Published in final edited form as:

Nat Genet. 2013 January ; 45(1): 18–24. doi:10.1038/ng.2495.

Recurrent *SETBP1* mutations in atypical chronic myeloid leukemia

Rocco Piazza^{1,16}, Simona Valletta^{1,16}, Nils Winkelmann^{2,3}, Sara Redaelli^{1,16}, Roberta Spinelli¹, Alessandra Pirola¹, Laura Antolini¹, Luca Mologni¹, Carla Donadoni¹, Elli Papaemmanuil⁴, Susanne Schnittger⁵, Dong-Wook Kim⁶, Jacqueline Boultonwood⁷, Fabio Rossi⁸, Giuseppe Gaipa⁹, Greta P De Martini¹, Paola Francia di Celle¹⁰, Hyun Gyung Jang^{11,15}, Valeria Fantin^{11,15}, Graham R Bignell⁴, Vera Magistroni¹, Torsten Haferlach⁵, Enrico Maria Pogliani^{1,12}, Peter J Campbell⁴, Andrew J Chase^{2,13}, William J Tapper^{2,13}, Nicholas C P Cross^{2,13}, and Carlo Gambacorti-Passerini^{1,14}

¹Department of Health Sciences, University of Milano–Bicocca, Monza, Italy.

²Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, UK.

³Klinik für Innere Medizin II, Universitätsklinikum Jena, Jena, Germany.

⁴Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

⁵Munich Leukemia Laboratory (MLL), Munich, Germany.

⁶Department of Hematology, Catholic University, Seoul, Korea.

⁷Leukaemia & Lymphoma Research Molecular Hematology Unit, John Radcliffe Hospital, Oxford, UK.

⁸Immunotransfusal Unit, San Gerardo Hospital, Monza, Italy.

⁹M Tettamanti Research Center, Pediatric Clinic University of Milano–Bicocca, Monza, Italy.

¹⁰Anatomia Patologica II, Laboratorio Oncologia Molecolare, Centro di Ricerca in Medicina Sperimentale (CeRMS), Azienda Ospedaliera San Giovanni Battista di Torino, Torino, Italy.

¹¹Agiros Pharmaceuticals, Cambridge, Massachusetts, USA.

¹²Hematology Unit, San Gerardo Hospital, Monza, Italy.

© 2013 Nature America, Inc. All rights reserved.

Correspondence should be addressed to C.G.-P. (carlo.gambacorti@unimib.it).

¹⁵Present addresses: Ariad Pharmaceutical, Cambridge, Massachusetts, USA (H.G.J.) and Tumor Cell Biology, Pfizer, La Jolla, California, USA (V.F.).

¹⁶These authors contributed equally to this work.

URLs. GeneRanker, <http://cbio.mskcc.org/tcga-generanker/index.jsp>; epestfind, <http://emboss.bioinformatics.nl/cgi-bin/emboss/epestfind>; Ingenuity, <http://www.ingenuity.com/>; Ensembl annotation file, ftp://ftp.ensembl.org/pub/release-54/gtf/homo_sapiens/; dChip, <http://biosun1.harvard.edu/~cli/complab/dchip/>.

Accession codes. High-throughput sequencing data have been deposited in the Sequence Read Archive (SRA) under accession SRA061202 and in the Gene Expression Omnibus (GEO) under accession GSE42146.

Note: Supplementary information is available in the online version of the paper.

AUTHOR CONTRIBUTIONS R.P., S.V., N.W., S.R., R.S., A.P., L.M., C.D., E.P., P.F.d.C., H.G.J., V.F., G.R.B., V.M., P.J.C. and A.J.C. performed the experiments. R.P., S.V., S.R., R.S., L.A., F.R., A.J.C., W.J.T. and N.C.P.C. performed data analysis. S.S., D.-W.K., J.B., G.G., G.P.D.M., T.H., P.J.C., E.M.P. and N.C.P.C. contributed reagents and analysis tools. R.P. and C.G.-P. wrote the first draft of the manuscript. L.A. performed statistical analysis. N.C.P.C. and C.G.-P. supervised research. C.G.-P. initiated the project. All coauthors contributed to the final version of the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

¹³Faculty of Medicine, University of Southampton, Southampton, UK.

¹⁴Hematology and Clinical Research Unit, San Gerardo Hospital, Monza, Italy.

Abstract

Atypical chronic myeloid leukemia (aCML) shares clinical and laboratory features with CML, but it lacks the *BCR-ABL1* fusion. We performed exome sequencing of eight aCMLs and identified somatic alterations of *SETBP1* (encoding a p.Gly870Ser alteration) in two cases. Targeted resequencing of 70 aCMLs, 574 diverse hematological malignancies and 344 cancer cell lines identified *SETBP1* mutations in 24 cases, including 17 of 70 aCMLs (24.3%; 95% confidence interval (CI) = 16–35%). Most mutations (92%) were located between codons 858 and 871 and were identical to changes seen in individuals with Schinzel-Giedion syndrome. Individuals with mutations had higher white blood cell counts ($P = 0.008$) and worse prognosis ($P = 0.01$). The p.Gly870Ser alteration abrogated a site for ubiquitination, and cells exogenously expressing this mutant exhibited higher amounts of SETBP1 and SET protein, lower PP2A activity and higher proliferation rates relative to those expressing the wild-type protein. In summary, mutated *SETBP1* represents a newly discovered oncogene present in aCML and closely related diseases.

aCML¹ is a heterogeneous disorder belonging to the group of myelodysplastic/myeloproliferative (MDS/MPN) syndromes. In aCML, many clinical features (splenomegaly and myeloid predominance in the bone marrow, with some dysplastic features but without a differentiation block) and abnormalities in the laboratory (myeloid proliferation and low leukocyte alkaline phosphatase values) suggest diagnosis with CML. However, lack of the pathognomonic Philadelphia chromosome² and of the resulting *BCR-ABL1* fusion point to a different pathogenetic process. Because no specific recurrent genomic or karyotypic abnormalities have been identified in aCML, the molecular pathogenesis of this disease has remained elusive and the outcome dismal (median survival of 37 months after diagnosis)³, with no improvement over the last 20 years. This prognosis sharply contrasts with the outcome for CML, for which the prognosis was markedly improved by the development of imatinib as a specific inhibitor of the BCR-ABL1 protein^{4–7}.

High-throughput sequencing has proven to be a powerful tool to identify recurrent, specific genetic abnormalities in solid cancers and leukemias^{8–10}. Although the genetic heterogeneity of cancer necessitates some caution in the interpretation of the results and in their application¹¹, high-throughput sequencing remains a powerful instrument to improve knowledge of the molecular pathogenesis of malignancies¹² and to potentially refine cancer diagnosis and treatment¹³. We applied a high-throughput sequencing strategy to aCML, including both exome sequencing and RNA sequencing (RNA-seq), with the aim of identifying new recurrent driver mutations. We present here the results of this combined approach and the identification of mutated *SETBP1* as a new oncogene.

RESULTS

Exome sequencing of aCML

We used exome sequencing technology to identify somatically acquired mutations in eight individuals with aCML by comparing DNA from leukocytes and constitutive DNA extracted from lymphocytes. Each read of a massively parallel sequencing run is clonal and therefore derives from a single molecule of genomic DNA. Thus, the proportion of sequencing reads reporting a variant allele provides a quantitative estimate of the proportion of cells in the DNA sample carrying that mutation, assuming adequate coverage of the investigated gene.

To minimize the detection of subclonal variation, only mutations with a frequency of at least 35% were considered (Online Methods). We identified 84 exonic mutations, of which 63 (75%, range of 5 to 14 mutations per case) were nonsynonymous (Supplementary Table 1), and 21 were synonymous. Transitions accounted for 73% (46 of 63) of the nonsynonymous mutations identified (Supplementary Fig. 1). The median absolute coverage at positions where mutations were identified was 84× (with a range from 20× to 232×). Four mutations were nonsense substitutions, including one in the *ASXL1* gene. The frequency of mutant reads over total reads ranged between 35% and 98% (median of 47%). All nonsynonymous mutations identified by high-throughput sequencing were subjected to standard sequencing (Supplementary Fig. 2 and Supplementary Table 2), and the validation rate was 96%.

In the case with an *IDH2* alteration (subject 1), the levels of 2-hydroxyglutarate in leukemic cells were >10 times higher than in autologous normal cells or in other cases (Supplementary Fig. 3). We also found two recurrently mutated genes: *EZH2* (subjects 4 and 8) and *SETBP1* (subjects 3 and 5). No additional recurrent mutation was observed, even when lowering the accepted frequency below 35%. *EZH2* encodes a histone methyltransferase involved in the epigenetic control of gene expression. *EZH2* mutations were previously identified as a recurrent abnormality in myeloid neoplasias, including aCML¹⁴. The second recurring alteration affected *SETBP1*. The same mutation (encoding a p.Gly870Ser alteration) was identified in both cases, with frequencies of 53% in subject 3 (coverage of 38) and 47% in subject 5 (coverage of 72).

Some of the genes identified as mutated in one of these eight cases (*IDH2*, *MTA2*, *EPHB3*, *ETNK1*, *GATA2* and *IRAK4*) and having a score of 1 in the oncogenic gene ranking score (GeneRanker; see URLs) were resequenced in a cohort of 40 aCML cases (15 with *SETBP1* mutations and 25 without). With the exception of *IDH2*, no gene was found to be mutated in any case apart from the index case (Supplementary Table 3).

Recurrent *SETBP1* mutations

The presence of an identical mutation not previously involved in cancer in two different aCML cases prompted us to resequence *SETBP1* in samples from additional subjects with aCML or other hematological malignancies and in cell lines representative of the most common human solid cancers. In this analysis, 17 of 70 aCML cases (24.3%, 95% CI = 16–35%) tested positive for *SETBP1* mutation (Table 1). Constitutive DNA was available from four of these additional *SETBP1*-mutated aCML cases, the analysis of which showed that mutations encoding p.Glu858Lys, p.Asp868Asn, p.Gly870Ser and p.Ile871Thr alterations were somatically acquired. Sequencing of 112 healthy donors and the inspection of SNP databases allowed us to identify variants encoding p.Arg1321His and p.Val1377Leu, identified in two aCML cases without available constitutive DNA, as rare polymorphisms (the variant encoding p.Arg1321His was found in SNP databases, and the variant encoding p.Val1377Leu was found in both SNP databases and healthy donor samples) and therefore to discard them. *SETBP1* mutations were also present in the closely related disorders unclassified MDS/MPN (3 of 30, 10%) and chronic myelomonocytic leukemia (CMML; 3 of 82, 4%) and in 1 of 4 cases of chronic neutrophilic leukemia (CNL). In all cases with *SETBP1* mutation, the mutations were heterozygous. *SETBP1* mutations seem to be enriched in aCML and closely related disorders, as no mutations were found in 458 individuals with other hematological malignancies nor in 344 cell lines representing lymphomas and the most common non-hematological malignancies (Table 1 and Supplementary Tables 4 and 5).

Of the 24 *SETBP1* alterations identified, 22 (92%) were located in a short stretch of 14 residues spanning Glu858 to Ile871 (Fig. 1) within the SKI homologous region, so called because of limited homology to the SKI oncoprotein¹⁵. The most frequently observed

alterations (p.Glu858Lys, p.Asp868Asn, p.Ser869Gly, p.Gly870Ser and p.Ile871Thr) were analyzed by SIFT and PolyPhen-2 software^{16,17} for the predicted change to the protein structure. All five changes generated the maximum score predicting alteration of normal function (SIFT score = 0), with a median information content (MIC) of 2.87.

All mutations were heterozygous at the genomic level. We verified the relative expression levels of the two alleles by deep sequencing in three aCML cases. Coverage of the mutated bases in the three cases was 905, 440 and 523. The frequency of the mutated allele in cDNA was 79%, 45% and 38%, respectively, which is compatible with a somatic heterozygous status without substantial imbalance in allelic expression. These results, together with the absence of nonsense and frameshift *SETBP1* mutations, strongly suggest that mutant *SETBP1* has a dominant and presumably altered biological activity.

To test the relationship between *SETBP1* variants and mutations in oncogenes known to be involved in myeloid malignancies, we also evaluated mutations in *ASXL1*, *TET2*, *IDH1*, *IDH2*, *EZH2*, *CBL*, *NRAS*, *KRAS*, *SUZ12*, *SF3B1*, *RUNX1*, *JARID2*, *JAK2*, *EED*, *DNMT3A*, *CEBPA*, *RBBP4*, *NPM1* and *FLT3* in a population of 61 aCML cases (14 with *SETBP1* mutations and 47 without). The results are shown in Figure 2 (see also Supplementary Table 3). No significant association or mutual exclusion with *SETBP1* mutations was observed. *ASXL1* mutations were present more frequently in cases with *SETBP1* mutations (36% versus 19%, respectively), whereas *TET2* mutations were more prevalent in cases with wild-type *SETBP1* (28% versus 14%, respectively); however, further analysis of larger collections of aCML cases will be necessary to determine whether these differences are significant.

To investigate the possibility of chimeric fusion genes as a result of cryptic chromosomal rearrangements, we performed RNA-seq (Online Methods) in seven *SETBP1*-mutated aCML cases and six aCML cases with wild-type *SETBP1*, analyzing the results using our in-house software FusionAnalyser¹⁸. No fusion genes were detected in any cases. We also used the RNA-seq data to analyze the expression of the mutated gene and confirmed transcription of *SETBP1*, with transcript levels in mutated cases (1.12 ± 0.4 fragments per kilobase of exon model per million mapped reads (FPKM; $\pm$ s.e.m.)) being similar to ones without *SETBP1* mutation (1.92 ± 0.5 FPKM). We also investigated the presence of exonic copy-number alterations using exome sequencing data (Supplementary Fig. 4) and dedicated software (CEQer, Comparative Exome Quantification analyzer; R.P. *et al.*, unpublished data) but found no recurrent alterations.

Clinical course of aCML cases according to *SETBP1* mutation status

Clinical information was available for 38 aCML cases, including 14 with *SETBP1* mutations and 24 with wild-type *SETBP1*. We analyzed the two groups by univariate analysis, considering sex, age, white blood cell count, hemoglobin concentration and platelet number at diagnosis, the percentage of peripheral blood blasts and overall survival. *SETBP1*-mutated cases showed worse prognosis (median survival = 22 versus 77 months, $P = 0.01$, hazard ratio = 2.27; Fig. 3a) and presented with higher white blood cell counts at diagnosis (median of 81.0 versus 38.5×10^9 cells/l, $P = 0.008$; Fig. 3b) compared to cases with wild-type *SETBP1*. We observed no significant differences for the number of peripheral blood blasts, age, hemoglobin concentration or platelet counts (Fig. 3c,d), and no difference was observed in sex distribution between the two groups. The negative effect of *SETBP1* mutations on survival was maintained ($P = 0.035$) after adjustment for the effects of age and white blood cell count or of sex, percentage of peripheral blood blasts, hemoglobin concentration or platelet number.

Biological effects of the SETBP1 p.Gly870Ser alteration

SETBP1 is a poorly characterized protein that is believed to inhibit PP2A phosphatase activity through SET stabilization¹⁵. The SETBP1 region where the somatic alterations cluster is highly conserved among vertebrates, which suggests that it might have an important yet unknown biological role. According to the Eukaryotic Linear Motif (ELM) server¹⁹, this region represents a virtually perfect degron (a specific sequence of amino acids in a protein that directs the initial step of degradation), containing the consensus binding region (DpSGXXpS/pT, where pS and pT represent phosphorylated residues) for β -TrCP1, the substrate recognition subunit of the E3 ubiquitin ligase (amino acids 868–873; Fig. 4a)²⁰. This degron includes a PEST domain (amino acids 860–884, HSEETIPSDSGIGTDNNSTSDQAEK), a sequence associated with proteins that have a short intracellular half-life. Therefore, this region might be critical for ubiquitin binding and for subsequent protein degradation. This hypothesis was experimentally confirmed using biotinylated phosphorylated peptides encompassing amino acids 859–879: whereas the wild-type peptide, incubated in the presence of TF1 cell lysate, could efficiently bind β -TrCP1 as predicted, a peptide with the p.Gly870Ser alteration was incapable of binding this E3 ligase subunit, indicating a possible difference in SETBP1 protein stability caused by this alteration (Fig. 4b). A critical requirement in the β -TrCP1 degron is the presence of one phosphorylated serine and one phosphorylated threonine within the core consensus region. To confirm the specificity of the interaction we observed, we repeated the same experiment using dephosphorylated peptides and purified recombinant β -TrCP1: in the absence of phosphorylation, the wild-type peptide did not interact with β -TrCP1 (Fig. 4c).

These experiments indicated a possible difference in SETBP1 protein stability caused by the p.Gly870Ser alteration. To further test this idea, TF1 cells transduced with viruses expressing wild-type SETBP1 or SETBP1 Gly870Ser and expressing similar levels of *SETBP1* mRNA (51.9 and 35.8 FPKM for wild-type SETBP1 and SETBP1 Gly870Ser, respectively) were assayed for the expression of SETBP1 protein using a specific antibody. In cells expressing wild-type SETBP1, SETBP1 protein was barely detectable, in line with its expected short half-life. By contrast, cells expressing SETBP1 Gly870Ser showed higher levels of SETBP1 protein, recognized as a band of approximately 250 kDa, comigrating with the positive control (Fig. 5a, results representative of three experiments). The amount of SET protein was also higher (Fig. 5b), although *SET* mRNA levels were similar between the two cell lines (140.4 and 162.9 FPKM for wild-type SETBP1 and SETBP1 Gly870Ser, respectively). We also observed significantly reduced PP2A activity in the cell line expressing SETBP1 Gly870Ser (Fig. 5c), as well as greater PP2A phosphorylation at position Tyr307, a well-known marker of PP2A inactivation (Fig. 5b). Cells expressing SETBP1 Gly870Ser also had a higher proliferation rate compared to cells expressing wild-type SETBP1 or to cells transfected with empty vector, when cultured at standard granulocyte-macrophage colony-stimulating factor (GM-CSF) concentrations (Fig. 5d).

To study the intracellular localization of SETBP1, wild-type protein and SETBP1 Gly870Ser forms fused with GFP were introduced into the 293T human cell line and sorted to express similar level of protein (Supplementary Fig. 5). We then examined the intracellular localization of mutated and wild-type SETBP1 protein by confocal microscopy (Supplementary Fig. 6). SETBP1 Gly870Ser maintained a mostly nuclear localization. These data exclude a gross alteration in the intracellular distribution of the protein, although localization in cells expressing SETBP1 Gly870Ser showed a more punctate appearance than in cells overexpressing wild-type SETBP1.

To determine whether *SETBP1* mutations are associated with a specific gene expression signature, we analyzed RNA-seq data from 13 aCML cases and found a total of 197 differentially expressed genes (Supplementary Table 6) in cases with mutated *SETBP1*.

(encoding p.Glu858Lys, p.Asp868Asn (2 cases), p.Ser869Gly, p.Gly870Ser (2 cases) and p.Ile871Thr) and cases with wild-type *SETBP1* (Supplementary Fig. 7). Of the 197 differentially expressed genes, 14 (7.1%, 95% CI = 3.5–10.7%) belonged to the group transcriptionally controlled by TGF- β 1 (Ingenuity Systems; Supplementary Fig. 8). This value represents an enrichment of approximately 4-fold compared to the number of TGF- β -related (TGFB1) genes present in the reference genome used for RNA-seq (399/20,907 = 1.9%, 95% CI = 1.7–2.1%). This difference is highly statistically significant ($P = 1.6 \times 10^{-7}$ by χ -square test).

DISCUSSION

SETBP1 represents the first gene shown to be enriched and recurrently mutated in aCML, a disease currently defined only by negative characteristics (for example, by not having the *BCR-ABL1* fusion). Thus, it may constitute a valuable diagnostic tool in the differential diagnosis of MDS/MPN syndromes and in their prognosis, as individuals with *SETBP1* mutations had a worse prognosis than cases with wild-type *SETBP1*. The presence of *SETBP1* mutations in approximately one-quarter of aCML cases, as well as the type of mutations identified, strongly point to a causal role of this gene in the pathogenesis of aCML. However, given the lack of information on the physiological role of *SETBP1*, extensive additional work will be necessary to clarify the mechanistic consequences of *SETBP1* mutations. *SETBP1* mutations were also found in CMML, a disease considered to be very similar to aCML that has overlapping diagnostic criteria, but with a prevalence almost seven times lower, demonstrating for the first time a biological difference between these two entities. In aCML, mutations of known oncogenes such as *NRAS*, *KRAS*, *TET2*, *EZH2* and *CBL* have been described²¹; it will be important to study the relationship between *SETBP1* mutations and these additional genetic alterations in larger cohorts of affected individuals. Although we were unable to identify *SETBP1* mutations in other cancers, more extensive analysis will be necessary to fully characterize the oncogenic potential of mutated *SETBP1*.

SETBP1 has been reported to be fused to *NUP98* in a single subject with T-cell acute lymphoblastic leukemia and to be overexpressed as a consequence of a translocation involving *ETV6* in acute myeloid leukemia^{22,23}. In these reports, no mutations or structural alterations in the coding portion of the gene were reported, but these rare cases are consistent with the possibility that overexpression of *SETBP1* may be oncogenic.

The identification of *SETBP1* mutations in aCML also represents the first time that recurrent point mutations of this gene have been shown to occur in cancer. Although the Catalogue of Somatic Mutations in Cancer (COSMIC) database contains 12 *SETBP1* somatic mutations, only one of these was validated by Sanger sequencing²⁴. *SETBP1* is located at chromosome 18q21.1 and codes for a protein of 1,596 residues (NM_015559.2, long isoform) with a predicted molecular weight of 170 kDa and a predominantly nuclear localization that is expressed in hematopoietic stem/progenitor cells and also in committed progenitors²⁵. Our experimental data, although confirming the nuclear localization of SETBP1, indicated an observed size larger than the predicted molecular weight of 170 kDa. The reason for this is unknown but may be related to post-translational modifications.

The only known interactions of SETBP1 are with the *HOXA9* and *HOXA10* promoters²⁶ and with SET through its SET-binding domain¹⁵. The resulting stabilization of SET can alter histone acetylation, or SET may directly bind and inhibit the PP2A phosphatase²⁷. PP2A activity is known to be inhibited in CML as a consequence of a BCR-ABL1-dependent increase in SET expression²⁷, and, for this reason, we tested whether expression of SETBP1 Gly870Ser could result in SET stabilization and PP2A inhibition. In addition,

the expression of LYN, a SRC family kinase known to be transcriptionally inhibited by PP2A²⁸, was higher in the presence of SETBP1 Gly870Ser, both in aCML cases (mean FPKM of 235.3 versus 80.4) and in transfected TF1 cells (mean value of triplicate experiments of 11.5 versus 6.4, $P = 0.02$). A similar upregulation of PTGS2, another transcriptional target of PP2A, was also observed in aCML cases (mean FPKM of 251 versus 20.0) and in TF1 cells (mean value of triplicate experiments of 0.045 (relative normalized units) versus 0.032, $P = 0.04$). These data suggest that inhibition of PP2A might be a common feature of SETBP1-mutated aCML. Additional unknown mechanisms are probably operative in this setting, as SETBP1 is a predominantly nuclear protein, whereas PP2A is also located inside the cytoplasm.

The dysregulation of SETBP1 protein levels and activity can be explained, at least in part, by the removal of a degron in the mutant SETBP1 protein, leading to decreased degradation of SETBP1 that might be functionally equivalent to overexpression. Although we tested only one of the SETBP1 alterations we identified, the proximity of the alterations and their presence inside the degron suggest a common mechanism of action.

The results from RNA-seq also suggest that some TGF- β target genes are differentially expressed in aCML cells with mutated and wild-type *SETBP1*. This is consistent with the known activity of SKI (and possibly of the SKI homology domain of SETBP1) on TGF- β via its interaction with SMADs²⁹. Further studies will be required to unravel both the physiological role of SETBP1 and its mechanistic role in the leukemogenic process.

Several germline *SETBP1* mutations have been described previously, albeit with different relative frequencies, in Schinzel-Giedion syndrome (SGS), a rare congenital disorder characterized by multiple malformations, many of which arise as a consequence of aberrant bone formation³⁰. It is tempting to connect the SGS phenotype (and the pathogenesis of aCML) to alterations in the TGF- β pathway, given the essential role of this cytokine in bone formation and remodeling³¹, but further research will be needed to test this hypothesis. The removal of a degron in the region of SETBP1 encoded by the mutational hotspot with resulting protein overexpression could also be operative in SGS, given the almost identical *SETBP1* mutations identified in the two disorders. SGS is a severely debilitating condition, and many affected individuals die in the perinatal period. Of those who survive, some develop tumors, predominantly of neuroepithelial origin³². Predisposition to myeloid malignancy has not been described, but the number of cases reported is small, and follow-up is limited. Notably, *SETBP1* adds to a growing list of genes that are constitutionally mutated in developmental disorders and somatically in cancer³³.

In summary, we have shown in this report that *SETBP1* mutations are present in approximately one-quarter of aCML cases, where they confer a worse clinical course. Furthermore, this is the first description of recurrent, validated *SETBP1* mutations in cancer. Expression of mutant SETBP1 Gly870Ser in the TF1 cell line resulted in higher SETBP1 protein levels, SET protein stabilization, PP2A inhibition and higher proliferation rates. Our results increase the knowledge of the mechanisms by which malignancy arises and will have important consequences for the diagnosis, prognosis and treatment of aCML and diseases associated with *SETBP1* alterations.

ONLINE METHODS

Subjects

Diagnoses of aCML and related diseases were performed according to the 2008 World Health Organization classification system (WHO-2008)¹. The eight aCML cases studied by exome sequencing were enrolled between 2008 and 2011. Their white blood cell counts

ranged between 22.4 and 89×10^9 cells/l; their ages were 75, 57, 83, 49, 74, 75, 66 and 69 years. Seven were male, and five were smokers or former smokers.

Cells and cell lines

Bone marrow or peripheral blood samples were collected at diagnosis in individuals with aCML and other hematological malignancies after obtaining written informed consent approved by the local ethics committee. Bone marrow samples were used for all cases but subject 4, for whom peripheral blood-derived cells were used. Leukemic cells were obtained by separation on a Ficoll-Paque Plus gradient (GE Healthcare). Surface markers were evaluated by fluorescence-activated cell sorting (FACS) analysis, and myeloid cells (positive for CD33, CD13 or CD117 staining) made up >80% of the total cells. As a source of normal cells, we used lymphocytes obtained by culturing cells with 2.5 µg/ml Phytohemagglutinin-M (PHA-M, Roche) and 200 International Units/ml interleukin-2 (IL-2, Aldesleukin, Novartis) for 3–4 d and then incubating cells for 2–3 weeks with IL-2 only. Phenotype was evaluated by FACS analysis, and lymphoid cells (positive for CD3, CD4, CD5, CD8 or CD19 staining) made up >80% of the total cells. After separation, cells were pelleted by centrifugation and lysed. The polyclonality of these populations was assessed by TCRExpress (BioMed Immunotech).

The TF1 human erythroleukemia cell line was purchased from DSMZ and maintained in RPMI 1640 medium (Lonza Cambrex) supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin G, 80 µg/ml gentamicin, 20 mM HEPES and 2 ng/ml human GM-CSF (Life Technology).

The 293T human embryonic kidney cell line was maintained in DMEM supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin G, 80 µg/ml gentamicin and 20 mM HEPES.

Exome sequencing

All exome libraries were generated from 1 µg of genomic DNA extracted with the Invitrogen PureLink Genomic DNA (gDNA) kit (Life Technology). Genomic DNA was fragmented to a size of 500 bp and then processed according to the standard protocol for the Illumina TruSeq DNA Sample Preparation kit (FC-121-1001), with selection of fragments of 200–300 bp in size on 2% agarose gels. Multiplexed genomic libraries were then enriched with the Illumina TruSeq Exome Enrichment kit (FC-121-1008). Libraries were subsequently sequenced on an Illumina Genome Analyzer IIx with 76-bp paired-end reads using Illumina TruSeq SBS kit v5 (FC-104-5001).

Bioinformatics

Image processing and base calling were performed using Illumina Real Time Analysis Software RTA v1.9.35. Qseq files were deindexed and converted to the Sanger FastQ file format using in-house scripts. FastQ sequences were aligned to the human genome database (NCBI Build 36/hg18) using the Burrows-Wheeler-based BWA alignment tool³⁴ within the Galaxy framework^{35–37}. The percentage of reads matching the reference human genome was over 90%, with mean exon coverage of >70-fold and the percentage of exons with a mean coverage of 20× over 90% for both the leukemic and control samples. The percentages of nucleotides targeting exonic regions and exonic regions plus the surrounding 100 bp were 48% and 68%, respectively, with an overall 28-fold enrichment for exonic versus non-exonic regions. The alignment files in the SAM format were processed using SAMtools alignment processing utilities³⁸: they were initially filtered by proper-pair, then converted into the binary BAM alignment format. Removal of duplicates was performed using the SAMtools rmdup command. Unique BAM files were then converted to the Pileup

format. Pileup data generated from paired cancer and control samples were cross-matched using a dedicated in-house software tool. This software initially analyzes each data set, extracting the information pertaining to each mismatch, either single nucleotide or indel, together with the corresponding read and mapping quality, read coverage of the mutated locus and specific coverage of each mutation. This intermediate information is stored in a condensed-Pileup format. The two condensed data sets are subsequently cross-matched and further filtered according to the following parameters: absolute coverage of each position (>20), relative coverage of each variant (>0.35), mapping quality (Phred mapping quality threshold = 30) and read quality (Phred read quality threshold = 30). Finally, a dedicated statistical model taking into account the coverage of each variant and the overall coverage in the cancer and control samples as well as the sequence of the reference genome is built to perform variant calling. Variants are then stored and further processed to predict the effect of nucleotide changes on protein function¹⁷. Taking into consideration the minimum relative coverage of each variant (0.35) and the percentage of leukemic cells in our preparations ($>80\%$) and assuming the mutations to be present as heterozygous alterations, the detection of a mutation in 35% of reads corresponds to its presence in 70–87.5% of leukemic cells.

Deep sequencing

We amplified 200 ng of cDNA with the Expand High-Fidelity PCR System (Roche). Each amplicon was then gel purified (3% agarose gel) using the QIAquick Gel Extraction kit (Qiagen). The purified amplicon was then directly processed according to the standard protocol for the Illumina TruSeq DNA Sample Preparation kit. Libraries were sequenced on an Illumina Genome Analyzer IIx with 76-bp paired-end reads using the Illumina TruSeq SBS kit v5.

Evaluation of copy-number alterations

Mean exonic coverage was calculated for all exons in the Consensus Coding Sequence (CCDS) exonic database in case and control samples. This information was initially used to calculate the median whole-exome coverage in cases and in controls. The mean exonic coverage of each exon was subsequently normalized accordingly. The mean normalized exon coverage was further modified by adding an arbitrary factor (20) to smoothen the effect of very-low-coverage values. Individual case-control $\log_2$ ratios were then calculated for all the exons in the data set and plotted. The presence of copy-number alterations was detected using a combined approach involving a set of statistical Wilcoxon signed-rank tests performed on sliding exonic windows combined with dedicated heuristic algorithms.

Analysis of recurrent mutations

Sanger sequencing of *NRAS* (exons 2 and 3), *KRAS* (exons 2 and 3), *TET2*, *EZH2*, *CBL* (exons 8 and 9), *ASXL1*, *IDH1* (Arg140 codon), *IDH2* (Arg132 codon), *WT1*, *SUZ12*, *RUNX1*, *RBBP4*, *NPM1*, *JARID2* (exons 1–18), *JAK2* (Val617 codon), *EED* (exons 2–12), *DNMT3A* (exon 23) and *CEBPA* was performed as described previously^{14,39–41}. Sequencing of *ETNK1* (exon 3), *EPHB3* (exons 3, 6–8, 10 and 11), *GATA2* (exons 5–7), *IRAK4* (exons 8–10), *MTA2* (exons 4–6, 8, 9, 14 and 15) and *SF3B1* (exons 14 and 15) was performed using the primers listed in Supplementary Table 3.

Validation of mutations

Genomic DNA was extracted from the peripheral blood or bone marrow samples of each subject using the PureLink Genomic DNA Mini kit (Invitrogen, Life Technology). CMML, aCML, CNL, JMML and unclassified MDS/MPN samples were amplified and sequenced with the primers listed in Supplementary Table 4 to cover the complete *SETBP1* coding sequence. The entire region encoding the SKI homologous domain was sequenced in all

samples using primers SETBP1_E_for, SETBP1_E_rev, SETBP1_F_for and SETBP1_F_rev. PCR amplification was performed using FastStart Taq DNA polymerase (Roche) with 100 ng of genomic DNA as template. All the mutations found in aCML samples were validated by PCR amplification followed by Sanger sequencing.

Polymorphism analysis

All the variants identified with either exome or Sanger sequencing were searched for in the dbSNP135 database to identify the presence of potential SNPs. All the variants present in the dbSNP database were discarded. To further test *SETBP1* variants in order to discriminate between real somatic mutations and previously unreported or rare SNPs, the exons of *SETBP1* were sequenced in a total of 112 healthy donors. All the variants previously identified in affected individuals and subsequently reported in the healthy donors were considered to be real SNPs and were therefore discarded.

Statistical analysis

Statistical analysis was performed using two-sided methodologies with a significance level of $\alpha = 0.05$. Continuous variables were described according to group classification (defined by wild-type and mutated *SETBP1*) by mean, median, 95% CI, mean standard error, and minimum and maximum values and compared across groups by the Wilcoxon test⁴². Categorical variables were described according to groups by the proportion of subjects falling into each category. Proportions were compared across groups by the χ -square test⁴². Bar graphs were used to describe continuous variables according to groups. Survival probabilities were estimated according to group classification by the Kaplan-Meier method⁴³. The null hypothesis of equality for the survival function across groups was tested by the log-rank test⁴³. The ratio of instantaneous hazards between groups was estimated resorting to the univariate Cox model⁴³. The multivariable Cox model was used to adjust the possible effect of *SETBP1* mutation for confounders, such as age, sex, white blood cell count, hemoglobin concentration, platelet number and the percentage of peripheral blood blasts⁴³.

2-hydroxyglutarate quantification

Cells (2×10^6) were suspended in 80% methanol, centrifuged, dried and stored at -80°C . 2-hydroxyglutarate levels were determined by ion-paired reverse-phase liquid chromatography coupled with negative-mode electrospray triple-quadrupole mass spectrometry, and integrated elution peaks were compared with 2-hydroxyglutarate standard curves for absolute quantification⁴⁴.

SETBP1 cloning and transfection experiments

A plasmid encoding the long isoform of *SETBP1* cDNA (SC114671, Origene) was used as a substrate for PCR amplification (Expand High Fidelity PCR System, Roche). Clone SC114671 (NM_015559.1) codes for the SETBP1 variant lacking 54 amino acids at the N terminus compared to the longest SETBP1 variant (NM_015559.2). The Gly870 codon in the NM_015559.2 isoform thus corresponds to the Gly816 codon in NM_015559.1; however, to keep consistency with previously published papers³⁰, coordinates are given with respect to the NM_015559.2 isoform. Two primers spanning the whole *SETBP1* cDNA and introducing artificial KpnI and XhoI sites at the 5' and 3' ends of the coding region (respectively) were used to perform the amplification. The amplicon was cloned into the p-EntrI Gateway entry vector (Life Technology). *SETBP1* was then subcloned into the pcDNA6.2/N-EmGFP-DEST destination vector using the Gateway clonase system (Life Technology).

We transfected 293T cells with 10 µg of plasmid DNA using Fugene Transfection Reagent (Roche) and selected cells with 10 µg/ml blasticidin. Cells expressing wild-type SETBP1 or SETBP1 Gly870Ser fused with GFP were sorted using a FACS Aria (BD Biosciences) flow cytometer (Supplementary Fig. 6).

Mutagenesis

SETBP1 Gly870Ser was generated using the following protocol. Specific primers (Supplementary Table 4) were designed and used to mutagenize the entry vector with the Pfu Ultra High Fidelity enzyme (Agilent). The product was digested with DpnI (Roche), and 2 µl was used to transform the competent TOP10 bacterial strain (Life Technology). The presence of the mutation encoding p.Gly870Ser was subsequently confirmed by Sanger sequencing.

Retroviral infection

The sequences encoding wild-type SETBP1 and SETBP1 Gly870Ser were excised from pENTR1A using the SalI and XhoI restriction enzymes and cloned into the MIGR1-EGFP vector⁴⁵ using the XhoI restriction site. Phoenix packaging cells were transfected with 10 µg of MIGR1-*SETBP1* (encoding wild-type or Gly870Ser protein) or with empty MIGR1 vector using FuGENE6 (Promega), and retroviruses were collected after 3 d of culture. To generate TF1 cells stably infected with retroviruses, we transduced 5×10^4 cells by spin infection in retroviral supernatants supplemented with 4 µg/ml polybrene (Sigma-Aldrich) and 20% RPMI 1640. After 48 h, the cells expressing wild-type SETBP1 and SETBP1 Gly870Ser were resuspended in complete medium.

Immunoblotting

TF1 cells (1×10^7) were used for immunoblotting analysis. Cells were washed with PBS and resuspended in lysis buffer (0.025 M Tris (pH 8.0), 0.15 M NaCl, 1% NP-40, 0.01 M NaF, 1 mM EDTA, 1 mM DTT, 1 mM sodium orthovanadate and protease inhibitors). Cell lysates were centrifuged for 20 min at 18,000 *g*. Equal amounts of total protein were separated by 10% SDS-PAGE and probed with selected antibodies, including to SETBP1 (ab98222) and phosphorylated PP2A (Tyr307, clone E155) (Abcam), SET (clone F-9, Santa Cruz Biotechnology), PP2A (C subunit, clone 1D6, Millipore) and actin (a2066, Sigma-Aldrich).

Peptide pulldown assay

Biotinylated phosphorylated peptides encompassing the SETBP1 region (amino acids 859–879) of either wild-type or Gly870Ser protein were synthesized by Innovagen. In the pulldown experiments, 0.5 mg of streptavidin magnetic beads (Pierce Biotechnology) was washed using a magnetic stand, according to the manufacturer's instructions, and resuspended in 100 µl of TBS with 2% BSA. Each peptide (10 µg) was bound to the beads for 1 h at room temperature on a rotating device. One sample with no peptide was used as a control for unspecific binding. Beads were washed with wash buffer (0.1% Tween-20 in TBS) and resuspended in lysis buffer, and TF1 cell lysate (600 µg) was added. Samples were incubated for 2 h at 4 °C on a rotating device. Alternatively, 75 ng of recombinant SKP1 CUL1 (SCF) complexed to β-TrCP1 (Millipore) was used. The unbound fraction was collected as a loading control. Elution of bound proteins was performed with 30 µl of Laemmli buffer. Peptides were dephosphorylated by adding 20 U calf intestinal phosphatase (NEB) for 2 h at 37 °C before binding to streptavidin beads. Immunoblotting analysis was performed with antibody to β-TrCP1 (clone H-85, Santa Cruz Biotechnology).

PP2A activity assays

PP2A phosphatase assays were carried out using the PP2A IP Phosphatase Assay kit (Millipore) according to the manufacturer's protocol on 5×10^6 cells.

Proliferation assay

TF1 cells transduced with viruses expressing GFP fused with wild-type SETBP1 or SETBP1 Gly870Ser were seeded at a concentration of 5,000 cells per well in 96-well round-bottom cell culture plates with complete medium. Cell proliferation was measured at different time points with the tritiated thymidine incorporation assay as described previously⁴⁶. Each test was performed in quadruplicate and was repeated at least twice.

RNA sequencing

All RNA libraries were generated from 2 µg of total RNA extracted with TRIzol (Life Technology) using the standard protocol. RNA was processed according to the protocol for the Illumina TruSeq RNA Sample Preparation kit (FC-122-1001) with a modification in the fragmentation time: mRNA was sheared for 1 min at 94 °C, and, after ligation of the adapters, fragments of 400–500 bp were selected on 2% agarose gels. Libraries were sequenced on an Illumina Genome Analyzer IIx with 76-bp paired-end reads using Illumina TruSeq SBS kit v5.

Transcriptome profiling

Image processing and base calling were performed using Illumina Real Time Analysis Software RTA v1.9.35. Qseq files were deindexed and converted into the Sanger FastQ file format using in-house scripts. FastQ sequences were aligned to the human genome database (NCBI Build 36/hg18) using TopHat⁴⁷ (version 1.2.0) with default parameters. Reads were mapped using the gene and splice-junction models provided in the Human Ensembl annotation file (Homo_Sapiens.NCBI36.54.GTF). TopHat aligns the RNA-seq reads across the genome using the Bowtie⁴⁸ algorithm and then maps the initially unmappable reads to the known splice-junction sequences supplied by the annotation GTF file. A splice-junction map for cases with wild-type and mutated *SETBP1* was inferred by TopHat. Visual inspection of exon junction maps in the *SETBP1* gene by Integrated Genomic Viewer⁴⁹ confirmed that both the mutated and wild-type samples expressed the longer isoform of the gene, which encodes the SKI homologous region (Ensembl release 54, May 2009).

The quantitative gene expression profile was estimated by SAMMATE⁵⁰ (version 2.6.1) using Human Ensembl annotation file version 54 and the default parameters. Gene expression values for paired-end data were measured in FPKM⁵¹, which is a normalized measure of exonic read density and a measure of the concentration of a transcript^{50,51}. SAMMATE calculates the FPKM expression values for each gene, taking into account the reads mapped on exons or on exon-exon junctions. The Human Ensembl gene annotation file version 54 was used to infer gene expression for coding and non-coding transcripts. Starting from the read alignment information, stored in the BAM format, a matrix of FPKM expression values and read counts for 36,655 unique Ensembl genes was obtained by SAMMATE. To focus on the gene expression profile of known coding transcripts, a data set of 20,907 protein-coding Ensembl genes was selected from the whole transcriptome.

Differential expression profiles between samples with wild-type and mutant *SETBP1* were obtained using the DESeq⁵² algorithm (R package, version 1.8.1) and NOISeq (R method, version last modified 29 April 2011)⁵³: the FPKM values ranged from 0 to 75,497; 37% of protein-coding Ensembl genes showed FPKM values between 0 and 1. To avoid differentially expressed genes biased from very-low-count data or expression values, we filtered out the genes for which the maximum FPKM value (mean value) of the two groups

was less than 1, obtaining 13,106 FPKM values. To focus on differentially expressed genes, we selected the protein-coding genes with fold change of ≥ 3 . Out of these 1,465 genes, 197 showed differential expression in samples with wild-type and mutated *SETBP1* (false discovery rate < 0.1 or probability of differential expression ≥ 0.8). The list of differentially expressed genes (Supplementary Table 6) was annotated using the Ensembl gene annotation file (version 54) and IPA (Ingenuity Systems; see URLs). The list of TGF- β -related genes (Supplementary Table 6) was based on the Ingenuity Knowledge database. Finally, the FPKM expression values of the differentially expressed gene list were plotted in a heatmap using dChip software (see URLs). The 60th percentile value was used as the upper bound of FPKM expression values, and each FPKM was converted into a color scale from -1.5 (lower limit) to 1.5 (upper limit).

RNA-seq data were also used to investigate the possible presence of gene fusions. This was accomplished using FusionAnalyser software¹⁸.

Immunofluorescence

We seeded 293T cells expressing wild-type SETBP1 or SETBP1 Gly870Ser on glass coverslips in a six-well plate. After 24 h, cells were washed, fixed with 4% paraformaldehyde, incubated at room temperature for 30 min and treated with buffer containing 0.1 M glycine in PBS (pH 7.4) and 0.3% Triton X-100. Cells were stained with Alexa Fluor 546-conjugated phalloidin (Invitrogen) for 1 h at room temperature. TOTO-3 iodide (642/660, Invitrogen) was used for nuclear staining. Confocal microscopy was carried out on a Radiance 2100 laser scanning confocal microscope (Bio-Rad) equipped with a krypton/argon laser and a red laser diode.

Acknowledgments

We kindly acknowledge the contributions of S. Mori in the preparation of this manuscript, M. Viltadi for technical help, G. Cazzaniga (Tettamanti Foundation, San Gerardo Hospital) for the MIGR1-EGFP plasmid and C. Cecchetti, Z. Sortino and C. Rizzo for clinical sample management. We thank M. Vogel for her critical reading of the manuscript. This work was supported by Associazione Italiana per la Ricerca sul Cancro (AIRC) 2010 (IG-10092 to C.G.-P.); Programmi di ricerca di Rilevante Interesse Nazionale (PRIN) program (20084XBENM_004 to R.P.); Fondazione Cariplo (2009-2667 to C.G.-P.); the Lombardy Region (ID-16871 and ID14546A to C.G.-P. and FSE Dote Ricercatori 16-AR to S.R.); Leukaemia and Lymphoma Research (UK) grants (to N.C.P.C. and J.B.); the Basic Research Program of the Korea Research Foundation (R21-2007-000-10041-0 to D.-W.K.) (2007); and Mildred Scheel Stiftung fuer Krebsforschung (Deutsche Krebshilfe, Germany, grant 109590 to N.W.).

References

1. Vardiman JW, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood*. 2009; 114:937–951. [PubMed: 19357394]
2. Nowell PC, Hungerford DA. A minute chromosome in human chronic granulocytic leukemia. *Science*. 1960; 132:1497.
3. Kurzrock R, et al. *BCR* rearrangement-negative chronic myelogenous leukemia revisited. *J. Clin. Oncol.* 2001; 19:2915–2926. [PubMed: 11387365]
4. Druker BJ, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* 2001; 344:1031–1037. [PubMed: 11287972]
5. Rebora P, et al. Are chronic myeloid leukemia patients more at risk for second malignancies? A population-based study. *Am. J. Epidemiol.* 2010; 172:1028–1033. [PubMed: 20861143]
6. Gambacorti-Passerini C, et al. Multicenter independent assessment of outcomes in chronic myeloid leukemia patients treated with imatinib. *J. Natl. Cancer Inst.* 2011; 103:553–561. [PubMed: 21422402]
7. Goldman JM. Chronic myeloid leukemia: a historical perspective. *Semin. Hematol.* 2010; 47:302–311. [PubMed: 20875546]

8. Shah SP, et al. Mutation of *FOXL2* in granulosa-cell tumors of the ovary. *N. Engl. J. Med.* 2009; 360:2719–2729. [PubMed: 19516027]
9. Tiacci E, et al. *BRAF* mutations in hairy-cell leukemia. *N. Engl. J. Med.* 2011; 364:2305–2315. [PubMed: 21663470]
10. Mardis ER, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* 2009; 361:1058–1066. [PubMed: 19657110]
11. Longo DL. Tumor heterogeneity and personalized medicine. *N. Engl. J. Med.* 2012; 366:956–957. [PubMed: 22397658]
12. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature.* 2009; 458:719–724. [PubMed: 19360079]
13. Godley LA. Profiles in leukemia. *N. Engl. J. Med.* 2012; 366:1152–1153. [PubMed: 22417205]
14. Ernst T, et al. Inactivating mutations of the histone methyltransferase gene *EZH2* in myeloid disorders. *Nat. Genet.* 2010; 42:722–726. [PubMed: 20601953]
15. Minakuchi M, et al. Identification and characterization of SEB, a novel protein that binds to the acute undifferentiated leukemia-associated protein SET. *Eur. J. Biochem.* 2001; 268:1340–1351. [PubMed: 11231286]
16. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. *Nat. Methods.* 2010; 7:248–249. [PubMed: 20354512]
17. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 2009; 4:1073–1081. [PubMed: 19561590]
18. Piazza R, et al. FusionAnalyser: a new graphical, event-driven tool for fusion rearrangements discovery. *Nucleic Acids Res.* 2012; 40:e123. [PubMed: 22570408]
19. Dinkel H, et al. ELM—the database of eukaryotic linear motifs. *Nucleic Acids Res.* 2012; 40:D242–D251. [PubMed: 22110040]
20. Winston JT, et al. The SCF β -TRCP-ubiquitin ligase complex associates specifically with phosphorylated destruction motifs in I κ B α and β -catenin and stimulates I κ B α ubiquitination *in vitro*. *Genes Dev.* 1999; 13:270–283. [PubMed: 9990852]
21. Reiter A, Invernizzi R, Cross NCP, Cazzola M. Molecular basis of myelodysplastic/myeloproliferative neoplasms. *Haematologica.* 2009; 94:1634–1638. [PubMed: 19996113]
22. Cristóbal I, et al. SETBP1 overexpression is a novel leukemogenic mechanism that predicts adverse outcome in elderly patients with acute myeloid leukemia. *Blood.* 2010; 115:615–625. [PubMed: 19965692]
23. Panagopoulos I, et al. Fusion of *NUP98* and the SET binding protein 1 (*SETBP1*) gene in a paediatric acute T cell lymphoblastic leukaemia with t(11;18)(p15;q12). *Br. J. Haematol.* 2007; 136:294–296. [PubMed: 17233820]
24. Shah SP, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature.* 2012; 486:395–399. [PubMed: 22495314]
25. Manfredini R, et al. The kinetic status of hematopoietic stem cell subpopulations underlies a differential expression of genes involved in self-renewal, commitment, and engraftment. *Stem Cells.* 2005; 23:496–506. [PubMed: 15790771]
26. Oakley K, et al. Setbp1 promotes the self-renewal of murine myeloid progenitors via activation of *Hoxa9* and *Hoxa10*. *Blood.* 2012; 119:6099–6108. [PubMed: 22566606]
27. Neviani P, et al. The tumor suppressor PP2A is functionally inactivated in blast crisis CML through the inhibitory activity of the BCR/ABL-regulated SET protein. *Cancer Cell.* 2005; 8:355–368. [PubMed: 16286244]
28. Janssens V, Goris J, Van Hoof C. PP2A: the expected tumor suppressor. *Curr. Opin. Genet. Dev.* 2005; 15:34–41. [PubMed: 15661531]
29. Liu X, Sun Y, Weinberg RA, Lodish HF. Ski/Sno and TGF- β signaling. *Cytokine Growth Factor Rev.* 2001; 12:1–8. [PubMed: 11312113]
30. Hoischen A, et al. *De novo* mutations of *SETBP1* cause Schinzel-Giedion syndrome. *Nat. Genet.* 2010; 42:483–485. [PubMed: 20436468]
31. Mundy GR. The effects of Tgf- β on bone. *Ciba Found. Symp.* 1991; 157:137–143. [PubMed: 2070682]

32. Lehman AM, et al. Schinzel-Giedion syndrome: report of splenopancreatic fusion and proposed diagnostic criteria. *Am. J. Med. Genet. A.* 2008; 146A:1299–1306. [PubMed: 18398855]
33. Cross NC. Histone modification defects in developmental disorders and cancer. *Oncotarget.* 2012; 3:3–4. [PubMed: 22287508]
34. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25:1754–1760. [PubMed: 19451168]
35. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 2010; 11:R86. [PubMed: 20738864]
36. Blankenberg D, et al. Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* 2010;10.1–21. chapter 19 Unit 19. [PubMed: 20069535]
37. Giardine B, et al. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 2005; 15:1451–1455. [PubMed: 16169926]
38. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009; 25:2078–2079. [PubMed: 19505943]
39. Score J, et al. Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms. *Blood.* 2012; 119:1208–1213. [PubMed: 22053108]
40. Grand FH, et al. Frequent *CBL* mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms. *Blood.* 2009; 113:6182–6192. [PubMed: 19387008]
41. Ernst T, et al. Transcription factor mutations in myelodysplastic/myeloproliferative neoplasms. *Haematologica.* 2010; 95:1473–1480. [PubMed: 20421268]
42. Randles, RH.; Wolfe, DA. *Introduction to the Theory of Nonparametric Statistics.* Wiley; New York: 1979.
43. Marubini, E.; Valsecchi, MG. *Analyzing Survival Data from Clinical Trials and Observational Studies.* John Wiley & Sons; Chichester, UK: 1995.
44. Dang L, et al. Cancer-associated *IDH1* mutations produce 2-hydroxyglutarate. *Nature.* 2009; 462:739–744. [PubMed: 19935646]
45. Pear WS, et al. Efficient and rapid induction of a chronic myelogenous leukemia-like myeloproliferative disease in mice receiving P210 bcr/abl-transduced bone marrow. *Blood.* 1998; 92:3780–3792. [PubMed: 9808572]
46. le Coutre P, et al. *In vivo* eradication of human BCR/ABL-positive leukemia cells with an ABL kinase inhibitor. *J. Natl. Cancer Inst.* 1999; 91:163–168. [PubMed: 9923858]
47. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009; 25:1105–1111. [PubMed: 19289445]
48. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009; 10:R25. [PubMed: 19261174]
49. Robinson JT, et al. Integrative genomics viewer. *Nat. Biotechnol.* 2011; 29:24–26. [PubMed: 21221095]
50. Xu G, et al. SAMMate: a GUI tool for processing short read alignments in SAM/BAM format. *Source Code Biol. Med.* 2011; 6:2. [PubMed: 21232146]
51. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods.* 2008; 5:621–628. [PubMed: 18516045]
52. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010; 11:R106. [PubMed: 20979621]
53. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. *Genome Res.* 2011; 21:2213–2223. [PubMed: 21903743]

Figure 1.

Distribution of alterations on the SETBP1 protein. Five exons (blue bars) encode isoform A of the protein (1,596 amino acids). The SETBP1 sequence contains three AT hook domains (amino acids 584–596, 1,016–1,028, 1,451–1,463), a SKI homologous region (amino acids 706–917), a SET-binding domain (amino acids 1,292–1,488) and a repeat domain (amino acids 1,520–1,543). Altered amino acids identified in our analysis are highlighted: black circles represent alterations found in aCML samples, and green circles represent alterations found in other diseases. Variants confirmed as somatic are indicated in bold. SETBP1 numbering refers to the NCBI reference sequence NM_015559.2.

Figure 2.

Mutation profile of 61 aCML cases for a panel of 15 genes. (a) Total numbers of mutations for each case and each gene are reported. The different kinds of mutations are indicated by color. Total numbers of mutations are given on the right. (b) Distribution of mutations in aCML. For each gene, the percentage of mutations associated with either wild-type or mutated *SETBP1* is reported. *IDH1*, *RBBP4*, *NPM1*, *JAK2*, *FLT3* and *DNMT3A* were also analyzed, but no mutations were identified.

Figure 3.

Clinical findings in cases with wild-type and mutated *SETBP1*. (a–d) Overall survival ($P=0.01$) (a), white blood cell count ($P=0.008$) (b), hemoglobin concentration ($P=0.44$) (c) and platelet number ($P=0.16$) (d) in 14 aCML cases with mutated *SETBP1* and 24 cases with wild-type *SETBP1* (WT). Values are shown as median (horizontal line), 25th and 75th percentiles (boxes), and maximum–minimum ranges (dotted lines). Error bars, s.e.m.

Figure 4.

Interaction between β -TrCP1 and SETBP1. **(a)** The β -TrCP1 degron motif (amino acids 868–873) is highlighted in red on the SETBP1 protein schematic. The sequences of biotinylated phosphorylated peptides (amino acids 859–879) used in the experiments are given. Black circles represent alterations found in aCML samples; green circles represent alterations found in other diseases. **(b)** Peptide pulldown experiment performed using TF1 total cell lysate and phosphorylated peptides. Beads with no peptides were used to control for nonspecific binding. Immunoblotting for β -TrCP1 was performed on the bound fractions. Immunoblotting for actin on the unbound fractions was used as a loading control. **(c)** Peptide pulldown experiment using recombinant SCF- β -TrCP1 complex on phosphorylated (+ P) and dephosphorylated (– P) peptides representing either wild-type SETBP1 or SETBP1 Gly870Ser. Beads with no peptide were used to control for nonspecific binding. Immunoblotting for β -TrCP1 was performed on bound as well as unbound (control) fractions.

Figure 5.

Effects of the SETBP1 Gly870Ser alteration on SETBP1 and SET protein expression, PP2A activity and cell growth. (a–d) TF1 cells were transfected with empty vector (pMIGR1, EV) or vector expressing wild-type SETBP1 or SETBP1 Gly870Ser. (a) Immunoblotting for SETBP1 on whole-cell lysates. Whole normal fetal stomach lysate was used as a control for SETBP1 protein. Immunoblotting for actin was used as a loading control. Lanes were derived from the same gel and were juxtaposed. (b) Immunoblotting for SET, phosphorylated PP2A (pPP2A), PP2A and actin on whole-cell lysates. Immunoblotting for actin was used as a loading control. Densitometric analysis of the amount of SET protein normalized over actin signal is shown in the bar graph. Mean and s.e.m. values from three independent experiments are plotted. *** $P < 0.0001$ compared to cells expressing wild-type SETBP1. (c) Lysates were used to assess the activity of PP2A. The activity relative to cells expressing wild-type SETBP1 is reported. Mean and s.e.m. values from three independent experiments are plotted. *** $P < 0.0001$. (d) Growth rate of the cells as measured by tritiated thymidine incorporation. Each curve was normalized to its value at time 0. Mean and s.e.m. values from two independent experiments are plotted. *** $P < 0.0001$ compared to cells expressing wild-type SETBP1.

Table 1Frequency of *SETBP1* mutations in 644 patient samples and 344 cancer cell lines

Tumor type	Number of samples	Number of mutated samples	Percent mutated samples
AML	106	0	0
ALL	62	0	0
CLL	32	0	0
MDS	100	0	0
MPN			
CML	42	0	0
PMF	33	0	0
PV	42	0	0
ET	36	0	0
CNL	4	1	25
MDS/MPN			
aCML	70	17	24
Unclassified MDS/MPN	30	3	10
CMML	82	3	4
JMML	5	0	0
Cell lines			
Brain/CNS	11	0	0
Breast	24	0	0
Colorectal	28	0	0
Esophagus/stomach	20	0	0
Head and neck	8	0	0
Liver/bladder	15	0	0
NSCLC	103	0	0
SCLC	17	0	0
Melanoma/skin	25	0	0
Mesothelioma	15	0	0
Non-Hodgkin lymphoma	12	0	0
Other (kidney, thyroid, testis)	9	0	0
Ovarian	8	0	0
Pancreas	11	0	0
Prostate	7	0	0
Sarcoma	18	0	0
Uterus	13	0	0
Total	988	24	

Diseases were grouped according to World Health Organization 2008 classification. AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; CLL, chronic lymphocytic leukemia; MDS, myelodysplastic syndromes (RA, 31; RARS, 30; RAEB-1, 24; RAEB-2, 10; MDS-U, 5); MPN, myeloproliferative neoplasm; PMF, primary myelofibrosis; PV, polycythemia vera; ET, essential thrombocythemia; CNL, chronic

neutrophilic leukemia; MDS/MPN, myelodysplastic/myeloproliferative neoplasm; JMML, juvenile myelomonocytic leukemia; NSCLC non-small-cell lung cancer; SCLC, small-cell lung cancer. A detailed list of cell lines is available in supplementary table 5.