cancers

MDPI

Article

Novel and Rare Fusion Transcripts Involving Transcription Factors and Tumor Suppressor Genes in Acute Myeloid Leukemia

Antonella Padella [1,†], Giorgia Simonetti [2,†], Giulia Paciello [3], George Giotopoulos [4,5], Carmen Baldazzi [6], Simona Righi [1], Martina Ghetti [2], Anna Stengel [7], Viviana Guadagnuolo [1], Rossella De Tommaso [1], Cristina Papayannidis [1], Valentina Robustelli [1], Eugenia Franchini [2], Andrea Ghelli Luserna di Rorà [2], Anna Ferrari [2], Maria Chiara Fontana [1], Samantha Bruno [1], Emanuela Ottaviani [1], Simona Soverini [1], Clelia Tiziana Storlazzi [8], Claudia Haferlach [7], Elena Sabattini [1], Nicoletta Testoni [1], Ilaria Iacobucci [9], Brian J.P. Huntly [4,5], Elisa Ficarra [3] and Giovanni Martinelli [2,*]

1 Department of Experimental, Diagnostic and Speciality Medicine, University of Bologna, 40138 Bologna, Italy; antonella.padella@irst.emr.it (A.P.); simona.righi5@unibo.it (S.R.); guadagnuoloviviana@gmail.com (V.G.); rossella.detommaso@live.it (R.D.T.); cristina.papayannidis@gmail.com (C.P.); valentina.robustell2@unibo.it (V.R.); mariachiara.fontana4@unibo.it (M.F.); samantha.bruno2@unibo.it (S.B.); emanuela.ottaviani@aosp.bo.it (E.O.); simona.soverini@unibo.it (S.S.); elena.sabattini@aosp.bo.it (E.S.); nicoletta.testoni@unibo.it (N.T.)

2 Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST) IRCCS, 47014 Meldola (FC), Italy; giorgia.simonetti@irst.emr.it (G.S.); martina.ghetti@irst.emr.it (M.G.); eugenia.franchini@irst.emr.it (E.F.); andrea.ghellilusernadirora@irst.emr.it (A.G.L.d.R.); anna.ferrari@irst.emr.it (A.F.); giovanni.martinelli@irst.emr.it (G.M.)

3 Department of Control and Computer Engineering DAUIN, Politecnico di Torino, 10129 Turin, Italy; giulia.paciello00@gmail.com (G.P.); elisa.ficarra@polito.it (E.F.)

4 Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge CB2 1TN, England, UK; gg320@cam.ac.uk (G.G.); bjph2@cam.ac.uk (B.J.P.H.)

5 Department of Haematology, Cambridge Institute for Medical Research and Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 0XY, UK

6 Institute of Hematology "L. and A. Seràgnoli", Sant'Orsola-Malpighi University Hospital, 40138 Bologna, Italy; carmen.baldazzi2@unibo.it

7 MLL-Munich Leukemia Laboratory, 81377 Munich, Germany; anna.stengel@mll.com (A.S.); claudia.haferlach@mll.com (C.H.)

8 Department of Biology, University of Bari Aldo Moro, 70125 Bari, Italy; cleliatiziana.storlazzi@uniba.it

9 Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA; ilaria.iacobucci@stjude.org

* Correspondence: giovanni.martinelli@irst.emr.it

† A.P. and G.S. contributed equally to this study.

Received: 16 October 2019; Accepted: 2 December 2019; Published: 5 December 2019

Abstract: Approximately 18% of acute myeloid leukemia (AML) cases express a fusion transcript. However, few fusions are recurrent across AML and the identification of these rare chimeras is of interest to characterize AML patients. Here, we studied the transcriptome of 8 adult AML patients with poorly described chromosomal translocation(s), with the aim of identifying novel and rare fusion transcripts. We integrated RNA-sequencing data with multiple approaches including computational analysis, Sanger sequencing, fluorescence in situ hybridization and in vitro studies to assess the oncogenic potential of the *ZEB2-BCL11B* chimera. We detected 7 different fusions with partner genes involving transcription factors (*OAZ-MAFK*, *ZEB2-BCL11B*), tumor suppressors (*SAV1-GYPB*, *PUF60-TYW1*, *CNOT2-WT1*) and rearrangements associated with the loss of *NF1* (*CPD-PXT1*, *UTP6-CRLF3*). Notably, *ZEB2-BCL11B* rearrangements co-occurred with *FLT3* mutations and were associated with a poorly differentiated or mixed phenotype leukemia.

Although the fusion alone did not transform murine c-Kit+ bone marrow cells, 45.4% of 14q32 non-rearranged AML cases were also BCL11B-positive, suggesting a more general and complex mechanism of leukemogenesis associated with BCL11B expression. Overall, by combining different approaches, we described rare fusion events contributing to the complexity of AML and we linked the expression of some chimeras to genomic alterations hitting known genes in AML.

Keywords: acute myeloid leukemia; rare fusion genes; *ZEB2-BCL11B*

1. Introduction

Fusion genes represent a major criterion of diagnosis and prognostic risk stratification in the European Leukemia Net 2017 classification of AML [1], where approximately 18% of cases are characterized by the presence of a known fusion genes as the main driver event [2].

The chromosomal translocation t(15;17), leads to the expression of a *PML-RARA* chimera and characterizes patients with acute promyelocytic leukemia, who generally have favourable prognosis. AMLs expressing the transcripts *RUNX1-RUNXT1* and *CBFβ-MYH11,* associate with the t(8;21) and inv(16), respectively, are also known to confer a favourable prognosis. However, the t(6;9), inv(3)/t(3;3), t(v;11q23.3) and t(9;22) abnormalities result in the expression of *DEK-NUP214*, *GATA2/MECOM* fusions, *KMT2A*-fusions and *BCR-ABL1*, respectively, all of which correlate with a poor outcome [1].

Moreover, fusion genes resulting from chromosomal translocations are common features of other haematological cancers and, due to their unique presence in cancer tissues, they represent extremely attractive therapeutic targets. The paradigm is *BCR-ABL1* in chronic myeloid leukemia (CML) and Philadelphia-positive acute lymphoblastic leukemia (ALL), which drives leukemogenesis and can be targeted by a specific therapy capable of reversing the leukemic phenotype [3,4].

Whole genome sequencing and RNA sequencing (RNA-seq) approaches have allowed the identification of several novel fusions in acute leukemia that remained cryptic by routine cytogenetic analysis. The Cancer Genome Atlas Research Network identified 118 fusions in 179 AML patients by RNA-seq, with an average of 1.5 fusions per patient [5]. Moreover, it has been shown that normal karyotype AMLs are characterized by the presence of several chimeras, mainly deriving from adjacent genes located on the same chromosome and with complex patterns of partner gene orientation [6]. Previous studies discovered the *NUP98-PHF23* fusion gene in paediatric cytogenetically normal AML carrying a cryptic chromosomal translocation between chromosomes 11 and 17 [7,8]. Chromosomal translocations leading to the expression of fusion transcripts are also an hallmark of ALL and the detection of such aberrations is an example of how genomic analysis can dramatically improve the sub-classification of patients [9].

Hence, the identification of fusion events, even when shared by a small subgroup of poorly characterized patents, may be of clinical significance. We thus performed RNA-seq on samples from eight AML patients characterized by the presence of a rare or poorly described chromosomal translocation(s) to identify novel fusion transcripts with a potential leukemogenic/pathogenetic role. We also combined different approaches including cytogenetic, RNA-seq, bioinformatics analysis and literature mining to help in understating the pathogenetic role of the identified novel and rare fusion events. We validated the presence of nine fusion genes involving either transcription factors, tumor suppressors, or associated with a loss event of candidate genes in AML. We found that the landscape of alterations in AML is not limited to known genes, and that fusion genes, albeit rare, may play an important role in the disease development.

2. Results

2.1. RNA-Seq Cohort Selection

We screened the biobank of AML biological samples collected at our Institution between 2010 and 2015. We identified 46 patients (<1% of total cases) carrying a rare chromosomal translocation (i.e., individual incidence <1% [1]) as the sole alteration (13%) or in association with other chromosomal abnormalities (87%). Based on the availability of biological material, eight samples collected at diagnosis or relapse were selected for RNA-seq (Table 1). According to the 2016 revision of WHO classification of myeloid malignancies [10], our cohort included one AML with inv(16)(p13q22) (sample #84), one AML with mutated *NPM1* (sample #63569), one AML without maturation (sample #59810), one AML with maturation (sample #20), one AML with mutated *RUNX1* (sample #21) and three AML cases with myelodysplasia-related changes (samples #32, #68187 and #125). Patients had an average of three mutations per case (range: 1–5). Recurrently mutated genes in our cohort included *DNMT3A* ($n = 2$), *FLT3* ($n = 3$), *IDH2* ($n = 2$), *KDM6A* ($n = 2$) and *TET2* ($n = 2$). All the molecular alterations in myeloid-related genes are listed in Table S2.

2.2. Identification and Validation of Fusions Genes

Among fusions detected by the RNA-seq analysis, we selected 19 for further validation by RT-PCR and Sanger sequencing (Table S3). Of these, 10 were successfully validated, including the known chimera *CBFβ-MYH11* (53% of selected fusions, Figure 1, Tables 2,S2). No chimeras were detected and/or confirmed in samples #32 and #63569. The biological information on the putative function of the novel chimeric proteins is described in Table 2. Specifically, a new in-frame fusion gene was identified in sample #20: *CPD-PXT1*[11] (tier 1, Figure 1), which is hypothesized to be the reciprocal fusion product of a t(6;17)(p21;q11) translocation (Figure S1A). *CPD* encodes for a metallocarboxypeptidase and it maps in chromosome 17q11, approximately 625 Kb upstream *NF1*. Copy number analysis from SNParray data revealed that *CPD* had complex rearrangements including a copy number loss of approximately 2 Mb, from chr17:2872554, which maps in the intron 2-3 of *CPD*, to chr17: 30768221, including the entire *NF1* gene (chr17:29419945-29706695, Figure S2A).

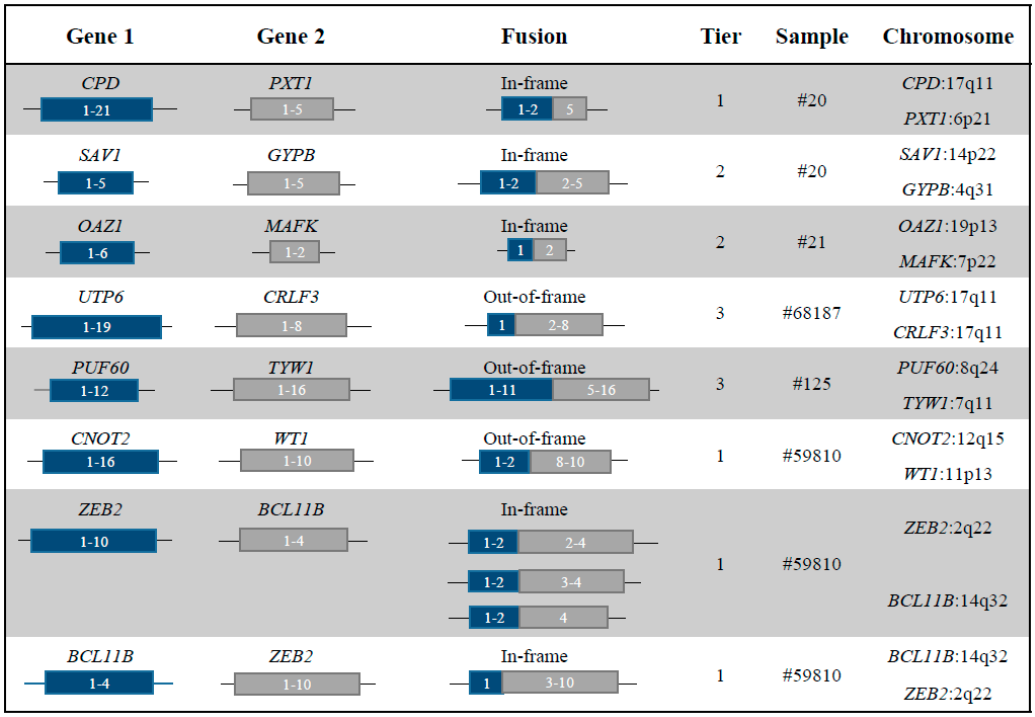

Figure 1. Schematic representation of validated fusion genes. Reading frames, tier, samples in which they were detected and chromosomal location of partner genes are reported. For the *ZEB2-BCL11B* transcript, we detected three splicing isoforms and the reciprocal transcript *BCL11B-ZEB2*.

Sample #20 was also characterized by the in-frame transcript *SAV1-GYPB*, which remained cryptic at cytogenetic analysis. The driver score (DS) predicted by Pegasus (DS=0.87) identified the

chimera as a potential driver of leukemogenesis [12,13] (tier 2, Figure 2A and S1B). In sample #21 we identified a novel fusion event between chromosomes 19 and 7, involving the genes *OAZ1* [14] and *MAFK* [15] (tier 2, Figures 1,2B,S1C).

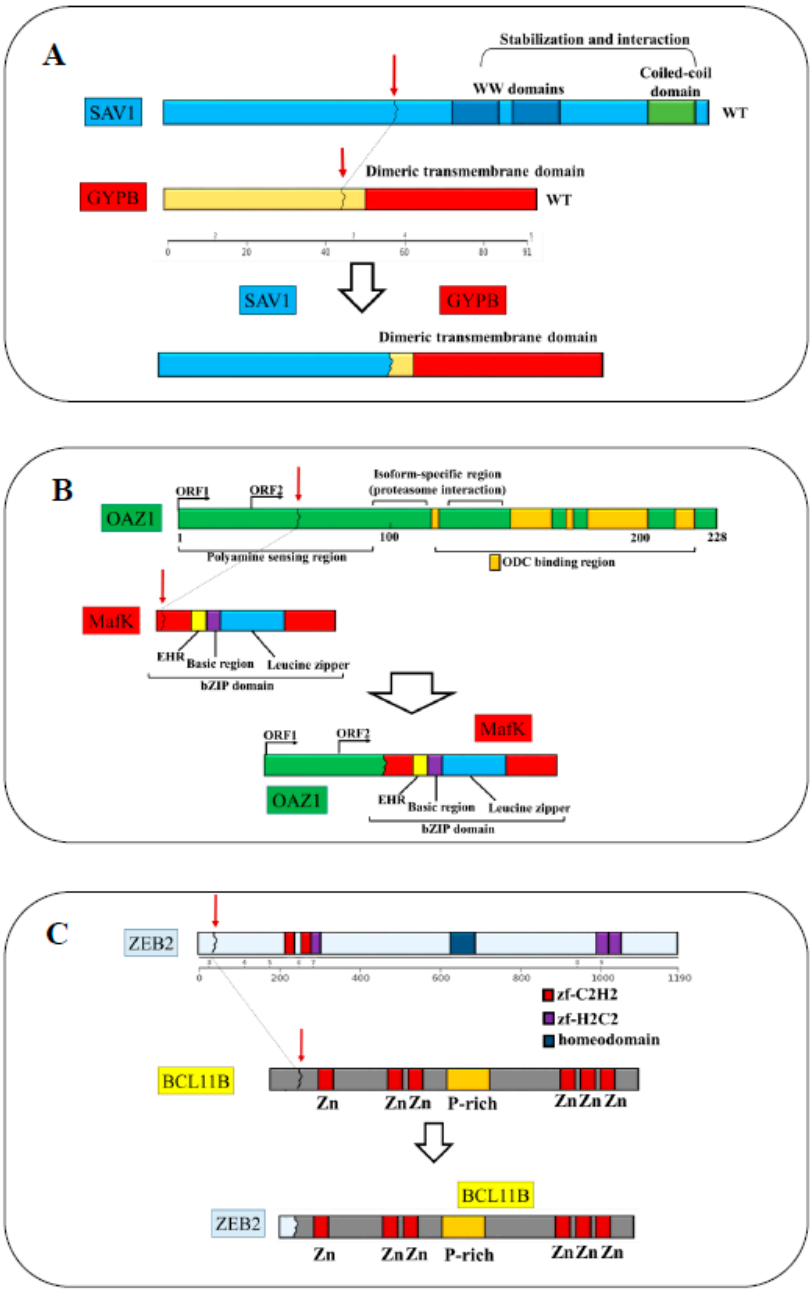

Figure 2. Representation of the domains of the in-frame fusion genes. (**A**) The breakpoint of *SAV1-GYPB* mapped on chromosome 14p22, exon 2 of *SAV1* (NM_021818) and chromosome 4q31, exon 2 of *GYPB* (NM_002100, Figure S1B). In the putative fusion protein, *SAV1* lost the stabilization and interaction domains including the WW domain and the coiled-coil domain, while *GYBP* lost the N-terminal domains and retained the dimeric transmembrane domain. (**B**) The breakpoint of *OAZ1-MAFK* mapped in exon 1 of *OAZ1* (NM_004152), which encodes for a polyamine sensing region and a proteasome interaction domain. The breakpoint at 3′ mapped in exon 2 of *MAFK* (NM_002360), which, together with exon 3, encodes for the bZIP domain. The putative chimeric protein was formed by the sensing regions of polyamine that normally controls the transcription of *OAZ1*, and the bZIP domain of MAFK. (**C**) The breakpoint of the fusion *ZEB2-BCL11B* mapped in exon 2 of *ZEB2* (NM_014795) and exon 2 of *BCL11B* (NM_00128223). Twenty-four residues of ZEB2 and 803 out of 823 residues of BCL11B formed the fusion protein. The codon 20 of BCL11B was the first involved in the fusion and it encoded for an alanine instead of a proline, due to a single nucleotide substitution at the breakpoints junctions (yellow dot).

We validated the out-of-frame fusion *UTP6-CRLF3* [16,17] in sample #68187 (tier 3 Figure 1,S1D). *UTP6* and *CRLF3* mapped on the minus strand of chromosome 17q11 (chr17:30188190-30230729 and chr17:29107702-29153778, respectively). These genes flanked the *NF1* locus and the rearrangement suggested the presence of a 1 Mb copy number loss, which encompasses the *NF1* gene. We also confirmed the presence of the out-of-frame fusion *PUF60-TYW1* [18–20] in sample #125 (tier 3 Figures 1,S1E). Sample #59810 showed the *CNOT2-WT1* [2,21–23] chimera, which is a novel out-of-frame fusion (tier 1, Figures 1,S1F) related to t(11;12)(p15;q22) translocation, identified by cytogenetic analysis. The breakpoint mapped in the forward strand of chromosome 12 and the reverse strand of chromosome 11. We also detected a variant that mapped in exon 3 of a non-coding transcript of *CNOT2* (NR_037615) . The partner genes mapped at opposite strands, the *CNOT2* and *WT1* sequence thus displayed a conserved and inverted sequence orientation, respectively.

In addition to the *CNOT2-WT1* rearrangement, sample #59810 carried the fusion transcript *ZEB2-BCL11B* [24–26] (tier 1, Figures 1,2C,S1G), which is an in-frame fusion and a rare event in AML associated with t(2;14)(q22.3;q32.2)18. Of note, we identified three splicing isoforms (Figure S1H–I), two of which have never been reported before. The type 1 isoform was the full-length chimera that retained all exons involved in the translocation. The type 2 isoform was formed by fusion of the junction of exon 2 of *ZEB2* and exon 3 of *BCL11B*. In the type 3 isoform, exon 2 and 3 of BCL11B were removed, resulting in a smaller transcript encoded by exon 2 of *ZEB2* and exon 4 of *BCL11B*. The reciprocal fusion transcript, formed by exon 1 of *BCL11B* and exon 3 to 10 of *ZEB2*, was also detected and validated (Figure S1J). Details for each chimera are reported in Figure 1, Tables 2,S3.

2.3. Expression of Genes Involved in Fusions and Frequency of Rearrangments Across Cancers

We evaluated the expression of each gene involved in the fusions by comparing its expression to the mean expression of the same gene in wild-type patients of the cohort (Figure S3A). The genes with the most variable expression between fused and wild-type patients were *CRLF3*, *CNOT2* and *WT1*. However, due to the limited number of samples, we could not perform additional statistical analysis to test the significance of our data.

To define the transcriptional program associated with AML carrying the fusion genes, we selected the 1000 most variable genes (based on median absolute deviation values) and we performed unsupervised clustering analysis. Figure S3B showed three clusters, one of which was defined by the *ZEB2-BCL11B* rearranged case alone. The first group was characterized by the presence of the *CBFB-MYH1*, *OAZ1-MAFK* rearranged cases (sample #84 and sample #21, respectively) and a patient without fusions (sample #32). The second cluster included cases characterized by *PUF60-TYW1*, *CPD-PXT1*, *SAV1-GYPB* and *UTP6-CRLF3* fusions. Notably, patients carrying *CPD-PXT1* and *UTP6-CRLF3*, which were associated with *NF1* loss, clustered in this group. This cluster showed a heterogeneous transcriptional profile.

Differentially up-regulated genes ($n = 434$, logFC > 1.5) in the first cluster were enriched for genes involved in the protein processing in endoplasmic reticulum pathway, spliceosome, RNA transport and mRNA surveillance pathway (Table S4). There were no significantly down-regulated genes in the first group one compared to the second one. However, larger cohorts would be required to confirm our signature.

To collect more patients information, we downloaded data from the TCGA Tumour Fusion Gene Data Portal (https://www.tumorfusions.org/[27]) and the Mitelman Database Chromosome Aberrations and Gene Fusions in Cancer (https://mitelmandatabase.isb-cgc.org/). We found that *ZEB2-BCL11B* (as also reported in the manuscript) and *OAZ1-MAFK* fusions were previously annotated in two AML and one multiple myeloma, respectively. Moreover, we analysed the TCGA cancer data looking for genomic rearrangements (and relative frequency) of the genes involved in the 7 fusions we detected in AML. We identified 12 genes that formed chimeras with other partners in different tumour types, namely *CPD*, *PXT1*, *SAV1*, *OAZ1*, *MAFK*, *UTP6*, *CRLF3*, *TYW1*, *CNOT2*, *WT1*, *ZEB2* and *BCL11B*. Moreover, to better understand the role of these genes in AML, we investigated their expression level in the TCGA AML cohort trough the cBio data portal (http://www.cbioportal.org/, Table S5).

Table 1. Patient characteristics and number of validated fusions per patient.

ID	Karyotype main clone	Karyotype second clone	Karyotype other clones	Blasts	WHO classification	Other genetic abnormalities	Phase	Validated fusion(s)
59810	46,XX,t(2;14)(q21;q32),t(11;12)(p15;q22) [17]	46,XX[3]	NA	80%	AML NOS, without maturation	*FLT3, TET2*	Diagnosis	2
20	46,XY,t(6;17)(p21;q11) [20]	NA	NA	90%	AML NOS, with maturation	*NRAS, SRSF2, STAG2, TET2*	Diagnosis	2
21	46,XY,t(3;12)(p22;q24),+4,-15,+mar [19]	46,XY[1]	NA	80%	AML with mutated *RUNX1* (provisional entity)	*CBL, DNMT3A, IDH2, KDM6A, RUNX1*	Relapse	1
32	45,XY,der(12)t(12;18)(p13;q12),-18 [12]	45,XY,t(4;16)(q31;q22),der(12)t(12;18)(p13;q12),-18[4]	45,XY,der(6)t(6;12;18)(p21;p13,q12),-18 [3]/46,XY [1]	80%	AML with MRC	*FLT3, WT1*	Relapse	0
84	47,XX,+8,del(11)(p11p15),t(15;17)(q24q25), inv(16)(p13q22) [20]	NA	NA	80%	AML with inv(16)(p13.1q22)	*CUX1, NOTCH1*	Diagnosis	1
68187	46,XX,add(8)(p23),der(16)t(1;16)(q11;q11) [18]	46,XX[2]	NA	70%	AML with MRC	*ETV6, KDM6A*	Diagnosis	1
63569	46,XY [20]	46,XY,add(10)(p15)[9]	46,XY,add(10)(p15),t(1;8)(p36;q13)[2]	70%	AML with mutated *NPM1*	*DNMT3A, FLT3, IDH2, NPM1*	Relaspe	0
125	46,XX [11]	44~47,XX,t(4;17)(p15;q21),del(5)(q13q33),-7,-18,der(X),+1~3mar[9]	NA	50%	AML with MRC	*TP53*	Diagnosis	1

Sample #84: positive control; NOS= not otherwise specified; MRC = myelodysplasia-related changes. NA = not available. Numbers in squared brackets indicates the number of cells with the relative karyotype.

Table 2. Biological function of genes affected by a fusion event and their potential role in leukemogenesis.

Sample	Fusion	Gene function	Category	Fusion protein putative function
20	*CPD-PXT1*	*CPD* encodes for a metallocarboxypeptidase[11] The role of *PXT1* is unknown	*NF1* loss	The breakpoint in *CPD* was associated with a complex rearrangements that involved the loss of *NF1*. The sample was also characterized by a mutation in *NF1* detected by WES.
20	*SAV1-GYPB*	SAV1 is a tumor suppressor of the Hippo pathway [12] GYBP is a sialoglycoproteins of the human erythrocyte membrane [13]	Tumor suppressor	Loss of function of SAV1.
21	*OAZ-MAFK*	*OAZ1* is an Ornithine decarboxylase (*ODC*) antizyme protein that negatively regulates *ODC* activity [14] *MAFK* is a transcriptional regulator with bZIP domains [15]	Transcription factor	The chimera may alter the cellular transcriptional program.
68187	*UTP6-CRLF3*	*UTP6* is involved in nucleolar processing of pre-18S ribosomal RNA and centriole duplication[16] *CRLF3* is a cytokine receptor-like factor that may negatively regulate cell cycle progression at the G0/G1 phase[28]	*NF1* loss	The rearrangement led to a CN loss involving *NF1*, which maps in the forward strand of chromosome 17: 29421945-29709134 (GRCh37).
125	*PUF60-TYW1*	*PUF60* participates in the splicing machinery[18,20] *TYW1* may be a component of the wybutosine biosynthesis pathway[19]	Tumor suppressor	*PUF60* haploinsufficiency was involved in *TP53*-dependent progression of a T-cell acute lymphoblastic leukaemia[20].
59810	*CNOT2-WT1*	*CNOT2* encodes for a subunit of the multi-component CCR4-NOT complex, which is involved in transcriptional regulation and mRNA degradation [21–23] *WT1* is a transcription factor and it is recurrently altered in haematological malignancies, including AML[2]	Tumor suppressor	The translocation was associated to a deletion at 5′ of *WT1*, which lead to its CN loss.
59810	*ZEB2-BCL11B* and *BCL11B-ZEB2*	*ZEB2* is a transcriptional factor involved in normal and malignant haematopoiesis[24,25] *BCL11B* is a transcription factor and key regulator of both differentiation and survival of T-lymphocytes during thymocyte development[26]	Transcription factor	The chimera may activate an aberrant transcriptional programme.

Table 3. Characteristics of patients carrying the *ZEB2-BCL11B* rearrangement and confirmed by FISH.

Case number	Gender	Age	WHO classification	Karyotype	FISH	T-cell markers	BCR	TCR
11942	male	58	AML NOS	46,XY,t(2;14)(q23;q32)	POSITIVE	NA	no clonality detected	no clonality detected
11954	male	85	AML with mutated *RUNX1* (provisional entity)	46,XY,t(2;14)(q14;q32)	POSITIVE	NA	no clonality detected	no clonality detected
11944	male	79	AUL	46,XY,t(2;14)(q21;q32)	POSITIVE	CD2+; CD7+; TdT+	clonal	no clonality detected
11945	male	59	T/myeloid MPAL	46,XY,t(2;14)(q22;q32)	POSITIVE	CD3+; CD7+; CD2+; TdT+	no clonality detected	clonal
59810	female	40	AML NOS, without maturation	46,XX,t(2;14)(q21;q32),t(11;12)(p15;q22)	POSITIVE	negative	no clonality detected	no clonality detected

The data showed that the some of the candidate genes form chimeras with a variety of partners in different tumor types and the most frequently rearranged genes were CPD and CNOT2. On the other hand, ZEB2-BCL11B was the only recurrent fusion in acute leukemias, suggesting a pro-tumorigenic function in the hematopoietic compartment.

2.4. Relative Frequency of ZEB2-BCL11B Chimera in Acute Leukemia

The fusion protein *ZEB2-BCL11B* was previously described in AML [29] and mixed phenotype acute leukemias [30]. To investigate the frequency of the t(2;14)(q22.3;q32.2) translocation in AML, we interrogated the Mitelman Database (last update on May 21, 2018, Table S6 [31]) and found four AML cases [29,32–34]. Moreover, while the 14q32 region and *BCL11B* are known to be frequently altered in hematological malignancies [35], we found only three additional cases of lymphoid malignancies carrying the t(2;14)(q21;q32) translocation, including biphenotypic leukemia [36] and acute lymphoblastic leukemia [37–39]. However, *ZEB2* and *BCL11B* involvement was confirmed only in one case of AML present in the database. In order to extend the screening to AML patients who are potential candidates on the basis of their the cytogenetic data, we performed FISH on four additional cases carrying t(2;14)(q14-q23;q32) and we confirmed the presence of the *ZEB2-BCL11B* fusion gene in all samples (Figure 3A). Notably, the presence of the fusion was confirmed by RNA-seq in the sample #11945 [40]. At a genomic level, the breakpoint mapped at coordinates chr2:145231055-145231058 and chr14:99736728-99736731. However, it was not possible to locate the exact position of the breakpoint due to the presence of 3 cytosines in the region of the breakpoint, which could belong to either *ZEB2* or *BCL11B* (Supplementary Figure S4A). Overall, *ZEB2-BCL11B* expressing patients (n = 5), were characterized by a median age at diagnosis of 59 years old and by poorly differentiated morphology (Table 3). The immunophenotypic analysis was performed in three patients and two of them expressed T-cell markers. In particular, patient #11944 expressed CD2, CD7 and TdT in 94%, 82% and 26% of cells, respectively, while patient #11945 was also positive for CD3 cytoplasmatic expression, TdT and MPO, with a diagnosis of T/myeloid mixed phenotype acute leukemia (T/M MPAL). Of note, patient #11944 had a diagnosis of acute undifferentiated leukemia (AUL) and patient #59810 was positive only for myeloid markers (CD13 and CD117).

2.5. Specific Pattern of Mutations in Patients Carrying the ZEB2-BCL11B Chimera

We performed targeted next-generation sequencing (NGS) on a panel of genes known to be involved in myeloid malignancies to characterize the mutational landscape of patients carrying the *ZEB2-BCL11B* chimera. *FLT3* alterations were present in 4/5 (80%) patients considered (Table 4): two (40%) were characterized by the internal tandem duplication (ITD) alone with an allelic frequency > 0.5 and two (40%) had point mutations in the tyrosine kinase domain (TKD, one and two point mutations, respectively) and the ITD alteration, but with an allelic frequency <0.5 (40%). Moreover, mutations co-occurring with the *ZEB2-BCL11B* transcript and the *FLT3* alterations targeted *TET2*, *DNMT3A*, *GATA2*, *JAK2*, *RUNX1* and *SRSF2*. Notably, we did not detect any mutation of the screened genes in the patient #11942, who was also negative for *FLT3* aberrations. In addition, Immunoglobulin (IG) and T cell receptor (TCR) molecular analysis showed a clonal rearrangement in the IG heavy chain (IGH) locus, mapping at 14q32 in sample #11944 (AUL), which had a previous history of diffuse large B cell lymphoma, and a TCR rearrangement in sample #11945 (T/myeloid MPAL).

Table 4. Mutational status of myeloid-related genes screened by NGS.

A											
	ASXL1	**BCOR**	**CALR**	**CBL**	**CSF3R**	**CSNK1A1**	**DNMT3A**	**ETNK1**	**ETV6**	**EZH2**	**FLT3-TKD**
#11942	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
#11944	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
#11954	NEG	NA	NA	NA	NA	NA	NEG	NA	NA	NA	NEG
#11945	NEG	NEG	NEG	NEG	NEG	NEG	POS	NEG	NEG	NEG	POS
#59810	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	POS
B											
FLT3-TKD mutation and VAF					**FLT3-ITD**	**FLT3-ITD VAF**	**GATA1**	**GATA2**	**IDH1**	**IDH2**	**JAK2**
					NEG		NEG	NEG	NEG	NEG	NEG
					POS	>0,5	NEG	NEG	NEG	NEG	NEG
					POS	>0,5	NA	NA	NEG	NEG	POS
c.2516A>G, c.2503G>T; p.Asp839Gly, p.Asp835Tyr; 4%, 8%					POS	<0,5	NEG	POS	NEG	NEG	NEG
c.2516A>G, p.Asp839Gly 34%					POS	<0,5	NEG	NEG	NEG	NEG	NEG
C											
KIT	**KRAS**	**MPL**	**NPM1**	**NRAS**	**PHF6**	**PTPN11**	**RUNX1**	**SETBP1**	**SF3B1**	**SRSF2**	**STAG2**	**STAT3**
NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
NEG	NEG	NEG	NEG	NEG	NEG	NEG	VARIANTE	VARIANTE	NEG	POS	NA	NEG
NA	NEG	NA	NEG	NEG	NA	NA	POS	NA	NEG	NA	NA	NA
NEG	NEG	NEG	NEG	NEG	NEG	NEG	VARIANTE	NEG	NEG	NEG	NEG	NA
NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
D											
STAT5B		**TET2**		**TP53**		**U2AF1**	**WT1**	**ZRSR2**			
NEG		VARIANTE		NEG		NEG	NEG	NEG			
NEG		POS		NEG		NEG	NEG	NEG			
NA		NEG		NEG		NA	NA	NA			
NA		NEG		NEG		NEG	NEG	NEG			
NEG		POS		NEG		NEG	NEG	NEG			

NEG: negative; POS: positive; VAF: variant allele frequency; NA: data not available.

2.6. BCL11B Protein Expression in AML and Its Transcriptional Signature

The pro-tumorigenic role of the *ZEB2-BCL11B* fusion has been previously linked to the overexpression of *BCL11B* [29,41]. Paraffin-embedded tissue was available for one of the patients carrying the chimera (#59810) and *BCL11B* expression was confirmed at protein level by immunohistochemistry (Figure 3B).

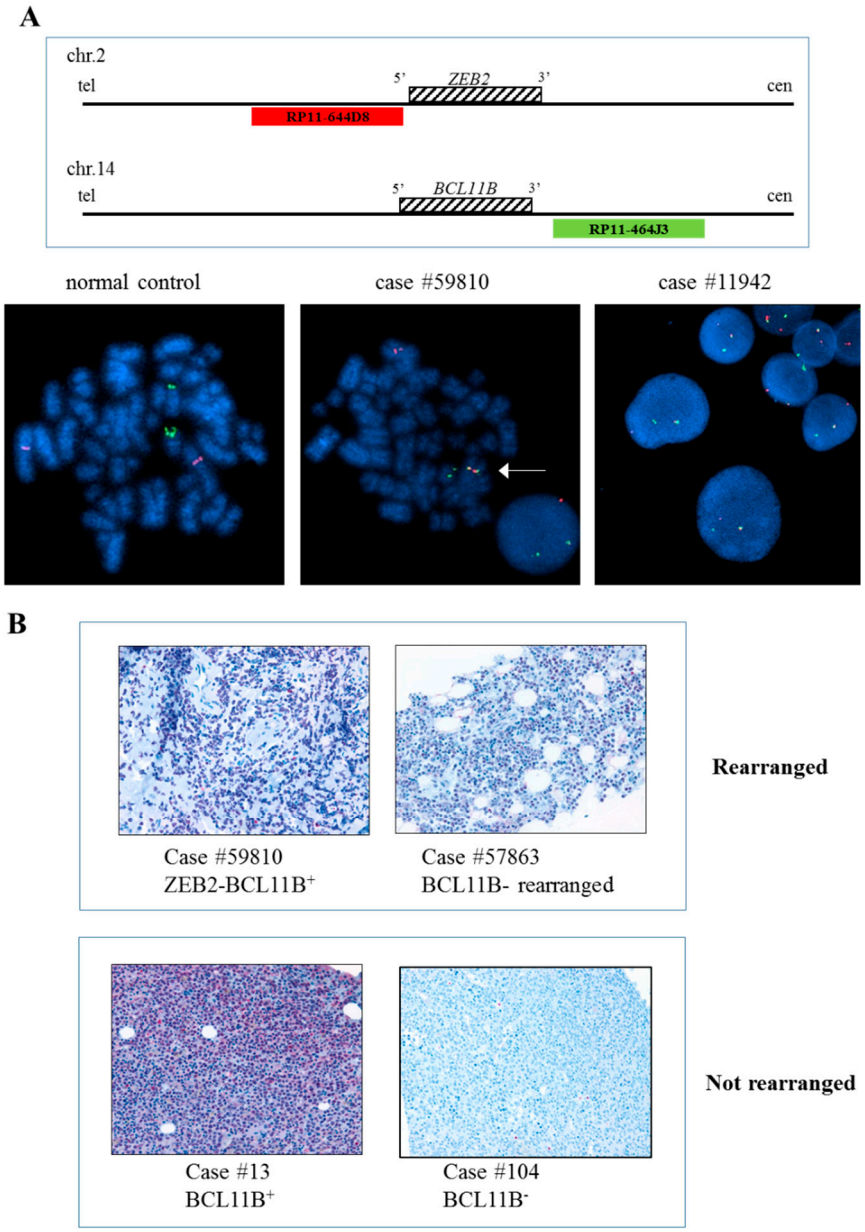

Figure 3. Immunohistochemistry and FISH of AML cases. (**A**) FISH analysis using specific probe for *ZEB2* and *BCL11B* flanking regions. Schematic representation of RP11-644D8 BAC probe in Spectrum Orange covering the 5′ region of *ZEB2* and RP11-464J3 BAC probe in Spectrum Green covering the 3′ region of *BCL11B*, is shown at the top. FISH performed on metaphase spread of case #59810 showing an abnormal fusion pattern (1 fusion, 1 orange and 1 green) with the fusion signal indicating *ZEB2-BCL11B* fusion gene on der(14) (bottom, central) and the same abnormal FISH pattern observed in interphase nuclei of case #11942 (bottom, right). A normal FISH pattern (2 red and 2 green signals) in a normal control case is shown (bottom, left). (**B**) Immunohistochemistry analysis of BCL11B-expressing AML samples #59810, #57863 and #13 carrying t(2;14)(q22.3;q32.3), t(7;14)(q21q32) and no 14q32 alteration, respectively. BCL11B expression was detected in samples regardless of the presence of the 14q32 alterations. The expression was limited to the nucleus and the percentage of positive neoplastic cells was always ≥ 50%.

To understand whether BCL11B expression is a more general feature of AML, we performed immunohistochemistry analysis of 21 additional cases of newly-diagnosed AML not carrying the fusion genes. We detected CD34 expression in 14/21 samples and aberrant nuclear and cytoplasmic nucleophosmin expression in 7/21 biopsies. BCL11B positivity was detected in 9/21 (40.9%) cases of AML (Table S7). BCL11B protein expression in leukemic blasts was limited to the nucleus and varied in strength from weak to moderate. Scattered cells with a stronger positivity could occasionally be seen. In positive cases, the percentage of positive neoplastic cells was always ≥50%. No significant association was found between BCL11B expression and AML immunohistochemical phenotype.

In addition, BCL11B stained positive in one T/M MPAL (CD34, MPO positive, CD3 positive) and one AUL (CD34 positive, CD117 positive and CD2 positive) corresponding to cases #193 and #57863 carrying BCL11B rearrangements and associated to t(6;14)(q25;q32) and t(7;14)(q21;q32), respectively (Table S7, Figures S4B–D).

To identify the transcriptional signature associated with BCL11B expression in AML, we studied the gene expression profile (GEP) data of patients. In our cohort (n = 22), 5% of AML patients had higher expression of *BCL11B* mRNA, however 10 (45.5%) and 12 (54.5%) cases either expressed BCL11B protein or did not, respectively. Of note, no significant difference was observed in terms of mRNA expression between $BCL11B^+$ and $BCL11B^-$ patients (mRNA data from array and qPCR, Supplementary Figure S5A), indicating the lack of association between *BCL11B* mRNA and protein levels. When comparing GEP according to protein expression, we identified 152 differentially expressed genes ($p < 0.05$), of which 36 and 116 were ≥ 2-fold upregulated and downregulated, respectively. Notably, $BCL11B^+$ patients were enriched for downregulated genes involved in the innate immune response (ES = 6.3; $p = 1 \times 10^{-8}$), inflammatory response (ES = 5; $p = 2.5 \times 10^{-5}$), leukocyte migration (enrichment score ES = 9.1; $p = 1.1 \times 10^{-4}$), cell adhesion (ES = 3.4; $p = 0.002$), leukotriene metabolic process (ES = 31.7; $p = 0.004$) and response to oxidative stress (ES = 5.8; $p = 0.03$, Table S8). Of note, among genes deregulated in the leukotriene pathway we identified *ALOX5* (fold change = −4.36) and *ALOX5AP* (fold change = −2), where the loss of *ALOX5* has been reported to impair leukemic stem cells and prevent the onset of chronic myeloid leukemia in mice [42].

2.7. ZEB2-BCL11B Expression Failed to Sustain Self-Renewal of Murine Hematopoietic Stem and Progenitor Cells

We assessed the leukemogenic potential of the *ZEB2-BCL11B* fusion by analyzing its ability to sustain self-renewal of murine hematopoietic progenitor cells. Bone marrow (BM) c-Kit+ cells expressing the full-length chimera were used in colony forming unit assays. In addition, cells were kept in liquid culture to monitor GFP expression and *ZEB2-BCL11B* mRNA levels over time: GFP expression increased from 2.5% at day 1–43% GFP+ cells at day 14, while mRNA levels were 30-fold and 800-fold higher than those of the negative control at day 6 and 13, respectively, highlighting low but specific expression of the chimeric transcript (Figures S5B–D). No differences in term of clonogenic capacity were detected between cells transduced with the empty vector (negative control) or the *ZEB2-BCL11B* transcript. Moreover, regarding self-renewal capacity, no colonies were detected at day 14 (second re-plating) in either the negative control or cells expressing the chimera, whereas *MLL-AF9* transduced cells (positive control) showed self-renewal capacity.

3. Discussion

Several studies have described a heterogeneous landscape of chimeras in AML [30,40,43,44], where very few fusions and genes were recurrently rearranged or altered. Here we analysed a cohort of AML patients characterized by the presence of a rare or never before reported chromosomal translocation with the aim of detecting the putative fusion gene correlated with the translocation. We identified novel and rare fusion events with an expected pathogenic role in adult AML patients.

The advantages of RNA-seq in detecting fusion events rely not only on the ability to systematically identify fusions whose partner genes are unknown, but also to detect those rearrangements that remain cryptic at cytogenetic analysis (small deletions, inversions or duplications). In the past years, several bioinformatics tools have been established for the detection

of fusion events in RNA-seq data. However, the output of these software is represented by a high number of false positive predictions. This is mainly due to systematic errors including read-through artefacts, reverse transcriptase template switching events or mapping biases. Moreover, fusions identification tools provide no information regarding the oncogenic relevance of the output fusions. These features make the systematic experimental validation of gene fusion lists obtained from in silico pipelines unfeasible. To overcome this limitation, we exploited the "downstream" tool FuGePrior to reduce the number of events to those highly reliable and with a putative biological function. FuGePrior combines results from state of the art bioinformatic tools for chimeric transcripts identification and prioritization, several filtering and processing steps designed on up-to-date literature on gene fusions and analysis of the potential functionality of the fusion according to its structure. This allowed us to conduct the experimental validation on a manageable list of candidates.

Five fusion genes associated with the known cytogenetic translocations and four fusions that remained cryptic at the level of cytogenetic analysis were closely studied. The fusions associated with balanced rearrangements were: (i) two isoforms of *ZEB2-BCL11B* and its reciprocal *BCL11B-ZEB2* chimeric transcript associated with the translocation t(2;14)(q21-q23;q32); (ii) *CNOT2-WT1* which derived from the translocation t(11;12); (iii) *CPD-PXT1* related to the t(6;17) aberration (Figures 1,2). Further cryptic fusions included *UTP6-CRLF3*, *PUF60-TYW1*, *SAV1-GYPB* and *OAZ1-MAFK* (Figures 1,2). The fusions *ZEB2-BCL11B*, *BCL11B-ZEB2* and *OAZ1-MAFK* involved genes encoding for transcription factors and we speculated that the putative mechanism of action of the fusion proteins may be linked to alterations of the transcriptional program.

We selected the chimera *ZEB2-BCL11B* for functional studies due to its frequency in acute leukemia. The remaining fusion events were not further investigated. However, we speculate on their potential activity in leukemic cells according to known features of partner genes involved in the translocations.

We associated the expression of fusion events involving genes on chromosomes 17, such as *UTP6-CRLF3* and *CPD-PXT1*, to the loss of *NF1.* The detection of these "hidden" alterations required the integration of different layers of genomic data (mutation analysis and copy number alterations), highlighting the complexity of the genomic alterations in AML and the importance of an accurate characterization of each patient´s alterations to permit a personalized medicine approach. The consequences of the out of frame fusions *CNOT2-WT1* and *PUF60-TYW1* is more difficult to speculate on but may be related to the loss of function of *WT1* (data not shown) and *PUF60*, respectively. Genomic alterations of *WT1* including point mutations and small insertions and deletions have been reported in 5% of AML cases[2,44] and the haploinsufficiency of *PUF60* has been associated with the progression of T-ALL in a mouse model with homozygous deletion of *TP53* [20]. However, functional studies are needed to elucidate *PUF60* role in AML. The fusion gene *SAV1-GYPB* may be of interest due to the role of the tumor suppressor *SAV1* [45]. SAV1 interacts with two kinases MST1 and MST2 to form an active protein complex and promotes cell-cycle exit. The ability of SAV1 to binds MST1/MST2 is limited to the functionality of its coiled-coil domain. In this scenario, the identified translocation impaired the coiled-coil domain, suggesting the loss of stability of the SAV1-MST1-MT2 complex [46].

Data from the TCGA Fusion Gene Database showed that the some of the candidate genes form chimeras with a variety of partners in different tumor types, suggesting that they might locate in genomic regions prone to chromosomal rearrangements [47,48] and/or have a role in carcinogenesis. The most frequently altered genes were *CPD* and *CNOT2*, whose overexpression was associated with survival, inhibition of apoptosis and angiogenesis in different cancer types [22,49–52]. Regarding the other genes that were rarely rearranged across cancer, they might participate to the leukemic phenotype, even though not being the driver of transformation. Our AML cohort was characterized by mutations in genes with a known pathogenic role in leukemia and the identified chimeras contributed to the disease complexity, as demonstrated by the involvement of genes such as *WT1* or copy-number loss of *NF1.*

Finally, we detected three isoforms of the rare fusion transcript *ZEB2-BCL11B* (sample #59810) and its reciprocal *BCL11B-ZEB2*. Interestingly, the fusion protein ZEB2-BCL11B was previously identified in two adult AML cases [29,40] and three paediatric T/M MPAL cases [30], suggesting a putative role in leukemogenesis. We described the characterization of five cases carrying the

t(2;14)(q22.3;q32.2) translocation involving the rearrangement of *ZEB2* and *BCL11B*. In two of the three patients with immunophentoypic characterization, leukemic cells co-expressed T-cells markers such as CD3, CD2 and CD7, and one additional case was diagnosed as AUL. Molecular profiling revealed that four out of five rearranged patients harboured *FLT3*-ITD internal tandem duplication, and two of these had an allelic fraction < 0.5 and carried a co-occurring alteration in the tyrosine kinase domain. These data suggested that *FLT3* alterations might arise as a secondary event. *In vitro* expression of the full-length *ZEB2-BCL11B* transcript in murine c-Kit$^+$ cells did not show evidence of transforming ability. This evidence suggests that as for other fusions, additional alterations are required for malignant transformation [53,54] and, based on our data, *FLT3* alterations might be the most promising candidates. The elucidation of the mechanism(s) of leukemogenesis driven by the t(2;14)(q22.3;q32.2) translocation deserves further investigation. Recent studies have shed light on the role of *ZEB2* in normal and malignant haematopoiesis [24,25], suggesting its loss of function or aberrant function may also contribute to neoplastic transformation.

Interestingly, by immunohistochemistry we showed that BLC11B is expressed in the t(2;14)(q22.3;q32.2)-rearranged leukemic blasts (patient #59810), but also in nine non-rearranged AML cases and two T/M MPAL or AUL with 14q32 rearrangement. This suggests that *BCL11B* may have a role in leukemogenesis. The comparison of gene expression profile from BCL11B$^+$ and BCL11B$^-$ patients revealed downregulation of genes involved in the innate immune response, inflammatory response, leukocyte migration and cell adhesion, leukotriene metabolic pathways and response to oxidative stress in BCL11B$^+$ AML patients. Abbas and colleagues showed that *BCL11B* overexpression in 32D myeloid cell line resulted in a decreased proliferation, less maturation toward granulocyte and more undifferentiated blast cells [41], but did not detect a transforming ability of *BCL11B*. Thus, further studies are needed to clarify the role of and interplay between the chimeric protein and co-occurring alterations in acute leukemia in an effort to identify potential therapeutic targets for these patients.

4. Materials and Methods

4.1. Patients and Samples

The study was approved by the Institutional Ethical Committee (protocol number 253/2013/O/Tess and 112/2014/U/Tess) of Sant'Orsola-Malpighi Polyclinic (Bologna, Italy) and the Internal Review Board of MLL Munich Leukemia Laboratory and was carried out in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. Samples from adult patients with primary adult AML were obtained after informed consent.

Leukocytes were enriched by separation on Ficoll density gradient and lysed in RLT buffer. Genomic DNA and RNA were extracted by column purification (AllPrep DNA/RNA/Protein Mini Kit and QIAcube, or RNeasy Mini Kit, Qiagen, Hilden, Germany).

4.2. Chromosome Banding Analysis (CBA)

CBA was performed as previously described [55]. Karyotypes were examined after GAW or GAG banding technique and described according to International System for Human Cytogenomic Nomenclature (ISCN 2016) [56].

4.3. Fluorescent In Situ Hybridization (FISH)

FISH analysis was carried out on fixed nuclei obtained using the CBA technique according to the manufacturer's instructions. Dual color breakapart FISH probes created with the BAC clones RP11-644D8 and RP11-360D1 (covering up- and down-stream regions of the *ZEB2* gene) and with RP11-1147k11 and RP11-464J3 (covering the up- and down-stream regions of the *BCL11B* gene), was used to identify *ZEB2* and *BCL11B* rearrangements, respectively. To identify the specific *ZEB2-BCL11B* fusion gene, a dual color single fusion was obtained using RP11-644D8 and RP11-464J3 clones. BAC clones were provided already marked in Spectrum Orange or Spectrum Green (Empire Genomics, New York, NY, USA). The slides were counterstained with DAPI and analysed using

fluorescent-microscopes equipped with FITC/TRITC/AQUA/DAPI filter sets and the Genikon imaging system software (Nikon Instruments, Tokyo, Japan). At least 100 nuclei were analysed for each sample.

4.4. Sequencing and Fusion Detection

Libraries for RNA-seq were prepared with the TruSeq stranded mRNA kit (Illumina, San Diego, CA, USA) following manufacturer's instructions. RNA-seq libraries were subjected to 2 × 75 bp paired-end sequencing and run on a HiSeq 2500 or 1000 instrument (Illumina), and following manufacturer's specifications. An average of 50 million reads *per* sample was obtained. Targeted DNA sequencing of myeloid-related genes was performed using the TruSight Myeloid Sequencing Panel (Illumina) and run on a MiSeq instrument (Illumina). Variants with a total read depth > 500 and falling into exonic regions and splice sites were retained. Targeted sequencing of *ZEB2-BCL11B* rearranged patients was performed as previously described [40].

Fusion genes were detected on RNA-seq data by applying FuGePrior pipeline to the gene fusion lists provided by ChimeraScan [57] and deFuse [58] tools. According to FuGePrior workflow [59], fusions with the following features were removed: (i) not supported by split reads (i.e., reads harboring the fusion breakpoint); (ii) involving at least one unannotated partner gene; (iii) shared by healthy samples; (iv) characterized by a non-reliable structure; (v) having at least the driver score probability lower than 0.7. The DS score was a measure of the probability of the fusion being an oncogenic event, according to Pegasus [60] and Oncofuse [61].

Firstly, we screened the putative fusions list to identify chimeras originating from chromosomal translocations detected by the cytogenetic analysis (tier 1). Secondly, to identify cryptic fusions and to reduce the number of false-positive predictions, we implemented additional filters to remove: (i) recurrently fused genes showing a large diversity among partner genes (including *HBB*, *HBA*, *HBD*, *MPO*, *DLG2*) [62]; (ii) conjoined genes; (iii) fusions recurring in more than one sample in our cohort. We added the latter criteria as we assumed it was not likely to found a recurrent fusion in such a small and heterogeneous cohort. Then, in order to identify cryptic but relevant fusions, we prioritized chimeras according to the probability of the transcript being an oncogenic event (tier 2). Finally, we rescued out-of-frame fusions (DS < 0.7) involving tumor suppressor genes (tier 3) to identify loss of function alterations in key genes. The recurrent gene fusion *CBFB-MYH11* was identified in the positive control (sample #84), thus confirming the reliability of our bioinformatic analysis. The dataset supporting the conclusions of this article is available in the NGS-PTL repository, at the following link: https://ngs-ptl.unibo.it:5006.

For expression analysis, raw data were aligned to the reference genome and read counts were normalized using the DESeq2 package and the rlog transformation for data normalization [63]. Differentially expressed genes, median absolute deviation calculations, unsupervised clustering and expression plots were performed using R packages limma [64], DescTools, ComplexHeatmap [65] and ggplot2, respectively. Enrichment pathway analysis was performed with Enrichr [66].

4.5. RT-PCR, PCR, qPCR and Sanger Sequencing

cDNA synthesis was performed using M-MLV Reverse Transcriptase for primary AML samples and Random Hexamers (Invitrogen, Thermo Fisher, Waltham, MA, USA) or the SuperScript III First-Strand Synthesis System (Invitrogen) for RNA extracted from transduced c-Kit$^+$ cells. Polymerase chain reaction (PCR) primers were designed to amplify fragments containing the fusion boundary detected by RNA-seq using Primer3 (http://primer3.ut.ee/, Table S1). Quantitative PCR (qPCR) was performed using Brilliant III Ultra-Fast QPCR Master Mix (Agilent Technologies, Santa Clara, CA, USA) on an Mx3000p qPCR system (Agilent Technologies) and standard cycling set-up (Table S1). TaqMan gene expression for *BCL11B* mRNA (Hs01102259_m1) was performed on BM cells from AML patients (blasts ≥ 80%, n = 10) and peripheral blood mononuclear cells from healthy controls (n = 3), using *GAPDH* (Hs02786624_g1) as reference gene, on the Applied Biosystems 7500 Real-Time PCR System (Thermo Fischer Scientific). Gene expression was quantified by the $2^{-\Delta\Delta Ct}$ method, using the average of healthy controls as reference sample. Long-distance PCR were performed with LA Taq DNA Polymerase (Takara Bio, Shiga, Japan) following manufacturer instructions for human genomic

DNA. Fast Start Taq DNA Polymerase (Roche, Basel, Switzerland) was used for standard PCR reactions. Products were purified with the QIAquick PCR purification kit (Qiagen) or conventional agarose gel electrophoresis and extraction of specific bands with the QIAquick Gel Extraction kit (Qiagen). PCR products were sequenced by Sanger Sequencing using an ABI PRISM 3730 automated DNA sequencer (Applied Biosystems) and the Big Dye Terminator DNA sequencing kit (Applied Biosystems, Foster City, CA, USA). Fusion detection was performed using NCBI Blast alignment and BLAT software tool (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) to reference genome GRCh37/hg19.

BCR and TCR clonality assay was performed as described by the BIOMED-2 study[67].

4.6. Immunohistochemistry

BM specimens were fixed in B5 solution for 2 hours, decalcified with EDTA-based solution for 3 hours and paraffin embedded. Histological stainings were examined (Hematoxylin&Eosin, Giemsa, Gomori silver impregnation) and 3 μm-thick sections were cut for immunohistochemistry. The antigen retrieval methods used were heat-based Pt-Link (Agilent Technologies, PT100/PT101) and EnVision Flex Target Retrieval Solution High pH (Agilent Technologies, K8004) at 92 °C or 82 °C. All samples were stained for the following molecules: CD34 (mouse monoclonal, clone END, NCL-L-END,1:100, Microsystems, Newcastle, UK), myeloperoxidase (rabbit polyclonal, A0398, 1:5000, Agilent Technologies), CD68 (mouse monoclonal, clone PGM1, 1:5, kindly provided by Prof. Falini, Perugia, Italy), BCL11B (rabbit polyclonal, NB100-2600, 1:200, Novus Biologicals Centennial, CO, USA). The BCL11B antibody was validated on reactive bone marrow and nodal follicular hyperplasia. The staining panels on the AML cases were performed using positive (the same sample for validation) and negative controls (slides with exclusion of the primary antibody). The analysis of CD34 and CD68 antibodies were performed according to long standing previously settled procedures.

The reaction detection was performed by using the Dako Real Detection Systems Alkaline Phosphatase/RED Rabbit/Mouse Kit (K 5005, Agilent Technologies). Overall, 24 BM biopsies were analysed. One BM biopsy referred to case #59810 with t(2;14), 21 BM biopsies referred to 21 AML patients without t(2;1) and/or 14q32 rearrangement, 2 BM biopsies referred to patients with 14q32 rearrangements (Table S6).

4.7. Gene Expression Profiling (GEP) and SNP-Array

We analysed gene expression and copy number data from a previously obtained internal cohort [55]. Gene expression raw data were processed by Expression Console software with Signal Space Transformation Robust Multi-Array average (sst-RMA) normalization. Supervised data analysis was carried out with Transcriptome Analysis Console v4.0 software (Affymetrix, Thermo Fisher). Functional annotation clustering and enrichment analysis was performed using David Bioinformatics Resources 6.8 (National Institute of Allergy and Infectious Diseases, NIH) [68]. CEL files from SNP-array raw intensities were processed using Rawcopy[69].

4.8. Retroviral Transduction Assays

The TY1-tagged full length transcripts *ZEB2-BCL11B* was subcloned into a retroviral vector using EcoRI restriction sites. The resulting plasmid's sequence was verified by Sanger sequencing. Murine stem cell virus–based (MSCV-based) retroviral constructs carrying the tagged *ZEB2-BCL11B* sequence upstream of an internal ribosomal entry site–green fluorescent protein (IRES-GFP) cassette were generated using 293T packaging cell line. Vectors containing the fusion gene (*ZEB2-BCL11B*), the *MLL-AF9* fusion (acting as positive control) or the empty vector (negative control) were used to transduce mouse c-Kit$^+$ BM cells. Mouse whole BM was positively selected with the CD117 (c-Kit) MicroBeads and the LS MACS column according manufacturer´s instructions (Miltenyi Biotec, Bergisch Gladbach, Germany). Retroviral transduction was performed as previously described[70].

4.9. Serial Colony Replating Assay

Colony forming unit assay was performed in duplicates by seeding 1000 c-Kit$^+$ transduced cells in Methocult M3434 methylcellulose medium (StemCell Technologies, Vancouver, BC, Canada). Cells were plated in duplicate and after 7–12 days colonies were scored, pooled and identical numbers of cells were re-plated under the same conditions.

4.10. Flow Cytometry Analysis

Multiparameter flow cytometry (MFC) and sample processing was carried out as described previously[71]. MFC analyses were performed using FC500 or Navios flow cytometers (Beckman Coulter, Miami, FL, USA). List mode files were analyzed using CXP Software version 2.0 and Kaluza version 1.0 (Beckman Coulter, Brea, CA, USA). Diagnoses were assigned according to EGIL and WHO classifications [10,72]. Single cell suspensions of transduced c-Kit$^+$ cells were prepared as described elsewehere [15]. Dead cells were excluded by gating on 7AAD (Miltenyi Biotec)-negative cells. Flow cytometry analysis were performed on an LSR Fortessa cell analyser (BD Biosciences, San Jose, CA, USA) and data were analysed with FlowJo software v 10 (BD, Franklin Lakes, NJ, USA).

4.11. Immunoblotting

Whole-cell lysates were prepared from 10^7 cells in 6× Laemmli buffer. Lysates were run on SDS–PAGE gels and transferred to PVDF membranes (Millipore). Membranes were probed with the anti-Gapdh (Abcam, Cambridge, UK), anti-TY1 (Thermo Fisher Scientific) and anti-BCL11B (Abcam) primary antibodies at 1:10000, 1:2000 and 1:10000 dilutions, respectively. Membranes were probed with secondary antibodies conjugated to IRDye 680RD or IRDye 800 CW (LI-COR Biosciences Ltd. Lincoln, NE, USA) at 1:10000 dilution and proteins were detected using the Odyssey Infrared Imaging System (LI-COR Biosciences Ltd). Restore Western Blot Stripping Buffer (Thermo Fisher Scientific) were used to remove primary and secondary antibodies from PVDF membrane in order to reprobe with the anti-BCL11B antibody.

5. Conclusions

Fusion genes are frequently detected in cancer and they are often the result of chromosomal rearrangements such as translocations, inversions and deletions, all of which may involve a single chromosome or different chromosomes. Here we reported the identification of novel gene fusion events in AML. Although the pathogenic role and functional properties of these alterations will require additional functional studies, here we demonstrated that *ZEB2-BCL11B* rearrangement is recurrent and associated with distinct immune-clinico characteristics.

Supplementary Materials: Table S1. Primers used for validation by RT-PCR and Sanger sequencing. Table S2. Mutational status of myeloid-related genes screened by NGS. Table S3. List of 19 fusions selected for validation with RT-PCR and Sanger Sequencing and their relative annotation, driver scores (according to Pegasus and Oncofuse), tier and cDNA breakpoints for validated chimera. Table S4. Enrichment pathway analysis of differentially expressed genes among groups identified by unsupervised clustering. Table S5. List of annotated cases in the TCGA Fusion Portal and their expression level as reported in the TCGA AML cohort in the cBio data portal. Table S6. List of annotated cases in the Mitelman database affected by haematological malignancies and characterized by the presence of a translocation between chromosome 2q21-23 and the 14q32 region. Figure S1. Electropherogram of fusion junctions. Figure S2. Genomic localization of copy number loss linked to the *CPD-PXT1* fusion. Figure S3. Expression analysis. Figure S4. Characterization of 14q32 genomic breakpoint A. Sequence and chromatogram of the genomic breakpoint. Figure S5. Expression of *BCL11B* mRNA and *ZEB2-BCL11B* in transduced cells.

Author Contributions: Conceptualization, A.P., I.I., B.J.P.H., E.F. and G.M.; Data curation, G.S., G.P., V.G., R.D.T., C.P., M.C.F. and S.B.; Formal analysis, G.P. and C.T.S.; Funding acquisition, G.S., I.I. and G.M.; Investigation, A.P., G.S., G.G., C.B., S.R., M.G., A.S., V.R., E.F. and A.F.; Methodology, A.P., G.S., G.P., G.G. and E.F.; Project administration, I.I. and G.M.; Resources, A.S., V.G., R.D.T., C.P., M.C.F., S.B., E.O., S.S., C.H., E.S., N.T., B.J.P.H., E.F. and G.M.; Software, G.P. and E.F.; Supervision, C.T.S., I.I., B.J.P.H., E.F. and G.M.; Validation, A.P., G.G., M.G., A.S. and C.H. Visualization, A.P., C.B., S.R. and A.G.L.Di.R.; Writing – original draft, A.P., G.S.,

C.B., S.R., E.S., E.F. and G.M.; Writing – review & editing, G.P., G.G., M.G., A.S., V.G., R.D.T., C.P., V.R., E.F., A.G.L.D.R, A.F., M.C.F., S.B., E.O., S.S., C.T.S., C.H., N.T., I.I. and B.J. P.H.

Funding: The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n° 306242-NGS-PTL. Funding for this project was provided in part by an EHA Research Fellowship award granted by the European Hematology Association (to Giorgia Simonetti) and by Associazione Italiana per la Ricerca sul Cancro, AIRC-IG n.19226 to Giovanni Martinelli.

Acknowledgments: We thank the Next Generation Sequencing Platform for Targeted Personalized Therapy of Leukemia consortium and Marco Sazzini for discussion and funding acquisition.

Conflicts of Interest: GM has competing interests with Incyte, Celgene, Pfizer, Daiichi Sankyo. AS is employed by MLL Munich Leukemia Laboratory. CH has equity ownership of MLL Munich Leukemia Laboratory.

References

1. Döhner, H.; Estey, E.; Grimwade, D.; Amadori, S.; Appelbaum, F.R.; Büchner, T.; Dombret, H.; Ebert, B.L.; Fenaux, P.; Larson, R.A.; et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **2017**, *129*, 424–447.
2. Papaemmanuil, E.; Gerstung, M.; Bullinger, L.; Gaidzik, V.I.; Paschka, P.; Roberts, N.D.; Potter, N.E.; Heuser, M.; Thol, F.; Bolli, N.; et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N. Engl. J. Med.* **2016**, *374*, 2209–2221.
3. Martens, J.H.A.; Stunnenberg, H.G. The molecular signature of oncofusion proteins in acute myeloid leukemia. *FEBS Lett.* **2010**, *584*, 2662–2669.
4. Soverini, S.; De Benedittis, C.; Mancini, M.; Martinelli, G. Best Practices in Chronic Myeloid Leukemia Monitoring and Management. *Oncologist* **2016**, *21*, 626–33.
5. Ley, T.J.; Miller, C.; Ding, L.; Raphael, B.J.; Mungall, A.J.; Robertson, G.; Hoadley, K.; Triche, T.J.; Laird, P.W.; Baty, J.D.; et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **2013**, *368*, 2059–2074.
6. Wen, H.; Li, Y.; Malek, S.N.; Kim, Y.C.; Xu, J.; Chen, P.; Xiao, F.; Huang, X.; Zhou, X.; Xuan, Z.; et al. New Fusion Transcripts Identified in Normal Karyotype Acute Myeloid Leukemia. *PLoS One* **2012**, *7*, e51203.
7. Gough, S.M.; Lee, F.; Yang, F.; Walker, R.L.; Zhu, Y.J.; Pineda, M.; Onozawa, M.; Chung, Y.J.; Bilke, S.; Wagner, E.K.; et al. NUP98-PHF23 Is a Chromatin-Modifying Oncoprotein That Causes a Wide Array of Leukemias Sensitive to Inhibition of PHD Histone Reader Function. *Cancer Discov.* **2014**, *4*, 564–577.
8. Togni, M.; Masetti, R.; Pigazzi, M.; Astolfi, A.; Zama, D.; Indio, V.; Serravalle, S.; Manara, E.; Bisio, V.; Rizzari, C.; et al. Identification of the NUP98-PHF23 fusion gene in pediatric cytogenetically normal acute myeloid leukemia by whole-transcriptome sequencing. *J. Hematol. Oncol.* **2015**, *8*, 69, doi: 10.1186/s13045-015-0167-8.
9. Iacobucci, I.; Mullighan, C.G. Genetic Basis of Acute Lymphoblastic Leukemia. *J. Clin. Oncol.* **2017**, *35*, 975.
10. Swerdlow, S.H.; World Health Organization; International Agency for Research on Cancer *WHO classification of tumours of haematopoietic and lymphoid tissues*; ISBN 9789283244943.
11. Riley, D.A.; Tan, F.; Miletich, D.J.; Skidgel, R.A. Chromosomal Localization of the Genes for Human Carboxypeptidase D (CPD) and the Active 50-Kilodalton Subunit of Human Carboxypeptidase N (CPN1). *Genomics* **1998**, *50*, 105–108.
12. Matsuura, K.; Nakada, C.; Mashio, M.; Narimatsu, T.; Yoshimoto, T.; Tanigawa, M.; Tsukamoto, Y.; Hijiya, N.; Takeuchi, I.; Nomura, T.; et al. Downregulation of SAV1 plays a role in pathogenesis of high-grade clear cell renal cell carcinoma. *BMC Cancer* **2011**, *11*, 523, doi: 10.1186/1471-2407-11-523..
13. Kudo, S.; Fukuda, M. Structural organization of glycophorin A and B genes: glycophorin B gene evolved by homologous recombination at Alu repeat sequences. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 4619–23.
14. Wu, H.-Y.; Chen, S.-F.; Hsieh, J.-Y.; Chou, F.; Wang, Y.-H.; Lin, W.-T.; Lee, P.-Y.; Yu, Y.-J.; Lin, L.-Y.; Lin, T.-S.; et al. Structural basis of antizyme-mediated regulation of polyamine homeostasis. *Proc. Natl. Acad. Sci.* **2015**, *112*, 11229–11234.
15. Katsuoka, F.; Yamamoto, M. Small Maf proteins (MafF, MafG, MafK): History, structure and function. *Gene* **2016**, *586*, 197–205.

16. Bonnart, C.; Gérus, M.; Hoareau-Aveilla, C.; Kiss, T.; Caizergues-Ferrer, M.; Henry, Y.; Henras, A.K. Mammalian HCA66 protein is required for both ribosome synthesis and centriole duplication. *Nucleic Acids Res.* **2012**, *40*, 6270–89.
17. Yang, F.; Xu, Y.-P.; Li, J.; Duan, S.-S.; Fu, Y.-J.; Zhang, Y.; Zhao, Y.; Qiao, W.-T.; Chen, Q.-M.; Geng, Y.-Q.; et al. Cloning and characterization of a novel intracellular protein p48.2 that negatively regulates cell cycle progression. *Int. J. Biochem. Cell Biol.* **2009**, *41*, 2240–2250.
18. Hastings, M.L.; Allemand, E.; Duelli, D.M.; Myers, M.P.; Krainer, A.R. Control of Pre-mRNA Splicing by the General Splicing Factors PUF60 and U2AF65. *PLoS One* **2007**, *2*, e538, doi: 10.1371/journal.pone.0000538.
19. Waas, W.F.; de Crécy-Lagard, V.; Schimmel, P. Discovery of a gene family critical to wyosine base formation in a subset of phenylalanine-specific transfer RNAs. *J. Biol. Chem.* **2005**, *280*, 37616–22.
20. Matsushita, K.; Kitamura, K.; Rahmutulla, B.; Tanaka, N.; Ishige, T.; Satoh, M.; Hoshino, T.; Miyagi, S.; Mori, T.; Itoga, S.; et al. Haploinsufficiency of the c-myc transcriptional repressor FIR as a dominant negative-alternative splicing model, promoted p53-dependent T-cell acute lymphoblastic leukemia progression by activating Notch1. *Oncotarget* **2015**, *6*, 5102–5117.
21. Jayne, S.; Zwartjes, C.G.M.; Van Schaik, F.M.A.; Timmers, H.T.M. Involvement of the SMRT/NCoR–HDAC3 complex in transcriptional repression by the CNOT2 subunit of the human Ccr4–Not complex. *Biochem. J.* **2006**, *398*, 461–467.
22. Ito, K.; Inoue, T.; Yokoyama, K.; Morita, M.; Suzuki, T.; Yamamoto, T. CNOT2 depletion disrupts and inhibits the CCR4-NOT deadenylase complex and induces apoptotic cell death. *Genes to Cells* **2011**, *16*, 368–379.
23. Zwartjes, C.G.M.; Jayne, S.; van den Berg, D.L.C.; Timmers, H.T.M. Repression of Promoter Activity by CNOT2, a Subunit of the Transcription Regulatory Ccr4-Not Complex. *J. Biol. Chem.* **2004**, *279*, 10848–10854.
24. Li, H.; Mar, B.G.; Zhang, H.; Puram, R. V.; Vazquez, F.; Weir, B.A.; Hahn, W.C.; Ebert, B.; Pellman, D. The EMT regulator ZEB2 is a novel dependency of human and murine acute myeloid leukemia. *Blood* **2017**, *129*.
25. Li, J.; Riedt, T.; Goossens, S.; Carrillo García, C.; Szczepanski, S.; Brandes, M.; Pieters, T.; Dobrosch, L.; Gütgemann, I.; Farla, N.; et al. The EMT transcription factor Zeb2 controls adult murine hematopoietic differentiation by regulating cytokine signaling. *Blood* **2017**, *129*.
26. Ha, V.L.; Luong, A.; Li, F.; Casero, D.; Malvar, J.; Kim, Y.M.; Bhatia, R.; Crooks, G.M.; Parekh, C. The T-ALL related gene BCL11B regulates the initial stages of human T-cell differentiation. *Leukemia* **2017**.
27. Hu, X.; Wang, Q.; Tang, M.; Barthel, F.; Amin, S.; Yoshihara, K.; Lang, F.M.; Martinez-Ledesma, E.; Lee, S.H.; Zheng, S.; et al. TumorFusions: An integrative resource for cancer-associated transcript fusions. *Nucleic Acids Res.* **2018**.
28. Yang, F.; Xu, Y.-P.; Li, J.; Duan, S.-S.; Fu, Y.-J.; Zhang, Y.; Zhao, Y.; Qiao, W.-T.; Chen, Q.-M.; Geng, Y.-Q.; et al. Cloning and characterization of a novel intracellular protein p48.2 that negatively regulates cell cycle progression. *Int. J. Biochem. Cell Biol.* **2009**, *41*, 2240–50.
29. Torkildsen, S.; Gorunova, L.; Beiske, K.; Tjønnfjord, G.E.; Heim, S.; Panagopoulos, I. Novel ZEB2-BCL11B Fusion Gene Identified by RNA-Sequencing in Acute Myeloid Leukemia with t(2;14)(q22;q32). *PLoS One* **2015**, *10*, e0132736.
30. Alexander, T.B.; Gu, Z.; Iacobucci, I.; Dickerson, K.; Choi, J.K.; Xu, B.; Payne-Turner, D.; Yoshihara, H.; Loh, M.L.; Horan, J.; et al. The genetic basis and cell of origin of mixed phenotype acute leukaemia. *Nature* **2018**, *562*, 373–379.
31. Mitelman F, J.B. and M.F. Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer Available online: https://cgap.nci.nih.gov/Chromosomes/Mitelman.
32. Palka, G.; Calabrese, G.; Fioritoni, G.; Stuppia, L.; Guanciali Franchi, P.; Marino, M.; Antonucci, A.; Spadano, A.; Torlontano, G. Cytogenetic survey of 80 patients with acute nonlymphocytic leukemia. *Cancer Genet. Cytogenet.* **1992**, *59*, 45–50.
33. Gmidène, A.; Sennana, H.; Wahchi, I.; Youssef, Y. Ben; Jeddi, R.; Elloumi, M.; Saad, A. Cytogenetic profile of a large cohort of Tunisian de novo acute myeloid leukemia. *Hematology* **2012**, *17*, 9–14.
34. Columbano-Green, L.M.; Romain, D.R.; Carter, J.; Crossen, P.E. t(2;14)(q23;q32.3) as the sole abnormality in a patient with acute nonlymphocytic leukemia (FAB-M4). *Cancer Genet. Cytogenet.* **1990**, *48*, 255–7.
35. Huang, X.; Du, X.; Li, Y. The role of BCL11B in hematological malignancy. *Exp. Hematol. Oncol.* **2012**, *1*, 22, doi: 10.1186/2162-3619-1-22.

36. Rubnitz, J.E.; Onciu, M.; Pounds, S.; Shurtleff, S.; Cao, X.; Raimondi, S.C.; Behm, F.G.; Campana, D.; Razzouk, B.I.; Ribeiro, R.C.; et al. Acute mixed lineage leukemia in children: The experience of St Jude Children's Research Hospital. *Blood* **2009**, *113*, 5083–5089.
37. Gu, Z.; Churchman, M.; Roberts, K.; Li, Y.; Liu, Y.; Harvey, R.C.; McCastlain, K.; Reshmi, S.C.; Payne-Turner, D.; Iacobucci, I.; et al. Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia. *Nat. Commun.* **2016**, *7*, 13331.
38. Aventín, A.; Sánchez, J.; Nomdedéu, J.F.; Estany, C.; Forcada, P.; La Starza, R.; Mecucci, C. Novel IGHα translocations, t(2;14)(q14.3;q32) and t(14;17)(q32;q21), in B-cell precursor acute lymphoblastic leukemia. *Cancer Genet. Cytogenet.* **2008**, *185*, 57–59.
39. Inaba, T.; Oku, N.; Gotoh, H.; Murakami, S.; Oku, N.; Itoh, K.; Ura, Y.; Nakanishi, S.; Shimazaki, C.; Nakagawa, M. Philadelphia chromosome positive precursor B-cell acute lymphoblastic leukemia with a translocation t(2;14)(p13;q32). *Leukemia* **1991**, *5*, 719–22.
40. Stengel, A.; Nadarajah, N.; Haferlach, T.; Dicker, F.; Kern, W.; Meggendorfer, M.; Haferlach, C. Detection of recurrent and of novel fusion transcripts in myeloid malignancies by targeted RNA sequencing. *Leukemia* **2018**, *32*, 1229–1238.
41. Abbas, S.; Sanders, M.A.; Zeilemaker, A.; Geertsma-Kleinekoort, W.M.C.; Koenders, J.E.; Kavelaars, F.G.; Abbas, Z.G.; Mahamoud, S.; Chu, I.W.T.; Hoogenboezem, R.; et al. Integrated genome-wide genotyping and gene expression profiling reveals BCL11B as a putative oncogene in acute myeloid leukemia with 14q32 aberrations. *Haematologica* **2014**, *99*, 848–57.
42. Chen, Y.; Hu, Y.; Zhang, H.; Peng, C.; Li, S. Loss of the Alox5 gene impairs leukemia stem cells and prevents chronic myeloid leukemia. *Nat. Genet.* **2009**, *41*, 783–92.
43. Iacobucci, I.; Wen, J.; Meggendorfer, M.; Choi, J.K.; Shi, L.; Pounds, S.B.; Carmichael, C.L.; Masih, K.E.; Morris, S.M.; Lindsley, R.C.; et al. Genomic subtyping and therapeutic targeting of acute erythroleukemia. *Nat. Genet.* **2019**, *51*, 694–704.
44. Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, Hoadley K, Triche TJ Jr, Laird PW, Baty JD, Fulton LL, Fulton R, Heath SE, Kalicki-Veizer J, Kandoth C, Klco JM, Koboldt DC, Kanchi KL, Kulkarni S, L. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* **2013**, *368*, 2059–2074.
45. Mardin, B.R.; Lange, C.; Baxter, J.E.; Hardy, T.; Scholz, S.R.; Fry, A.M.; Schiebel, E. Components of the Hippo pathway cooperate with Nek2 kinase to regulate centrosome disjunction. *Nat. Cell Biol.* **2010**, *12*, 1166–1176.
46. Callus, B.A.; Verhagen, A.M.; Vaux, D.L. Association of mammalian sterile twenty kinases, Mst1 and Mst2, with hSalvador via C-terminal coiled-coil domains, leads to its stabilization and phosphorylation. *FEBS J.* **2006**, *273*, 4264–4276.
47. Lin, C.; Yang, L.; Rosenfeld, M.G. Molecular Logic Underlying Chromosomal Translocations, Random or Non-Random? *Adv. Cancer Res.* **2012**; *113*;241–279.
48. Shugay, M.; Ortiz de Mendíbil, I.; Vizmanos, J.L.; Novo, F.J. Genomic Hallmarks of Genes Involved in Chromosomal Translocations in Hematological Cancer. *PLoS Comput. Biol.* **2012**, *8*, e1002797.
49. Abdelmagid, S.A.; Too, C.K.L. Prolactin and estrogen up-regulate carboxypeptidase-D to promote nitric oxide production and survival of MCF-7 breast cancer cells. *Endocrinology* **2008,** *149*, 4821–4828.
50. Thomas, L.N.; Merrimen, J.; Bell, D.G.; Rendon, R.; Goffin, V.; Too, C.K.L. Carboxypeptidase-D is elevated in prostate cancer and its anti-apoptotic activity is abolished by combined androgen and prolactin receptor targeting. *Prostate* **2014,** *74*, 732–742.
51. Jin, T.; Fu, J.; Feng, X.J.; Wang, S.M.; Huang, X.; Zhu, M.H.; Zhang, S.H. SiRNA-targeted carboxypeptidase D inhibits hepatocellular carcinoma growth. *Cell Biol. Int.* **2013,** *37*, 929–939.
52. Sohn, E.J.; Jung, D.B.; Lee, H.J.; Han, I.; Lee, J.; Lee, H.; Kim, S.H. CNOT2 promotes proliferation and angiogenesis via VEGF signaling in MDA-MB-231 breast cancer cells. *Cancer Lett.* **2018,** *412*, 88–98.
53. Fenske, T.S.; Pengue, G.; Mathews, V.; Hanson, P.T.; Hamm, S.E.; Riaz, N.; Graubert, T.A. Stem cell expression of the AML1/ETO fusion protein induces a myeloproliferative disorder in mice. *Proc. Natl. Acad. Sci.* **2004**, *101*, 15184–15189.
54. Schessl, C.; Rawat, V.P.S.; Cusan, M.; Deshpande, A.; Kohl, T.M.; Rosten, P.M.; Spiekermann, K.; Humphries, R.K.; Schnittger, S.; Kern, W.; et al. The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice. *J. Clin. Invest.* **2005**, *115*, 2159–68.

55. Simonetti, G.; Padella, A.; do Valle, I.F.; Fontana, M.C.; Fonzi, E.; Bruno, S.; Baldazzi, C.; Guadagnuolo, V.; Manfrini, M.; Ferrari, A.; et al. Aneuploid acute myeloid leukemia exhibits a signature of genomic alterations in the cell cycle and protein degradation machinery. *Cancer* **2018**.
56. International Standing Committee on Human Cytogenomic Nomenclature; McGowan-Jordan, J.; Simons, A.; Schmid, M. (Michael) *ISCN : an international system for human cytogenomic nomenclature (2016)*; ISBN 9783318058574.
57. Iyer, M.K.; Chinnaiyan, A.M.; Maher, C.A. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics* **2011**, *27*, 2903–2904.
58. McPherson, A.; Hormozdiari, F.; Zayed, A.; Giuliany, R.; Ha, G.; Sun, M.G.F.; Griffith, M.; Heravi Moussavi, A.; Senz, J.; Melnyk, N.; et al. deFuse: An Algorithm for Gene Fusion Discovery in Tumor RNA-Seq Data. *PLoS Comput. Biol.* **2011**, *7*, e1001138.
59. Paciello, G.; Ficarra, E. FuGePrior: A novel gene fusion prioritization algorithm based on accurate fusion structure analysis in cancer RNA-seq samples. *BMC Bioinformatics* **2017**, *18*, 58.
60. Abate, F.; Zairis, S.; Ficarra, E.; Acquaviva, A.; Wiggins, C.H.; Frattini, V.; Lasorella, A.; Iavarone, A.; Inghirami, G.; Rabadan, R. Pegasus: a comprehensive annotation and prediction tool for detection of driver gene fusions in cancer. *BMC Syst. Biol.* **2014**, *8*, 97.
61. Shugay, M.; De Mend??bil, I.O.; Vizmanos, J.L.; Novo, F.J. Oncofuse: A computational framework for the prediction of the oncogenic potential of gene fusions. *Bioinformatics* **2013**, *29*, 2539–2546.
62. Yoshihara, K.; Wang, Q.; Torres-Garcia, W.; Zheng, S.; Vegesna, R.; Kim, H.; Verhaak, R.G.W. The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* **2015**, *34*, 4845–4854.
63. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**.
64. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015,** *43*, doi: 10.1093/nar/gkv007. Epub 2015 Jan 20.
65. Gu, Z.; Eils, R.; Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **2016**, *32*, 2847–2849.
66. Kuleshov, M. V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A.; et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **2016,** *8:44*, doi: 10.1093/nar/gkw377..
67. van Dongen, J.J.M.; Langerak, A.W.; Brüggemann, M.; Evans, P.A.S.; Hummel, M.; Lavender, F.L.; Delabesse, E.; Davi, F.; Schuuring, E.; García-Sanz, R.; et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 concerted action BMH4-CT98-3936. *Leukemia* **2003**, *17*, 2257–2317.
68. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57.
69. Mayrhofer, M.; Viklund, B.; Isaksson, A. Rawcopy: Improved copy number analysis with Affymetrix arrays. *Sci. Rep.* **2016**, *6*, 36158, doi: 10.1038/srep36158.
70. Giotopoulos, G.; van der Weyden, L.; Osaki, H.; Rust, A.G.; Gallipoli, P.; Meduri, E.; Horton, S.J.; Chan, W.-I.; Foster, D.; Prinjha, R.K.; et al. A novel mouse model identifies cooperating mutations and therapeutic targets critical for chronic myeloid leukemia progression. *J. Exp. Med.* **2015**, *212*, 1551–1569.
71. Kern, W.; Voskova, D.; Schoch, C.; Hiddemann, W.; Schnittger, S.; Haferlach, T.; Fonatsch, C.; Haase, D.; Schoch, C.; Hossfeld, D.; et al. Determination of relapse risk based on assessment of minimal residual disease during complete remission by multiparameter flow cytometry in unselected patients with acute myeloid leukemia. *Blood* **2004**, *104*, 3078–85.
72. Bene, M.C.; Castoldi, G.; Knapp, W.; Ludwig, W.D.; Matutes, E.; Orfao, A.; van't Veer, M.B. Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL). *Leukemia* **1995**, *9*, 1783–6.