Here we report whole-exome sequencing of individuals with various myeloid malignancies and identify recurrent somatic mutations in SETBP1, consistent with a recent report on atypical chronic myeloid leukemia (aCML). Closely positioned somatic SETBP1 mutations encoding changes in Asp868, Ser869, Gly870, Ile871 and Asp880, which match germline mutations in Schinzel-Giedion syndrome (SGS), were detected in 17% of secondary acute myeloid leukemias (sAML) and 15% of chronic myelomonocytic leukemia (CMML) cases. These results from deep sequencing demonstrate a higher mutational detection rate than reported with conventional sequencing methodology. Mutant cases were associated with advanced age and monosomy 7/deletion 7q (–7/del(7q)) constituting poor prognostic factors. Analysis of serially collected samples indicated that SETBP1 mutations were acquired during leukemic evolution. Transduction with mutant Setbp1 led to the immortalization of mouse myeloid progenitors that showed enhanced proliferative capacity compared to cells transduced with wild-type Setbp1. Somatic mutations of SETBP1 seem to cause gain of function, are associated with myeloid leukemic transformation and convey poor prognosis in myelodysplastic syndromes (MDS) and CMML.

During the past decade, substantial progress has been made in the understanding of the pathogenic gene mutations driving myeloid malignancies. Following the early identification of mutations in RUNX1 (ref. 6), JAK2 (ref. 7) and RAS (refs 8,9), SNP array karyotyping led to the discovery of mutations in CBL (ref. 10), TET2 (ref. 11) and EZH2 (ref. 12). More recently, new sequencing technologies have enabled exhaustive screening of somatic mutations in myeloid malignancies, leading to the discovery of unexpected mutational targets, such as DNMT3A (ref. 13), IDH1 (ref. 14) and spliceosomal genes (refs 15–17). Insights into the progression to sAML constitute an important goal of biomedical investigations, now augmented by the availability of next-generation sequencing technologies.

We performed whole-exome sequencing of 20 index cases with myeloid malignancies (Supplementary Table 1) and identified 38 non-silent somatic mutations that were subsequently confirmed by Sanger sequencing and targeted deep sequencing. We found that seven genes were recurrently mutated in multiple samples (Supplementary Tables 2–4). Of these, we identified a new recurrent somatic mutation in SETBP1 (encoding a p.Asp868Asn alteration) in two cases with refractory anemia with excess blasts (RAEB) (Fig. 1 and Supplementary Tables 1–3 and 5), which were confirmed using DNA from both tumor and CD3+ T cells.

SETBP1 was initially identified as a 170-kDa nuclear protein that binds to SET20,21 and is activated to support the recovery of granulopoiesis in chronic granulomatous disease22. Mutations in SETBP1 are causative in SGS, a congenital disease characterized by a higher than normal prevalence of tumors, typically neuroepithelial neoplasia23,24. Notably, the mutations identified in our cohort exactly corresponded with the recurrent de novo germline mutations responsible for SGS, which prompted us to investigate SETBP1 mutations in a large cohort of 727 cases with various myeloid malignancies (Supplementary Table 6).

SETBP1 mutations were found in 52 of 727 cases (7.2%). Consistent with recent reports (refs 13–15,25,26), p.Asp868Asn (n = 28), p.Gly870Ser (n = 15) and p.Ile871Thr (n = 5) alterations were more frequent than p.Asp868Tyr, p.Ser869Asn, p.Asp880Asn and p.Asp880Glu alterations (n = 1 for each) (Fig. 1 and Supplementary Tables 1 and 7). All these alterations were located in the SKI homology region, which is highly
conserved between species (Supplementary Fig. 1). Comparable expression of mutant and wild-type alleles was confirmed for the p.Asp868Asn and p.Gly870Ser alterations by allele-specific PCR using genomic DNA and cDNA (Supplementary Fig. 2). SETBP1 mutations were significantly associated with advanced age (P = 0.01) and −7/del(7q) (P = 0.01) and were frequently found in sAML (19 of 113 cases; 16.8%; P < 0.001) and CMML (22 of 152 cases; 14.5%; P = 0.002), whereas they were less frequent in primary AML (1 of 145 cases; <1%; P = 0.002) (Table 1 and Supplementary Fig. 3a). The lack of apparent segmental allelic imbalance involving the SETBP1 locus (18q12.3) in SNP array karyotyping in all mutated cases (Supplementary Fig. 4), together with no more than 50% mutant allele frequencies in deep sequencing and allele-specific PCR, suggested the presence of heterozygous mutations (Fig. 1b and Supplementary Fig. 2). Medical history and physical findings did not support clinical diagnosis with SGS in any of these cases, and formal confirmation of the somatic origin of all types of mutation found was carried out using germline SGS in any of these cases, and formal confirmation of the somatic origin of all types of mutation found was carried out using germline SGS and normal (CD3+ T cell) DNA from a case with RAEB (whole exome 4), where red and blue bars indicate positive and negative strands, respectively. Mutated nucleotides (c.2602G>A) are shown in green. Black rectangles highlight the codon affected by mutation. A small amount of tumor cell contamination caused occasional mutant reads in the CD3+ T cell sample, where the presence of multiple single-nucleotide variants (SNVs) of similar frequencies precluded the possibility of somatic mosaicism. (b) Top, allele frequencies in paired bone marrow and CD3+ T cell samples in two RAEB cases and one sAML case and in paired refractory cytopenia with multilineage dysplasia (RCMD)-sAML and RCMD-CMML samples from the same individual as measured by deep sequencing. Bottom, depth of coverage of independent reads. In paired tumor-normal samples, small tumor contaminations were detected in CD3+ T cells. Paired RCMD-sAML and RCMD-CMML samples show very small numbers of mutant reads in the initial MDS presentation (RCMD), indicating the presence of a minor SETBP1-mutated clone, which evolved later into more aggressive disease. *P < 0.001.

leukemogenesis were compared in the cases with SETBP1 mutations and in cases with wild-type SETBP1 (Fig. 2c.d and Supplementary Table 8). Only CBL mutations were significantly associated with SETBP1 mutations (P = 0.002; Supplementary Table 9). Notably, mutations of FLT3 and NPM1 were not found in cases with SETBP1 mutation. Coexisting SETBP1 and CBL mutations were found in 12 cases, of which 6 were subjected to deep sequencing, and CBL-mutated clones were significantly smaller than SETBP1-mutated clones, suggesting that CBL mutations were acquired by a subclone with SETBP1 mutation (Supplementary Fig. 5). The significant association of CBL and SETBP1 mutations suggests their potential cooperation in leukemia progression. Although direct physical interaction between mutant Setbp1 and CBL proteins was not detected (Supplementary Fig. 7), it is possible that CBL mutations cooperate with SETBP1 mutations indirectly by reducing the cytokine dependence of leukemia cells10,27. SETBP1 mutations were also found in aCML1 and juvenile chronic myelomonocytic leukemia28, characterized by RAS pathway defects, including CBL mutations.

Analysis of the expression patterns of SETBP1 mRNA in normal hematopoietic tissues showed relatively low levels of this transcript in myeloid and/or monocytic cells as well as in CD34+ cells (Supplementary Fig. 8). In contrast, SETBP1–mutant cases showed significantly higher expression levels than samples with wild-type SETBP1 (P = 0.03; Supplementary Fig. 9). When SETBP1 expression was also evaluated using expression array data in the cases with different subtypes of myeloid neoplasm (Supplementary Fig. 10), SETBP1 was found to be overexpressed in cases with non–core binding factor (CBF) primary AML, including MDS, whereas CBF leukemias showed normal levels of the corresponding mRNA. In particular, SETBP1
expression was significantly higher in cases with loss of chromosome 7 (P = 0.03) and complex karyotype (P < 0.001) (Supplementary Fig. 3). Clustering analysis of gene expression profiles suggested that SETBP1-mutant cases had a similar expression pattern to that of cases with overexpression of wild-type SETBP1, including overexpression of TCF4, BCL11A and DNTT (Supplementary Fig. 10 and Supplementary Table 10). Methylation array analysis showed that relative hypomethylation of the CpG site located in proximity to the SETBP1 coding region was associated with higher expression and mutation of SETBP1 (Supplementary Fig. 11). It remains unclear what factors drive the increase in SETBP1 mRNA levels in these leukemias; however, these mechanisms may involve aberrant hypomethylation of the SETBP1 promoter or activation of upstream regulators such as MECOM (22,29).

Within the entire cohort, SETBP1-mutated cases were significantly associated with shorter overall survival time (hazards ratio (HR) = 2.27, 95% confidence interval (CI) = 1.56–3.21; P < 0.001), with this association especially prominent in the younger age group (<60 years; HR = 4.92, 95% CI = 2.32–9.46; P < 0.001). The presence of SETBP1 mutations was also associated with compromised survival in the cohort with normal karyotype (HR = 3.13, 95% CI = 1.66–5.41; P = 0.002) (Fig. 3). Multivariate analysis confirmed that SETBP1 mutation was an independent prognostic factor (HR = 2.90, 95% CI = 1.71–4.83; P < 0.001) together with male sex, advanced age and the presence of ASXL1, CBL and DNMT3A mutations. −7/del(7q) was associated with shorter length of survival in univariate analysis but did not remain an independent risk factor after multivariate analysis (Supplementary Table 11). The multivariate analysis in the subgroup of MDS and CML cases (with white blood cell (WBC) counts of <12,000 cells/µl), in which the International Prognostic Scoring System (IPSS) score was applicable (30), also showed that SETBP1 mutation was an independent prognostic factor (HR = 1.83, 95% CI = 1.04–3.12; P = 0.04), whereas the impact of the IPSS score was not significant (HR = 1.04–3.12; P = 0.04).

Clinical characteristics of myeloid malignancies with or without SETBP1 mutation

| Characteristic          | Wild-type SETBP1 | Mutant SETBP1 | P* |
|-------------------------|------------------|---------------|----|
| Number                  | 675              | 52            |    |
| Age at study entry (years), mean ± s.d. | 61 ± 15          | 67 ± 12       | 0.01b |
| Age range (years)       | 16–91            | 26–83         |    |
| Ancestry, number        |                  |               |    |
| Caucasian               | 222              | 29            | 0.27 |
| African American        | 10               | 0             |    |
| Asian                   | 298              | 23            |    |
| Other                   | 2                | 0             |    |
| Male sex, number        | 376              | 29            | 0.23 |
| Increased (≥10%) bone marrow blasts, number | 376              | 33            | 0.31 |
| Diagnosis, number       |                  |               |    |
| 5q− syndrome            | 7                | 1             | 1.00 |
| RCMD                    | 52               | 2             | 1.00 |
| RAEB                    | 86               | 4             | 1.00 |
| sAML                    | 94               | 19            | <0.001 |
| CMML                    | 130              | 22            | 0.002 |
| CML BP                  | 25               | 2             | 1.00 |
| PMF                     | 25               | 1             | 1.00 |
| pAML                    | 144              | 1             | 0.002 |
| Cytogenetics, number    |                  |               |    |
| Normal                  | 208              | 17            | 1.00 |
| −5,del(5q)              | 39               | 1             | 1.00 |
| −7,del(7q)              | 72               | 15            | 0.01 |
| −Y only                 | 9                | 0             | 1.00 |
| −20,del(20q)            | 18               | 1             | 1.00 |
| +8                      | 45               | 2             | 1.00 |
| Complex (≥3)            | 69               | 2             | 1.00 |

CML BP, chronic myelogenous leukemia blast phase; PMF, primary myelofibrosis.

* A Fisher’s exact test was used to determine P values, except where otherwise indicated. P values in multiple comparisons were evaluated by Bonferroni correction, and statistically significant P values are indicated with bold font. *A Wilcoxon test was used to calculate the P value.

**Table 1** Clinical characteristics of myeloid malignancies with or without SETBP1 mutation

---

Figure 2 The relationship of SETBP1 mutations with other common mutations. (a,b) Clonal profiles of gene mutations in two representative cases with MDS that transformed to RAEB (a) and CMLM (b). Initially, hypocellular MDS (RCMD) was diagnosed on the basis of hypocellular bone marrow with normal karyotype in both cases. (c) Coexisting mutations in the SETBP1-mutated cohort are shown in a matrix. 36 of 52 cases (69%) were positive for other somatic concomitant mutations tested by Sanger sequencing. Sequenced genes are listed in [Supplementary Table 8](#). CMLM1 and CMLM2 were discriminated by the number of blasts plus promonocytes in the peripheral blood and bone marrow. PV, polycythemia vera; pAML, primary AML. (d) Circos plots illustrating coexisting mutations in the selected 12 genes in the whole cohort. No mutations that occurred in a mutually exclusive manner were observed.

---
dissipated after multivariate analysis (Supplementary Tables 11 and 12). Next, because comprehensive mutational screening identified a significant association between SETBP1 and CBL mutations, we compared overall length of survival in cases with either of these mutations or with these mutations in combination (Supplementary Figs. 12 and 13 and Supplementary Table 13). Overall length of survival was shorter in cases with mutation in both SETBP1 and CBL compared to those with the wild-type forms of these genes, and the combination of these mutations was also unfavorable in an isolated CMML cohort in which either of these mutations alone did not affect survival (Fig. 3 and Supplementary Fig. 13). However, no impact of these mutations was found in a sAML cohort, probably owing to the already very poor prognosis in this subset of individuals (Supplementary Figs. 12 and 14).

Previous studies demonstrated that overexpression of Setbp1 can effectively immortalize mouse myeloid precursors31. Expression of Setbp1 mutants (either Asp868Asn or Ile871Thr) also caused efficient immortalization of mouse myeloid progenitors with similar phenotypes (Fig. 4a,b and Supplementary Fig. 15). Moreover, although having similar levels of Setbp1 protein expression as cells immortalized with wild-type Setbp1, cells immortalized with mutant Setbp1 showed significantly more efficient colony formation and faster proliferation (Fig. 4c,d and Supplementary Figs. 16 and 17). This observation is consistent with the gain of leukemogenic function due to SETBP1 mutation. As with overexpressed wild-type Setbp1, homeobox genes Hoxa9 and Hoxa10 represent critical targets of Setbp1 mutants, as cells immortalized by wild-type or mutant Setbp1 expressed comparable levels of the corresponding mRNAs, and knockdown of either caused a marked reduction in colony-forming potential (Supplementary Figs. 18 and 19). In agreement with these findings, SETBP1-mutant leukemias (n = 14) showed significantly higher Hoxa9 and Hoxa10 expression levels compared to wild-type cases without SETBP1 overexpression (n = 9; P = 0.03 and 0.03, respectively), supporting the notion that Hoxa9 and Hoxa10 are likely functional targets of mutated SETBP1 in myeloid neoplasms (Supplementary Fig. 20).

Multiple mechanisms could contribute to the enhanced oncogenic properties of SETBP1 mutations. For instance, mutation could increase protein stability (Supplementary Fig. 21), resulting in greater protein amounts (analogous to upmodulation of SETBP1 mRNA), in agreement with a previously reported observation1. However, we also showed that SETBP1 mRNA overexpression in vitro was associated with the immortalization of progenitors and that there were primary cases of sAML with and without mutations of SETBP1 and high levels of wild-type mRNA. Thus, although plausible, the mechanisms underlying increased SETBP1 expression and its proto-oncogenic role may be more complicated. It is also possible that interaction of Ski and/or SnoN with SETBP1 through the SKI homology region could be affected by mutations, leading to transformation20,32. SETBP1 was shown to regulate PP2A activity via binding to SET20, and decreased PP2A activity has been described in AML21,33. In fact, we observed that mutant Setbp1–immortalized myeloid progenitors had increased tyrosine phosphorylation of Ppp2ca compared to myeloid progenitors immortalized with wild-type Setbp1 (Supplementary Fig. 22), suggesting that SETBP1 mutations could cause further PP2A inhibition.

In summary, recurrent somatic SETBP1 mutations are new lesions that interact with previously defined pathways underlying poor prognosis and provide new insights into the process of leukemic evolution. The apparent association of SETBP1 mutations with poor clinical outcome observed here provides an important focal point for future mechanistic studies as well as a goal for therapeutic targeting.
Whole-exome sequencing results have been deposited in the Sequence Read Archive (SRA; BioProject accession PRJNA203580).

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

ACKNOWLEDGMENTS
We thank T. Yamaguchi (The University of Tokyo) for providing CS-Ubc lentivirus vector. This work was supported by US National Institutes of Health (NIH) grants R01 HL-082983 (J.P.M.), U54 RR019391 (J.P.M.), K24 HL-077522 (J.P.M.) and R01 CA-141933 (Y.D.), by a grant from the AA & MDS International Foundation, by the Robert Duggan CARES grant (H. Makishima) and by Grants-in-Aid from the Ministry of Health, Labor and Welfare of Japan and KAKENHI (23249052, 22134006 and 21799097; S.O.), the project for the development of innovative research on cancer therapies (p-direct; S.O.), the Japan Society for the Promotion of Science (JSPS) through the Funding Program for World-Leading Innovative R&D on Science and Technology, initiated by the Council for Science and Technology Policy (CSTP; S.O.), NHRI-EX100-10035NI Taiwan (L.-Y.S.) and Uniformed Services University of the Health Sciences Pediatrics grant KM86GI (Y.D.). The results presented here are partly based on data generated by The Cancer Genome Atlas (TCGA) pilot project established by the National Cancer Institute and the National Human Genome Research Institute. Information about TCGA and the investigators and institutions that constitute the TCGA research network can be found at http://cancergenome.nih.gov/.

AUTHOR CONTRIBUTIONS
H. Makishima and K.Y. designed research, performed research, collected data, performed statistical analysis and wrote the manuscript. Y.O., N.N., K.P. and H. Makishima and K.Y. designed research, performed research, collected data, analyzed and interpreted data, and wrote the manuscript. Y.D., S.O. and J.P.M. designed research, contributed analytical tools, collected data, analyzed and interpreted data, and wrote 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.

1. Piazza, R. et al. Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. Nat. Genet. 45, 18–24 (2013).

2. Hoischen, A. et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat. Genet. 42, 483–485 (2010).

3. Damm, F. et al. SETBP1 mutations in 658 patients with myelodysplastic syndromes, chronic myelomonocytic leukemia and secondary acute myeloid leukemias. Leukemia 27, 1401–1403 (2013).

4. Laborde, R.R. et al. SETBP1 mutations in 415 patients with primary myelofibrosis or chronic myelomonocytic leukemia (CML): independent prognostic impact in CML. Leukemia published online; doi:10.1038/leu.2013.97 (5 April 2013).

5. Thol, F. et al. SETBP1 mutation analysis in 944 patients with MDS and AML. Leukemia published online; doi:10.1038/leu.2013.145 (7 May 2013).

6. Osato, M. et al. Biallelic and heterozygous point mutations in the runt domain of the AML1/PEBP2α gene associated with myeloblastic leukemias. Blood 93, 1817–1824 (1999).

7. Farr, C.J., Saiki, R.K., Erlich, H.A., McCormick, F. & Marshall, C.J. Analysis of RAS gene mutations in acute myeloid leukemia by polymerase chain reaction sequencing. Cancer Cell 7, 387–397 (2005).

8. Fehr, S.A., Mardis, E.R. et al. Somatic mutations in non-small cell lung cancer. Cancer Cell 7, 23–29 (2005).

9. Lyons, J., Janssen, J.W., Bartram, C., Layton, M. & Mufti, G.J. Mutation of K-myc and N-ras oncogenes in myelodysplastic syndromes. Blood 71, 1707–1712 (1988).

10. Makishima, H. et al. Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. Nature 460, 904–908 (2009).

11. Delhommeau, F. et al. Mutation in TET2 in myeloid cancers. N. Engl. J. Med. 360, 2289–2301 (2009).

12. Ernst, T. et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat. Genet. 42, 722–726 (2010).

13. Ley, T.J. et al. DNMT3A mutations in acute myeloid leukemia. N. Engl. J. Med. 363, 2424–2433 (2010).

14. Mardis, E.R. et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N. Engl. J. Med. 361, 1058–1066 (2009).

15. Yoshida, K. et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478, 64–69 (2011).

16. Papaemmanuil, E. et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N. Engl. J. Med. 365, 1384–1395 (2011).

17. Graubert, T.A. et al. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat. Genet. 44, 53–57 (2012).

18. Walter, M.J. et al. Clonal architecture of secondary acute myeloid leukemia. N. Engl. J. Med. 366, 1090–1092 (2012).

19. Walter, M.J. et al. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. Leukemia 27, 1275–1282 (2013).

20. Minakuchi, M. et al. Identification and characterization of SEB, a novel protein that may be involved in the acute undifferentiated leukemia–associated protein SET. Eur. J. Biochem. 268, 1340–1351 (2001).

21. Cristóbal, 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).

22. Ott, M.G. et al. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat. Med. 12, 401–409 (2006).

23. Schinzel, A. & Giedion, A. A syndrome of severe midface retraction, multiple skull anomalies, clubfeet, and cardiac and renal malformations in sibs. Am. J. Med. Genet. 1, 361–375 (1978).

24. Rodriguez, J.L., Jimenez-Heffernan, J.A. & Leal, J. Schinzel-Giedion syndrome: autopsy report and additional clinical manifestations. Am. J. Med. Genet. 53, 374–377 (1994).

25. Pardanani, A. et al. CSF3R T618I is a highly prevalent and specific mutation in chronic neutrophilic leukemia. Leukemia published online; doi:10.1038/ leu.2013.122 (2 April 2013).

26. Meggendorfer, M. et al. SETBP1 mutations occur in 9% of MDS/MPN and in 4% of MPN cases and are strongly associated with atypical CML, monosomy 7, isochromosome 17(q)10, ASXL1 and CBL mutations. Leukemia published online; doi:10.1038/leu.2013.133 (30 April 2013).

27. Makishima, H. et al. CBL mutation-related patterns of phosphorylation and sensitivity to tyrosine kinase inhibitors. Leukemia 26, 1547–1554 (2012).

28. Sakaguchi, H. et al. Exome sequencing identifies secondary mutations of SETBP1 and JAK3 in juvenile myelomonocytic leukemia. Nat. Genet. published online; doi:10.1038/ng.2698 (7 July 2013).

29. Goyama, S. et al. Evi-1 is a critical regulator for hematopoietic stem cells and transformed leukemic cells. Cell Stem Cell 3, 207–220 (2008).

30. Greenberg, P. et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood 89, 2079–2088 (1997).

31. Oakley, K. et al. SETBP1 promotes the self-renewal of murine myeloid progenitors via activation of Hoxa9 and Hoxa10. Blood 119, 6099–6108 (2012).

32. Cohen, S.B., Zheng, G., Heyman, H.C. & Stawnezer, E. Heterodimers of the SnoN and Ski oncoproteins form preferentially over homodimers and are more potent transforming agents. Nucleic Acids Res. 27, 1006–1014 (1999).

33. Cristóbal, I. et al. PP2A impaired activity is a common event in acute myeloid leukemia and its activation by forskolin has a potent anti-leukemic effect. Leukemia 25, 606–614 (2011).
ONLINE METHODS

Subject population. Bone marrow aspirates or blood samples were collected from 727 individuals with various myeloid malignancies seen at the Cleveland Clinic, the University of Tokyo, the University of California, Los Angeles, the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Chang Gung University and Showa University (Supplementary Table 6). Informed consent for sample collection was obtained according to protocols approved by the institutional review board at each participating institute and in accordance with the Declaration of Helsinki. Diagnosis was confirmed and assigned according to World Health Organization (WHO) classification criteria 34. Prognostic risk assessment was assigned according to the International Scoring Criteria for individuals with MDS and chronic myelomonocytic leukemia with a white cell count of <12,000 cells/μl 35. For the purpose of this study, low-risk MDS was defined as having ≤5% myeloblasts. Individuals with ≥5% myeloblasts constituted those with higher risk disease. Serial samples were obtained for 12 individuals with SETBP1 mutations. As a source of germline controls, immunoselected CD3 T lymphocytes were used in an additional nine cases. Cytogenetic analysis was performed according to standard banding techniques on the basis of 20 metaphases, if available. Clinical parameters studied included age, sex, overall survival, bone marrow blast counts and metaphase cytogenetics.

Cytogenetics and SNP arrays. Technical details regarding sample processing for SNP array assays were previously described 35–36. The Gene Chip Mapping 250K Assay kit and the Genome-Wide Human SNP Array 6.0 (Affymetrix) were used. A stringent algorithm was applied for the identification of lesions using SNP arrays. Individuals with lesions identified by SNP array concordant with those identified in metaphase cytogenetics or typical lesions known to be recurrent required no further analysis. Changes reported in our internal or publicly available (Database of Genomic Variants; see URLs) copy number variation (CNV) databases were considered non-somatic and were excluded. Results were analyzed using CGHQC (v3.0) 37 or Genotyping Console (Affymetrix). All other lesions were confirmed as somatic or germline by analysis of CD3-sorted cells 38.

Whole-exome sequencing. Whole-exome sequencing was performed as previously reported 15. Briefly, tumor DNA was extracted from bone marrow or peripheral blood mononuclear cells from affected individuals. For germline controls, DNA was obtained from paired CD3 T cells. Whole-exome capture was accomplished using liquid-phase hybridization of sonicated genomic DNA with mean length of 150–200 bp to the bait cDNA library synthesized on magnetic beads (SureSelect, Agilent Technologies) according to the manufacturer’s protocol. The SureSelect Human All Exon 50Mb kit was used for 20 cases (Supplementary Table 1). Captured targets were subjected to massive sequencing using the Illumina HiSeq 2000 platform with the paired-end 75- to 108-bp read option, according to the manufacturer’s instructions. Raw sequence data generated from HiSeq 2000 sequencers were processed through the in-house pipeline constructed for the whole-exome analysis of paired cancer genomes at the Human Genome Center, Institute of Medical Science, University of Tokyo, which is summarized in a previous report 45. Data processing is divided into two steps: (i) generation of a BAM file (using SAMtools) for paired normal and tumor samples for each case and (ii) detection of somatic SNVs and indels by comparing normal and tumor BAM files. Alignment of sequencing reads on the hg19 reference genome was visualized using Integrative Genomics Viewer (IGV) software 46.

For all candidate somatic mutations, the accuracy of the prediction of these SNVs and indels by whole-exome sequencing was tested by validation of 65 genes (80 events) by Sanger sequencing and targeted deep sequencing. Prediction had a true positive rate of 47% (39% for missense mutation, 75% for nonsense mutations and 75% for indels). It is of note that prediction of known somatic mutations (for example, in TET2 (n = 9), CBL (n = 2), SETBP1 (n = 2) and ASXL1 (n = 2)) showed accuracy of 100% (Supplementary Tables 2–4).

Targeted deep sequencing. To detect allelic frequencies for mutations or SNPs, we applied deep sequencing to targeted exons as previously described 15. Briefly, we screened for possible mutations of SETBP1 and other genes that were concomitantly mutated in the cases with SETBP1 mutation (U2AF1, DNMT3A, NRAS, ASXL1, SRSF2, CBL, IDH1, IDH2, SRSF2, TET2, PTPN11 and RUNX1). Each targeted exon was amplified with NotI linker attached to each primer as previously described 45. After digestion with NotI, amplicons were ligated with T4 DNA ligase and sonicated into fragments that were on average up to 200 bp in size using Covaris. Sequencing libraries were generated according to an Illumina paired-end library protocol and were subjected to deep sequencing on the Illumina Genome Analyzer IIx or HiSeq 2000 sequencers according to the standard protocol.

Sanger sequencing and allele-specific PCR. Exons of selected genes were amplified and underwent direct genomic sequencing by standard techniques on the ABI 3730xl DNA analyzer (Applied Biosystems) as previously described 40–42. Coding and sequenced exons are shown in Supplementary Table 8. All mutations were detected by bidirectional sequencing and were scored as pathogenic if not present in non-clonal paired DNA from CD3-selected cells. When a mutant allele with small burden was not confirmed by Sanger sequencing, cloning and sequencing of individual colonies (TOPO TA cloning, Invitrogen) was performed for validation. The allelic presence of p.Asp686Asn and p.Gly870Ser alterations was determined by allele-specific PCR. Primer sequences for SETBP1 sequencing and SETBP1 allele-specific PCR are provided in Supplementary Table 14.

Quantitative RT-PCR using TaqMan probes. Total RNA was extracted from bone marrow mononuclear cells and cell lines. cDNA was synthesized from 500 ng of total RNA using the iScript cDNA synthesis kit (Bio-Rad). Quantitative gene expression levels were detected using RT-PCR with the ABI PRISM 7500 Fast Sequence Detection System and FAM dye–labeled TaqMan MGB probes (Applied Biosystems). TaqMan probes for all genes analyzed were gene expression assay products purchased from Applied Biosystems (SETBP1, Hs00210209_m1; HOXA9, Hs00365956_m1; HOXA10, Hs00172012_m1; GAPDH, Hs99999905_m1). Expression levels of target genes were normalized to GAPDH mRNA levels.

Retrovirus generation. pMYS-Setbp1 retrovirus expressing 3× Flag–tagged wild-type Setbp1 protein and green fluorescent protein (GFP) marker was described previously 31. Point mutations of Setbp1 (encoding p.Asp866Asn and p.Ile871Thr alterations) were generated using the same construct and the QuickChange II Site-Directed Mutagenesis kit (Agilent Technologies). Virus was produced by transient transfection of Plat-E cells (Cell Biolabs) using FuGene 6 (Roche). Viral titers were calculated by infecting NIH3T3 cells with serially diluted viral stock and counting GFP-positive colonies 48 h after infection.

Immortalization of myeloid progenitors. Immortalization of myeloid progenitors was performed as described according to protocols approved by the Institutional Animal Care and Use Committee of the Uniformed Services University of the Health Sciences 31. Briefly, whole–bone marrow cells harvested from three young C57BL/6 mice were first cultured in StemSpan medium (Stemcell Technologies) with 10 ng/ml mouse SCF, 20 ng/ml mouse TPO, 20 ng/ml mouse IGF-2 (all from R&D Systems) and 10 ng/ml human FGF-1 (Invitrogen) for 6 d to expand primitive stem and progenitor cells. Myeloid differentiation was subsequently induced by growing the expanded cells in IMDM supplemented with 20% heat-inactivated horse serum with 100 ng/ml mouse SCF (PeproTech) and 10 ng/ml mouse IL-3 for 4 d. Resulting cells (5 × 105) were infected with retrovirus (1 × 105 colony-forming units (CFUs)) on plates coated with Retronectin (Takara) for 48 h. Infected cells were then continuously passaged at a 1:10 ratio every 3 d for 4 weeks to test whether transduction caused immortalization of the myeloid progenitors. In the absence of immortalization, transduced cultures generally ceased expanding in 2 weeks.

Methylation analysis. The DNA methylation status of bisulfite-treated genomic DNA was probed at 27,578 CpG dinucleotides using the Illumina Infinium HumanMethylation 27k BeadChip assay as previously described 43. Briefly, methylation status was calculated from the ratio of methylation-specific and demethylation-specific fluorophores (β value) using the BeadStudio Methylation Module (Illumina).
Resistance of SETBP1 protein degradation associated with SETBP1 mutation.

Full-length wild-type human SETBP1 cDNA encoding 3× HA–tagged protein was cloned from peripheral blood mononuclear cells. Mutagenesis of SETBP1 (to introduce mutations encoding the p.As868Asn and p.Ile871Thr alterations) was performed using the PrimeSTAR kit (Takara Bio). Wild-type and mutant cDNA constructs were cloned into the CS-Ubc lentivirus vector (a kind gift of T. Yamaguchi). Vectors were cotransfected with packaging vector and with vectors expressing VSV-G and Rev into 293T cells, and lentiviral particles were harvested. Protein blotting experiments on whole lysates from Jurkat cell line stably transduced with viruses expressing wild-type and mutant SETBP1 were carried out with antibodies for HA at a 1:2,000 dilution (MMS-101R, Covance) and actin at a 1:1,000 dilution (sc-1616, Santa Cruz Biotechnology). Both cell lines were obtained from ATCC. For proteasomal inhibition, cell lines were treated with 0.5 µM lactacystin (Peptide Institute) and 0.25 µM bafilomycin A1 (Peptide Institute) for 2 h.

Statistical analysis. The Kaplan–Meier method was used to analyze survival outcomes (overall survival) by the log-rank test. Pairwise comparisons were performed by Wilcoxon test for continuous variables and by two-sided Fisher’s exact test for categorical variables. Paired data were analyzed by Wilcoxon signed-rank test. For multivariate analyses, a Cox proportional hazards model was conducted for overall survival. Variables considered for model inclusion were IPSS risk group, age, sex and gene mutation status. Variables with \( P < 0.05 \) in univariate analyses were included in the model. Statistical analyses were performed with JMP9 software (SAS). Significance was determined at a two-sided \( \alpha \) level of 0.05, except for \( P \) values in multiple comparisons, in which Bonferroni correction was applied.

34. Shaffer, L.G. & Tommerup, N. ISCN 2009. An International System for Human Cytogenetics Nomenclature (Karger, Basel, Switzerland, 2009).
35. Maciejewski, J.P., Tiu, R.V. & O’Keefe, C. Application of array-based whole genome scanning technologies as a cytogenetic tool in haematological malignancies. Br. J. Haematol. 146, 479–488 (2009).
36. Gondek, L.P. et al. Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML. Blood 111, 1534–1542 (2008).
37. Nannya, Y. et al. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res. 65, 6071–6079 (2005).
38. Tiu, R.V. et al. New lesions detected by single nucleotide polymorphism array-based chromosomal analysis have important clinical impact in acute myeloid leukemia. J. Clin. Oncol. 27, 5219–5226 (2009).
39. Robinson, J.T. et al. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26 (2011).
40. Dunbar, A.J. et al. 250K single nucleotide polymorphism array karyotyping identifies acquired uniparental disomy and homozygous mutations, including novel missense substitutions of c-Cbl, in myeloid malignancies. Cancer Res. 68, 10349–10357 (2008).
41. Jankowska, A.M. et al. Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. Blood 113, 6403–6410 (2009).
42. Makishima, H. et al. CBL, CBLB, TET2, ASXL1, and IDH1/2 mutations and additional chromosomal aberrations constitute molecular events in chronic myelogenous leukemia. Blood 117, e198–e206 (2011).
43. Ko, M. et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 468, 839–843 (2010).