OPEN

Leukemia (2013) **27**, 2165–2176

www.nature.com/leu

**ORIGINAL ARTICLE**

# The *MLL* recombinome of acute leukemias in 2013

C Meyer[1], J Hofmann[1], T Burmeister[2], D Gröger[2], TS Park[3], M Emerenciano[4], M Pombo de Oliveira[4], A Renneville[5], P Villarese[6], E Macintyre[6], H Cavé[7], E Clappier[7], K Mass-Malo[7], J Zuna[8], J Trka[8], E De Braekeleer[9], M De Braekeleer[9], SH Oh[10], G Tsaur[11], L Fechina[11], VHJ van der Velden[12], JJM van Dongen[12], E Delabesse[13], R Binato[14], MLM Silva[15], A Kustanovich[16], O Aleinikova[16], MH Harris[17], T Lund-Aho[18], V Juvonen[19], O Heidenreich[20], J Vormoor[21], WWL Choi[22], M Jarosova[23], A Kolenova[24], C Bueno[25], P Menendez[25], S Wehner[26], C Eckert[27], P Talmant[28], S Tondeur[29], E Lippert[30], E Launay[31], C Henry[31], P Ballerini[32], H Lapillone[32], MB Callanan[33], JM Cayuela[34], C Herbaux[35], G Cazzaniga[36], PM Kakadiya[37], S Bohlander[37], M Ahlmann[38], JR Choi[39], P Gameiro[40], DS Lee[41], J Krauter[42], P Cornillet-Lefebvre[43], G Te Kronnie[44], BW Schäfer[45], S Kubetzko[45], CN Alonso[46], U zur Stadt[47], R Sutton[48], NC Venn[48], S Izraeli[49], L Trakhtenbrot[49], HO Madsen[50], P Archer[51], J Hancock[51], N Cerveira[52], MR Teixeira[52], L Lo Nigro[53], A Möricke[54], M Stanulla[54], M Schrappe[54], L Sedék[55], T Szczepański[55], CM Zwaan[56], EA Coenen[56], MM van den Heuvel-Eibrink[56], S Strehl[57], M Dworzak[57], R Panzer-Grümayer[57], T Dingermann[1], T Klingebiel[26] and R Marschalek[1]

Chromosomal rearrangements of the human *MLL* (mixed lineage leukemia) gene are associated with high-risk infant, pediatric, adult and therapy-induced acute leukemias. We used long-distance inverse-polymerase chain reaction to characterize the chromosomal rearrangement of individual acute leukemia patients. We present data of the molecular characterization of 1590 *MLL*-rearranged biopsy samples obtained from acute leukemia patients. The precise localization of genomic breakpoints within the *MLL* gene and the involved translocation partner genes (TPGs) were determined and novel TPGs identified. All patients were classified according to their gender (852 females and 745 males), age at diagnosis (558 infant, 416 pediatric and 616 adult leukemia patients) and other clinical criteria. Combined data of our study and recently published data revealed a total of 121 different *MLL* rearrangements, of which 79 TPGs are now characterized at the molecular level. However, only seven rearrangements seem to be predominantly associated with illegitimate recombinations of the *MLL* gene (~90%): *AFF1/AF4*, *MLLT3/AF9*, *MLLT1/ENL*,

[1]Department of Biochemistry, Chemistry and Pharmacy, Institute of Pharmaceutical Biology/ZAFES/Diagnostic Center of Acute Leukemia (DCAL), Goethe-University of Frankfurt, Frankfurt/Main, Germany; [2]Charité-Department of Hematology, Oncology and Tumor Immunology, Berlin, Germany; [3]Department of Laboratory Medicine, School of Medicine, Kyung Hee University, Seoul, Korea; [4]Pediatric Hematology-Oncology Program-Research Center, Instituto Nacional de Cancer Rio de Janeiro, Rio de Janeiro, Brazil; [5]Laboratory of Hematology, Biology and Pathology Center, CHRU of Lille; INSERM-U837, Team 3, Cancer Research Institute of Lille, Lille, France; [6]Biological Hematology, AP-HP Necker-Enfants Malades, Université Paris-Descartes, Paris, France; [7]Department of Genetics, AP-HP Robert Debré, Paris Diderot University, Paris, France; [8]CLIP, Department of Paediatric Haematology/Oncology, Charles University Prague, Second Faculty of Medicine, Prague, Czech Republic; [9]Université de Bretagne Occidentale, Faculté de Médecine et des Sciences de la Santé, Laboratoire d'Histologie, Embryologie et Cytogénétique and INSERM-U1078, Brest, France; [10]Department of Laboratory Medicine, Inje University College of Medicine, Busan, Korea; [11]Regional Children Hospital 1, Research Institute of Medical Cell Technologies, Pediatric Oncology and Hematology Center, Ekaterinburg, Russia; [12]Erasmus MC, Department of Immunology, Rotterdam, The Netherlands; [13]CHU Purpan, Laboratoire d'Hématologie, Toulouse, France; [14]Lab. Célula tronco-CEMO-INCA, Rio de Janeiro, Brazil; [15]Lab. Citogenética-CEMO-INCA, Rio de Janeiro, Brazil; [16]Belarusian Research Center for Pediatric Oncology, Hematology and Immunology, Minsk, Republic of Belarus; [17]Departments of Pathology and Laboratory Medicine, Boston Children's Hospital, Boston, MA, USA; [18]Laboratory of Clinical Genetics, Fimlab Laboratories, Tampere, Finland; [19]Department of Clinical Chemistry and TYKSLAB, University of Turku and Turku University Central Hospital, Turku, Finland; [20]Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne, UK; [21]Northern Institute for Cancer Research, Newcastle University and the Great North Children's Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK; [22]Department of Pathology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; [23]Department of Hemato-Oncology, Faculty of Medicine and Dentistry, Palacky University Olomouc, Olomouc, Czech Republic; [24]Department of Pediatric Hematology and Oncology, University Childrens' Hospital and Medical School of Comenius University, Bratislava, Slovakia; [25]GENyO, Centre for Genomics and Oncological Research: Pfizer, Universidad de Granada, Junta de Andalucia, Granada and Josep Carreras Leukemia Research Institute/Cell Therapy Program University of Barcelona, Barcelona, Spain; [26]Pediatric Hematology and Oncology, University of Frankfurt, Frankfurt, Germany; [27]Charité-Department of Pediatric Oncology and Hematology, Berlin, Germany; [28]Department of Hematology, Centre Hospitalier Universitaire, Nantes, France; [29]CHU Montpellier, Institute for Research in Biotherapy, Laboratory of Hematology, Hôpital Saint-Eloi and NSERM-U847, Montpellier, France; [30]Laboratoire d'Hématologie, CHU de Bordeaux, Bordeaux, France; [31]Service de Cytogénétique et de Biologie Cellulaire, CHU de Rennes, Hôpital Pontchaillou, Rennes, France; [32]Biological Hematology, AP-HP A Trousseau, Pierre et Marie Curie University, Paris, France; [33]INSERM-U823, Oncogenic Pathways in the Haematological Malignancies, Institut Albert Bonniot, Grenoble, France; [34]Laboratoire d'Hématologie, AP-HP Saint-Louis, Paris Diderot University, Paris, France; [35]Service d'Hématologie Immunologie Cytogénétique, Centre Hospitalier de Valenciennes, Valenciennes, France; [36]Centro Ricerca Tettamanti, Clinica Pediatrica Univ. Milano Bicocca, Monza, Italy; [37]Center for Human Genetics, Philipps University Marburg, Marburg, Germany; [38]University Childrens Hospital Muenster, Pediatric Hematology and Oncology, Muenster, Germany; [39]Department of Laboratory Medicine, Yonsei University College of Medicine, Seoul, Korea; [40]Hemato-Oncology Laboratory, UIPM, Portuguese Institute of Oncology of Lisbon, Lisbon, Portugal; [41]Department of Laboratory Medicine, Seoul National University College of Medicine, Seoul, Korea; [42]Hannover Medical School, Clinic for Hematology, Hemostasis, Oncology and Stem Cell Transplantation, Hannover, Germany; [43]Laboratoire d'Hématologie, Hôpital Robert-Debré, Reims, France; [44]Department of Women's and Children's Health, University of Padova, Padova, Italy; [45]University Children's Hospital Zurich, Department of Oncology, Zurich, Switzerland; [46]Hospital Nacional de Pediatría Professor Dr JP Garrahan, Servcio de Hemato-Oncología, Buenos Aires, Argentina; [47]Center for Diagnostic, University Medical Center Hamburg Eppendorf, Hamburg, Germany; [48]Children's Cancer Institute Australia, University of New South Wales, Sydney, New South Wales, Australia; [49]The Chaim Sheba Medical Center, Department of Pediatric Hemato-Oncology and the Cancer Research Center, and Sackler Medical School Tel Aviv University, Tel Aviv, Israel; [50]Department of Clinical Immunology, University Hospital Rigshospitalet, Copenhagen, Denmark; [51]Bristol Genetics Laboratory, Pathology Sciences, Southmead Hospital, North Bristol NHS Trust, Bristol, UK; [52]Department of Genetics, Portuguese Oncology Institute-Porto, and Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal; [53]Center of Pediatric Hematology Oncology, University of Catania, Catania, Italy; [54]Department of Pediatrics, University Medical Centre Schleswig-Holstein, Kiel, Germany; [55]Department of Pediatric Hematology and Oncology, Medical University of Silesia, Zabrze, Poland; [56]Erasmus MC, Sophia Children's Hospital, Department of Pediatric Oncology/Hematology, Rotterdam, The Netherlands and [57]Children's Cancer Research Institute and Medical University of Vienna, Vienna, Austria. Correspondence: Professor Dr R Marschalek, Department of Biochemistry, Chemistry and Pharmacy, Institute of Pharmaceutical Biology/ZAFES/Diagnostic Center of Acute Leukemia (DCAL), Goethe-University of Frankfurt, Marie-Curie Strasse 9, Frankfurt/Main 60439, Germany. E-mail: Rolf.Marschalek@em.uni-frankfurt.de
Received 25 March 2013; revised 23 April 2013; accepted 25 April 2013; accepted article preview online 30 April 2013; advance online publication, 17 May 2013

*MLLT10/AF10*, *ELL*, partial tandem duplications (*MLL* PTDs) and *MLLT4/AF6*, respectively. The *MLL* breakpoint distributions for all clinical relevant subtypes (gender, disease type, age at diagnosis, reciprocal, complex and therapy-induced translocations) are presented. Finally, we present the extending network of reciprocal *MLL* fusions deriving from complex rearrangements.

*Leukemia* (2013) **27**, 2165–2176; doi:10.1038/leu.2013.135

**Keywords:** *MLL*; chromosomal translocations; translocation partner genes; acute leukemia; ALL; AML

## INTRODUCTION

Chromosomal rearrangements involving the human *MLL* (mixed lineage leukemia) gene are recurrently associated with the disease phenotype of acute leukemias.[1,2] The presence of distinct *MLL* rearrangements is an independent dismal prognostic factor, while very few *MLL* rearrangements display either a good or intermediate outcome.[3,4] It became also clear from recent studies that the follow-up of patients during therapy by minimal residual disease (MRD) monitoring has a very strong impact on outcome.[5–7] For this purpose, we established a diagnostic network that allowed different study groups and clinical centers to obtain genomic *MLL* breakpoint sequences that can be directly used for quantifying MRD levels in patients. The current work flow to identify *MLL* rearrangements includes a prescreening step (cytogenetic analyses,[8,9] split-signal fluorescence *in situ* hybridization[10–12] or reverse transcription-polymerase chain reaction (PCR) in combination with long-distance inverse-PCR that was performed on small amounts ($\sim$1 μg) of isolated genomic DNA.[13] This allowed us to identify readily reciprocal translocations, complex chromosomal rearrangements, gene-internal duplications, deletions or inversions on chromosome 11q, and *MLL* gene insertions into other chromosomes, or *vice versa*, the insertion of chromatin material into the *MLL* gene.

To gain insight into the frequency of distinct *MLL* rearrangements, all prescreened samples of infant, pediatric and adult leukemia patients was sent for analysis to the Frankfurt Diagnostic Center of Acute Leukemia (DCAL). Prescreening tests were performed at different European centers (Aarhus, Berlin, Bordeaux, Bratislava, Brest, Bristol, Catania, Copenhagen, Frankfurt, Giessen, Granada, Graz, Grenoble, Haifa, Hamburg, Hanover, Heidelberg, Jena, Jerusalem, Kiel, Lille, Lisbon, Madrid, Minsk, Montpellier, Monza, Munster, Munich, Nancy, Nantes, Newcastle upon Tyne, Olomouc, Padua, Paris, Porto, Prague, Reims, Rotterdam, Tampere, Tel Hashomer, Toulouse, Turku, Tubingen, Vienna, Yekaterinburg, Zabrze and Zurich) and centers located outside of Europe (Boston, Buenos Aires, Hong Kong, Houston, Rio de Janeiro, Seoul, Sydney and Tohoku), where acute leukemia patients are enrolled in different study groups. All prescreened *MLL* rearrangements were successfully analyzed at the Frankfurt DCAL and patient-specific *MLL* fusion sequences for MRD monitoring were obtained.

On the basis of the results obtained in this and previous studies,[13–15] a total of 79 direct translocation partner genes (TPGs) and their specific breakpoint regions have now been identified. Seven additional loci have been cloned where the 5′-portion of *MLL* was not fused to another gene. In 19 other cases, we were not able to identify a der(11) fusion gene. This could be either attributed to a technical problem (such as a too long genomic fragment) or to the fact that no der(11) exists in these few patients. However, in all of these 19 cases, we successfully identified a reciprocal *MLL* fusion allele. The latter subgroup was allocated to the group of 'complex *MLL* rearrangements' ($n = 182$) because of the extending class of 'reciprocal *MLL* fusion genes' (63 loci, 119 fusion genes). Finally, there were still 35 chromosomal translocations of the human *MLL* gene that were characterized in the past by cytogenetic methods, but that were never analyzed at the molecular level. Thus, the *MLL* recombinome presently comprises 121 different 'direct TPGs' (decoding the MLL N terminus), whereas the 182 'reciprocal TPGs' (decoding the MLL C terminus) derive from complex rearrangements that involved already known 'direct TPGs'. It is worth noting that in nearly all of the investigated cases the 3′-*MLL* gene portion was not lost, except the very few cases ($n = 4$ out of 1622) that were interstitial deletions at 11q23 causing a direct fusion of the 5′-*MLL* gene portion with a gene portion localized telomeric to *MLL*, or where we were able to demonstrate that only an *MLL* spliced fusion exists ($n = 3$ out of 1622). Besides the number of direct and reciprocal *MLL* fusions, we tried to analyze all available patient data for interesting association between age, sex, disease type, secondary leukemia and breakpoint localization. All these data and their analyses is here presented and discussed.

## PATIENTS AND METHODS

### Patient material

Genomic DNA was isolated from bone marrow and/or peripheral blood samples of leukemia patients and sent to the DCAL (Frankfurt/Main, Germany). Patient samples were obtained from study groups (the AMLCG-study group, Munich; the GMALL study group, Berlin; Polish Pediatric Leukemia and Lymphoma Study Group; Zabrze; I-BFM network) or other diagnostic centers (Aarhus, Berlin, Bordeaux, Boston, Bratislava, Brest, Bristol, Buenos Aires, Catania, Copenhagen, Frankfurt, Giessen, Granada, Graz, Grenoble, Haifa, Hamburg, Hanover, Heidelberg, Hong Kong, Houston, Jena, Jerusalem, Kiel, Lille, Lisbon, Madrid, Minsk, Montpellier, Monza, Munster, Munich, Nancy, Nantes, Newcastle upon Tyne, Olomouc, Padua, Paris, Porto, Prague, Reims, Rio de Janeiro, Rotterdam, Seoul, Sydney, Tampere, Tel Hashomer, Tohoku, Toulouse, Turku, Tubingen, Vienna, Yekaterinburg, Zabrze and Zurich). Informed consent was obtained from all patients or patients' parents/legal guardians and control individuals.

### Long distance inverse-PCR experiments

All DNA samples were treated and analyzed as described.[13–15] Briefly, 1 μg genomic patient DNA was digested with restriction enzymes and religated to form DNA circles before long-distance inverse-PCR analyses. Restriction polymorphic PCR amplimers were isolated from the gel and subjected to DNA sequence analyses to obtain the patient-specific fusion sequences. This genomic DNA fusion sequence is idiosyncratic for each leukemia patient and was made available to the sender of the DNA sample. The average processing time was around five working days.

### Data evaluation and statistical analyses

All clinical and experimental patient data were implemented into a database program (FileMaker Pro) for further analysis. Information about all individual patients was used to compare all defined subgroups and to perform statistical analyses to retrieve important information or significant correlations. $\chi^2$ Tests were performed to identify significant deviations from mean values.

## RESULTS

### The study cohort

To analyze the recombinome of the human *MLL* gene, 1622 prescreened acute leukemia samples were obtained from the above-mentioned centers over a period of one decade (2003–2013). Successful analysis of the direct *MLL* fusion could be performed for all patient samples except 19 cases, where only a reciprocal *MLL* fusion allele could be characterized. In these cases we identified only the reciprocal *MLL* fusion allele to guarantee

MRD experiments. Of those 1622 cases, 1590 entered this study because we obtained all the critical information that was necessary for data processing (gender, age at diagnosis, disease type and subtype or information about *de novo* or secondary leukemia). A total of 32 cases was excluded from our study because relevant information about these patients were missing; they had the following *MLL* rearrangements: 9 × *MLL-MLLT3/AF9*; 5 × *MLL-AFF1/AF4*; 4 × *MLL-MLLT1/ENL*; 4 × *MLLMLLT10/AF10*, 3 × *MLL-MLLT4/AF6*, 2 × *MLL-MLLT6/AF17*, 1 × *MLL-GAS7*, 1 × *MLL-EPS15*, 1 × *MLL-LOC100128568*, 1 × *LOC387646-MLL* and 1 × *MLL*-partial tandem duplication (*PTD*). The exclusion of these 32 patients did not interfere with the general conclusions made in this study.

### Age distribution according to clinical subtypes

We first analyzed our cohort according to the age at diagnosis. As displayed in Figure 1, the age distribution is quite similar to the expected age distribution known from different cancer registries. Acute lymphocytic leukemia (ALL) incidence has a peak at the age of 2–3 years, and then decreases with age and increases again in older adults. Acute myeloid leukemia (AML) patients display a small peak at 2 years, decline and then steadily increases with age. For the purpose of our study, we separated our cohort into an 'infant acute leukemia group' (0–12 months; $n = 558$: 440 ALL, 105 AML, 13 N/A), a 'pediatric acute leukemia group' (13 months–18 years; $n = 416$: 205 ALL, 202 AML, 9 N/A) and an 'adult acute leukemia patient' group (>18 years; $n = 616$: 333 ALL, 272 AML, 11 N/A). As shown in Figure 1, we also added information about therapy-induced leukemia (TIL; $n = 77$). Thirty-three patients could not be simply categorized into 'ALL' or 'AML' because they received other diagnoses (MLL = 18; myelodysplastic syndrome = 5, primary myelofibrosis = 1; lymphoma = 2) or because we had simply no informations from the corresponding center (unknown disease type = 7).

### Identification of *MLL* rearrangements and their distribution in clinical subgroups

The most frequent *MLL* rearrangements in these six subgroups were summarized in Figure 2. Infant ALL patients ($n = 440$) displayed 216 t(4;11)(q21;q23) involving the *AFF1/AF4* gene, 73 t(9;11)(p22;q23) involving the *MLLT3/AF9* gene, 96 t(11;19)(q23;p13.3) involving the *MLLT1/ENL* gene, 22 t(10;11)(p12;q23) involving the *MLLT10/AF10* gene, 1 t(6;11)(q27;q23) involving the *MLLT4/AF6* gene, 12 t(1;11)(p32;q23) involving the *EPS15* gene and 20 other *MLL* rearrangements (*9p13.3*, *9p22*, *AFF4/AF5*, *DCP1A/SACM1L*, *AFF3/LAF4* (2 ×), *BTBD18*, *N/A (9 ×)*, *PICALM*, *PRPF19*, *EEFSEC* and *TRNC18*).

Infant AML patients ($n = 105$) displayed 2 t(4;11)(q21;q23) involving the *AFF1/AF4* gene, 23 t(9;11)(p22;q23) involving the *MLLT3/AF9* gene, 1 t(11;19)(q23;p13.3) involving the *MLLT1/ENL* gene, 28 t(10;11)(p12;q23) involving the *MLLT10/AF10* gene, 18 t(11;19)(q23;p13.1) involving the *ELL* gene, 3 t(6;11)(q27;q23) involving the *MLLT4/AF6* gene, 1 t(1;11)(p32;q23) involving the *EPS15* gene and 29 other *MLL* rearrangements (*11q24*, *ABI1*, *ABI2*, *MLLT11/AF1Q (7 ×)*, *FLNA (2 ×)*, *FNBP1*, *GAS7*, *KIAA1524*, *MYO1F (3 ×)*, *N/A (3 ×)*, *NEBL*, *NRIP3*, *PICALM*, *SEPT6 (3 ×)* and *SEPT9 (2 ×)*).

Pediatric ALL patients ($n = 205$) displayed 97 t(4;11)(q21;q23) involving the *AFF1/AF4* gene, 37 t(9;11)(p22;q23) involving the *MLLT3/AF9* gene, 40 t(11;19)(q23;p13.3) involving the *MLLT1/ENL* gene, 4 t(10;11)(p12;q23) involving the *MLLT10/AF10* gene, 5 t(6;11)(q27;q23) involving the *MLLT4/AF6* gene, 4 t(1;11)(p32;q23) involving the *EPS15* gene and 18 other *MLL* rearrangements (*1p32*, *21q22*, *MLLT6/AF17*, *BCL9L*, *FOXO3 (2 ×)*, *AFF3/LAF4 (3 ×)*, *MAML2 (2 ×)*, *N/A (2 ×)*, *PICALM*, *RUNDC3B*, *SEPT5*, *SEPT11* and *TRNC18*).

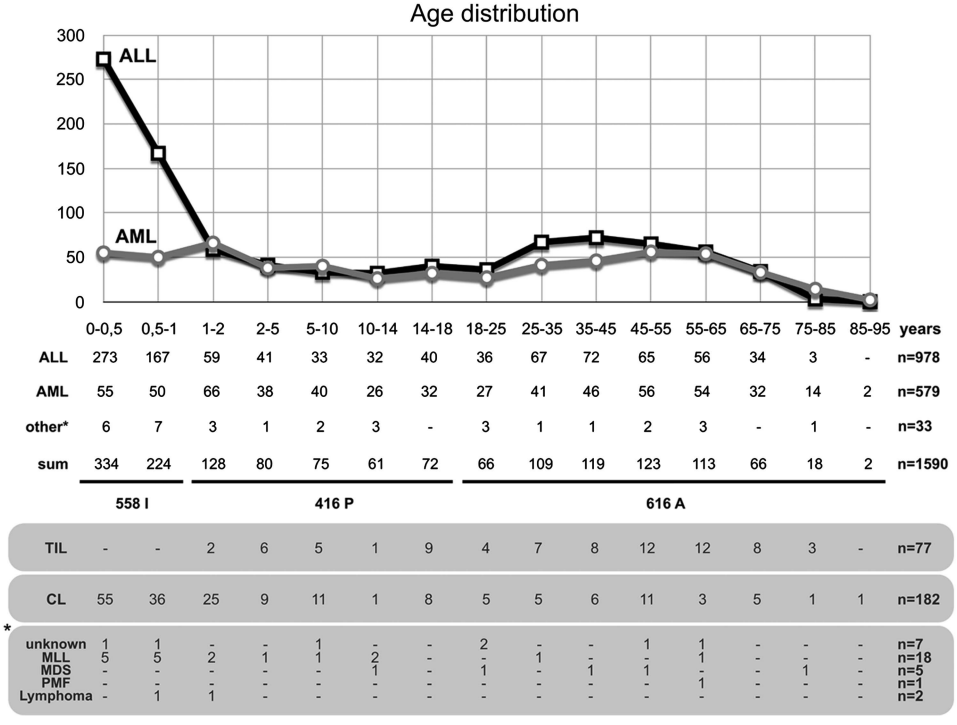


**Figure 1.** Age distribution of investigated patients. The age distribution of all analyzed patients ($n = 1690$) is summarized. (Upper part) Diagram displaying ALL and AML patients. Age at diagnosis was for infants (0–1 year), pediatric (1–18 years) and adult patients (>18 years). The number of ALL, AML and other patients is listed below. We also added the information about TIL patients, the number of complex *MLL* rearrangements (CL) and specified the 'Non-ALL' and 'Non-AML' patients (MLL, myelodysplastic syndrome (MDS), primary myelofibrosis (PMF) and unknown) in more detail for each age group. The precise number of patient cases summarized on the right.

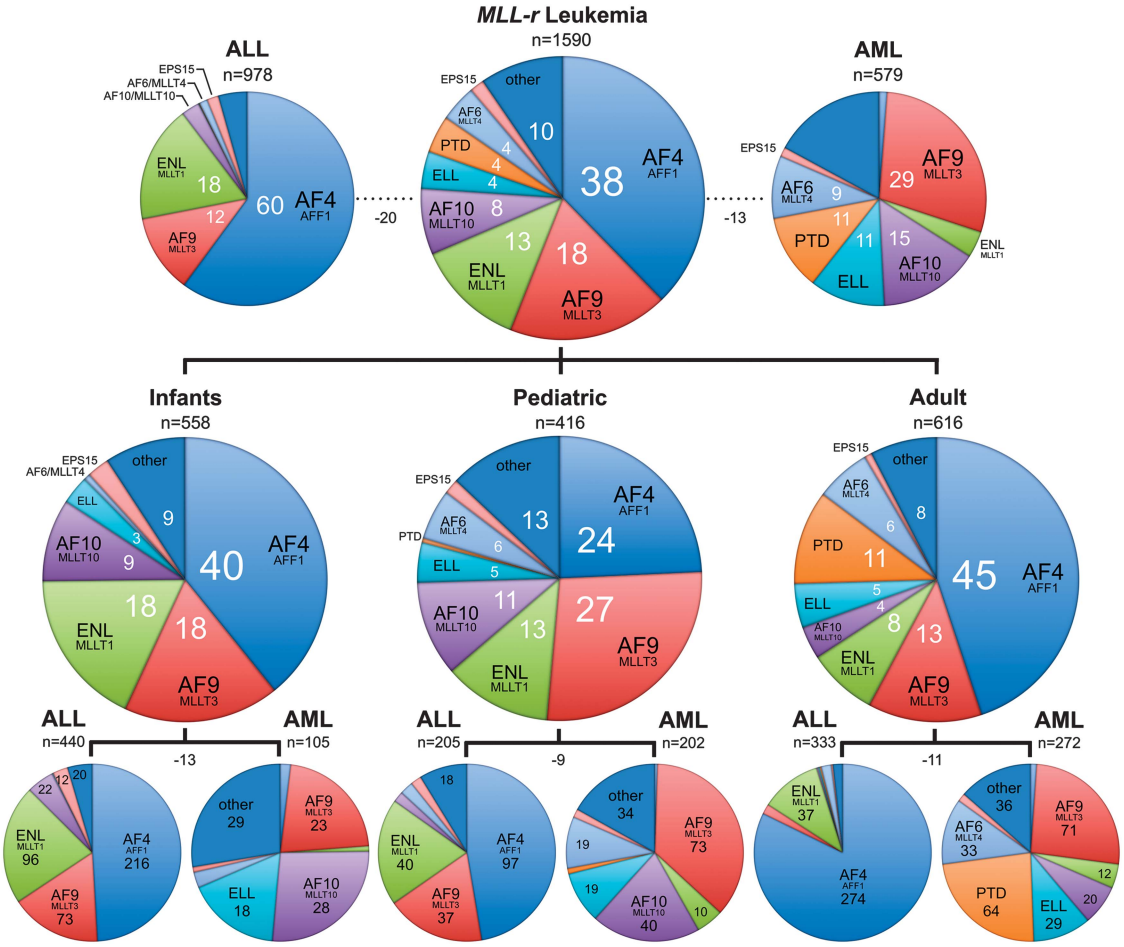


**Figure 2.** Classification of patients according to age classes and disease type. (Top) Frequency of most frequent TPGs in the investigated patient cohort of MLL-rearranged acute leukemia patients ($n = 1590$). This patient cohort was divided into ALL (left) and AML patients (right). Gene names are written in black, and percentages are indicated as white numbers. Thirty-three patients could not be classified into the ALL or the AML disease types, respectively. (Middle) TPG frequencies for the infant, pediatric and adult patient group. (Bottom) Subdivision of all three age groups into ALL and AML patients. Negative numbers confer again to the number of patients who were neither classified to the 'ALL' nor to the 'AML' subgroup.

Pediatric AML patients ($n = 202$) displayed 2 t(4;11)(q21;q23) involving the *AFF1/AF4* gene, 73 t(9;11)(p22;q23) involving the *MLLT3/AF9* gene, 10 t(11;19)(q23;p13.3) involving the *MLLT1/ENL* gene, 40 t(10;11)(p12;q23) involving the *MLLT10/AF10* gene, 19 t(11;19)(q23;p13.1) involving the *ELL* gene, 2 *MLL* PTDs, 19 t(6;11)(q27;q23) involving the *MLLT4/AF6* gene, 3 t(1;11)(p32;q23) involving the *EPS15* gene and 34 other *MLL* rearrangements (*11q23.3*, *ABI1* (2×), *ACACA*, *ACTN4*, *MLLT6/AF17* (2×), *MLLT11/AF1Q* (4×), *ARHGEF17*, *BUD13*, *CASC5*, *LAMC3*, *NA* (3×), *SEPT2*, *SEPT5*, *SEPT6* (6×), *SEPT9* (5×), *SEPT11*, *TET1* and *VAV1*).

Adult ALL patients ($n = 333$) displayed 274 t(4;11)(q21;q23) involving the *AFF1/AF4* gene, 6 t(9;11)(p22;q23) involving the *MLLT3/AF9* gene, 37 t(11;19)(q23;p13.3) involving the *MLLT1/ENL* gene, 1 t(10;11)(p12;q23) involving the *MLLT10/AF10* gene, 1 t(11;19)(q23;p13.1) involving the *ELL* gene, 1 *MLL* PTD, 6 t(6;11)(q27;q23) involving the *MLLT4/AF6* gene, 1 t(1;11)(p32;q23) involving the *EPS15* gene, and 6 other *MLL* rearrangements (*11q23* (2×), *ACTN4*, *CEP164* and *TET1* (2×)).

Adult AML patients ($n = 272$) displayed 3 t(4;11)(q21;q23) involving the *AFF1/AF4* gene, 71 t(9;11)(p22;q23) involving the *MLLT3/AF9* gene, 12 t(11;19)(q23;p13.3) involving the *MLLT1/ENL* gene, 20 t(10;11)(p12;q23) involving the *MLLT10/AF10* gene, 29 t(11;19)(q23;p13.1) involving the *ELL* gene, 64 *MLL* PTDs, 33 t(6;11)(q27;q23) involving the *MLLT4/AF6* gene, 4 t(1;11)(p32;q23) involving the *EPS15* gene and 36 other *MLL* rearrangements (*MLLT6/AF17* (7×), *MLLT11/AF1Q* (2×), *AKAP13*, *AP2A2*, *ARHGEF12*, *C2CD3*, *CASP8AP2*, *CBL*, *DCPS*, *GMPS*, *CEP170B* (2×), *ME2*, *MYH11*, *NA*, *PDS5A*, *PICALM*, *SEPT5*, *SEPT6* (2×), *SEPT9* (5×), *SMAP1*, *TET1* (2×) and *TOP3A*). All these data are summarized in Table 1.

On the basis of the above distribution, about 95% of all ALL patients ($n = 978$) were characterized by the fusion genes *MLL-AFF1/AF4* (~60.0%), *MLL-MLLT1/ENL* (~17.7%), *MLL-MLLT3/AF9* (~11.9%), *MLL-MLLT10/AF10* (~2.8%), *MLL-EPS15* (~1.7%) and *MLL-MLLT4/AF6* (~1.2%), respectively. About 84% of all AML patients ($n = 579$) were characterized by the fusion genes *MLL-MLLT3/AF9* (~28.8%), *MLL-MLLT10/AF10* (~15.2%), *MLL-ELL* (~11.4%), *MLL* PTDs (~11.4%), *MLL-MLLT4/AF6* (~9.5%), *MLL-MLLT1/ENL* (~4.0%), *MLL-SEPT6* (~1.9%) and *MLL-MLLT6/AF17* (~1.6%), respectively. This updates recently published data on the frequency and distribution of different *MLL* fusion partner genes.[15–17]

## Breakpoint distribution according to clinical subtypes

We also investigated the distribution of chromosomal breakpoints within the *MLL* breakpoint cluster region in all investigated clinical subgroups. Briefly, the breakpoint cluster region is localizing between *MLL* exon 9 and *MLL* intron 11, where the majority of

**Table 1.** Overview about all investigated TPGs

| *Direct TPG* | *Infant* | | | *Pediatric* | | | *Adult* | | | *Total* |
|---|---|---|---|---|---|---|---|---|---|---|
| | *ALL* | *AML* | *Other* | *ALL* | *AML* | *Other* | *ALL* | *AML* | *Other* | |
| *AFF1/AF4* | 216 | 2 | 4 | 97 | 2 | 2 | 274 | 3 | — | 600 |
| *MLLT3/AF9* | 73 | 23 | 2 | 37 | 73 | 3 | 6 | 71 | 3 | 291 |
| *MLLT1/ENL* | 96 | 1 | 2 | 40 | 10 | 1 | 37 | 12 | — | 199 |
| *MLLT10/AF10* | 22 | 28 | 2 | 4 | 40 | 1 | 1 | 20 | 2 | 120 |
| *ELL* | — | 18 | — | — | 19 | — | 1 | 29 | 1 | 68 |
| *PTD* | — | — | — | — | 2 | — | 1 | 64 | 1 | 68 |
| *MLLT4/AF6* | 1 | 3 | — | 5 | 19 | — | 6 | 33 | — | 67 |
| *EPS15* | 12 | 1 | 1 | 4 | 3 | — | 1 | 4 | — | 26 |
| *MLLT11/AF1Q* | — | 7 | — | — | 4 | — | — | 2 | — | 13 |
| *SEPT9* | — | 2 | — | — | 5 | — | — | 5 | — | 12 |
| *SEPT6* | — | 3 | — | — | 6 | — | — | 2 | — | 11 |
| *MLLT6/AF17* | — | — | — | 1 | 2 | — | — | 7 | — | 10 |
| *NA* | 9 | 3 | 1 | 2 | 3 | | | 1 | | 19 |
| *AFF3/LAF4* | 2 | — | — | 3 | — | — | — | — | — | 5 |
| *TET1* | — | — | — | — | 1 | — | 2 | 2 | — | 5 |
| *PICALM* | 1 | 1 | — | 1 | — | — | — | 1 | — | 4 |
| *ABI1* | — | 1 | — | — | 2 | — | — | — | — | 3 |
| *CASC5* | — | — | — | — | 1 | — | — | — | 2 | 3 |
| *MYO1F* | — | 3 | — | — | — | — | — | — | — | 3 |
| *SEPT5* | — | — | — | 1 | 1 | — | — | 1 | — | 3 |
| *ACTN4* | — | — | — | — | 1 | — | 1 | — | — | 2 |
| *FLNA* | — | 2 | — | — | — | — | — | — | — | 2 |
| *FOXO3* | — | — | — | 2 | — | — | — | — | — | 2 |
| *CEP170B* | — | — | — | — | — | — | — | 2 | — | 2 |
| *MAML2* | — | — | — | 2 | — | — | — | — | — | 2 |
| *SEPT11* | — | — | — | 1 | 1 | — | — | — | — | 2 |
| *TNRC18* | 1 | — | — | 1 | — | — | — | — | — | 2 |
| *ABI2* | — | 1 | — | — | — | — | — | — | — | 1 |
| *ACACA* | — | — | — | — | 1 | — | — | — | — | 1 |
| *AFF4/AF5* | 1 | — | — | — | — | — | — | — | — | 1 |
| *AKAP13* | — | — | — | — | — | — | — | 1 | — | 1 |
| *AP2A2* | — | — | — | — | — | — | – | 1 | — | 1 |
| *ARHGEF12* | — | — | – | — | – | — | — | 1 | — | 1 |
| *ARHGEF17* | — | — | — | — | 1 | — | — | — | — | 1 |
| *BCL9L* | — | — | — | 1 | — | — | — | — | — | 1 |
| *BUD13* | — | — | — | — | 1 | —— | — | — | — | 1 |
| *C2CD3* | — | — | — | — | — | — | — | 1 | — | 1 |
| *CASP8AP2* | — | — | — | — | — | — | — | 1 | — | 1 |
| *CBL* | — | — | — | — | — | — | — | 1 | — | 1 |
| *CEP164* | — | — | — | — | — | — | 1 | — | — | 1 |
| *CREBBP* | — | — | — | — | — | 1 | — | — | — | 1 |
| *DCP1A* | 1 | — | — | — | — | — | — | — | — | 1 |
| *DCPS* | — | — | — | — | — | — | — | 1 | — | 1 |
| *FNBP1* | — | 1 | — | — | — | — | — | — | — | 1 |
| *GAS7* | — | 1 | — | — | — | — | — | — | — | 1 |
| *GMPS* | — | — | — | — | — | — | — | 1 | — | 1 |
| *KIAA1524* | — | 1 | — | — | — | — | — | — | — | 1 |
| *LAMC3* | — | — | — | — | 1 | — | — | — | — | 1 |
| *LOC100131626* | — | — | — | — | — | — | — | — | 1 | 1 |
| *BTBD18* | 1 | — | — | — | — | — | — | — | — | 1 |
| *ME2* | — | — | — | — | — | — | — | 1 | — | 1 |
| *MYH11* | — | — | — | — | — | — | — | 1 | — | 1 |
| *NEBL* | — | 1 | — | — | — | — | — | — | — | 1 |
| *NRIP3* | — | 1 | — | — | — | — | — | — | — | 1 |
| *PDS5A* | — | — | — | — | — | — | — | 1 | — | 1 |
| *PRPF19* | 1 | — | — | — | — | — | — | — | — | 1 |
| *RUNDC3B* | — | — | — | 1 | — | — | — | — | — | 1 |
| *EEFSEC/SELB* | 1 | — | — | — | — | — | — | — | — | 1 |
| *SEPT2* | — | — | — | — | 1 | — | — | — | — | 1 |
| *SMAP1* | — | — | — | — | — | — | — | 1 | — | 1 |
| *TOP3A* | — | — | — | — | — | — | — | 1 | — | 1 |
| *VAV1* | — | — | — | — | 1 | — | — | — | — | 1 |
| *1p13.1* | — | — | — | — | — | — | — | — | 1 | 1 |
| *1p32 (EPS15)* | — | — | 1 | 1 | — | — | — | — | — | 2 |
| *9p13.3* | 1 | — | — | — | — | — | — | — | — | 1 |
| *9p22 (MLLT3/AF9)* | 1 | — | — | — | — | — | — | — | — | 1 |

**Table 1.** (Continued)

| *Direct TPG* | *Infant* | | | *Pediatric* | | | *Adult* | | | *Total* |
|---|---|---|---|---|---|---|---|---|---|---|
| | *ALL* | *AML* | *Other* | *ALL* | *AML* | *Other* | *ALL* | *AML* | *Other* | |
| *11q23* | — | — | — | — | — | — | 2 | — | — | 2 |
| *11q23.3* | — | — | — | — | 1 | — | — | — | — | 1 |
| *11q24* | — | 1 | — | — | — | — | — | — | — | 1 |
| *21q22* | — | — | — | 1 | — | — | — | — | — | 1 |
| *Xq26.3 (CT45A2)* | — | — | — | — | — | 1 | — | — | — | 1 |
| Sum | 440 | 105 | 13 | 205 | 202 | 9 | 333 | 272 | 11 | 1590 |

Abbreviations: ALL, acute lymphocytic leukemia; AML, acute myeloid leukemia; DCAL, Diagnostic Center of Acute Leukemia. All fusion genes that have been analyzed at the DCAL and their distribution between infant, pediatric and adult leukemia patients are shown. Total numbers are given for each patient group separate in ALL, AML and other diseases. The most frequent fusion partner genes were separated from the gene that has been isolated less frequently.

patients had their individual breakpoints ($n = 1530$). Only sixty patients (3.8%) had their breakpoint outside of the major breakpoint cluster region (see Supplementary Table S1).

Of interest, recently published clinical studies put a new focus on chromosomal breakpoint localization: the distribution of chromosomal breakpoints within the *MLL* breakpoint cluster region was correlated with the outcome of *MLL*-rearranged leukemia patients.[18] Basically, the outcome of leukemia patients with breakpoint in *MLL* intron 11 was worse compared to those patients with upstream breakpoints. A rational explanation for this observation came from the PHD1–3 domain, which is encoded by *MLL* exons 11–16. This domain confers oligomerization[19] and was described to bind to the CYP33/PPIE protein.[20,21] In addition, the PHD3 domain binds either to CYP33/PPIE or to methylated lysine-4 residues of histone H3.[22] Binding of PHD3 to $H3K4_{me2/3}$ peptides is greatly enhanced by the adjacent bromo-domain,[23] but CYP33/PPIE represents a prolyl-peptidyl isomerase and performs a *cis–trans* isomerization of the proline-1665 residue. This *cis-to-trans* conversion is mutual exclusive with $H3K4_{me2/3}$ binding by the PHD3 domain. By contrast, a CYP33/PPIE-bound PHD3 enables binding to BMI1 and associated repressor proteins (HDAC/CBX4/KDM5B), and thus switches the human MLL protein from a transcriptional activator/maintenance factor to a transcriptional repressor. It is worth noting that the adjacent bromo-domain binds to ASB2 and triggers the degradation of MLL.[24] Similarly, a recent publication demonstrated that the PHD2 domain also binds another E3 ligase, named CDC34, which controls again the steady-state stability of the MLL protein.[25]

Breakpoints upstream of *MLL* exon 11 will not alter the domain structure and the associated functions of the PHD1–3 domain, whereas breakpoints within *MLL* exon 11 or intron 11 will definitively destroy this cysteine–histidine-rich domain, most likely because of an alternative protein fold.[18] This will have several effects on the functions of the resulting fusion proteins, like for example, losing the oligomerization capacity, an increased fusion protein stability or losing the ability to switch into a transcriptional repressor (CYP33 → BMI1/HDAC/CBX4/KDM5B).[26] As this should impact cancer biology and clinical behavior, we started to analyze the breakpoint distribution for all clinical subgroups and compared them with the mean distribution observed for all 1590 patients. We decided not to use a random distribution of breakpoints because this will be based only on the length of each DNA region, but will not take into account that the specific chromatin features of *MLL* intron 11 that is highly sensitive against cytotoxic drugs, exhibits a DNase1 hypersensitive site,[27] an apoptotic cleavage site,[28] an RNA polymerase II binding site[29] and several topoisomerase II binding sites.[30]

For our analyses, we subdivided the *MLL* breakpoint cluster region into three subregions: (A) exon 9–intron 9 = 1761 bp; (B) exon 10–intron 10 = 679 bp; and (C) intron 11–intron 12 = 4929 bp. The observed 'mean breakpoint frequencies' for these three regions were A = 38.5%, B = 19.5% and C = 38.7% for all 1530 patients listed in Supplementary Table S1.

As shown in Figure 3, we first subcategorized all patient cases according to their origin. We had 70 samples from North and South American states, 1403 samples from European countries and 117 cases from Russia, Asian countries or the Australian continent. When analyzing the breakpoint frequencies for A–C, it became obvious that the majority of patients in Europe display a breakpoint distribution that was nearly identical to the mean breakpoint frequencies mentioned above. The South American patient group was very young and displayed a nonsignificant tendency to *MLL* intron 11 breakpoints (43.5% vs 37.4%), whereas the Russian/Asian/Australian group displayed a shift towards breakpoints localizing within *MLL* intron 11 (50.43% vs 37.4%, $P = 0.138$). This could neither be attributed to the mean age nor to a higher rate for secondary malignancies (6% vs 5% in Europe). Of interest, all 77 cases of our cohort that were classified as therapy-induced leukemia (TIL) displayed a breakpoint distribution of A = 33.8%, B = 9.5% and C = 54.1%. Thus, even when a controlled exposition to drugs was causing an *MLL* rearrangement, only a maximum of 54% *MLL* intron 11 breaks could be reached. As this is the first description of such a phenomenon and we are missing demographic controls, we cannot draw any conclusions about a putative environmental or maternal exposition during pregnancy that would explain such a shift towards *MLL* intron 11 recombinations. However, when we analyzed this phenomenon in more detail (see Supplementary Table S2), we realized some remarkable differences in certain countries that are even gender specific. Currently, we have no explanations for the observed differences, but future research may help to unravel this phenomenon.

Another observation concerning the breakpoints localization became obvious, when we analyzed breakpoint distributions together with TPGs. As shown in Supplementary Table S3, recombinations affecting *MLLT4/AF6* and *MLLT10/AF10* display a tendency for *MLL* intron 9 breaks rather than *MLL* intron 11 breaks (*MLLT4/AF6*, $P < 0.0001$; *MLLT10/AF10*, $P = 0.006$). This was quite different for *AFF1/AF4* and *MLLT1/ENL* recombinations where *MLL* intron 11 breaks seem to be favored ($P \leqslant 0.0001$). As already described above, the biological properties of the MLL PHD1–3 domain depends on the *MLL* breakpoint. Thus, all fusions occurring within *MLL* introns 9 and 10 will result in fusion proteins that are still able to oligomerize and to be controlled in its steady-state abundance like the wild-type MLL protein. *Vice versa*, recombination within *MLL* intron 11 will result in fusion proteins that could neither be degraded efficiently nor can be switched into transcriptional repressor proteins.

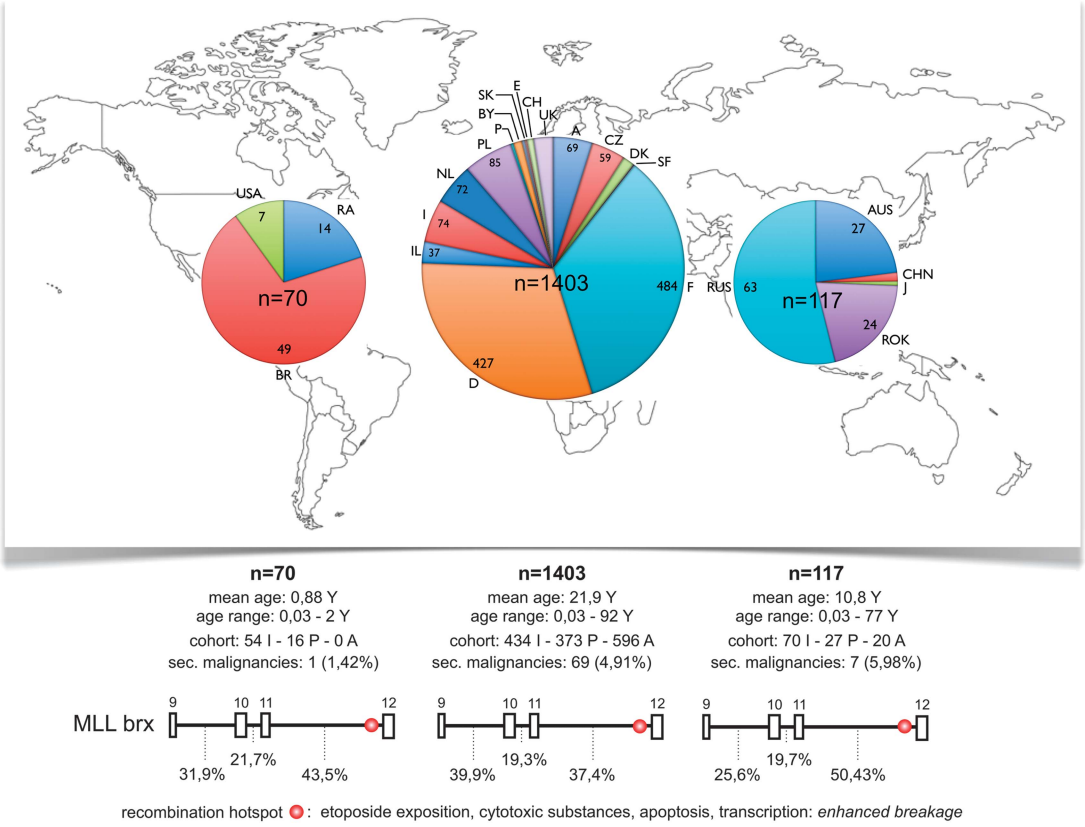


**Figure 3.** World distribution of patients. (Top) Worldmap grossly dividing the investigated patients into three distinct subgroups: American, European and Asian countries. The number of investigated patients is shown and the contribution of individual countries is given in patient numbers. Each country is indicated by its international country code. (Below) Information about the patient cohort. Mean age, age range and the amount of infants (I), pediatric (P) and adult patients (A) is indicated. In addition, we added the amount of therapy-induced malignancies in number and percentage. The breakpoint distribution for each subgroup within MLL exon 9/intron 9, MLL exon 10/intron 10 and MLL exon 11/intron 11 is displayed. Red mark in *MLL* intron 11: fragile site within *MLL* that is sensible to exogenous drug exposure.

These findings also suggest that oligomerization capacity or binding to certain PHD domain-interacting proteins may be quite important for the oncogenic function exerted by MLL fusion proteins. In addition, the breakpoint distribution in infant and adult patients changes significantly: infants display a higher rate of *MLL* intron 11 breakpoints ($P<0.0001$), whereas adults display a higher rate of *MLL* intron 9 breakpoints ($P=0.009$). These findings could not be attributed to the number of cases with secondary malignancies (TIL) or any other parameter, which we listed. These data underscore the importance of the precise breakpoint localization that may—dependent on the involved fusion partner gene—influence even the outcome of patients.[18]

## Novel TPGs

Apart from the many new *MLL* fusion genes that have already been discovered at the DCAL and published in the past years (see Supplementary Table S4; $n=26$), we present additional eight novel TPGs: *RUNDC3B* (Run domain-containing protein 3B; 483 amino acids), *AP2A2* (adaptor protein complex AP-2 subunit α-2; 939 amino acids), *PRPF19* (pre-mRNA processing factor; 504 amino acids); *BUD13* (619 amino acids), *CEP164* (centrosomal protein; 1460 amino acids), *AKAP13* (A kinase-anchoring protein (PKA associated), ARHGEF13; 2813 amino acids), *MYH11* (myosin heavy chain 11; 1938 amino acids) and *ME2* (malic enzyme 2, NAD(+)-dependent, mitochondrial (malate to pyruvate conversion); 584 amino acids).

The RUNDC3B protein has been described to bind to RAP2,[31] a RAS adaptor protein, which has distinct roles in cell adhesion and cell migration. AP2A2 interacts with the mutant form of Huntingtin and alters the kinetic of aggregate formation, thereby functioning as chaperone.[32] PRPF19, also named PRP19 or SNEV, was described to be part of large protein complexes involved in pre-mRNA processing,[33] DNA repair,[34] regulation of proteasomal degradation[35] and was also described as 'senescence evasion factor'.[36] For BUD13 no functional data are available. CEP164 is a centrosomal protein that binds to XPA and is required for UV-dependent DNA repair.[37] Upon DNA damage, CEP164 becomes phosphorylated by ATM/ATR at the serine-186 residue.[38] AKAP13, also known as AKAP-Lbc, represents a Rho-GEF that is regulated by LC3/MAP1LC3A, an important protein for autophagy.[39] It has been described to be involved into the signal pathway from TLR2 to NFKB1[40] and to enhance the cAMP-controlled activation of ERK1/2.[41] *MYH11* is a smooth muscle myosin gene that has been identified through chromosomal rearrangements with CBFB. These inv(16) AML patients express the CBFB–MYH11 fusion protein that is highly oncogenic.[42] Finally, ME2 is a nuclear-encoded mitochondrial enzyme that converts malate into pyruvate.

## The *MLL* recombinome

Within the past 22 years, many genetic aberrations involving the human *MLL* gene located on chromosome 11 band q23 have been described. Seventy-nine TPGs out of 121 are now characterized at the molecular level (see Supplementary Table S4 and Table 1). Forty-five *MLL* fusion genes have been described by others, whereas 34 TPGs have been first identified at the Frankfurt DCAL.

Additional seven loci are presented here, where neither a direct fusion partner gene nor a 'spliced fusion' could be identified. Spliced fusions have been described in cases where the 5′-portion of the *MLL* gene (exons 1–9) is fused with the upstream region of another intact gene. In most of these cases, the last *MLL* exon splices to the second exon of this downstream located gene. Examples for this type of mechanism have already been described,[15] but will also be discussed below. Finally, additional 35 genetic loci were identified by cytogenetics but not further characterized. All yet characterized TPGs and the appropriate citation references were summarized in Supplementary Table S4.

## Genetic alterations resulting in genetic rearrangements of the human *MLL* gene

In general, human *MLL* rearrangements are initiated by a DNA damage situation, which induces DNA repair via the non-homologous-end-joining DNA repair pathway.[43,44] Genetic recombinations involving the human *MLL* gene are predominantly the result of 'reciprocal chromosomal translocations' ($n = 51$; see Figure 4). On the basis of our analyses and the literature, reciprocal recombinations lead to fusions of the 5′-*MLL* gene portion with the following TPGs: *ABI1*, *ABI2*, *ACTN4*, *AFF1/AF4*, *AKAP13*, *ARHGAP26*, *ARHGEF17*, *ASAH3*, *CASC5/AF15Q14*, *CASP8AP2*, *CEP170B*, *CREBBP*, *DAB2IP*, *DCP1A/SACM1L*, *EEFSEC/SELB*, *ELL*, *EP300*, *EPS15*, *FOXO3*, *FOXO4*, *FRYL*, *GAS7*, *GMPS*, *GPHN*, *KIAA1524*, *LAMC3*, *LASP1*, *LPP*, *MAPRE1*, *ME2*, *MLLT1/ENL*, *MLLT3/AF9*, *MLLT4/AF6*, *MLLT6/AF17*, *MLLT11/AF1Q*, *MYO1F*, *MYH11*, *NCKIPSD*, *NEBL*, *PDS5A*, *RUNDC3B*, *SACM1L*, *SEPT2*, *SEPT5/PNUTL*, *SEPT9*, *SEPT11*, *SH3GL1*, *SMAP1*, *TET1/LCX*, *TNRC18* and *TOP3A*, respectively.

Gene-internal PTDs of specific *MLL* gene portions (duplication of *MLL* gene segments coding either for introns 2–9, 2–11, 4–9, 4–11 or 3–8) are frequently observed in AML patients.[45] *MLL* PTDs mediate dimerization of the MLL N terminus, a process that seems to be sufficient to mediate leukemogenic transformation.[46] We have observed *MLL* PTDs in 2 patients within the group of pediatric AML, 1 patient within the group of adult ALL and 65 patients within the group of adult AML. This demonstrates that *MLL* PTDs are predominantly detected in adult AML patients, in line with previously published data.[47]

*MLL* recombinations involving only chromosome 11 are based on two independent DNA strand breaks that are accompanied either by inversions or deletions on 11p or 11q (Inv, Del). Several recombinations have been characterized that belong to these two groups. *MLL* fusions to *AP2A2*, *BTBD18*, *BUD13*, *C2CD3*, *LOC100131626*, *MAML2*, *NRIP3*, *PICALM* and *PRPF19* are based on the inversion of a chromatin portion of 11p or 11q, leading to reciprocal *MLL* fusions. By contrast, a deletion on chromosome 11 fuses the 5′-portion of *MLL* directly to another gene located further downstream (*ARHGEF12*, *BCL9L*, *CBL* and *CEP164)*. In few cases, we observed that the 3′-truncated *MLL* is located upstream of another, intact gene. In that case, we could demonstrate an '*MLL* spliced fusion', which means that the last exon of the *MLL* gene splices directly to the second exon of the further downstream gene. This has been observed for the *MLL-DCPS* fusion. Beside the above-mentioned *DCPS* gene, other genes have been identified that can transcriptionally fuse to 5′-*MLL* sequences. These were *ZFYVE19*, and also the *MLL* fusion partners like *AFF1/AF4*, *CT45A2*, *ELL*, *EPS15*, *MLLT3/AF9*, *MLLT4/AF6*, *MYO1F* and *SEPT5*. In case of *MLLT1/ENL*, about 50% of all recombination events were spliced fusions,[48] and for *MLL-EPS15* fusions about 30%. Spliced fusions to *AFF1/AF4*, *CT45A2*, *DCPS*, *ELL*, *MLLT3/AF9*, *MLLT4/AF6*, *MYO1F*, *SEPT5*, *ZFYVE19* and *SEPT5* represent very rare events.

Beside reciprocal chromosomal translocations of *MLL*, *MLL* PTDs and 11p/q rearrangements (Del and Inv), additional genetic rearrangements were identified in the genomic DNA of analyzed leukemia samples. While the previous rearrangements are based on two independent DNA strand breaks, all other genetic events observed for the *MLL* gene represent more complex rearrangements with at least three or more DNA double-strand breaks. In these cases, the expected reciprocal *MLL* fusion gene cannot be detected, because other sequences will be fused to the 3′-portion of the *MLL* gene.

Complex *MLL* rearrangements are best represented by 'three-way chromosomal translocations' involving three independent chromosomes and resulting in three different fusion genes. More complex is a mechanism that we referred to 'chromosomal fragment insertions'. Either a fragment of chromosome 11 (including portions of the *MLL* gene) is inserted into another chromosome (Ins1), or *vice versa*, a fragment of another chromosome (including portions of a TPG) is inserted into the breakpoint cluster region of the *MLL* gene (Ins2). An insertion mechanism is required in those cases where the transcriptional orientation of a given TPG is not identical to the transcriptional orientation of the *MLL* gene. The *MLL* gene is transcribed in telomeric direction. TPGs with a transcriptional orientation in direction to the centromer are predominantly recombining with *MLL* by such a chromatin insertion mechanism. These genes are *ACACA*, *AFF3/LAF4*, *AFF4/AF5*, *CENPK/FKSG14*, *FLNA*, *FNBP1*, *LOC100128568*, *MLLT10/AF10*, *SARNP*, *SEPT6*, *SORBS2/ARGBP2* and *VAV1*. In all these events at least three independent fusion genes will be generated. The most prominent gene frequently involved in the latter mechanism is the *MLLT10/AF10* gene (see below).

Finally, even more complex rearrangements may occur when 'chromothripsis' comes into play. Chromothripsis has been identified as novel mechanism that generates many fusion alleles in a single event upon a single-cell division (for a review see Holland and Cleveland[49]).

## Reciprocal *MLL* fusions

From two recent papers it became clear that reciprocal MLL fusion proteins may have an important role for cancer development.[50,51] Therefore, we also put emphasis on the analyses of complex *MLL* rearrangements. These 182 patient cases had three-way or four-way translocations resulting in more than two fusion alleles. From these 182 cases, 63 were identified to carry a single 3′-MLL gene portion that was not fused to any upstream gene (only non-coding loci were identified). By contrast, 119 reciprocal gene fusions were identified from which 80% were out-of-frame fusions. Only 24 reciprocal *MLL* fusion genes with in-frame fused exons were identified, being capable of expressing the C-terminal portion of the MLL protein under the control of promoters that derive from reciprocal fusion partner genes ($n = 24$; *ACER1*, *ADARB2*, *APBB1IP*, *ATG16L2*, *CEP164 (2 ×)*, *DENND4A*, *FLJ46266*, *GNA12*, *GPSN2*, *LOC10013227*, *LRRTM4*, , *MYO18A*, , *N-PAC*, *NFKB1*, *NKAIN2*, *PIUP4K2A*, *RABGAP1L*, *RNF115*, *SCAF8*, *SEPT8*, *SEPT5*, *TRIP4*, *UVRAG* and *WNK2*). In all other cases ($n = 158$), the 3′-MLL gene portion was fused either to no gene ($n = 63$; *1p36*, *1q25*, *3 × 1q32*, *2p12*, *2p13*, *2p16*, *2 × 2p21*, *2q11.2*, *3p23.3*, *4p14*, *2 × 4q12*, *4q13*, *6 × 4q21*, *4q22*, *4q27*, *2 × 4q28*, *5q23*, *6p21*, *6q27*, *7p14*, *7q22*, *8p21*, *9p13*, *9p21*, *9p23*, *10p12*, *10p15*, *11p11*, *11p15*, *11q12*, *11q13*, *2 × 11q14*, *11q21*, *3 × 11q22*, *9 × 11q23*, *12p13*, *15q13*, *17q11.2*, *19q12*, *20q11.2* and *2 × 22q13*) or to genes in an out-of-frame or a head-to-head manner ($n = 119$; *ADSS*, *ANTXR2*, *ARCN1*, *ARHGAP12*, *BMP2K*, *BTN3A1*, *BUD13*, *C18orf25*, *CACNA1B*, *CACNB2*, *CCDC33*, *CDK14*, *CMAH*, *CRLF1*, *CRTAC1*, *CUGBP1*, *DHX16*, *DLG2*, *DNAH6*, *DNAJA1*, *DNAJC1*, *DOCK5*, *DSCAML1*, *DSCAML1*, *ELF2*, *EPYC*, *ETV6*, *FCHSD2*, *FXYD2*, *FXYD6*, *GRIA4*, *GRIP1*, *GTDC1*, *HELQ*, *HK1*, *IKZF1*, *KDM2A*, *2 × KIAA0999*, *KIAA1239*, *LMO2*, *LOC100506746*, *LOC390877*, *LOC441179*, *LPXN*, *LRBA*, *MALAT1*, *MCL1*, *MDM1*, *MED1*, *MEF2A*, *MEF2C*, *MMP13*, *MPZL2*, *MPZL3*, *NCAM1*, *NDUFS3*, *NRG3*, *NT5C2*, *PARP14*, *PBRM1*, *PBX1*, *PDE6C*, *PHLDB1*, *PITPNA*, *PIWIL4*, *RDH5*, *RNF25*, *RPS3*, *SCGB1D1*, *SCN3B*, *SEC14L1*, *SFRS4*, *SGK1*, *SLC43A3*, *SNAPC3*, *SORL1*, *2 × SVIL*, *TCF12*, *TIMM44*, *TLN1*,

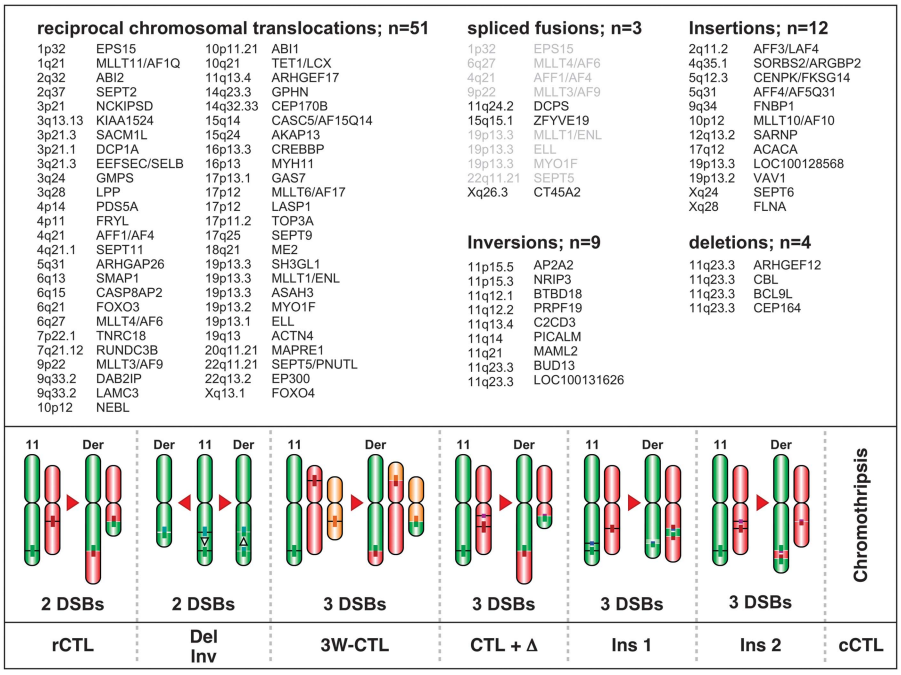


**Figure 4.** General recombination mechanism and associated TPGs. (Top) Genes are categorized either by reciprocal chromosomal translocation (rCTL; $n = 51$), spliced fusion (Spl; $n = 3$), inversions at 11p/q (Inv; $n = 9$), insertions (Ins1 and Ins2; $n = 12$) or 11q deletions (Del; $n = 4$). (Bottom) All identified recombination events, arranged according to the number of DNA double-strand breaks (DSBs) necessary to explain the recombination event. Green: Chromosome 11; red and orange: partner chromosomes involved in the recombination process. Green vertical bars: *MLL*; red, orange, blue and pink vertical bars: partner genes involved in recombination events; derivative 11 chromosomes is always depicted by 'Der'. Black and white horizontal lines: recombination sites on wild-type and derivative chromosomes. rCTL: reciprocal chromosomal translocation; Del/Inv: deletion/inversion; 3 W-CTL: three-way chromosomal translocation; CTL + Δ: chromosomal translocation including deletion(s); Ins1: chromosomal fragment including portions of the *MLL* gene is inserted into a partner chromosome; Ins2: chromosomal fragment including portions of a partner gene is inserted into the *MLL* gene; cCTL: complex chromosomal translocations, for example, by chromothripsis.

*TMEM123*, *TMEM135*, *TNRC6B*, *TNRC6C*, *TNXB*, *TPTE2P5*, *TUBGCP2*, *UBASH3B*, *UBE4A*, *UNC84A*, *USP20*, *WDTC1* and *ZNF57*).

As summarized in Supplementary Table S5, a total of 20 different genes were identified that were involved in these complex rearrangements (*ABI1 (1/3)*, *MLLT10/AF10 (41/120)*, *MLLT6/AF17 (1/10)*, *MLLT11/AF1Q (4/13)*, *AFF1/AF4 (49/600)*, *AFF4/AF5 (1/1)*, *MLLT4/AF6 (6/67)*, *MLLT3/ AF9 (25/291)*, *ELL (4/68)*, *MLLT1/ENL (16/199)*, *EPS15 (2/26)*, *AFF3/LAF4 (2/2)*, *LOC100131626 (1/1)*, *MYO1F (2/3)*, *PICALM (1/4)*, *SEPT6 (3/11)*, *SEPT9 (1/12)*, *TNRC18 (1/1) VAV1 (1/1)* and *Xq26 (1/1)*). The 3′-portion of these TPGs were regularly fused to the 5′-portion of *MLL*, whereas the above-mentioned 182 loci or genes were fused to the 3′-portion of the *MLL* gene. The latter fusions are termed 'reciprocal TPGs' and are summarized in Supplementary Table S5. In all cases where the 3′-portion of the *MLL* gene was fused either to a chromosomal locus (non-coding) or in an out-of-frame manner to another gene, one would argue that no transcript is being made. However, the 3′-portion of the *MLL* is by itself sufficient to produce its own mRNA (starting at the *MLL* intron 11 to exon12 borderline), which can be translated into the MLL* protein.[29] This MLL* protein starts at a *bona fide* AUG start codon encoded by *MLL* exon 18, which results in a protein beginning within the MLL BD domain and ending at the end of the SET domain. The MLL* protein is processed by Taspase1 and results in a 97 kDa MLL*-N and an MLL-C protein fragment. This shorter version of MLL (~235 kDa) loses all functions of the N-terminal portion, whereas functions of the C-terminal portion are retained (for example, H3K4 HMT activity).

Additional 19 *MLL* rearrangements have been characterized where we could not identify the direct *MLL* fusion partner gene. However, in all 19 cases we were able to isolate the reciprocal MLL fusion alleles (*1q25*, *1q32*, *7q22*, *9p21*, *11p11*, *11q21*, *11q23*, *CRTAC1*, *DNAJA1*, *DSCAML1*, *KDM2A*, *RNF115*, *RNF25*, *SEPT5*, *SORL1*, *USP20*, *WDTC1* and *ZNF57*). Only 2 of these 19 cases displayed an in-frame fusion to the 3′-MLL portion (RNF115-MLL and SEPT5-MLL), whereas all the others had solely the intact 3-portion of MLL left to express the MLL* protein (see Supplementary Table S5).

## DISCUSSION

Here, we present an update of the '*MLL* recombinome' associated with different hematologic malignancies, and in particular with acute leukemia (ALL and AML). All our analyses were performed by using small amounts of genomic DNA that were isolated from bone marrow or peripheral blood samples ($n = 1622$) of leukemia patients. In some cases, we analyzed cDNA from a given patient to validate the presence of *MLL* spliced fusions, or to investigate alternative splice products generated from the investigated *MLL* fusion genes. The results of this study allow to draw several conclusions.

The applied long-distance inverse-PCR technique allowed to identify direct and reciprocal *MLL* fusions, *MLL* gene-internal duplications, chromosome 11 inversions, chromosomal 11 deletions and the insertion of chromosome 11 material into other chromosomes, or *vice versa*, the insertion of chromatin material of other chromosomes into the *MLL* gene (see Figure 4). Moreover, we successfully extended our knowledge by analyzing more cases with complex *MLL* rearrangements. During the latter analyses, a large collection of reciprocal *MLL* fusions was identified. About 15% represent in-frame fusions that can be readily expressed into a reciprocal fusion protein. All other characterized reciprocal *MLL* alleles represented out-of-frame fusions with either a

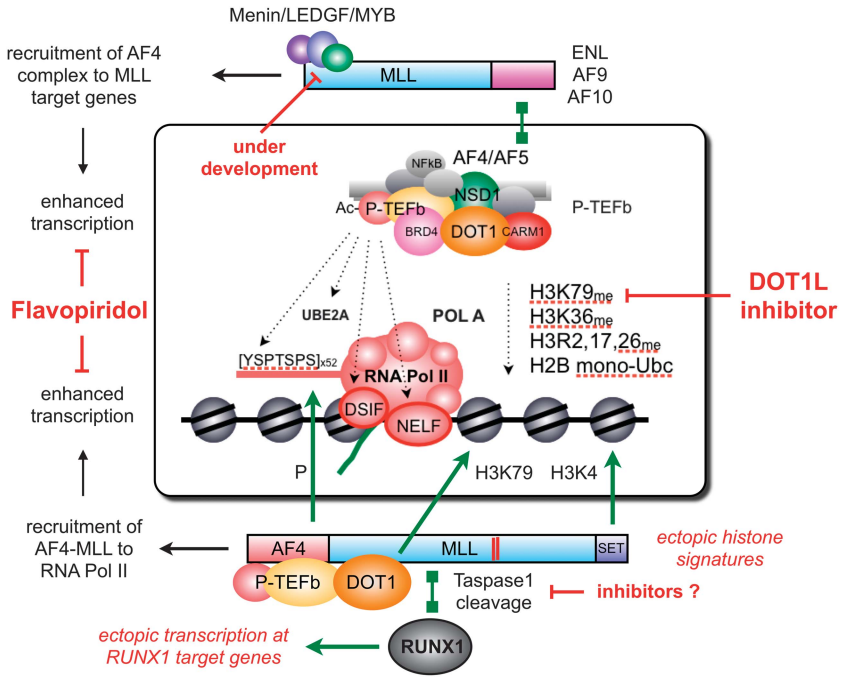


**Figure 5.** Common pathways of the most frequent MLL fusions. The four most frequent MLL fusions, MLL-ENL, MLL-AF9, MLL-AF10 and AF4-MLL, are either interacting directly with the AF4 complex or are mimicking the AF4 complex in case of AF4-MLL. The crucial components within the AF4 complex are the P-TEFb kinase and the H3K79 HMT DOT1L protein. Hyperactive AF4 or AF4-MLL is strongly enhances the transcriptional processes. In addition, changes in the steady-state AF4 complex stability is causing extended H3K79me2/3 signatures. Future inhibitory strategies are indicated in red.

chromosomal locus or a reciprocal TPG, but even these events allow to transcribe and express a 5′-truncated MLL protein, termed MLL*.[29] This shorter version of MLL has no ability to bind Menin1, LEDGF or MYB, but still carries all enzymatic functions necessary to carry out H4K16 acetylations by the associated MOF protein or H3K4 methylation by the SET domain complex.

The analysis of 1622 *MLL* fusion alleles led to the discovery of 34 novel TPGs in the past 10 years, of which 26 have already been described (see Supplementary Table S4). Eight TPGs are completely new and have not been published yet. Taken together with 45 MLL fusions that have been described by others (see Supplementary Table S4), we can present today a total of 79 'direct MLL fusions' that have been characterized at the molecular level. All these *MLL* fusions provide a rich source for future analyses of oncogenic MLL protein variants.

According to our data, the seven most frequent rearrangements of the *MLL* gene occur either with TPGs like *AFF1/AF4*, *MLLT3/AF9*, *MLLT1/ENL*, *MLLT10/AF10*, *ELL*, *MLLT4/AF6* or derive from gene-internal duplications (*MLL-PTDs*). Their occurrence differed significantly in the cohorts of infant, pediatric and adult leukemia patients. We also observed tendencies that correlate specific gene fusions with sex or age at diagnosis. Examples were that *MLLT3/AF9* ($P = 0.080$), *MLLT10/AF10* ($P = 0.019$) and *MLL-PTDs* ($P = 0.065$) occur more frequently in the male group of patients, whereas the female patients were more affected by *MLL-AFF1/AF4* fusions ($P = 0.015$). The most striking finding was that breakpoint distributions differ significantly when concerning distinct TPGs and age groups. It is well known that breakpoints in infants occur more frequently in *MLL* intron 11. We could validate this finding for *MLL-AFF1/AF4* and *MLL-MLLT1/ENL* fusions, but observed a completely contrary situation in case of *MLL-MLLT10/AF10* fusions. Quite surprising was the breakpoint distribution for *MLL-AF6* fusions that displayed a clear preference for *MLL* intron 9 recombinations. Again, these deviations from the observed mean breakpoint distribution are an argument for differences in the biology of the resulting fusion proteins with respect to oligomerization or factor binding dependency. This has to be investigated in more detail in the future to understand these observations.

An important translational aspect of this study is the establishment of patient-specific DNA sequences that can be used for monitoring MRD by quantitative PCR techniques. Owing to the fact that a given *MLL* fusion allele is genetically stable and a monoallelic marker for each tumor cell, a more reliable quantification and tracing of residual tumor cells becomes possible. For each of these 1622 acute leukemia patients at least one *MLL* fusion allele was identified and characterized by sequencing. Several prospective studies were already initiated and first published data verified the reliability of these genomic markers for MRD monitoring.[4] Therefore, the use of these MRD markers will contribute in the future to a better stratification of leukemia patients, which will help to further improve the outcome.

The analysis of the *MLL* recombinome allows to classify *MLL* fusion partner genes into functional categories. As discussed above, only very few TPGs are recurrently identified in different individuals, and moreover, with a significant frequency. On the basis of this study, these TPGs are *AFF1/AF4*, *MLLT3/AF9*, *MLLT1/ENL*, *MLLT10/AF10* and *MLLT4/AF6*. At least for the AFF1/AF4, MLLT3/AF9, MLLT1/ENL and MLLT10/AF10 protein exists a functional correlation, as all these proteins are organized within a protein complex (or different subcomplexes) that affect transcriptional elongation. AF4 is the docking platform for AF9 or ENL, which both interact (via MLLT10/AF10) to DOT1L.[52,53] DOT1L enable methylation of lysine-79 residues of histone H3 proteins, a prerequisite for the maintenance of RNA transcription.[54,55] AF4 binds with its N-terminal portion to the P-TEFb kinase that phosphorylates the largest subunit of RNA polymerase II, DSIF, the NELF complex and UBE2A. This converts RNA POL A into POL E and allows gene transcription.[56] As a result, increased and

extended H3K79 methylation signatures seem to accompany the presence of several fusion proteins (MLL-AFF1/AF4, AFF1/AF4-MLL, MLL-MLLT3/AF9, MLL-MLLT1/ENL, MLL-MLLT10/AF10 and MLL-MLLT4/AF6),[57] whereas an additional increase in H3K4 methylation was only demonstrated by the presence of the reciprocal AFF1/AF4-MLL[56] that causes pro-B ALL in C57Bl6 mice[50] and was shown to cooperate with the RUNX1 protein.[58] Thus, all the major MLL fusions share a common pathway, which is not only functionally related but offers new and interesting venues to develop new drugs against this leukemias, for example, by the development of DOT1L inhibitors.[59] This shared pathway and the effects of certain MLL fusion protein on basic transcription and on the epigenetic layer are summarized in Figure 5. The fusion proteins MLL-MLLT1/ENL, MLL-MLLT3/AF9 and MLL-MLLT10/AF10 recruit thereby the AFF1/AF4 complex, whereas the reciprocal AFF1/AF4-MLL fusion protein is able to perform exactly the same actions on RNA polymerase II and DOT1L. Thus, future therapies addressing either the inhibition of DOT1L, P-TEFb or blocking the interaction of the MLL N terminus with MENIN1/LEDGF/MYB are promising new ways to address these leukemias. In addition, the inhibition of Taspase1 would help to inactivate the AFF1/AF4-MLL fusion protein, as the uncleaved fusion protein is rapidly degraded by SIAH1 and SIAH2.[60]

In summary, *MLL* rearrangements are associated with poor outcome in pediatric and adult acute leukemia. As outlined above, the systematic analysis of the *MLL* recombinome allows one to draw conclusions on certain aspects of the hematomalignant transformation processes. We also present additional information as Supplementary data files (see Supplementary Tables S6–8), which contain general information about the investigated patient cohort, the analyzed T-ALL cases ($n = 36$) and the TIL cases ($n = 77$). Our efforts to analyze the *MLL* recombinome will be continued and provided as free-of-charge service to any collaborators.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We thank all local doctors and biologists who provided clinical information and material. This work was made possible by and conducted within the framework of the International BFM Study Group. This study was supported by Grant DKS 2011.09 from the German Children Cancer Aid to RM. TB was supported by Grants R 12/09 and R 10/37f from the German José Carreras Leukemia Foundation. PM is an ICREA Professor from the Catalunya Government.

## REFERENCES

1 Pui CH, Gaynon PS, Boyett JM, Chessells JM, Baruchel A, Kamps W *et al.* Outcome of treatment in childhood acute lymphoblastic leukaemia with rearrangements of the 11q23 chromosomal region. *Lancet* 2002; **359**: 1909–1915.
2 Pui CH, Chessells JM, Camitta B, Baruchel A, Biondi A, Boyett JM *et al.* Clinical heterogeneity in childhood acute lymphoblastic leukemia with 11q23 rearrangements. *Leukemia* 2003; **17**: 700–706.
3 Balgobind BV, Raimondi SC, Harbott J, Zimmermann M, Alonzo TA, Auvrignon A *et al.* Novel prognostic subgroups in childhood 11q23/MLL-rearranged acute myeloid leukemia: results of an International Retrospective Sstudy. *Blood* 2009; **114**: 2489–2496.
4 Szczepański T, Harrison CJ, van Dongen JJ. Genetic aberrations in paediatric acute leukaemias and implications for management of patients. *Lancet Oncol* 2010; **11**: 880–889.
5 Burmeister T, Marschalek R, Schneider B, Meyer C, Gökbuget N, Schwartz S *et al.* Monitoring minimal residual disease by quantification of genomic chromosomal breakpoint sequences in acute leukemias with MLL aberrations. *Leukemia* 2006; **20**: 451–457.
6 van der Velden VH, Corral L, Valsecchi MG, Jansen MW, De Lorenzo P, Cazzaniga G *et al.* Prognostic significance of minimal residual disease in infants with acute lymphoblastic leukemia treated within the Interfant-99 protocol. *Leukemia* 2009; **23**: 1073–1079.
7 Yeoh AE, Ariffin H, Chai EL, Kwok CS, Chan YH, Ponnudurai K *et al.* Minimal residual disease-guided treatment deintensification for children with acute lymphoblastic leukemia: results from the Malaysia–Singapore acute lymphoblastic leukemia 2003 study. *J Clin Oncol* 2012; **30**: 2384–2392.
8 Johansson B, Moorman AV, Secker-Walker LM. Derivative chromosomes of 11q23-translocations in hematologic malignancies. European 11q23 Workshop participants. *Leukemia* 1998; **12**: 828–833.
9 Heerema NA, Sather HN, Ge J, Arthur DC, Hilden JM, Trigg ME *et al.* Cytogenetic studies of infant acute lymphoblastic leukemia: poor prognosis of infants with t(4;11)—a report of the Children's Cancer Group. *Leukemia* 1999; **13**: 679–686.
10 Van der Burg M, Beverloo HB, Langerak AW, Wijsman J, van Drunen E, Slater R *et al.* Rapid and sensitive detection of all types of MLL gene translocations with a single FISH probe set. *Leukemia* 1999; **13**: 2107–2113.
11 van der Burg M, Poulsen TS, Hunger SP, Beverloo HB, Smit EM, Vang-Nielsen K *et al.* Split-signal FISH for detection of chromosome aberrations in acute lymphoblastic leukemia. *Leukemia* 2004; **18**: 895–908.
12 Harrison CJ, Moorman AV, Barber KE, Broadfield ZJ, Cheung KL, Harris RL *et al.* Interphase molecular cytogenetic screening for chromosomal abnormalities of prognostic significance in childhood acute lymphoblastic leukaemia: a UK Cancer Cytogenetics Group Study. *Br J Haematol* 2005; **129**: 520–530.
13 Meyer C, Schneider B, Reichel M, Angermueller S, Strehl S, Schnittger S *et al.* Diagnostic tool for the identification of MLL rearrangements including unknown partner genes. *Proc Natl Acad Sci USA* 2005; **102**: 449–454.
14 Meyer C, Schneider B, Jakob S, Strehl S, Schnittger S, Schoch C *et al.* The MLL recombinome of acute leukemias. *Leukemia* 2006; **20**: 777–784.
15 Meyer C, Kowarz E, Hofmann J, Renneville A, Zuna J, Trka J *et al.* New insights into the MLL recombinome of acute leukemias. *Leukemia* 2009; **23**: 1490–1499.
16 Daser A, Rabbitts TH. The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis. *Semin Cancer Biol* 2005; **15**: 175–188.
17 Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer* 2007; **7**: 823–833.
18 Emerenciano M, Meyer C, Mansur MB, Marschalek R, Pombo-de-Oliveira MS. The Brazilian Collaborative Study Group of Infant Acute Leukaemia. The distribution of MLL breakpoints correlates with outcome in infant acute leukaemia. *Br J Haematol* 2013; **161**: 224–236.
19 Linder B, Newman R, Jones LK, Debernardi S, Young BD, Freemont P *et al.* Biochemical analyses of the AF10 protein: the extended LAP/PHD-finger mediates oligomerisation. *J Mol Biol* 2000; **299**: 369–378.
20 Fair K, Anderson M, Bulanova E, Mi H, Tropschug M, Diaz MO. Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. *Mol Cell Biol* 2001; **21**: 3589–3597.
21 Xia ZB, Anderson M, Diaz MO, Zeleznik-Le NJ. MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. *Proc Natl Acad Sci USA* 2003; **100**: 8342–8347.
22 Chang PY, Hom RA, Musselman CA, Zhu L, Kuo A, Gozani O *et al.* Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription. *J Mol Biol* 2010; **400**: 137–144.
23 Wang Z, Song J, Milne TA, Wang GG, Li H, Allis CD *et al.* Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. *Cell* 2010; **141**: 1183–1194.
24 Wang J, Muntean AG, Hess JL. ECSASB2 mediates MLL degradation during hematopoietic differentiation. *Blood* 2012; **119**: 1151–1161.
25 Wang J, Muntean AG, Wu L, Hess JL. A subset of mixed lineage leukemia proteins has plant homeodomain (PHD)-mediated E3 ligase activity. *J Biol Chem* 2012; **287**: 43410–43416.
26 Grow EJ, Wysocka J. Flipping MLL1's switch one proline at a time. *Cell* 2010; **141**: 1108–1010.
27 Strissel PL, Strick R, Rowley JD, Zeleznik-Le NJ. An *in vivo* topoisomerase II cleavage site and a DNase I hypersensitive site colocalize near exon 9 in the MLL breakpoint cluster region. *Blood* 1998; **92**: 3793–3803.
28 Stanulla M, Wang J, Chervinsk DS, Thandla S, Aplan PD. DNA cleavage within the MLL breakpoint cluster region is a specific event which occurs as part of higher-order chromatin fragmentation during the initial stages of apoptosis. *Mol Cell Biol* 1997; **17**: 4070–4079.
29 Scharf S, Zech J, Bursen A, Schraets D, Oliver PL, Kliem S *et al.* Transcription linked to recombination: a gene-internal promoter coincides with the recombination hot spot II of the human MLL gene. *Oncogene* 2007; **26**: 1361–1371.
30 Felix CA. Leukemias related to treatment with DNA topoisomerase II inhibitors. *Med Pediatr Oncol* 2001; **36**: 525–535.
31 Wang S, Zhang Z, Ying K, Chen JZ, Meng XF, Yang QS *et al.* Cloning, expression, and genomic structure of a novel human Rap2 interacting gene (RPIP9). *Biochem Genet* 2003; **41**: 13–25.

32 Raychaudhuri S, Sinha M, Mukhopadhyay D, Bhattacharyya NP. HYPK, a Huntingtin interacting protein, reduces aggregates and apoptosis induced by N-terminal Huntingtin with 40 glutamines in Neuro2a cells and exhibits chaperone-like activity. *Hum Mol Genet* 2008; **17**: 240–255.
33 Grote M, Wolf E, Will CL, Lemm I, Agafonov DE, Schomburg A *et al.* Molecular architecture of the human Prp19/CDC5L complex. *Mol Cell Biol* 2010; **30**: 2105–2119.
34 Vander Kooi CW, Ohi MD, Rosenberg JA, Oldham ML, Newcomer ME, Gould KL *et al.* The Prp19 U-box crystal structure suggests a common dimeric architecture for a class of oligomeric E3 ubiquitin ligases. *Biochemistry* 2006; **45**: 121–130.
35 Sihn CR, Cho SY, Lee JH, Lee TR, Kim SH. Mouse homologue of yeast Prp19 interacts with mouse SUG1, the regulatory subunit of 26S proteasome. *Biochem Biophys Res Commun* 2007; **356**: 175–180.
36 Löscher M, Fortschegger K, Ritter G, Wostry M, Voglauer R, Schmid JA *et al.* Interaction of U-box E3 ligase SNEV with PSMB4, the beta7 subunit of the 20S proteasome. *Biochem J* 2005; **388**: 593–603.
37 Pan YR, Lee EY. UV-dependent interaction between Cep164 and XPA mediates localization of Cep164 at sites of DNA damage and UV sensitivity. *Cell Cycle* 2009; **8**: 655–664.
38 Sivasubramaniam S, Sun X, Pan YR, Wang S, Lee EY. Cep164 is a mediator protein required for the maintenance of genomic stability through modulation of MDC1, RPA, and CHK1. *Genes Dev* 2008; **22**: 587–600.
39 Baisamy L, Cavin S, Jurisch N, Diviani D. The ubiquitin-like protein LC3 regulates the Rho-GEF activity of AKAP-Lbc. *J Biol Chem* 2009; **284**: 28232–28242.
40 Shibolet O, Giallourakis C, Rosenberg I, Mueller T, Xavier RJ, Podolsky DK. AKAP13, a RhoA GTPase-specific guanine exchange factor, is a novel regulator of TLR2 signaling. *J Biol Chem* 2007; **282**: 35308–35317.
41 Smith FD, Langeberg LK, Cellurale C, Pawson T, Morrison DK, Davis RJ *et al.* AKAP-Lbc enhances cyclic AMP control of the ERK1/2 cascade. *Nat Cell Biol* 2010; **12**: 1242–1249.
42 Liu P, Tarlé SA, Hajra A, Claxton DF, Marlton P, Freedman M *et al.* Fusion between transcription factor CBF beta/PEBP2 beta and a myosin heavy chain in acute myeloid leukemia. *Science* 1993; **261**: 1041–1044.
43 Reichel M, Gillert E, Nilson I, Siegler G, Greil J, Fey GH *et al.* Fine structure of translocation breakpoints in leukemic blasts with chromosomal translocation t(4;11): the DNA damage-repair model of translocation. *Oncogene* 1998; **17**: 3035–3044.
44 Richardson C, Jasin M. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* 2000; **405**: 697–700.
45 Schichman SA, Caligiuri MA, Gu Y, Strout MP, Canaani E, Bloomfield CD *et al.* ALL-1 partial duplication in acute leukemia. *Proc Natl Acad Sci USA* 1994; **91**: 6236–6239.
46 Martin ME, Milne TA, Bloyer S, Galoian K, Shen W, Gibbs D *et al.* Dimerization of MLL fusion proteins immortalizes hematopoietic cells. *Cancer Cell* 2003; **4**: 197–207.
47 Bäsecke J, Whelan JT, Griesinger F, Bertrand FE. The MLL partial tandem duplication in acute myeloid leukaemia. *Br J Haematol* 2006; **135**: 438–449.
48 Meyer C, Burmeister T, Strehl S, Schneider B, Hubert D, Zach O *et al.* Spliced MLL fusions: a novel mechanism to generate functional chimeric MLL-MLLT1 transcripts in t(11;19)(q23;p13.3) leukemia. *Leukemia* 2007; **21**: 588–590.
49 Holland AJ, Cleveland DW. Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements. *Nat Med* 2012; **18**: 1630–1638.
50 Bursen A, Schwabe K, Rüster B, Henschler R, Ruthardt M, Dingermann T *et al.* AF4-MLL is capable of inducing ALL in mice without requirement of MLL-AF4. *Blood* 2010; **115**: 3570–3579.
51 Emerenciano M, Kowarz E, Karl K, de Almeida Lopes B, Scholz B, Bracharz S *et al.* Functional analysis of the two reciprocal fusion genes MLL-NEBL and NEBL-MLL reveal their oncogenic potential. *Cancer Lett* 2013; **332**: 30–34.
52 Zeisig DT, Bittner CB, Zeisig BB, García-Cuéllar MP, Hess JL, Slany RK. The eleven-nineteen-leukemia protein ENL connects nuclear MLL fusion partners with chromatin. *Oncogene* 2005; **24**: 5525–5532.
53 Bitoun E, Oliver PL, Davies KE. The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Hum Mol Genet* 2007; **16**: 192–106.
54 Okada Y, Feng Q, Lin Y, Jiang Q, Li Y, Coffield VM *et al.* hDOT1L links histone methylation to leukemogenesis. *Cell* 2005; **121**: 167–178.
55 Zhang Y, Reinberg D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev* 2001; **15**: 2343–2360.
56 Benedikt A, Baltruschat S, Scholz B, Bursen A, Arrey TN, Meyer B *et al.* The leukemogenic AF4-MLL fusion protein causes P-TEFb kinase activation and altered epigenetic signatures. *Leukemia* 2011; **25**: 135–144.
57 Krivtsov AV, Feng Z, Lemieux ME, Faber J, Vempati S, Sinha AU *et al.* H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer Cell* 2008; **14**: 355–368.
58 Wilkinson AC, Ballabio E, Geng H, North P, Tapia M, Kerry J *et al.* RUNX1 Is a key target in t(4;11) leukemias that contributes to gene activation through an AF4–MLL complex interaction. *Cell Rep* 2013; **3**: 116–127.
59 Daigle SR, Olhava EJ, Therkelsen CA, Majer CR, Sneeringer CJ, Song J *et al.* Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. *Cancer Cell* 2011; **20**: 53–65.
60 Bursen A, Moritz S, Gaussmann A, Dingermann T, Marschalek R. Interaction of AF4 wildtype and AF4•MLL fusion protein with SIAH proteins: indication for t(4;11) pathobiology? *Oncogene* 2004; **23**: 6237–6249.



Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu)