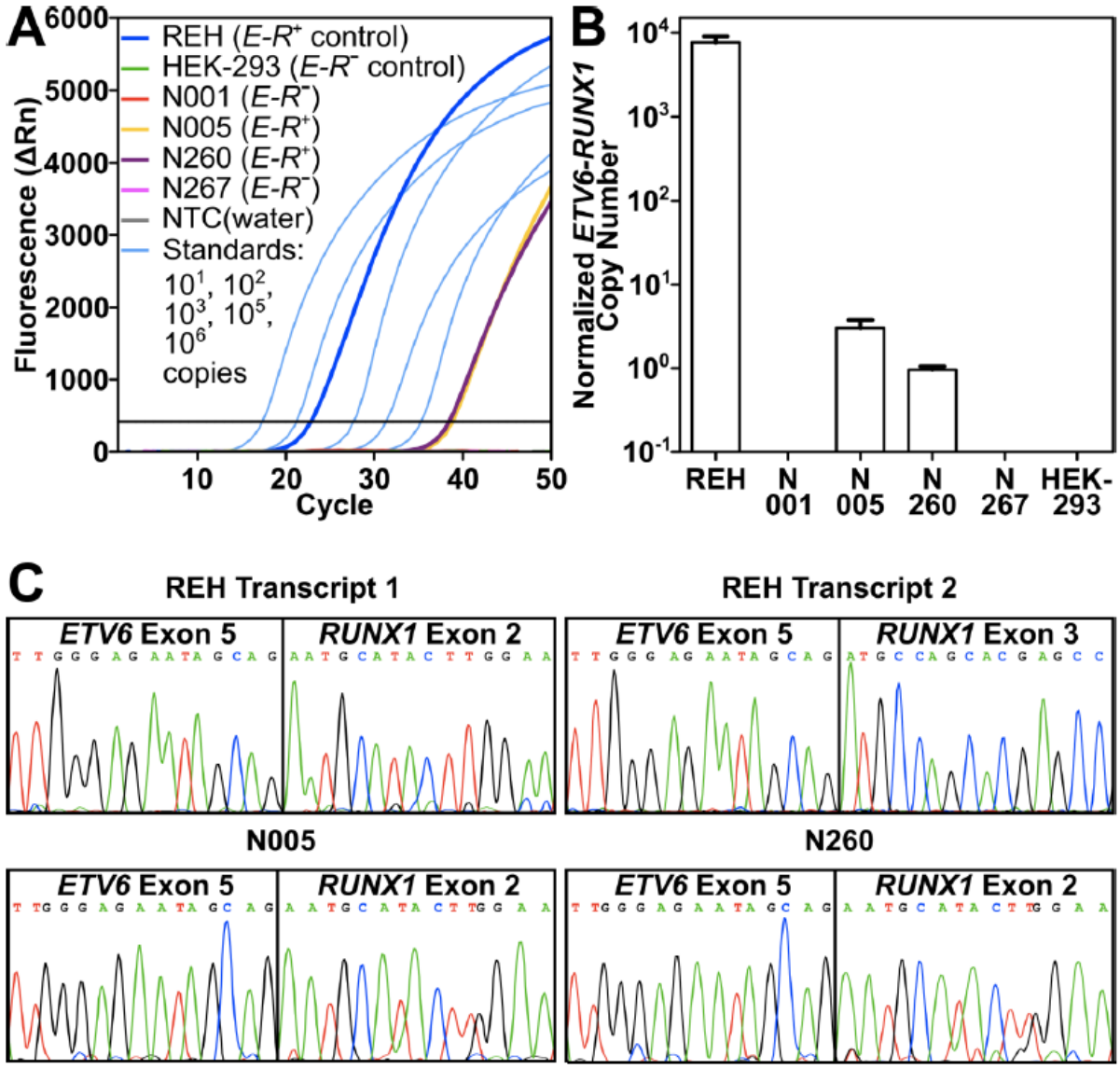
Supplemental Figure S3: Heatmap of the *ETV6* breakpoints detected by GIPFEL screening of umbilical cord blood samples of 1,000 healthy newborns. *Upper panel:* Schematic view of the breakpoint cluster region (BCR) of *ETV6* on chromosome 12. Black boxes mark exons and the blue area is the BCR. *SacI* restriction sites and the length of the resulting *SacI* fragments are given. The black arrows indicate the primers (two per site, one outer primer and one nested primer). *Lower panel:* The region covered by GIPFEL. The colors show how often the break occurred in the respective region during the screening of 1,000 cord blood samples. Primer names are indicated.



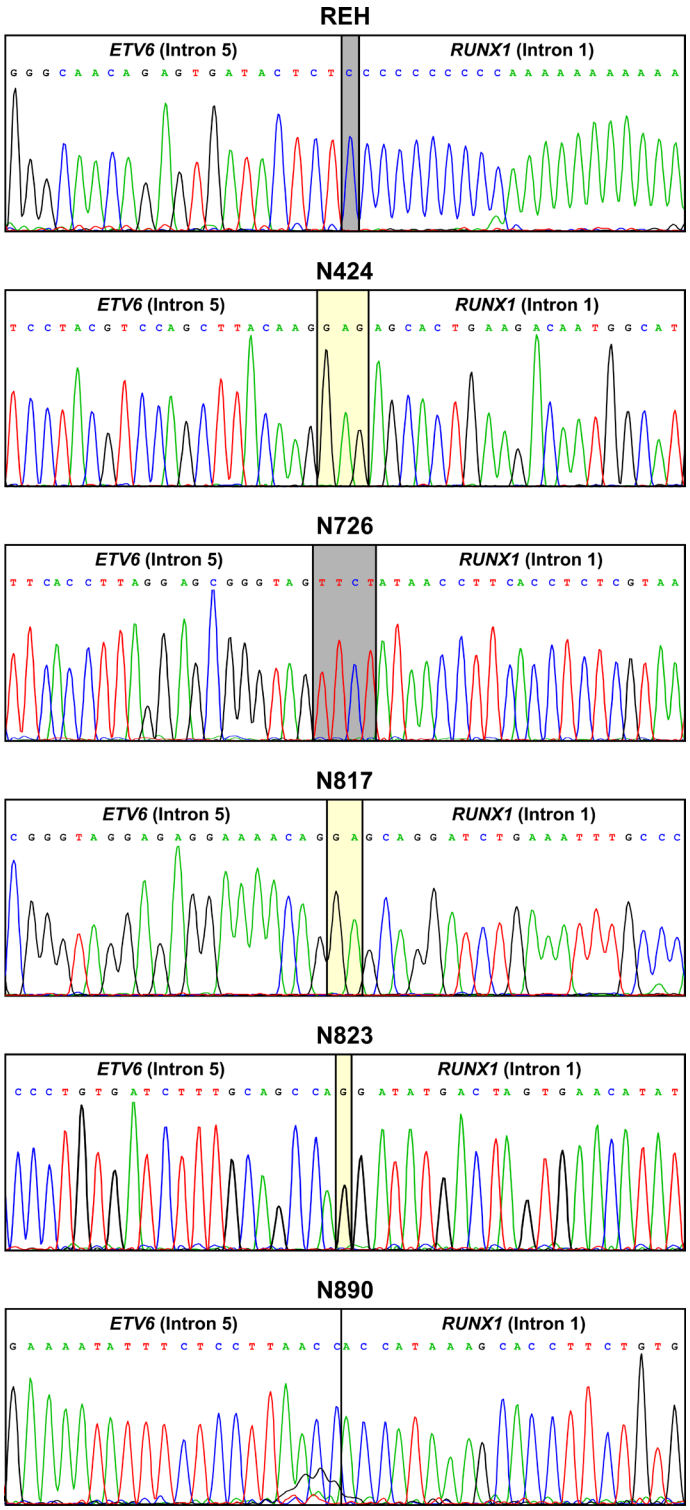
Supplemental Figure S4: Heatmap of the *RUNX1* breakpoints detected by GIPFEL screening of umbilical cord blood samples of 1,000 healthy newborns. *Upper panel:* Schematic view of the breakpoint cluster region (BCR) of *RUNX1* on chromosome 21. Black boxes mark exons and the blue area is the BCR. *SacI* restriction sites and the length of the resulting *SacI* fragments are given. The black arrows indicate the primers (one per site). *Lower panel:* The region covered by GIPFEL. The colors show how often the break occurred in the respective region during the screening of 1,000 cord blood samples. Primer numbers are indicated.



Supplemental Figure S5: Cord blood sample N505 harbors two different coexistent *ETV6-RUNX1*-fusions as revealed by GIPFEL screening. (A) Amplification plot of the second Real-Time PCR. The forward primers of bundle 2 were demultiplexed. Primer RUNX1-S11f was responsible for the amplification. **(B)** Agarose gel showing the PCR products of N505 with RUNX1-S11f and reverse primers, both pooled and demultiplexed. The pooled reverse primers and ETV6-S3r-n show the expected product of 166 bp. **(C)** Sanger sequencing result of cord blood N505 with RUNX1-S11f and ETV6-S3r-n. The expected sequences of *RUNX1* segment S11 and *ETV6* segment S3 were identified flanking the *SacI* ligation joint. **(D-F)** As in **(A-C),** presenting the second coexisting *ETV6-RUNX1* fusion. The primers RUNX1-S13f of bundle 4 and ETV6-S1r-n amplified a positive product. The expected product size was 159 bp.



Supplemental Figure S6: GIPFEL results can be confirmed by RT-qPCR. RNA was isolated from cord blood samples and reverse transcribed. Real-Time PCR with two *ETV6-RUNX1*-positive (*E-R*$^+$) and 50 *ETV6-RUNX1*-negative (*E-R*$^-$) cord blood samples was done with FAM labeled probes for *ETV6-RUNX1* and *ABL1*, which served as an internal control. The cell lines REH and HEK-293 served as positive and negative controls, respectively. Two positive and negative cord blood samples are shown exemplarily. **(A)** Amplification plot of the RT Real-Time PCR. Light blue lines indicate *ETV6-RUNX1* standards. The cell line REH and the cord blood samples N005 and N260 showed amplification of an *ETV6-RUNX1*-specific product. **(B)** The *ETV6-RUNX1* transcript copy number was normalized to 10^4 *ABL1* transcripts. Both positive cord blood samples show low transcription levels of *ETV6-RUNX1*. **(C)** Sanger sequencing results of the PCR products confirmed that both cord blood samples and the REH cell line harbor *ETV6* exon 5 fused to *RUNX1* exon 2 on RNA level. This is in accord with the results obtained from GIPFEL. The cell line REH additionally presented the fusion of ETV6 exon 5 to RUNX1 exon 3.



Supplemental Figure S7: *ETV6-RUNX1* fusions detected by GIPFEL screening can be confirmed by Sanger sequencing of patient specific genomic breakpoints. Amplification of DNA circles derived from the GIPFEL screening was carried out by *Phi29* DNA polymerase using specific primers hybridizing to the ligation region. Amplified DNA was purified and Sanger sequenced. Results of the Sanger sequencing of the breakpoint in the cord blood samples N424, N726, N817, N832 and N890 are shown. Vertical black lines indicate the fusion points. Microhomologies are highlighted in yellow, inserted non-template bases in gray. Note: the REH breakpoint sequence is identical to the published one,[14] but according to current reference sequences, one cytosine is an inserted non-template base.

References

1. Fueller E, Schaefer D, Fischer U, et al. Genomic inverse PCR for exploration of ligated breakpoints (GIPFEL), a new method to detect translocations in leukemia. *PLoS One*. 2014;9(8):e104419.
2. Dean FB, Nelson JR, Giesler TL, Lasken RS. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res*. 2001;11(6):1095-1099.
3. Eguchi-Ishimae M, Eguchi M, Ishii E, et al. Breakage and fusion of the TEL (ETV6) gene in immature B lymphocytes induced by apoptogenic signals. *Blood*. 2001;97(3):737-743.
4. Mori H, Colman SM, Xiao Z, et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci U S A*. 2002;99(12):8242-8247.
5. Olsen M, Madsen HO, Hjalgrim H, Gregers J, Rostgaard K, Schmiegelow K. 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):734-740.
6. Lausten-Thomsen U, Hjalgrim H, Marquart H, Lutterodt M, Petersen BL, Schmiegelow K. ETV6-RUNX1 transcript is not frequent in early human haematopoiesis. *Eur J Haematol*. 2008;81(2):161-162.
7. Lausten-Thomsen U, Madsen HO, Vestergaard TR, Hjalgrim H, Lando A, Schmiegelow K. Increased risk of ALL among premature infants is not explained by increased prevalence of pre-leukemic cell clones. *Blood Cells Mol Dis*. 2010;44(3):188-190.
8. Lausten-Thomsen U, Madsen HO, Vestergaard TR, Hjalgrim H, Nersting J, Schmiegelow K. Prevalence of t(12;21)[ETV6-RUNX1]-positive cells in healthy neonates. *Blood*. 2011;117(1):186-189.
9. Zuna J, Madzo J, Krejci O, et al. ETV6/RUNX1 (TEL/AML1) is a frequent prenatal first hit in childhood leukemia. *Blood*. 2011;117(1):368-369; author reply 370-361.
10. Olsen M, Hjalgrim H, Melbye M, Madsen HO, Schmiegelow K. 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):301-303.
11. Skorvaga M, Nikitina E, Kubes M, et al. Incidence of common preleukemic gene fusions in umbilical cord blood in Slovak population. *PLoS One*. 2014;9(3):e91116.
12. Ornelles DA, Gooding LR, Garnett-Benson C. Neonatal infection with species C adenoviruses confirmed in viable cord blood lymphocytes. *PLoS One*. 2015;10(3):e0119256.
13. Kosik P, Skorvaga M, Durdik M, et al. Low numbers of pre-leukemic fusion genes are frequently present in umbilical cord blood without affecting DNA damage response. *Oncotarget*. 2017;8(22):35824-35834.
14. Wiemels JL, Greaves M. Structure and possible mechanisms of TEL-AML1 gene fusions in childhood acute lymphoblastic leukemia. *Cancer Res*. 1999;59(16):4075-4082.