SETBP1 induces transcription of a network of development genes by acting as an epigenetic hub

Rocco Piazza1, Vera Magistroni1, Sara Redaelli1, Mario Mauri1, Luca Massimino1, Alessandro Sessa2, Marco Peronaci1, Maciej Lalowski3, Rabah Soliymani3, Caterina Mezzatesta1, Alessandra Pirola1, Federica Banfi2, Alicia Rubio2, Delphine Rea4, Fabio Stagno5, Emilio Usala6, Bruno Martino7, Leonardo Campiotti8, Michele Merli9, Francesco Passamonti10, Francesco Onida11, Alessandro Morotti12, Francesca Pavesi13, Marco Bregni14, Vania Broccoli2,15 Marc Baumann3 & Carlo Gambacorti-Passerini1

SETBP1 variants occur as somatic mutations in several hematological malignancies such as atypical chronic myeloid leukemia and as de novo germline mutations in the Schinzel-Giedion syndrome. Here we show that SETBP1 binds to gDNA in AT-rich promoter regions, causing activation of gene expression through recruitment of a HCF1/KMT2A/PHF8 epigenetic complex. Deletion of two AT-hooks abrogates the binding of SETBP1 to gDNA and impairs target gene upregulation. Genes controlled by SETBP1 such as MECOM are significantly upregulated in leukemias containing SETBP1 mutations. Gene ontology analysis of deregulated SETBP1 target genes indicates that they are also key controllers of visceral organ development and brain morphogenesis. In line with these findings, in utero brain electroporation of mutated SETBP1 causes impairment of mouse neurogenesis with a profound delay in neuronal migration. In summary, this work unveils a SETBP1 function that directly affects gene transcription and clarifies the mechanism operating in myeloid malignancies and in the Schinzel-Giedion syndrome caused by SETBP1 mutations.
Recently, we and others demonstrated the involvement of SETBP1 mutations in several hematological malignancies. In our work, we showed that SETBP1 somatic variants occur in a 4 amino acid mutational hotspot within the so-called SKI-homology domain. This mutational hotspot is part of a degron motif that specifies substrate recognition by the cognate SCF-β-TrCP E3 ubiquitin ligase. SETBP1 mutations cause a functional loss of the degron motif targeted by SCF-β-TrCP and responsible for the short half-life of the protein. Therefore, these mutations result in an increased half-life of the mutated SETBP1 protein causing its accumulation and inhibition of the PP2A phosphatase oncosuppressor plus the SETBP1–SET–PP2A axis.

SETBP1 mutations occurring in the same hotspot were previously found in a germline disease known as the Schinzel–Giedion syndrome (SGS). Despite the overlap between the mutations present in hematological disorders and in SGS, recent data suggest that somatic SETBP1 mutations found in leukemias are more disruptive to the degron than germline variants responsible for the onset of SGS. In SGS, germline SETBP1 mutations occur as de novo variants, causing a severe phenotype characterized by mental retardation associated with distorted neuronal layering, multi-organ development abnormalities, and higher than normal risk of tumors.

Although the first studies led to a reliable characterization of SETBP1 as an oncogene, several elements of SETBP1 activity were still unclear, in particular: (1) inhibition of PP2A phosphatase alone does not explain the SETBP1-dependent phenotype of SGS, and (2) SETBP1 possesses three conserved AT-hooks, therefore suggesting a role as a DNA-binding protein.

Preliminary evidence that murine Setbp1 is able to bind to genomic DNA (gDNA) was initially given by Oakley and colleagues by transducing murine bone marrow progenitors with high titer retrovirus expressing Setbp1 followed by chromatin immunoprecipitation (ChIP) experiments, the authors demonstrated binding of Setbp1 to Hoxa9/10 promoters and upregulation of the two genes. Using a similar murine model, Vishwakarma and colleagues showed binding of Setbp1 to the Runx1 promoter; this, however, was associated with down-modulation of RUNX1 expression.

In this study, we analyze the interaction between SETBP1 and gDNA on a global, unbiased scale and demonstrate that SETBP1 binds DNA in adenine-thymine (AT)-rich promoter regions, causing activation of gene expression through the recruitment of a SET1/KMT2A (MLL1) COMPASS-like complex.

**Results**

**SETBP1 as a DNA-binding protein.** Cristobal et al. showed a clear role for SETBP1 as a natural inhibitor of PP2A. However, SETBP1 possesses three conserved AT-hook domains, responsible for binding to the minor groove of AT-rich gDNA regions; thus their presence suggests a role for SETBP1 as a DNA-binding protein.

To test this hypothesis, we generated isogenic 293 FLP-In cell lines harboring wild-type (WT) and mutated (G870S) SETBP1 in fusion with the V5 tag. Other isogenic models, such as CRISPR-Cas9, could not be used given the absence of immunoprecipitation-grade anti-SETBP1 antibodies.

The G870S variant was chosen as it is the most frequent mutation found in myelodysplastic/myeloproliferative disorders, together with the D868N. The G870S line showed similar SETBP1 transcript levels but increased SETBP1 protein compared to the WT one (Supplementary Figure 1a, b), as expected. Anti-V5 ChIP-Seq experiments revealed a total of 3065 broad genomic regions as being bound by SETBP1-G870S (Supplementary Data 1). The presence of broad peaks occurred mainly in AT-rich regions (Fig. 1a; mean peaks A/T content: 65.8% vs. 59% for the whole human genome; p < 0.0001) in cells expressing both SETBP1-G870S and SETBP1-WT. De novo motif discovery identified a SETBP1 consensus binding site (Fig. 1b; AAAATAA/ T; p = 0.002) largely overlapping the AT-hook consensus motif of HMGA1. (AAAATA; http://hocomoco.autosome.ru/motif/HMGA1_HUMAN.H10MO.D), suggesting that SETBP1 binds gDNA through its AT-hook domains. Querying the Catalog of Inferred Sequence Binding Preferences (https://cisbp.cbr.utoronto.ca/index.php) for murine Setbp1 resulted in a very similar sequence, with an AAT trinucleotide as the core motif (http://cisbp.cbr.utoronto.ca/TFreport.php?searchTF=7008692_1.02), therefore suggesting that the DNA-binding motif for SETBP1/Sebp1 is conserved across evolution.

Independent ligand-fishing experiments confirmed the ability of SETBP1 to bind to gDNA (Fig. 1c).

Peak distribution analysis revealed enrichment around promoters (Fig. 1d; 58% and 65% of the total peak count reside at +/−50 Kb from each transcription start site (TSS) for WT and G870S, respectively). However, the binding of SETBP1 was not restricted to promoters but also present in enhancers, exonic, intronic, and intergenic regions (Fig. 1e). In line with the quantitative effect of SETBP1 mutations on SETBP1 protein stability, SETBP1-G870S and SETBP1-WT ChIP-Seq experiments revealed a global increase of SETBP1-G870S binding across all the regions tested (Fig. 1e; Jaccard p = 0.046).

Peak annotation identified ectopic binding of SETBP1-G870S to 277 genes (Supplementary Figure 2a-c; Supplementary Data 2) resulting in a strong functional enrichment for development-related biological processes such as nervous system, heart, and bone development (Fig. 1f). Notably, over one third of the binding regions lies within either evolutionary conserved regions, defined as 100 bp gDNA windows characterized by a human-mouse conservation ≥70% or DNase I hypersensitive clusters, thus suggesting that a dysregulation in their activation could lead to significant functional consequences (Fig. 1g).

**SETBP1 upregulates target genes at the transcriptional level.** To dissect the effect of SETBP1 at the transcriptional level, we generated RNA-Seq profiles for cells overexpressing SETBP1-G870S, SETBP1-WT, as well as an Empty control. Comparative analysis of cells overexpressing G870S vs. WT revealed a very similar profile (Fig. 2a, b). This finding supports the expected quantitative mechanism of action of SETBP1 mutations, as previously proposed.

This model is further supported by the evidence that virtually all the activating SETBP1 variants identified so far fall exactly within the PEST domain and >95% within 4 amino acids of the extremely short degron linear motif and that the disruption of the PEST domain causes subsequent impairment of the SETBP1–β-TrCP axis. To confirm this hypothesis in the context of transcriptional regulation, we analyzed the intersection between SETBP1 peaks generated in our ChIP experiments and ENCODE transcription factor binding sites (TFBSs). In the presence of mutated SETBP1, we detected an increase in the total number of shared TFBS; however, the putative binding partners did not change between SETBP1-WT and mutated lines (Supplementary Figure 3). These data highlight the presence of a quantitative rather than a qualitative effect for SETBP1 mutations, thus supporting the reduced degradation model.

Subsequent comparative RNA-Seq analysis of SETBP1-G870S vs. Empty showed the presence of 2687 differentially expressed genes (DEGs), 57% of which were downregulated and 43% upregulated (Fig. 2a, b; Supplementary Data 3). The intersection between genes bound by SETBP1 in promoter regions and DEGs
promoter occupancy and increase of H3K4me2 and H3K9ac being part of a nucleosome-remodeling complex involved in transcriptional activation (Fig. 3d; Supplementary Data 5). In line with the previous findings, correlation of SETBP1-binding regions as well as related epigenetic marks between SETBP1-G870S and SETBP1-WT was nearly perfect \((r = 0.985\) and \(p < 0.00001\) for anti-V5; \(r = 0.873\) and \(p < 0.00001\) for anti-H3K9Ac; Supplementary Figure 4).

Co-immunoprecipitation (Co-IP)/proteomics experiments directed against SETBP1 (Supplementary Data 6) revealed direct interaction of both WT and mutated SETBP1 with HCF1, a core protein of the SET1/KMT2A complex, responsible for H3K4 mono- and di-methylation\(^\text{22}\). These results were confirmed by independent immunoprecipitation/western blot experiments (Fig. 4a) and by acceptor photobleaching fluorescence resonance energy transfer (FRET) assays (Fig. 4b). In silico linear domain analysis\(^\text{23}\) revealed the presence of a putative HCF1-binding motif (HBM) occurring at position 991–994 of SETBP1 (Supplementary Figure 5). Deletion of this motif in WT and mutated cells (SETBP1ΔHBM and SETBP1-G870S-ΔHBM; Supplementary Figure 6) caused a complete and specific abrogation of SETBP1/HCF1 interaction in both lines (Fig. 4b); conversely, the known SETBP1-β-TrCP\(^\text{1}\) interaction present only in the WT protein was not affected by the deletion (Fig. 4c). To assess whether the loss of the SETBP1/HCF1 interaction could result in the impairment of the SETBP1 transcriptional machinery, we analyzed by Q-PCR the expression levels of a set of upregulated genes: the deletion of HBM resulted in a complete normalization of the expression level for all the tested genes (Fig. 4d; MECOM \(p = 0.01\); BMP5 \(p < 0.001\), PDE4D \(p < 0.001\), ERBB4 \(p = 0.036\).

SETBP1 is part of a multigenic epigenetic complex. AT-hook-containing proteins are often part of large chromatin remodeling complexes\(^\text{19–21}\). ChiP-Seq profiles of a set of histone marks associated with gene expression (H3K4me2, H3K4me3, H3K9ac, H3K27ac, and H3K36me3) revealed peak distribution enrichment around the promoter regions for all the tested marks with the exception of H3K36me3, which was enriched in gene bodies, as expected (Fig. 3a). Differential enrichment analysis in SETBP1-G870S vs. Empty showed a correlation between SETBP1 promoter occupancy and increase of H3K4me2 and H3K9ac (Fig. 3b, c; \(p = 1 \times 10^{-4}\) and \(p = 1.2 \times 10^{-16}\), respectively). A significant transcriptional upregulation was detected for the genes bound by SETBP1-G870S and displaying an increase in H3K4me2 and H3K9ac, corroborating the hypothesis of SETBP1 being part of a nucleosome-remodeling complex involved in transcriptional activation (Fig. 3d; Supplementary Data 5). In line with the previous findings, correlation of SETBP1-binding regions as well as related epigenetic marks between SETBP1-G870S and SETBP1-WT was nearly perfect \((r = 0.985\) and \(p < 0.00001\) for anti-V5; \(r = 0.873\) and \(p < 0.00001\) for anti-H3K9Ac; Supplementary Figure 4).

Co-immunoprecipitation (Co-IP)/proteomics experiments directed against SETBP1 (Supplementary Data 6) revealed direct interaction of both WT and mutated SETBP1 with HCF1, a core protein of the SET1/KMT2A complex, responsible for H3K4 mono- and di-methylation\(^\text{22}\). These results were confirmed by independent immunoprecipitation/western blot experiments (Fig. 4a) and by acceptor photobleaching fluorescence resonance energy transfer (FRET) assays (Fig. 4b). In silico linear domain analysis\(^\text{23}\) revealed the presence of a putative HCF1-binding motif (HBM) occurring at position 991–994 of SETBP1 (Supplementary Figure 5). Deletion of this motif in WT and mutated cells (SETBP1ΔHBM and SETBP1-G870S-ΔHBM; Supplementary Figure 6) caused a complete and specific abrogation of SETBP1/HCF1 interaction in both lines (Fig. 4b); conversely, the known SETBP1-β-TrCP\(^\text{1}\) interaction present only in the WT protein was not affected by the deletion (Fig. 4c). To assess whether the loss of the SETBP1/HCF1 interaction could result in the impairment of the SETBP1 transcriptional machinery, we analyzed by Q-PCR the expression levels of a set of upregulated genes: the deletion of HBM resulted in a complete normalization of the expression level for all the tested genes (Fig. 4d; MECOM \(p = 0.01\); BMP5 \(p < 0.001\), PDE4D \(p < 0.001\), ERBB4 \(p = 0.036\).
Finally, a direct interaction between SETBP1 and KMT2A was confirmed by Co-IP experiments (Fig. 4e).

Plant homeodomain (PHD) fingers interact with methylated histone tails and are typically found in proteins responsible for the epigenetic modulation of gene expression. Interestingly, germline mutations in two PHD family members, PHF8 and PHF6, have been identified as the cause of two X-linked mental retardation syndromes, namely Siderius X-linked mental retardation and Borjeson–Forsman–Lehmann syndrome, and somatic PHF6 mutations have been found in T cell acute...
also responsible for brain and craniofacial development in PHD members are part of the SETBP1 complex (Fig. 4g, h). The interaction of SETBP1 with PHF8 and PHF6, indicating that both possess H4K20 demethylase activity. Co-IP experiments confirmed the recruitment of SETBP1 on tri-methylated H3K4 in the context of KMT2A complexes through its N-terminal PHD finger24, exerting its demethylating activity on H4 Lysine 20. ChIP against H4K20me1 revealed a significant decrease in H4K20 mono-methylation in cells expressing mutated SETBP1 for all the tested SETBP1 target genes (Fig. 4f), suggesting that the SETBP1 complex possesses H4K20 demethylase activity. Co-IP experiments confirmed the interaction of SETBP1 with PHF8 and PHF6, indicating that both PHD members are part of the SETBP1 complex (Fig. 4g, h).

Taken globally, these data suggest a link between altered PHD activity and SGS phenotype.

**SETBP1 epigenetic machinery requires functional AT-hooks.** The identification of an AAAATAA/T consensus binding motif closely resembling that of HMG1 suggests that SETBP1 may bind to gDNA thanks to the presence of its AT-hook domains. To test this hypothesis, we generated new isogenic lines carrying deletions of the first (SETBP1-G870SΔAT1), second (SETBP1-G870SΔAT2), and both (SETBP1-G870SΔAT1,2) AT-hook domains (Supplementary Figure 7). The third AT-hook was not deleted as it resides within the SET-binding domain. Combined disruption of AT-hooks 1 and 2 resulted in a marked reduction of mRNA

lymphoblastic27 and acute myeloid leukemias (AMLs)28. PHF8 is also responsible for brain and craniofacial development in zebrafish29. While the role of PHF6 as an epigenetic modulator is less clear, PHF8 acts as a lysine demethylase. It can bind di- and tri-methylated H3K4 in the context of KMT2A complexes through its N-terminal PHD finger24, exerting its demethylating activity on H4 Lysine 20. ChIP against H4K20me1 revealed a significant decrease in H4K20 mono-methylation in cells expressing mutated SETBP1 for all the tested SETBP1 target genes (Fig. 4f), suggesting that the SETBP1 complex possesses H4K20 demethylase activity. Co-IP experiments confirmed the interaction of SETBP1 with PHF8 and PHF6, indicating that both PHD members are part of the SETBP1 complex (Fig. 4g, h).
**Fig. 4** SETBP1 interacts with the SET1/KMT2A COMPASS-like complex. a Co-immunoprecipitation was performed against the HCF1 protein (left) or the V5 flag (right) and blotted with an anti-V5 or HCF1 antibody. b, c FRET analysis showing physical interaction between SETBP1 and HCF1. Positive FRET signal was recorded in both couples of HCF1 and SETBP1 WT or G870S (b, green and blue lines), conversely no FRET signal was recorded for HCF1 and SETBP1ΔHBM or G870SΔHBM (b, red and orange lines). FRET between β-TrCP and SETBP1 variants was assayed to demonstrate that ΔHBM did not modify the known SETBP1-β-TrCP interaction (c). Acceptor photobleaching was performed after the third acquired frame and indicated with gray bars in both the graphs. Bars represent the standard error of three experiments. d Relative expression of SETBP1 target genes as assessed by Q-PCR in empty (black), SETBP1-G870S (orange), and SETBP1-G870S-ΔHBM (light orange) lines. e Co-immunoprecipitation was performed against the V5 flag and blotted with an anti-KMT2A antibody. f ChIP against H4K20me1 followed by Q-PCR on a set of SETBP1 target genes performed on Empty (black bars) and SETBP1-G870S cells (orange bars). g Co-immunoprecipitation was performed against the V5 flag and blotted with anti-PHF8 and anti-PHF6 antibodies. h Proposed model for SETBP1 epigenetic network. i Relative expression of SETBP1 target genes in cells transduced with empty vector, SETBP1-G870S, or SETBP1-G870S carrying deletion of the first (ΔATH1), second (ΔATH2), and both (ΔATH1,2) AT-hooks. j ChIP against SETBP1-G870S in cells transduced with empty vector, SETBP1-G870S, or SETBP1-G870S carrying deletion of the first (ΔATH1), second (ΔATH2), or both (ΔATH1,2) AT-hooks, followed by Q-PCR on a set of SETBP1 target genes. In panels d, f, i and j, statistical analysis was performed using t-test. Bars represent the standard error of three experiments. *p < 0.05; **p < 0.01; ***p < 0.001
expression for all the target genes under analysis (Fig. 4i; MECOM \( p = 0.014 \); BMP5 \( p < 0.001 \); PDE4D \( p < 0.001 \); ERBB4-G870S\[^{\Delta AT1} \] \( p < 0.001 \); ERBB4-G870S\[^{\Delta AT1,2} \] \( p = 0.1 \)). ChIP performed on target genes followed by Q-PCR confirmed the same effect (MECOM \( p < 0.001 \); BMP5 \( p < 0.001 \); PDE4D \( p < 0.001 \); ERBB4 \( p = 0.003 \)), indicating that SETBP1 binding to gDNA depends on the presence of functional AT-hooks (Fig. 4j). The fact that SETBP1 is commonly found in DNase I hypersensitive cluster regions together with its ability to interact with an epigenetic activator complex containing H3K4
methyltransferase activity suggests that SETBP1 should be able to control gene expression through modulation of chromatin accessibility. To test this hypothesis, we generated ATAC-Seq (assay for transposase-accessible chromatin using sequencing) experiments on the isogenic 293 FLP-In Empty, SETBP1-WT, and SETBP1-G870S cell lines. We assessed the ATAC-Seq signal in a region comprised between −2000 bp and +10,000 bp from the TSS of all the human genes. In line with the expected high accessibility of TSS regions, ATAC-Seq analysis revealed a peak at TSS in all the lines under test (Supplementary Figure 8a, b; Supplementary Data 7–9). Subsequently, we investigated whether a correlation could be found between the intensity of the relative ATAC-Seq signal at TSS and RNA expression for both SETBP1-G870S and SETBP1-WT. No clear relationship between differential ATAC-Seq signal at TSS and gene expression was observed. We next identified genes that were significantly up- or downregulated in the presence of either WT or mutated SETBP1 and that were associated with the peak detected by ATAC-Seq analysis. These experiments were performed throughout a set of iterations (50) characterized by progressively more stringent differential expression filtering criteria. A significant linear relationship between differential ATAC-Seq signal at TSS and RNA expression for both SETBP1-G870S and SETBP1-WT was observed (Supplementary Figure 8c). Taken globally, these data indicate that genes that are significantly up- or downregulated in the presence of either WT or mutated SETBP1 exhibit an increased chromosomal accessibility in their TSS regions as well as in gene bodies.

**MECOM is a direct transcriptional target of SETBP1.** In 2013, Makishima and colleagues reported that the presence of mutated SETBP1 was associated with upregulation of MECOM expression; however, a clear mechanistic explanation of this finding was missing. MECOM, located on the long arm of chromosome 3, is a transcription factor able to recruit both coactivators and corepressors. It is expressed in hematopoietic stem cells, playing an important role in hematopoiesis and in hematopoietic stem cell self-renewal. MECOM is often overexpressed, as a result of chromosomal translocations, in myelodysplastic syndromes and in approximately 10% of AML cases, being a strong negative prognostic marker for therapy response and survival.

In line with Makishima’s report, differential expression RNA-Seq (Fig. 5a; p = 0.04 and p = 0.08 for SETBP1-WT and SETBP1-G870S vs. Empty, respectively) and Q-PCR analyses (Fig. 5b; p = 0.01 and p = 0.001 for SETBP1-WT and SETBP1-G870S vs. Empty, respectively) showed that MECOM is upregulated in the vast majority of cells electroporated with SETBP1-G870S. A similar upregulation was detected in the human myeloid TF-1 cell line transduced with SETBP1-G870S (Supplementary Figure 9; p = 0.0001). ChIP-Seq data showed that both SETBP1-WT and SETBP1-G870S bind to the MECOM promoter, highlighting MECOM as a direct target of SETBP1 transcriptional activity (Fig. 5c). MECOM modulates the expression of a significant number of genes involved in hematopoietic stem cell proliferation and myeloid differentiation (Supplementary Data 10). In accordance with its previously observed transcriptional activity, we show here that MECOM target genes are also differentially expressed (p < 0.0001, Fig. 5d) in cells expressing SETBP1-G870S. To confirm these results in fresh leukemic cells, RNA-Seq and Q-PCR analyses were performed in 32 atypical chronic myeloid leukemia (aCML) cases (11 positive and 21 negative for SETBP1 somatic mutations). The results show that SETBP1-positive patients express higher levels of MECOM (Fig. 5e, f; p = 0.0002) and MECOM target genes (Fig. 5g, p < 0.0001), with excellent correlation between RNA-Seq and Q-PCR data (Fig. 5f).

**Mutated SETBP1 delays neuronal migration to the cortex.** Upon their general developmental delay, Schinzel–Giedion patients present severe neurological impairment characterized by microcephaly, altered neuronal layering, underdeveloped corpus callosum, ventriculomegaly, cortical atrophy, or dysplasia that cause seizures and severe mental retardation. Additionally, the mRNA of SETBP1 gene is expressed in both germline (high level of expression) and differentiated area (low level) of different brain regions including cerebral cortex (Supplementary Figure 10). To gain insight into the function of the mutated form of SETBP1 in nervous system development, we transduced radial glial progenitors of cerebral cortices of E13.5 mouse embryos with expression plasmids encoding SETBP1-WT, SETBP1-G870S, SETBP1-G870SΔAT1,2, and SETBP1-G870S-ΔHBM using an in utero electroporation system (Fig. 6a)36. The transduced cells and their progeny can be easily traced along their development thanks to the green fluorescent protein (GFP) expression. Two days after the procedure, almost the totality of the SETBP1-G870S electroporated cells remained stacked in the deep part of the developing cortical wall while many control cells (transduced with GFP only) were already migrating in the outward cortical region where postmitotic neurons reside (Fig. 6b; Supplementary Movie 1 and 2). In addition, in SETBP1-G870S-overexpressing tissue, the apical domain of the cortical layer where the neural progenitors are located was severely disorganized as shown by aberrant localization of TBR2+ intermediate progenitors that lost their typical strip arrangement in the basal part of the ventricular zone (Fig. 6c). Five days after surgery, many control GFP+ neurons were detected in the mature cerebral mantle zone, contributing to the fiber tract of the corpus callosum (Fig. 6d, arrow), while the vast majority of cells electroporated with SETBP1-G870S were incorrectly located at the deepest cortical tissue (65 vs. 32% of the control condition; t-test p < 0.001 in bins 1–3) with only few neurons projecting to the contralateral hemisphere through the corpus callosum (Fig. 6d). In line with the proposed quantitative model for SETBP1 mutations, overexpression of SETBP1-WT showed similar, albeit reduced, effects (Fig. 6d). Overexpression of mutated SETBP1 carrying either AT-hooks 1 and 2 double

---

**Fig. 5** Analysis of MECOM expression and downstream targets. **a** Box plot showing the RNA-Seq differential expression analysis of MECOM in the 293 FLP-In Empty, SETBP1-WT, and SETBP1-G870S cell models. The top and bottom of each box represent the first and third quartile, respectively; the internal line represents the median; the dot represents the mean. Experiments were performed in triplicate. **b** Q-PCR analysis of MECOM expression in the 293 FLP-In Empty, SETBP1-WT, and SETBP1-G870S cell models. The top and bottom of each box represent the first and third quartile, respectively; the internal line represents the median; the square represents the mean. Experiments were performed in triplicate; statistical analysis was performed using t-test. **c** SETBP1 ChIP-Seq coverage track and peak alignment to the hg19 reference genome (blue tracks) are superimposed to the different histone methylation tracks and KMT2A (MLL) ChIP-seq coverage track within the MECOM locus. The boxed histogram represents an independent ChIP experiment performed against the V5 flag in the FLP-In cells followed by a Q-PCR directed against the predicted SETBP1-G870S-binding locus on the MECOM promoter. ChIP was performed in triplicate; statistical analysis was performed using t-test. **d** Gene expression heatmap of MECOM target genes in three Empty/SETBP1-G870S FLP-In clones. **e** Differential MECOM expression as read counts per million of mapped reads (RPM) in 32 aCML patients carrying WT (21) or mutated (11) SETBP1. The top and bottom of the box represent the first and third quartile, respectively; the internal line represents the median. Statistical analysis was performed using t-test. **f** Linear correlation of MECOM expression as assessed by RNA-Seq (x axis) and Q-PCR (y axis). r represents the Pearson linear correlation coefficient. **g** Gene expression heatmap of MECOM target genes in 32 aCML patients carrying WT (21) or mutated SETBP1 (11). Error bars represent the standard error. *p < 0.05; **p < 0.01; ***p < 0.001.
we focused on the potential role of SETBP1 as a transcriptional gene promoter regions, as SETBP1 has also been detected in neurons as con...modulator, showing that SETBP1 interacts with gDNA through its AT-hook domains, forming a multiprotein complex including HCF1, KMT2A, PHF8, and PHF6, which results in increased chromatin accessibility, as assessed by ATAC-Seq (Supplementary Figure 8), and transcriptional activation. The altered transcription caused by the increased levels of SETBP1-G870S was also shown to be involved in the pathogenesis of SGS, aCML, and related myeloid malignancies, as revealed by in utero electroporation of SETBP1-G870S in ventricular central nervous system cells and by the analysis of aCML samples.

Oakley et al. previously showed that, in murine bone marrow, progenitors transduced with high-titer retrovirus-expressing Setbp1 are able to bind to Hoxa9/10 and Myb promoters and to upregulate these genes. Vishwakarma and colleagues, using a similar murine model, demonstrated binding of Setbp1 to the Runx1 promoter, which, however, was associated with Runx1 downmodulation. Interestingly, in our human 293 FLP-In model we found a similar downregulation (Supplementary Figure 11a). ChIP experiments, however, demonstrated only a very weak binding of SETBP1 to the human RUNX1 promoter (Supplementary Figure 11b), likely due to the expression of SETBP1 being much lower in our model than in traditional high-titer retroviral transduction systems or to differences in species specificity. To test this hypothesis, we repeated ChIP experiments...
using a transient transfection model, achieving a 4.2- and 6.9-fold increase in SETBP1-WT and G870S expression, respectively, when compared with FLP-In (Supplementary Figure 11c). Indeed, ChIP experiments performed using transient SETBP1 transfectants showed a significant increase in the binding to RUNX1 promoter (Supplementary Figure 11d). This, however, was accompanied by an increase in RUNX1 expression (Supplementary Figure 11e), confirming that SETBP1 promotes upregulation of gene expression and suggesting that RUNX1 down-modulation is not a direct effect of SETBP1. Further studies will be required to assess whether the differences between our model and the murine model developed by Vishwakarma and colleagues are species specific. DEG analysis between SETBP1-G870S and Empty cells in our FLP-In model and in aCML patients (11 positive and 21 negative for SETBP1 somatic mutations) failed to reveal significant changes in the expression of RUNX1, HOX9/10, or MYB (Supplementary Figure 12), corroborating the hypothesis that the effect of SETBP1 as a transcriptional activator complex. The exact explanation of this phenomenon together with the increase in H3K9Ac and the direct interaction of SETBP1 with PHF8 demethylase reveal that SETBP1 is part of a transcriptional activator complex. The exact explanation of this complex histone pattern is still unclear, given also our limited knowledge about the functions of the H3K4me2 mark. Nevertheless, several lines of evidence suggest that H3K4me2 is strongly enriched in lineage-specific gene promoters and the abundance of genes associated with multi-organ development among those regulated by SETBP1 is particularly interesting in the context of SGS, whose hallmark is the presence of multi-organ development abnormalities. Indeed, analysis of the Gene Ontologies associated with DEGs in SETBP1-G870S cells highlights the presence of a much stronger association with SGS than with myeloid malignancies, especially related to HGAM, which is one of the genes whose expression is most constitutive in SGS, somatic SETBP1 mutations are typically present in conjunction with several other somatic events in myeloid neoplasms, in which SETBP1 mutations usually represent late events1, 2.

In summary, we showed here that SETBP1 is able to recruit a HCF1/KMT2A/PHF8/PHF6 transcriptional activator complex, thus identifying SETBP1 as a transcriptional activator, beyond its original SET-binding activity. Further studies will be required to gain further insights into this complex network and to dissect the functional role of SETBP1–DNA interaction in intergenic regions.

**Methods**

**Patients.** Diagnosis of aCML and related diseases was performed according to the World Health Organization 2008 classification. All patients provided written informed consent, which was approved by the institutional ethics committee. This study was conducted in accordance with the Declaration of Helsinki.

**Cell-lines.** 293T, TF-1, and the 293 FLP-In cell lines were purchased from ATCC (Manassas, VA, USA), DSMZ (Braunschweig, Germany), and Thermo Fisher Scientific (Waltham, MA, USA), respectively, and maintained following the manufacturers’ instructions.

**Plasmids and transfections.** Stable 293 FLP-In Empty, 293 FLP-In SETBP1wt, and 293 FLP-In SETBP1-G870S cell lines were prepared cotransfecting pOG44 (Thermo Fisher Scientific) and pCGN1-SETBP1 vectors peptide sequence G870S, or empty vectors1 with FuGENE6 (Promega); retroviruses were collected after 3 days of culture. Transient 293T transfectants were prepared using the pDNA6.2-SETBP1wt, pDNA6.2-G870S, or empty vectors1 with FuGENE6 reagent (Promega, Madison, WI, USA) following standard protocols. The deletion of the HBREM site (aa991–994; NM_015559.2) within the SETBP1 coding sequence was performed with the following primers: SETBP1_HCF1del_for (CAGCTTCTATGATATCCCTCTCCTCTGAGATTCA), SETBP1_HCF1del_rev (CATACTGGATATATGGCAGCGAAATTATGCGAAGAGTAGCT), the deletion of the ATG870S, or empty vectors1 with FuGENE6 reagent (Promega, Madison, WI, USA) following standard protocols. The deletion of the HREM site was performed using the following primers as previously described42.

**ChIP sequencing (ChIP-Seq).** ChIP was performed as previously reported43 (GEO accession number: GSE86335). Briefly, proteins were crosslinked with 0.4% formaldehyde and cells were lysed. Chromatin was fragmented with a BioRuptor sonicator system (Diagenode, SA, USA) and subsequently immunoprecipitated with three monoclonal antibodies specific to histone H3K4me2 (ab8580, Abcam, UK), H3K4me3 (C15410035C, Diagenode), and H3K36me3 (ab9050, Abcam), H3K27ac (AbI729, Abcam), H3K9ac (39137, Active Motif, Carlsbad, CA, USA), and H3K20me1 (Mab1470–014, Diagenode) antibodies or anti-V5 Vero beads (Sigma-Aldrich). After immunoprecipitation, DNA was purified and libraries were prepared for sequencing following the Illumina ChIP-Seq protocol (TrueSeq ChIP library prep kit IP-202–1012) with an Illumina HiSeq2500 in single read mode (Galseq, Monza, Italy). Validation of ChIP-Seq data was performed amplifying with the immunoprecipitated DNA with SybrGreen Q-PCR. Input was used as a loading control. Primer sequences were SETBP1_Fw (CAACCGCTGTGGGAAAGAGAAC), BMP5_Rw (TCTATGCACTGATTAGGCCAACC), NFE2L2_Fw (AACCGAAGAATTTCAATCACTTACCAAT), NFE2L2_Rw (AAGAGAATTCTGCTGATCTTGGT), PDE4D_Fw (CCCTCAGCCCAATCTCTCCTC), PDE4D_Rw (CAACCGAAGAATTTCAATCACTTACCAAT)

**ARTICLE**

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-04462-8 | www.nature.com/naturecommunications

**Figure 11.** Inseparable regions.

**Figure 12.** A complex mechanism that may lead to the recognition of SETBP1 as a transcriptional activation complex.

**Figure 13.** The functional role of SETBP1 in the transcriptional activity of aCML and related disorders, the hyperactivation of the SETBP1–SET axis caused by the stabilization of mutated SETBP1 protein and leading to the inhibition of the PP2A oncosuppressor probably plays a major functional role in driving the leukemic phenotype, while the interaction with gDNA and the subsequent modulation of gene expression likely plays a critical role in the onset of SGS, in which the SETBP1-mediated inhibition of PP2A is probably insufficient to recapitulate the complex phenotype of the SGS syndrome. However, the transcriptional activity of SETBP1 likely contributes to the leukemic phenotype too. In 2013, Makishima and colleagues reported that cases with SETBP1 mutations were characterized by high expression of the MECOM oncogene. Here we show that MECOM is a direct target of SETBP1 transcriptional activity, which explains the original observation. Interestingly, Goya and colleagues previously showed that SETBP1 is one of the genes whose expression is most strongly reduced in MECOM−/− hematopoietic cells, suggesting that SETBP1 is also one of the MECOM downstream targets. Taken together, these findings suggest the presence of a positive feedforward loop occurring between MECOM and SETBP1, whose biological importance and effects will require further studies. These findings highlight a potential critical role for this transcriptional complex and its downstream effectors in the oncogenesis of SETBP1-positive myeloid disorders, therefore suggesting that a strict dichotomous model is too simplistic to fully recapitulate and explain both phenotypes. It is also important to note that, while de novo germline SETBP1 mutations constitute the only genetic alteration in SGS, somatic SETBP1 mutations are typically present in conjunction with several other somatic events in myeloid neoplasms, in which SETBP1 mutations usually represent late events1, 2.

**In summary, we showed here that SETBP1 is able to recruit a HCF1/KMT2A/PHF8/PHF6 transcriptional activator complex, thus identifying SETBP1 as a transcriptional activator, beyond its original SET-binding activity. Further studies will be required to gain further insights into this complex network and to dissect the functional role of SETBP1–DNA interaction in intergenic regions.**
narrow or broad (SETBP1 peak calling parameters).\textsuperscript{48} Transcription factor association strength and relative fold change were calculated as previously reported.\textsuperscript{47} Downstream statistics, such as multivariate analysis with linear correlation assumptions, were performed with the IBM SPSS statistical package.

**ATAC-Seq analysis.** Cells (100,000/sample) were washed once in cold phosphate-buffered saline (PBS) 1×, spun at 800 × g for 5’ at 4 °C and resuspended in cold PBS in the presence of proteasome inhibitors and incubated on ice for 10’. Cells were centrifuged at 800 × g at 4 °C for 10’ and resuspended in 2× TD buffer (25 µl), Tris Transpose (2.5 µl; Illumina), and water (up to 50 µl). The sample was incubated at 37 °C for 30’ and purified using the SPRI Ampure XP beads. Post-tagmentation amplification was performed using Nextera primers (Illumina) and Herculease II polymerase (Agilent) using a standard protocol. Quality of the ATAC-Seq libraries was assessed using a Tape Station (Agilent) and by agar gel. Quantification was performed using a Qubit Fluometer (ThermoFisher).

**Epigenetic marks correlation tool.** Anti-V5 and anti-H3K8ac 293T-SETBP1 WT and G870ts coverage files were used as input. Coverage was binned (BigWig) in 1000 bp regions for all the ChIP-Seq experiments. Individual bins for all the regions across the entire genome were paired and used for Pearson correlation analysis as well as for XY plots.

**Oligonucleotide pulldown assay.** Biotinylated oligonucleotides were synthesized by Metabion International AG (Steinkirchen, Germany). Target and non-target (unrelated) oligonucleotides were designed according to ChIP-Seq from SETBP1 bound (target-T, chr6:55,742,037-55,742,136; hg19) and unbound (unrelated-U, chr6:55,775,477-55,775,576; hg19) regions. In the pulldown experiment, 25 µl of beads was bound to the beads for 20 min at room temperature (RT) on a rotating device. One sample without probe was used as a control for unspecific binding. Beads were washed with TE buffer and resuspended in 200 µl of B/W buffer (10 mM Tris HCl, pH 7.5, 1 mM EDTA, 2 M NaCl). Each double-stranded biotinylated DNA probe (200 nM) was bound to the beads for 20 min at room temperature (RT) on a rotating device. One sample without probe was used as a control for unspecific binding. Beads were washed with TE buffer and resuspended in 100 µl of BS/THES buffer (22 mM Tris HCl, pH 7.5, 4.4 mM EDTA, 8.9% Sucrose, 62 mM NaCl, 10 mM Hepes, 5 mM CaCl2, 50 mM KCl, 12% Glycerol) supplemented with HALT protease inhibitor cocktail (Thermo Fisher Scientific). In all, 15 µg of nuclear extract from 293 FLP in transfection (target-T) or 1 µg of nuclear extract from 293 FLP in transfection (unrelated-U) was used for pulldown experiment, 25 µl of Pierce-streptavidin magnetic beads (Thermo Fisher Scientific) were washed using a magnetic stand and resuspended in 200 µl of B/W buffer (10 mM Tris HCl, pH 7.5, 1 mM EDTA, 2 M NaCl). Each double-stranded biotinylated DNA probe (200 nM) was bound to the beads for 20 min at room temperature (RT) on a rotating device. One sample without probe was used as a control for unspecific binding. Beads were washed with TE buffer and resuspended in 100 µl of BS/THES buffer (22 mM Tris HCl, pH 7.5, 4.4 mM EDTA, 8.9% Sucrose, 62 mM NaCl, 10 mM Hepes, 5 mM CaCl2, 50 mM KCl, 12% Glycerol) supplemented with HALT protease inhibitor cocktail (Thermo Fisher Scientific). In all, 15 µg of nuclear extract from 293 FLP in transfection (target-T) or 1 µg of nuclear extract from 293 FLP in transfection (unrelated-U) was used for pulldown experiment. Following washing twice with cold PBS, lysed in Buffer 1 (Pipes, pH 8.0, 5 mM, KCl 85 mM, NP-40 0.5%), supplemented with protease inhibitors (Halt Protease Inhibitor Cocktail, Thermo Fisher Scientific), kept for 10 min on ice, homogenized with douncer homogenizer (10000) and centrifuged at 2000 × g for 10 min. The pellet was then resuspended in Buffer 2 (Tris HCl, pH 8.0, 50 mM, sodium dodecyl sulfate 0.1%, deoxicholate 0.5%) supplemented with protease inhibitors, sonicated with Bioruptor Next Gen (Diagenode) (5 cycles 30 s ON, +30 s OFF) to promote gDNA disruption and centrifuged at 18000 × g for 10 min. The supernatant representing the soluble fraction was quaniticated with Bradford assay and Bradford fraction of protein was loaded on 100 µl V5-agarose beads (Sigma-Aldrich) and incubated under rotation overnight at 4 °C. Beads were then washed three times with +protease inhibitors and elution was performed with 7 M Urea, 2 M Thiourae, and 4% CHAPS for subsequent mass spectrometric analysis or with Laemml buffer for western blot analysis. All chemicals were purchased from Sigma Aldrich.

**Proteomics data analysis.** The protein samples were digested in Amicon Ultra-0.5 centrifugal filters using modified FASP method.\textsuperscript{49} The peptides were separated with the nanoAcquity UPLC system (Waters) equipped with a 5-µm Symmetry C18 trapping column, 180 µm×20 mm, reverse-phase (Waters), followed by an analytical 1.7-µm, 75 µm×250 mm BEH-C18 reversed-phase column (Waters), in a single-pump trapping mode. The parameters of the HD-MSE runs were described previously.\textsuperscript{50} Protein identifications were performed with Proteinlynx Global Server (PLGS v3.0) as described.\textsuperscript{51} Database searches were carried out against UniProt human protein database (release_97072015, 71097 entries) with Ion Accounting algorithm and using the following parameters: peptide and fragment tolerance: automatic, maximum protein mass: 500 kDa, minimum fragment ions matches per protein: 7, minimum fragment ions matches per peptide: 3, minimum peptide matches per protein: 1, primary digest reagent: trypsin, missed cleavages allowed: 2, fixed modification: carboxymethylation C, variable modifications: deamidation (N, Q), oxidation of Methionine (M) and FDR < 4%.

**Immunolab analysis.** Primary antibodies were V5 (ab27671 330 Abcam, Cambridge, UK; dilution 1:2000), HCF1 (A301-399A BetaLabs, Inc., Montgomery, TX, USA; dilution 1:1000), ML1 (14689 Cell Signaling Technology, Danvers, MA, USA; dilution 1:1000), PHF8 (A301-772A BetaLabs, Inc., Montgomery, TX, USA; dilution 1:500) and PHF6 (A301-451A BetaLabs Laboratories, Inc., Montgomery, TX, USA; dilution 1:900). Secondary antibody was anti-mouse Alexa Fluor 647 or anti-rabbit horseradish peroxidase (HRP) conjugated (both Sigma-Aldrich, USA). The uncropped scan of western blots related to PHF6 immunoprecipitation are shown in Supplementary Figure 13 as an example.

**Immunofluorescence microscopy.** Transfected HEK 293 cells, seeded on glass coverslips coated with poly-D-lysine (0.1 mg ml\(^{-1}\)), were washed twice with PBS and fixed for 10 min at 25 °C with 4% (v/v) paraformaldehyde in 0.12 M sodium phosphate buffer, pH 7.4, incubated for 1 h with primary antibodies in GDB buffer (0.02 M sodium phosphate buffer, pH 7.4, containing 0.45 M NaCl, 0.2% (v/v) bovine gelatin) followed by staining with conjugated secondary anti IgG antibodies for 1 h. After two washes with PBS, coverslips were mounted on glass slides with a 90% (v/v) glycerol/PBS solution.

**FRET.** FRET measurements were performed with the laser-induced acceptor photobleaching method.\textsuperscript{52,53} In our experimental set-up, FRET couples analyzed were made up of transfectantly transfected GFP-fused SETBP1 WT or mutated forms in combination with transiently transfected Flag-tagged β-TrCP or endogenous HCF1. Forty eight hours after transfection, HEK 293 cells were labeled with proper primary and secondary antibodies and imaged for FRET analysis. Used FRET couples were GFP signal as donor fluorochrome and Alexa Fluor 555-conjugated secondary antibodies as acceptor fluorochrome.\textsuperscript{56} Briefly, three images were cap- }
RNA-Seq. RNA libraries were generated starting from 1 µg of total RNA extracted from 5 × 10^6 cells using TRIzol (Invitrogen, Life Technologies). RNA quality was assessed using a Bioanalyzer instrument (Agilent). To avoid over-representation of 3′-ends, only high-quality RNA with a RNA Integrity Number 28 was used. RNA was processed according to the TruSeq Stranded mRNA Library Prep Kit protocol. The libraries were sequenced on an Illumina HiSeq 3000 with 76 bp paired-end reads using Illumina TruSeq technology (GEO accession number: GSE63355).

Image processing and basecall was performed using the Illumina Real Time Analysis Software. Fastq files were aligned to the human genome (GRCh38/hg39) by using STAR94, a splice junction mapper for RNA-Seq data, together with the following splicing junctions: Ensembl GTF annotation, using the following parameters:

```bash
--runThreadN 8 --outReadsUnmapped Fastx --outFilterType BySout --outSAMattrIBs NH HI AS nM MD --outFilterMultimapNmax 20 --alignInferNoverhangMin 8 --alignSJoverhangMin 1 --alignSJMismatchesNmax 999 --alignSO mismatchesNoverLink 0.04 --alignIntronMin 20 --alignIntronMax 10000000 --alignTranscriptsPerReadNmax 1000000 --quantMode TranscriptsSAM GeneCounts --limitBMsortRAM 1662057818 --outSAMtype BAM SortedByCoordinate --chimSegmentMin 20 --chimJunctionOverhangMin 10.
```

Quantitative real-time PCR. Total RNA was extracted with Trizol reagent following standard protocol (Thermo Fisher Scientific). One microgram of RNA was used to synthesize cDNA using Reverse Transcription Reagents (Thermo Fisher Scientific) after pretreatment with DNAseal to avoid contamination from gDNA. Real-time Q-PCR was performed using TaqMan® Brilliant II QPCR Master Mix (Agilent Technologies, Santa Clara, CA, USA) on a Stratagene-MX3005P (Agilent) following standard conditions. The housekeeping gene GUS was used as an internal reference.95 TaqMan® Gene Expression Assays (Thermo Fisher Scientific) were used: SKIDA1 (Hs01096520), NFE2L2 (Hs01597588), SETBP1 (Hs00210203), RUNX1 (Hs00975961), PDE4D (Hs03988495), FBXO8 (Hs00942619), CEP44 (Hs00955525), CDKN1B (Hs001957588), SETBP1-G870S (Hs00722035), RUNX1 (Hs01021971), MECOM (Hs00627925).

Differential expression analysis. The gene counts generated by the--quantMode GeneCounts parameter were used to calculate gene expression. All the subsequent steps were performed using R scripts. These scripts were automatically generated through metaprogramming, using a dedicated tool (StarCounts2DESeq2) developed by the authors using R and C as a meta-language. Differential expression analyses were performed using DESeq2. Analysis of DEGs characterized by the co-occurrence of both H3K4me2 and H3K9Ac marks was performed by multivariate regression analysis using the following filtering criteria: transcription factor and histone mark relative coverage enrichment >0.5 (log2 FC G870S/Empty); differential gene expression p-value < 0.05 (G870S vs. Empty, DESeq2).

In utero electroporation. SETBP1-G870S was cloned in the pCAG expression vector.96 In utero electroporation was carried out at E13.5 to target the expression vectors into the ventricular radial glial cells of the mouse embryos, as described previously.94,97 After electroporation, embryos were placed back into the abdomen, the cavity was filled with warm sterile PBS, and the abdominal muscle and skin incisions were closed with silk sutures. All experiments were conducted according to the guidelines of the Animal Care and Use Committee of the University of California, San Francisco.

Data availability. All relevant sequencing data are available at the NCBI GEO Server, accession code: GSE63355.

Received: 14 December 2016 Accepted: 27 April 2017
Published online: 06 June 2018

References
1. Piazza, R. et al. Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. Nat. Genet. 45, 18–24 (2013).
2. Sakaguchi, H. et al. Exome sequencing identifies secondary mutations of SETBP1 and JAK3 in juvenile myelomonocytic leukemia. Nat. Genet. 45, 937–941 (2013).
3. Makushima, H. et al. Somatic SETBP1 mutations in myeloid malignancies. Nat. Genet. 45, 942–950 (2013).
4. Meggendorfer, M. et al. SETBP1 mutations occur in 9% of MDS/MPN and in 4% of MNP cases and are strongly associated with atypical CML, monosomy 7, isochromosome i(17)(q10), ASXL1 and CBL mutations. Leukemia 27, 1852–1860 (2013).
5. Inoue, D. et al. SETBP1 mutations drive leukemic transformation in ASXL1-mutated MDS. Leukemia 29, 847–857 (2015).
6. Fabiani, E. et al. SETBP1 mutations in 106 patients with therapy-related myeloid neoplasms. Haematologica 99, 152–153 (2014).
7. Patnaik, M. M. et al. ASXL1 and SETBP1 mutations and their prognostic contribution in juvenile myelomonocytic leukemia: a two-center study of 466 patients. Leukemia 28, 2206–2212 (2014).
8. Theil, F. et al. SETBP1 mutation analysis in 944 patients with MDS and AML. Leukemia 27, 2072–2075 (2013).
9. Wang, X. et al. GATA2 and secondary mutations in familial myelodysplastic syndromes and pediatric myeloid malignancies. Haematologica 100, e398–401 (2015).
10. Houschen, A. et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat. Genet. 42, 483–485 (2010).
11. Acuna-Hidalgo, R. et al. Overlapping SETBP1 gain-of-function mutations in Schinzel-Giedion syndrome and hematologic malignancies. PLoS Genet. 13, e1006683 (2017).
12. Lach, B. & Arredondo, J. Cobblestone lissencephaly in Schinzel-Giedion syndrome. J. Child Neurol. 28, 259–263 (2013).
13. Oakley, K. et al. Setbp1 promotes the self-renewal of murine myeloid progenitors via activation of Hoxa9 and Hoxa10. Blood 119, 6099–6108 (2012).
14. Vishwakarma, B. A. et al. Runx1 repression by histone deacetylation is critical for Setbp1-induced mouse myeloid leukemia development. Leukemia 30, 200–208 (2016).
15. Cristobal, I. et al. SETBP1 overexpression is a novel leukemogenic mechanism that predicts adverse outcome in elderly patients with acute myeloid leukemia. Blood 115, 615–625 (2010).
16. Kulakovsky, I. V. et al. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Res. 44, D116–125 (2016).
17. Weirach, M. T. et al. Determination and inference of eucharyotic transcription factor sequence specificity. Cell 158, 1431–1443 (2014).
18. Ovcharenko, I., Nobrega, M. A., Loots, G. G. & Stubb, L. ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. Nucleic Acids Res. 32, W280–286 (2004).
19. Bourachot, B., Yaniv, M. & Muchardt, C. The activity of mammalian brm/SNF2alpha is dependent on a high-mobility-group protein I/Y-like DNA binding domain. Mol. Cell. Biol. 19, 3931–3939 (1999).
20. Cairns, B. R. et al. Two functionally distinct forms of the RSC nucleosome-remodeling complex, containing essential AT hook, BAH, and eukaryotic linear motif resource. Mol. Cell 4, 715–723 (1999).
21. Xiao, H. et al. Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interaction. Mol. Cell 8, 531–543 (2001).
22. Rao, R. C. & Dou, Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. Nat. Rev. Cancer 15, 334–346 (2015).
23. Dinkel, H. et al. ELM 2016—data update and new functionality of the eukaryotic linear motif resource. Nucleic Acids Res. 44, D294–300 (2016).
24. Sanchez, R. & Zhou, M. M. The PHD finger: a versatile epigenome reader. Trends Biochem. Sci. 36, 364–372 (2011).
25. Siderius, L. E. et al. X-linked mental retardation associated with clef lippalate maps to Xp11.3–q21.3. Am. J. Med. Genet. 85, 216–220 (1999).
26. Lower, K. M. et al. Mutations in PHF6 are associated with Forssman-Lehmann syndrome. Nat. Genet. 32, 661–665 (2002).
27. Van Vlerkenbergh, P. et al. PHF6 mutations in T-cell acute lymphoblastic leukemia. Nat. Genet. 42, 339–342 (2010).
28. Van Vlerkenbergh, P. et al. PHF6 mutations in adult acute myeloid leukemia. Leukemia 25, 130–134 (2011).
29. Qi, H. H. et al. Histone H4K20/H3K9 dimethylate PHF8 regulates zebrafish brain and craniofacial development. Nature 466, 503–507 (2010).
30. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218 (2013).
31. Wieser, R. The oncogenic and developmental regulator EVI1: expression, biochemical properties, and biological functions. Gene 396, 346–357 (2007).
32. Kataoka, K. & Kurokawa, M. Ecotropic viral integration site 1, stem cell self-renewal and leukemogenesis. Cancer Sci. 103, 1371–1377 (2012).
33. Barjesteh van Waalwijk van Doorn-Khosrovani, S. et al. High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients. Blood 101, 837–845 (2003).
34. Kataoka, K. & Kurokawa, M. Ecotropic viral integration site 1 gene expression predicts outcome in younger adult patients with acute myeloid leukemia and is associated with distinct cytogenetic abnormalities. J. Clin. Oncol. 28, 2101–2107 (2010).
35. Bard-Chapeau, E. A. et al. Ecotopic viral integration site 1 (EVI1) regulates biochemical properties, and biological functions.
36. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS).
37. Nguyen, N. et al. Myb expression is critical for myeloid leukemia development.
38. Pekowska, A., Benoukraf, T., Ferrier, P. & Spicuglia, S. A unique H3K4me2 enrichment using MACS.
39. Zhang, J., Parvin, J. & Huang, K. Redistribution of H3K4me2 on neural tissue.
40. Makishima, H. Somatic SETBP1 mutations in myeloid neoplasms.
41. Makela, J. et al. Peroxisome proliferator-activated receptor-gamma.
42. Scifo, E. et al. Proteomic analysis of the palmitoyl protein thioesterase 1.
43. Dobins, A. et al. STAR: ultrafast universal RNA-seq aligner.
44. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
45. Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene 108, 193–199 (1991).
46. Saito, T. & Nakatsuji, N. Efficient gene transfer into the embryonic mouse brain using in vivo electroporation. Dev. Biol. 240, 237–246 (2001).
47. Sessa, A., Mao, C. A., Hadjantonakis, A. K., Klein, W. H. & Broccoli, V. Thrb2 directs conversion of radial glia into basal precursors and guides neuronal amplification by indirect neurogenesis in the developing neocortex. Neuro 60, 56–69 (2008).
48. Pathan, M. et al. FunRich: an open access standalone functional enrichment and interaction network analysis tool. Proteomics 15, 2597–2601 (2015).
49. Zhu, L. et al. ASH1L links histone H3 lysine 36 dimethylation to MLL leukemia. Cancer Discov. 6, 779–783 (2016).

Acknowledgements
This work was supported by Associazione Italiana Ricerca sul Cancro 2013 (IG-14249) to C.G.P., Associazione Italiana Ricerca sul Cancro 2015 (IG-17727) to R.P., Telethon grant (GGP15096), and the Italian Ministry of Health Young investigator grant (# GR-2013-02355540) to A.S.

Author contributions
R.P.: conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper. V.M. and S.R.: performed experiments, analyzed the data, critically revised the manuscript. M.Mauri, M.P., and C.M.: performed experiments, critically revised the manuscript. L.M.: performed bioinformatics analyses, critically revised the manuscript. A.S., F.B., and A.R.: performed in utero electroporation experiments, critically revised the manuscript. M.L., R.S., and M.Baumann: performed mass-spectrometry experiments, critically revised the manuscript. A.P.: performed library preparation for NGS analyses, critically revised the manuscript. D.R., F.S., E.U., B.M., L.C., M.Merli, F. Passamonti, F.O., A.M., F.Pavesi, and M.Bregni: contributed reagents, critically revised the manuscript. C.G.P.: conceived and designed the experiments, wrote the paper.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04462-8.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018