Identification of genes that synergize with Cbfb-MYH11 in the pathogenesis of acute myeloid leukemia


*Program in Gene Function and Expression and ‡Laboratory of Cellular Oncology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

Communicated by Janet D. Rowley, University of Chicago Medical Center, Chicago, IL, February 17, 2004 (received for review December 2, 2003)

Acute myeloid leukemia subtype M4 with eosinophilia is associated with a chromosome 16 inversion that creates a fusion gene Cbfb-MYH11. We have previously shown that Cbfb-MYH11 is necessary but not sufficient for leukemogenesis. Here, we report the identification of genes that specifically cooperate with Cbfb-MYH11 in leukemogenesis. Neonatal injection of Cbfb-MYH11 knock-in chimeric mice with retrovirus 4070A led to the development of acute myeloid leukemia in 2–5 months. Each leukemia sample contained one or a few viral insertions, suggesting that alteration of one gene could be sufficient to synergize with Cbfb-MYH11. The chromosomal position of 67 independent retroviral insertion sites (RISs) was determined, and 90% of the RISs mapped within 10 kb of a flanking gene. In total, 54 candidate genes were identified; six of them were common insertion sites (CISs). CIS genes included members of a zinc finger transcription factors family, Plag1 and Plag2, with eight and two independent insertions, respectively. CIS genes also included Runx2, Myb, H2-T24, and D6Mm5e. Comparison of the remaining 48 genes with single insertion sites with known leukemia-associated RISs indicated that 18 coincide with known RISs. To our knowledge, this retroviral genetic screen is the first to identify genes that cooperate with a fusion gene important for human myeloid leukemia.

Materials and Methods

Retroviral Infection. Molecularly cloned amphotropic virus 4070A (19, 20) was propagated in NIH 3T3 cells in DMEM containing 10% FCS. The titer of the infectious virus in stocks from the medium of infected cells was determined in an assay by using PG-4 cat cell line (21). Briefly, PG-4 cells were infected with serial dilutions of virus in the presence of 8 μg/ml polybrene. After an overnight incubation, the virus-containing medium was removed and the cells were maintained in fresh medium for 5–6 days. After fixation in methanol and staining with carbol fuchsin, the paws, and dehydration. At the first signs of illness, peripheral blood was analyzed for cell number and the presence of imma-

Abbreviations: AML, acute myeloid leukemia; RIS, retroviral insertion site; CIS, common insertion site; SIS, single insertion site; SMMHC, smooth muscle myosin heavy chain; CBF, core binding factor; RIM, retroviral insertional mutagenesis; IPCKR, inverse PCR; MLV, murine leukemia virus; Cri, Cbfb-MYH11 retrovirus insertion.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. CL001593–CL001659).

AC...
tured cells. Peripheral blood was collected from the retroorbital sinus for smears, and fluorescence-activated cell sorting analysis was performed by using cell-surface markers Gr-1, CD11b, B220, CD3, Ter119, and CD117.

Leukemic cells were harvested from bone marrow or spleen of affected mice in RPMI medium 1640 (Life Technologies) and 10% FBS, and single cell suspension aliquots of $1 \times 10^6$ cells were injected intravenously in 4- to 6-week-old F1/129SvEv-C57BL6 recipients.

**Inverse PCR (iPCR).** One microgram of spleen DNA from leukemic mice was digested for 12–16 h with EcoRI or BamHI in a 40-μl final volume. DNA fragments were ligated in a 600-μl volume by using 4,000 units of T4-Ligase (M0202M, New England Biolabs) at 16°C for 12 h, and were resuspended in 20 μl of water after precipitation. Two sequential PCRs were performed by using nested primers specific for 4070A retrovirus. The PCR mix included 2 μl of template DNA, 400 μM dNTP, 0.4 μM of each primer, and 2.5 units of Taq enzyme (Expand Long Template PCR, Roche Applied Sciences, Indianapolis), in a 50-μl final volume. Primers for the first PCR included LTR3 (CCGAACTTCCATTCTCAGGTC) and gagF1 (GATCTGATTGAGACACAGA) for BamHI 5’ LTR amplification, LTR1 (CTCTTGGTCTGATCGGACT) and 5env1 (GTCAGGGTGTCCTCTATTCTCAGTGTAATT) for EcoRI 3’ LTR amplification, and LTR2 (CTCTTGGTCTGACTAC) and gagF1 (GATTCTGGTAGGACTGAC) for BamHI 3’ LTR amplification. The first PCR was performed with a denaturation step (5 min at 94°C), followed by 30 cycles of amplification (each cycle included 30 sec of denaturing at 94°C, 1 min annealing at 60°C, and 10 min extension at 68°C), and a final extension step (5 min at 68°C). In most cases, 1 μl of the first PCR was used as a template for the second PCR. In the case an ampiclon was identified after the first PCR when analyzed by electrophoretic separation, dilutions to 1:10 or 1:100 were used as template for the second PCR. The second PCR was performed by using nested primers LTR4 (CGGTTACTTAAGCTAGCTTGACT) and 5env1 (AGATTGACCGAGCTGGGAT) for EcoRI 3’ LTR, and LTR2 and env4R (AGTGAAGCAGGGTGAC) for BamHI 3’ LTR amplification. Amplicons from the second PCR were analyzed by electrophoretic separation in a 1% agarose gel, were cloned into XL- or TA-TOPO vectors (Invitrogen), and were sequenced using M13 vector primers.

**Southern Blot Analysis.** Southern blot analysis was performed following standard protocols. Briefly, spleen DNA from leukemic and control mice was cut with EcoRI, was separated in a 0.8% agarose gel, and was transferred onto a nylon membrane (Hybond-XL, Amersham Biosciences, Buckinghamshire, U.K.). The viral probe included a 1.2-kb EcoRI–ClaI fragment from the 4070A envelope sequence. Genomic probes were generated from clones isolated by iPCR, and included 100 μg/ml denatured salmon sperm DNA in the hybridization solutions. Probes were hybridized overnight at 45°C in formamide hybridization solution (Hybrisol-I, Serologicals, Norcross, GA).

**Database Analysis and URL Information.** Genomic sequence tags after the end of the proviral LTR sequence TCA were analyzed by using a combination of public (genome.ucsc.edu, and www.ensembl.org, February 2003 freeze) and private (Celera Discovery System, www.celera.com) databases. Each insertion site was assigned with a Cbfb-MYH11 retrovirus insertion (Cri) number, and the name of the proximal gene. Uncharacterized genes include ESTs, published expressed genes of unknown function, and predicted genes. These genes (and their human homologues) were annotated as they appear in public databases. In the case more than one candidate gene is located near a Cri, the gene symbol and accession number of both flanking genes are denoted. Of note is that this assignment is relative to the quality of the mouse genome sequence and gene annotation available at the time of study, which may need to be adjusted after new available information is discovered. Sequence tags were deposited into GenBank with accession nos. CL001593–CL001659, as shown in Table 2, which is published as supporting information on the PNAS web site. These tags can also be found in the Retroviral Tagged Cancer Gene Database (http://genome2.ncifcrf.gov/RTCGD). The chromosomal position of flanking genes was obtained from the Ensembl Mouse Genome Server (www.ensembl.org/Mus.musculus/blastview) and was adapted to the Standard Mouse ideogram (Fig. 4), as accepted by the International Mouse Nomenclature Committee. The reference Standard Ideogram/Anomaly Breakpoints of the Mouse can be accessed at www.mgu.har.mrc.ac.uk/anomaly/anomaly-intro.html.

**Results**

**Cbfb<sup>=MYH11</sup> Knock-in Chimeras Efficiently Develop AML After Infection with Amphotropic Retrovirus 4070A.** Amphotropic murine leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak inducer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak inducer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak inducer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak inducer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak inducer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak inducer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak inducer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak inducer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak inducer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak inducer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak induer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak induer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak induer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak induer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak induer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak induer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak induer of hematologic malignancies in the mouse. It can induce follicular lymphoma and myeloid leukemia virus (MLV) strain 4070A (known henceforth as 4070A) is a weak induer of hematologic malignancies in the mouse. It can induce folli...
disease-free chimeras probably escaped leukemia due to low contribution of Cbfb$^{+/\text{MYH11}}$ embryonic stem cells to the hematopoietic compartment or to an inefficient viral injection at birth. The 20 129Sv/Ev and 20 C57BL/6 wild-type controls did not develop leukemia 1 year after injection (experimental endpoint), indicating that 4070A-induced AML depended on Cbfb-MYH11 expression.

The clinical presentation and histological findings of the 4070A-induced leukemias in the Cbfb$^{+/\text{MYH11}}$ chimeras were similar to those induced by N-ethyl-N-nitrosourea. Peripheral blood analysis showed a progressive anemia (hematocrit between 18% and 35%) and the presence of immature cells in the circulation (Fig. 2A), which were predominantly myeloblasts and monocytic blasts by morphology (Fig. 2B). Infrquent immature erythroid, granulocytic, and apoptotic cells could also be detected in some samples. Fluorescence-activated cell sorting analysis consistently indicated that most of the leukemic cells in peripheral blood were CD117$^+$ (59.46 ± 14.07%, n = 5 and 1.9 ± 0.26%, n = 3 for the leukemic and control mice, respectively) and negative for lineage markers (CD3, CD45, CD11b, and Gr-1). In addition, some samples contained CD117$^+/\text{ter}119^+$ cells (5.2 to 24.1% of the total cell populations, n = 5), as detected previously in N-ethyl-N-nitrosourea studies (9). Spectral karyotyping analysis was performed on leukemic cells from two mice with 10 cells each being analyzed, which failed to identify chromosomal rearrangements, suggesting that AML arose with no evidence of genome instability (data not shown). Twenty of the 27 leukemic mice were killed at terminal stage for further analysis. Pathology analysis consistently revealed splenomegaly (a 10- to 35-fold increase in size), with massive infiltration of leukemic cells and disruption of splenic architecture (Fig. 2C). In addition, hepatomegaly was frequently observed with focal infiltration of leukemic cells around blood vessels (Fig. 2D). Thymus and lymph nodes were not affected. These results indicate that Cbfb$^{+/\text{MYH11}}$ chimeras readily develop AML after acquisition of additional genetic changes, and that the leukemic phenotype is independent of the mutagen used.

**Clonal Expansion of RIM Leukemic Samples.** The number of proviral insertions present in the leukemic samples was estimated by Southern blot analysis (Fig. 3A) of spleen DNA from moribund mice. The 4070A-envelope-specific probe identified one to three proviral integrations, with an average of 1.7 insertions per tumor. In tumor VC4 (Fig. 3A, lane 5), no viral integrations were identified. One insertion was later identified in this tumor by iPCR, using primers specific for the 5' end of the provirus. Sequence analysis of the viral integration site isolated by PCR revealed a truncated provirus 4070A lacking the envelope and 3' LTR sequences (data not shown).

To verify the functional significance of the insertions during leukemogenesis, bone marrow and spleen cell suspensions from affected mice were transplanted into isogenic wild-type mice. Recipient mice developed AML within 8 weeks of transplantation with a phenotype similar to the original retrovirus-infected chimeras. The clonality of the leukemia was assessed by analysis of the pattern of proviral integration in primary and transplanted tumors (Fig. 3B). One band (representing one viral insertion) from primary tumors (see asterisks in lanes 1 and 4) was always present in all transplanted samples (lanes 2 and 3 and lanes 5 and 6). As 4070A is replication-competent, additional bands were found in the transplanted samples. These results indicate that a discrete number of genetic alterations (one to three), caused by 4070A proviral insertion, can synergize with Cbfb-MYH11 to generate leukemia in the mouse. Furthermore, these leukemic cells expanded in a clonal fashion.

**Cloning of Retroviral Insertion Sites (RISs).** To identify candidate genes that cooperate with Cbfb-MYH11 in AML, iPCR was carried out on spleen DNA from 20 tumors. The iPCR primers matching the 4070A envelope region did not generate any background (false-positive amplicons) because they do not match with endogenous sequences in the mouse genome. In addition, the efficiency of cloning 4070A insertions was increased by independently using two restriction enzymes (EcoRI and BamHI), and using primers for both ends of the provirus. A total of 67 retroviral insertion tags were identified (Table 2), rendering an average of 3.3 tags per tumor. This tag per tumor ratio is higher than that detected by Southern blot hybridization described above, and this difference is probably due to integration of broken copies of the viral genome that were only
transcription initiation site and 16 mapped in the 5 kb upstream of the candidate genes is shown as a mouse standard ideogram in Fig. 3, suggesting that promoter activation was a frequent event. In addition, one Cri disrupted an exon and 21 others are located in introns of the protein coding region, which probably generated truncated proteins or altered expression of the full length transcript. Only seven Cris mapped >10 kb away from the flanking genes.

Candidate Genes That Synergize with Cbfb-MYH11 in Leukemogenesis.

Next, we analyzed the genes flanking the CIS (Table 3). Cri10 represents eight independent insertions near the transcription start site (introns 1 and 2) of the transcription factor gene Plag1, and the inserted viral genomes were in the same orientation as the Plag1 gene. These insertions were also detected by Southern blot analysis using genomic DNA probes (Fig. 3C), confirming that Plag1 insertions were predominant in the related tumors and probably participated in the transformation process. PLAG1 has been identified at the breakpoints of a t(3;8)(p21;q12) translocation, which is the most common recurrent chromosomal abnormality in human pleomorphic adenoma of the salivary gland, lipomas, and lipoblastomas (24–26). The translocation results in a promoter swapping between PLAG1 and CTNNB1 that led to the activation of PLAG1 expression. The data reported here is evidence of PLAG1 involvement in leukemia, and the likely consequence of retroviral insertions near the transcriptional start site of the Plag1 gene is also activation of expression. Plag2, a Plag1-related gene, has recently been proposed also to be a protooncogene (27), Two insertions (Cri07) were identified near the transcriptional start site of Plag2. Therefore, it is predicted that a promoter activation
event for Plagl2 could play a role in Cbfβ-MYH11-dependent leukemogenesis.

Two insertions (Cri30) mapped at a common site of MLV insertions in mouse chromosome 10, called Mml-3 locus (28), 80 and 83 kb upstream of the Myb transcription factor and 60 and 57 kb upstream of Hbs1 helicase. This region is a hot spot for MLV insertions (see Table 1 and references therein), but its actual effect on Myb or Hbs1 expression is poorly understood. Recent data, however, suggest that viral insertions at Mml loci may alter Myb expression (29).

The runt-related RUNX2 protein forms a heterodimeric complex with CBFβ, and plays a critical role in osteoblast and chondrocyte differentiation (30). Sustained expression of Runx2 alters T cell activation (31, 32), whereas MLV insertions upstream of Runx2 induce lymphomagenesis in cooperation with Myc (14, 33, 34). In our study, however, three insertions (Cri49) mapped within Runx2 intron 5. Such an event would predict expression of a truncated Runx2 including the runt domain (excluding the transactivation and repression domains) with a putative dominant-negative function. Alternatively, a yet unidentified regulatory sequence in intron 5 could up- or down-regulate expression of the full-length Runx2 transcript.

Two insertions (Cri44) were identified in introns 3 and 6 of the D6Mm5e gene in mouse chromosome 6. The function of this gene is unknown, and analysis of its amino acid sequence does not predict any known conserved domains. Alternatively, these insertions could affect expression of Dok1, a gene located 18 and 45 kb from these insertions toward the telomere. P62-Dok (encoded by Dok1) belongs to the Dok family of adaptor proteins that interacts with Ras-GAP, Csk, and Nck, and participates in a variety of cellular processes, such as migration (35) and activation of T cells (36), and the pathogenesis of cancer (37).

Two insertions (Cri46) mapped within the MHC Class 1 T gene cluster, 3 kb upstream and 7 kb downstream of the cell-surface glycoprotein gene H2T24, a histocompatibility 2 T region locus 24 antigen-related gene. The guanine nucleotide-binding protein gene Gna-rs1, which maps 25 and 40 kb from these insertions, could also be affected in these tumors.

In addition to the six CIs identified, six single insertion sites (SISs) have previously been identified as CISs with MLV insertions (Table 1), suggesting a common role in cell transformation. We identified SISs associated with regulation of IFN-related receptor signaling (receptors Ifngr2 and kinase Jak2), TGFβ pathway (type I receptor Acyl-1 and the transcription factors Dec1 and Smad1), and Ras/PI3K signaling pathway (adaptor Dap10, PI3Kcg, and Ship1). Finally, genes associated with cytoskeleton organization and cell adhesion that may have key transforming functions in AML were also targeted. Cytoskeleton-associated genes included structural genes (radixin and tensin genes), as well as regulatory genes (phosphatase gene Ptpn2, the Rho GDP-GTP exchange factor gene Lbc, and the Rho GTPase gene Rif). Genes associated with cell adhesion included the protocadherin gene Pcdh1, homeoprotein gene Prx2, and the tumor suppressor gene Tsc1.

**Discussion**

Fusion genes generated by chromosome translocations/inversions play important roles in leukemogenesis. Previously, we demonstrated that one of the fusion genes, Cbfβ-MYH11, generated by chromosome 16 inversion in human AML subtype M4 with eosinophilia, is necessary but not sufficient for leukemia development in our knock-in mouse model (9). In more recent years, it has been demonstrated that many other fusion genes are not sufficient to induce leukemia (38–40). In this manuscript, we identified specific candidate genes that can cooperate with Cbfβ-MYH11 for leukemogenesis in our knock-in mouse model. AML was efficiently induced by injecting retrovirus 4070A in newborn Cbfβ-MYH11 chimeras with a latency of 2–5 months. Risets were identified by Southern blot hybridization and were identified easily and efficiently by iPCC. The insertion sites were mapped to the genome and flanking genes were identified. We showed that as few as one additional alteration (a second genetic “hit”) could efficiently accelerate AML in the mouse. This finding was verified by the fact that one insertion per leukemia was retained after secondary transplantation, and the fact that no additional mutations were identified at cytogenetic level. However, we cannot rule out the presence of subtle mutations, such as point mutations, that may also play a role in the leukemogenic process. Because a similar type of myeloid leukemia was induced in this insertional mutagenesis model and in our previous N-ethyl-N-nitrosourea mutagenesis model (9), it seems that Cbfβ-MYH11 expression is an important factor in determining the morphology of the leukemic cells, whereas the secondary alterations are associated with proliferative or antiapoptotic events.

Of particular interest is the identification of CISs specific for Cbfβ-MYH11. This study identified, for the first time, the role of the C2H2 zinc finger transcription factors Plagl and Plag2 in AML. Both proteins recognize the same DNA-binding element and regulate IGFII expression (27, 41). PLAGL2 also regulates expression of the proapoptotic gene Nip3 in response to hypoxia or iron deficiency (42). However, little is known about the mechanism of action of these proteins in cancer formation.

It is intriguing that Myb and Runx2 were identified as CISs in Cbfβ-MYH11-induced AML. CBFβ/RUNX1 heterodimer regulation of normal proliferation and differentiation of myeloid and lymphoid lineages depend on the combined functions of Myb and CBF proteins (43, 44). Therefore, because CBFβ-SMMHC inhibits CBF function, it is surprising that Myb or Runx2 may play a synergistic effect in transformation. We speculate that Cbfβ-MYH11 expression may have an incomplete effect on CBF function that is sufficient to affect differentiation, although the residual CBF function may rescue the cells from transformation. Additional alterations on CBF regulation of target genes, due to Runx2 or Myb misexpression, may transform the cells to AML. Alternatively, these genes may have CBF-independent functions that cooperate with CBFβ-SMMHC in AML.

Many SIS genes in this study encode members of signaling pathways linked to proliferation, apoptosis, and cancer. These pathways include Ras/PI3K, TGFβ IFN signaling, and cytoskeleton organization. Whereas the SIS matching previously identified sites may represent genes with broad functions associated with transformation, the majority of SISs are candidates involved in a variety of cellular processes.
and may be specifically associated with AML subtype M4 with eosinophilia.

The efficiency of our approach is highlighted by the absence of leukemia in MLV-mutagenized 129Sv/Ev wild-type mice and in untreated Cbfb-MYH11 mice. In addition, each leukemia case harbored very few insertions (sometimes only one), which was probably attributed to the transformation predisposition provided by Cbfb-MYH11. The fact that the 4070A envelope sequence does not share homology with endogenous retroviruses contributes to the easy identification of the insertion sites by iPCR with relatively low backgrounds. As a result, unlike previous studies that show a background incidence of lympho-

ma/leukemia and numerous insertions per sample (16–18), most of the genes identified in this study are likely to play a role in Cbfb-MYH11-associated AML. Finally, the use of a systematic mutagenesis approach, such as RIM in transgenic mice expressing leukemia-associated fusion genes, will provide a wealth of genetic information for the study of human leukemia, including new directions for improved therapies in the future.

We thank Carolyn Wilson for advice on quantitation of replication competent retroviruses. I.H.C. is a Leukemia and Lymphoma Society Special Fellow. This work was supported in part by National Institutes of Health Grant RO1-CA096983 (to L.H.C.) and American Cancer Society Grant IRG 93-033.