Exome sequencing identifies secondary mutations of *SETBP1* and *JAK3* in juvenile myelomonocytic leukemia

エクソーム解析による若年性骨髄単球性白血病における 2 次変異としての SETBP1 および JAK3 遺伝子変異の同定

Hirotoshi Sakaguchi[†1], Yusuke Okuno[†2], Hideki Muramatsu[†1], Kenichi Yoshida[†2], Yuichi Shiraishi[4], Mariko Takahashi[2], Ayana Kon[2], Masashi Sanada[2,3], Kenichi Chiba[4], Hiroko Tanaka[5], Hideki Makishima[6], Xinan Wang[1], Yinyan Xu[1], Sayoko Doisaki[1], Asahito Hama[1], Koji Nakanishi[1], Yoshiyuki Takahashi[1], Nao Yoshida[7], Jaroslaw P. Maciejewski[6], Satoru Miyano[4,5], Seishi Ogawa*[2,3], and Seiji Kojima*[1]

坂口大俊、奥野友介、村松秀城、吉田健一、白石友一、高橋真理子、昆彩菜、真田昌、千葉健一、田中洋子、牧島秀樹、王稀楠、徐銀燕、土居崎小夜子、濱麻人、中西康詞、高橋義行、吉田奈央、ヤロスヲフ・マチェイェフスキー、宮野悟、小川誠司、小島勢二

[1]Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan, [2]Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, [3]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan, [4]Laboratory of DNA information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [5]Laboratory of Sequence Data Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [6]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio, USA, [7]Department of Hematology and Oncology, Children's Medical Center, Japanese Red Cross Nagoya First Hospital, Nagoya, Japan.

[†]These authors contributed equally to this work.
* co-senior and co-corresponding authors.
Introductory paragraph: 142/ 150
Word Count excluding introduction, tables, figure legends, and references: 1,500/ 1,500

Total Number of Figures: 2
Total Number of Tables: 2
Number of References: 26

Key word: JMML, whole-exome sequencing, *SETBP1*, *JAK3*
Running title: NEW SECONDARY MUTATIONS IN JMML

Corresponding authors:
Seishi Ogawa, M.D., Ph.D.
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-8655, Japan
Phone: +81-3-5800-9046
FAX: +81-3-5800-9046
E-mail: sogawa-tky@umin.ac.jp

Seiji Kojima, M.D., Ph.D.
65 Tsuruma-cho, Showa-ku
Nagoya 466-8550, Japan
Phone: +81-52-744-2294
FAX: +81-52-744-2974
E-mail: kojimas@med.nagoya-u.ac.jp

Juvenile myelomonocytic leukemia (JMML) is an intractable pediatric leukemia with poor prognosis,[1] whose molecular pathogenesis is poorly understood except for somatic/germline mutations of RAS pathway genes, including *PTPN11*, *NF1*, *N/KRAS,* and *c-CBL,* in the majority of cases.[2-4] To obtain a complete registry of gene mutations in JMML, whole-exome sequencing was performed for paired tumor-normal DNA from 13 JMML cases, followed by deep sequencing of the 8 target genes for 92 tumor samples. JMML was characterized by a paucity of gene mutations (0.85 non silent mutations per sample) with in 82 (89%) patients. somatic/germline RAS pathway involvement. *SETBP1* and *JAK3* were among common targets for secondary mutations. The latter mutations were often subclonal and supported to be involved in progression, rather than initiation of leukemia, and associated with poor clinical outcome. Our findings identified new insights into the pathogenesis and progression of JMML.

JMML is a rare myelodysplastic/myeloproliferative neoplasm unique to childhood, characterized by excessive proliferation of myelomonocytic cells and hypersensitivity to granulocyte-macrophage colony-stimulating factor.[1] A cardinal genetic feature of JMML is frequent somatic and/or germline mutation of RAS pathway genes, such as *NF1*, *N/KRAS*, *PTPN11* and *c-CBL*, which are mutated in more than 70% of JMML cases in a mutually exclusive manner.[2-4] However, it is still open to question whether RAS pathway mutations are sufficient for the development of JMML or whether secondary mutations play a role in the development and progression of JMML. To address these issues and better define the molecular pathogenesis of JMML, we performed whole-exome sequencing of paired tumor-normal DNA in 13 cases (Supplementary Table 1). The mean coverage of exome sequencing was x137 (tumor) and x143 (normal) was obtained (Supplementary Fig. 1). The Monte Carlo simulation indicated the study detected 88% of the existing somatic mutations (Supplementary methods, Supplementary Fig.2).

Sanger sequencing of 25 candidate non-silent somatic nucleotide alterations confirmed one nonsense and 10 missense mutations (Table 1 and Supplementary Fig. 3), where the low true positive rate was suggestive of the very low number of somatic mutations in JMML. The low true positive rate was thought to indicate the truly among the 11 somatic mutations, 6 involved the known RAS pathway genes and including 6 germline mutations/deletions, non-overlapping RAS pathway mutations were confirmed in 11 out of the 13 discovery cases (86%) (Table 1). For the remaining 2 cases that lacked documented RAS pathway mutations, we intensively searched for possible germline mutations that could be relevant to the development of JMML. In total, 179 and 167 candidate germline mutations were detected in patients #77 and #92, which, however, did not include known RAS pathway

genes or other cancer related genes, including the ones registered in the pathway databases (Supplementary Methods). A frameshift deletion of *MLL2* (p.V1670fs) was found in patient #92 who had been diagnosed as having Noonan syndrome based on typical features such as hypertelorism, webbed neck, and congenital heart disease (Supplementary Fig. 3), but lacked a distinctive facial appearance of Kabuki syndrome, which was shown to be caused by germline *MLL2* mutations.[5]

Five out of the 11 somatic mutations were non-RAS pathway mutations, involving *SETBP1* (3 p.D868N mutations), *JAK3* (p.R657Q), and *SH3BP1* (p.S277L), which had never been reported in JMML cases. SETBP1 was originally isolated as a 170 Kd nuclear protein that interact with SET, a small protein inhibitor of putative tumor suppressors, PP2A and NM23-H1.[6] Several lines of recent evidence suggest its role in leukemogenesis (Supplementary Fig. 4)[7-11]. *SETBP1* participates in translocations which result in an aberrant fusion gene (*NUP98/SETBP1*) and overexpression of *SETBP1* in T-cell acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML), respectively.[12,13] *SETBP1* is one of the downstream targets induced by Evi-1 oncoprotein[14] and together with *Evi1* and its homologue, *MEL1*, was reported to be activated through retrovirus integration and augment the recovery of granulopoiesis after gene therapies for chronic glanulomatous disease.[15] *SETBP1* overexpression is found in more than 27% in adult AML and associated with poor survival.[13] The discovery of recurrent hotspot mutations of *SETBP1* provides unequivocal evidence for the leukemogenic role of deregulated SETBP1 function. Conspicuously, the p.D868N mutation of *SETBP1* was identical to one of the *de novo* mutations reported to be causative for Schinzel-Giedion syndrome (SGS, MIM#269150), which is a highly recognizable congenital disease characterized by severe mental retardation, distinctive facial features, and multiple congenital malformations. It shows a higher-than-normal prevalence of tumors, notably neuroepithelial neoplasia,[16] although development of myeloid malignancies has not been reported so far.

To further validate our findings, we screened the entire cohort of 92 JMML cases for gene mutations in the newly identified three genes together with known RAS pathway targets using deep sequencing[17] (Supplementary Fig. 5).

RAS pathway mutations were found in 82 out of 92 cases (89%) in a mutually exclusive manner, where *PTPN11* mutations were predominant followed by *N/KRAS*, *c-CBL* and *NF1* mutations (Fig. 1a, Table 2). In accordance with previous reports, most of the *c-CBL* (8/14) and *NF1* (4/9) mutations were bi-allelic (Fig. 1a,b, Supplementary Table 2),[2,3,18] whereas the majority of the mutations of *PTPN11* and *N/KRAS* were heterozygous[4]. The patients without RAS pathway mutations (n=10) were vigorously investigated by whole-genome sequencing of tumor-normal paired samples (n=2) (Supplementary Fig. 6), or

by whole-exome sequencing of only tumor samples (n=8) (Supplementary Fig. 7). As anticipated, we found no known RAS pathway mutations.

On the other hand, 18 mutations were found in *SETBP1* (n=7) or *JAK3* (n=11) in 16 cases (Fig. 1a,b, Table 2 and Supplementary Table 2), which were more frequent in *PTPN11*-(and possibly *NF1*-) mutated cases than *N/KRAS*- or *c-CBL*-mutated cases (Fig. 2a). Mutations in *SH3BP1*, SH3-domain binding protein 1, were not recurrent event. All *SETBP1* mutations were heterozygous and within the SKI domain, of which 6 were identical to the *de novo* recurrent mutations reported in SGS and 5 were the identical D868N mutation (Fig. 1b). RT-PCR analysis revealed both wild-type and mutant alleles of *SETBP1* were equally expressed (Supplementary Fig. 8). Similarly, 8 out of the 11 *JAK3* mutations in 10 cases were the well-described activating mutation (p.R657Q) found in various hematological malignancies including Down syndrome-associated acute megakaryoblastic leukemia,[19-23] ALL,[24,25] and NK/T-cell lymphoma,[26] and the remaining 3 were also within the pseudo-kinase or kinase domain suggestive of gain-of-function.

Deep sequencing of the relevant mutant alleles enabled an accurate estimation of allele frequencies of individual mutations (Supplementary Fig. 9). As shown in Supplemental Fig. 10a, *SETBP1* and *JAK3* mutations showed lower allele frequencies (but not statistically significant in *SETBP1*) than did the corresponding RAS pathway mutations, indicating that the former mutations represented secondary genetic hits which contributed clonal evolution after the main tumor population was established (Supplemental Fig. 10b). Patients with secondary mutations had shorter survival compared to those without mutations; 5-year overall survival (HR=1.90 95%CI; 0.87-4.19) (Table 2, Fig. 2c). In addition, none of the JMML patients who survived without HSCT (n=26) harbored any of the secondary mutations (Fig. 2b), showing a significantly inferior 5-year transplant-free survival of patients with secondary mutations (HR=2.18 95%CI; 1.18-4.02) (Table 2, Fig. 2d).

JMML is characterized by a paucity of gene mutations. The average number of mutations per sample (0.85; 0-4) was surprisingly low compared to those reported in other human cancers (Supplementary Fig. 11); excluding common RAS pathway mutations, only 5 mutations were detected in 3 out of the 13 discovery cases. This small number of mutations is only comparable to the figure reported for retinoblastoma (mean 3.3; 0-5)[27] and shows a stark contrast to the abundance of gene mutations in chronic myelomonocytic leukemia (CMML) in adult cases, where the mean number of non-silent mutations was 12.4 per sample, of which 3.1 account for known driver changes[17] (ref#12 and unpublished data), underscoring distinct pathogenesis between the two neoplasms showing indistinguishable morphology. The impact of germline events would be underscored by the fact that 6 out of the 13 discovery cases harbored germline RAS pathway mutations and an additional case without known RAS

pathway mutations showed constitutive abnormalities similar to Noonan syndrome. Despite the central role of RAS pathway mutations, a small subset of patients have no documented RAS pathway mutations even after whole-exome analysis in the two RAS pathway mutation-negative cases, raising a possibility that the latter cases represent a genetically distinctive myeloproliferative neoplasm in childhood.

The other significant finding in the current study is the discovery of secondary mutations that involve *SETBP1* and *JAK3*. Detected only in a subpopulation of leukemic cells, most of these mutations were thought to be involved in the progression rather than the establishment of JMML and associated with a poor clinical outcome. *SETBP1* is a newly identified proto-oncogene, the identical mutations of which have recently been reported in 15-25% of adult cases with atypical CML,[28] CMML and secondary AML.[29] Affecting one of the highly conserved 3 amino-acid positions, *SETBP1* mutations have been shown to abolish the binding of an E3 ubiquitin ligase (β-TrCP1) to SETBP1, which prevents ubiquitination and subsequent degradation, leading to gain-of-function through the consequent increase in SETBP1 protein expression thought to result in a gain-of-function.[10,29] Although the precise leukemogenic mechanisms of *SETBP1* mutations are still unclear, we have demonstrated that mutant *SETBP1* alleles confers self-renewal capability to myeloid progenitors *in vitro*, which were associated with increases of *HOXA9* and *HOXA10*.[29] Recurrent *JAK3* mutations in JMML are also intriguing. The JAK/STAT pathway is a key component of normal hematopoiesis.[30] As is the case with other hematopoietic malignancies,[20] the p.R657Q mutation represents the most frequent mutation in JMML. This mutation confers IL-3-independence to Ba/F3 cells and induces STAT5 phosphorylation.[20] Targeting the JAK/STAT pathway with a pan-JAK inhibitor like CP-690550[31] could be a promising therapeutic possibility for *JAK3*-mutated JMML patients.

In conclusion, our whole-exome sequencing analysis identified the spectrum of gene mutations in JMML. Together with high frequency of RAS pathway mutations, the paucity of non-RAS pathway mutations is a prominent feature of JMML. Mutations of *SETBP1* and *JAK3* were common recurrent secondary events presumed to be involved in tumor progression and associated with poor clinical outcomes. Our finding provides an important clue to understand the pathogenesis of JMML that may help to develop novel diagnostics and therapeutics for this leukemia.

Acknowledgements

We thank the subjects and their parents for participating in this study. This work was supported by “Research on Measures for Intractable Diseases” Project from Ministry of Health

Labour and Welfare, Grant-in-Aids from the Ministry of Health, Labor and Welfare of Japan and KAKENHI (23249052, 22134006, and 21790907) and the Funding Program for World-Leading Innovative R&D on Science and Technology.

Conflict-of-interest disclosure

All authors declare that there are no competing financial interests.

Authorship

H.S., Y.O., H. Muramatsu, K.Y., M.T., A.K. and M.S. designed and performed the research and analyzed the data, and wrote the paper. Y.S., K.C., H.T. and S.M. performed bioinformatics analyses of the resequencing data. X.W. and Y.X. performed Sanger sequencing. S.D., A.H., K.N., Y.T. and N.Y. collected specimens and performed the research. H. Makishima and J.P.M. designed the research and analyzed the data. S.O. and S.K. led the entire project and wrote the paper.

Table 1. List of gene mutations identified by whole-exome sequencing

Patient No.	RAS pathway mutations								Other somatic mutations			
	Somatic				Germline							
	Gene	Changes at DNA level	Changes at protein level	VAF in tumor/ref**	Gene	Changes at DNA level	Changes at protein level	VAF in tumor/ref**	Gene	Changes at DNA level	Changes at protein level	VAF in tumor/ref**
11†	*NF1*	c.4537C>T	p.Arg1513X	40.1 / 24.2	*NF1*	c.5927delG	p.Trp1976fs	44.0 / 47.1	*SETBP1*	c.2602G>A	p.Asp868Asn	32.6 / 27.0
63	*KRAS*	c.38G>A	p.Gly13Asp	44.3 / 0.0	-	-	-	-	-	-	-	-
72	*PTPN11*	c.172A>T	p.Asn58Tyr	48.2 / 5.7	-	-	-	-	*SETBP1* *JAK3* *SH3BP1*	c.2602G>A c.1970G>A c.830C>T	p.Asp868Asn p.Arg657Gln p.Ser277Leu	45.9 / 2.5 30.5 / 2.2 47.8 / 5.1
77	-	-	-	-	-	-	-	-	*SETBP1*	c.2602G>A	p.Asp868Asn	33.4 / 2.1
78	*NRAS*	c.35G>C	p.Gly12Ala	45.5 / 9.5	-	-	-	-	-	-	-	-
82	-	-	-	-	*c-CBL*	c.1217del22	p.Thr406fs	34.7 / 38.9	-	-	-	-
83	-	-	-	-	*NF1*	c.4970A>G	p.Tyr1657Cys	50.0 / 51.0	-	-	-	-
84	-	-	-	-	*c-CBL*	c.1096-110del643	p.Glu366_Phe488del	NA / NA	-	-	-	-
85	*PTPN11*	c.226G>A	p.Glu76Lys	47.5 / 4.4	-	-	-	-	-	-	-	-
86	*KRAS*	c.38G>A	p.Gly13Asp	38.9 / 3.1	-	-	-	-	-	-	-	-
89*	-	-	-	-	*PTPN11*	c.1502T>G	p.Ser502Ala	50.0 / 49.9	-	-	-	-
91*	-	-	-	-	*PTPN11*	c.218C>T	p.Thr73Ile	49.0 / 48.0	-	-	-	-
92*	-	-	-	-	-	-	-	-	-	-	-	-

* Noonan syndrome associated-myeloproliferative disorder.
** Variant allele frequencies in tumor/reference, where reference was $CD3^+$ cells except for patient 63, in which umbilical cord was used as a reference.
† Substantial contamination of tumor cell components in CD3+ T cell reference.

Table 2. Patient characteristics

	Total cohort (N=92)	Secondary mutations		*P*-Value
		(+) (n=16)	(-) (n=76)	
Gender (Male / Female)	61 / 31	12 / 4	49 / 27	NS
Median age at diagnosis (m)	19 (1-160)	38 (2-160)	13 (1-79)	<0.001
Diagnosis				
JMML	85	16	69	
NS/MPD	7	0	7	
Genetic mutations in RAS pathway				
PTPN11	39	9	30	NS
NF1	9	5	4	0.001
RAS (NRAS / KRAS)	28 (15 / 13)	2 (1 / 1)	26 (14 / 12)	0.08
c-CBL	14	0	14	0.06
Without RAS pathway mutation	10	1	9	NS
Secondary genetic mutations				
SETBP1	7	7	0	
JAK3	10	10	0	
Cytogenetics				
Normal karyotype	77	12	65	NS
Monosomy 7	8	1	7	NS
Trisomy 8	4	2	2	NS
Other abnormalities	3	1	2	NS
WBC at diagnosis, $x10^9$/L, median (range)	30.0 (1.0-563)	29.6 (5.6-563)	30.0 (1.0-131)	NS
Monocyte at diagnosis, $x10^9$/L, median (range)	4.6 (0.2-31.6)	3.1 (0.5-15.2)	4.9 (0.2-31.6)	NS
HbF at diagnosis, %, median (range)	21 (0-68)	26 (9-55)	16 (0-68)	NS
PLT at diagnosis, $x10^9$/L, median (range)	61.0 (1.4-483)	47.5 (1.4-175)	65.0 (5.0-483)	NS
HSCT (+/-)	56/36	16/0	40/36	
Alive/Dead	62/30	7/9	55/21	
Probabilities of 5-year OS,% (95%CI)	60 (46-71)	33 (10-59)	65 (49-77)	0.10
Probabilities of 5-year transplantation free survival, % (95%CI)	15 (6-27)	0 (0-0)	18 (8-33)	0.007

Abbreviations; JMML, juvenile myelomonocytic leukemia; NS/MPD, Noonan syndrome associated-myeloproliferative disorder; WBC, white blood cell; HbF, hemoglobin F; HSCT, hematopoietic stem cell transplantation; OS, overall survival; CI, confidential interval; NS, not significant.

Figure Legends

Figure 1. Mutation profiles of 92 JMML cases

Mutation status of RAS pathway genes and two newly identified gene targets in a cohort of 92 JMML cases are summarized, in which homozygous and compound heterozygous mutations are indicated by asterisks and slashes, respectively **(a)**. Distribution of mutations is shown for each gene **(b)**. Open and filled circles indicate homozygous and heterozygous mutations, respectively. Red circles indicate the mutations found in NS/MPD patients.

Figure 2. Clinical features of JMML cases with or without secondary mutations

Frequency of secondary mutations in JMML patients depending on the type of RAS pathway mutations (left: *PTPN11*/*NF1*; right: other or no mutations) **(a)** and the status of hematopoietic stem cell transplantation **(b)**. *P*-values were calculated by two-sided Fisher's exact test. Impact of secondary mutations on overall and transplantation-free survivals are shown in Kaplan-Meier survival curves, where statistical significance was tested by log-rank test **(c, d)**.

Data Access: EGAS00001000521.

Figure 1

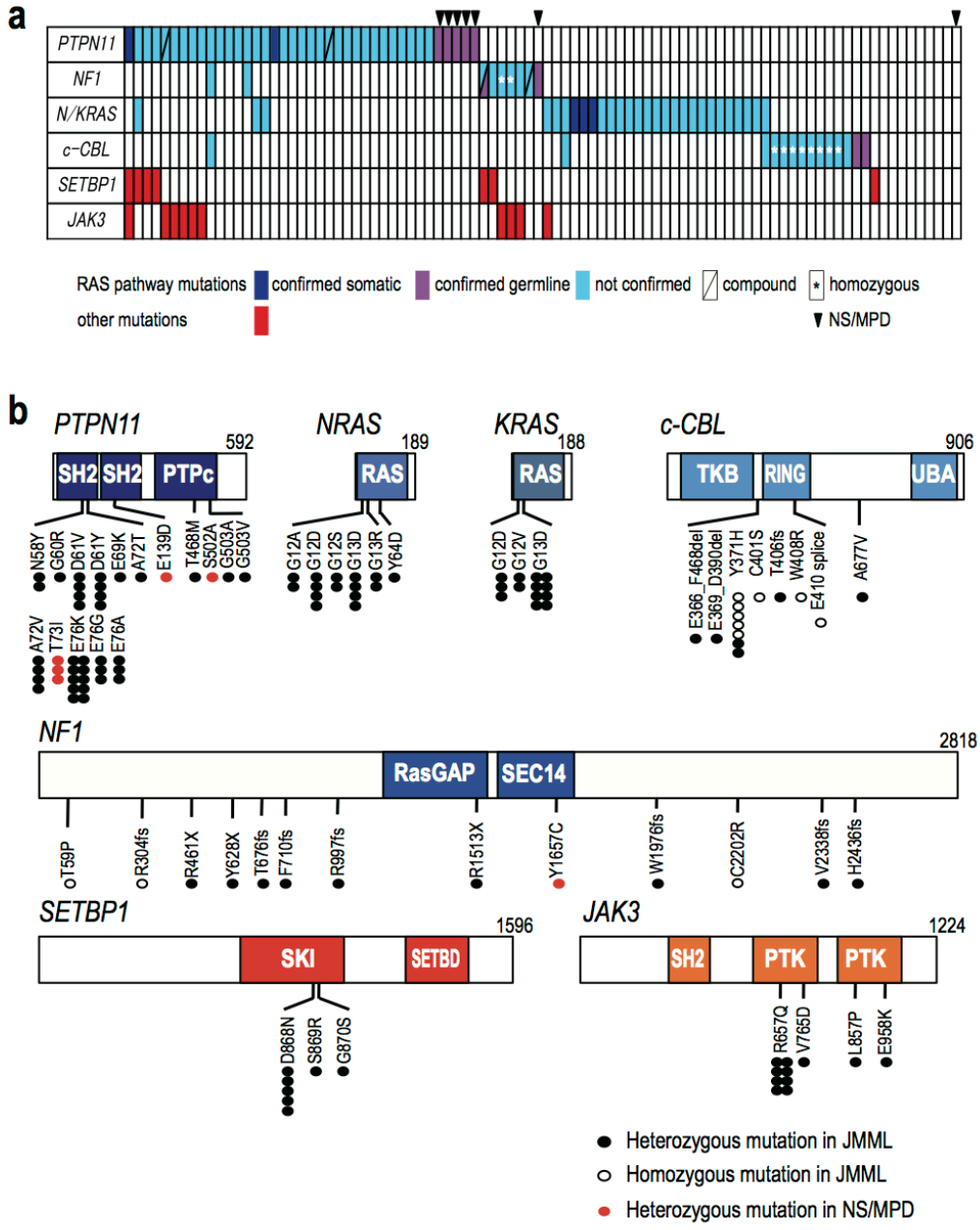
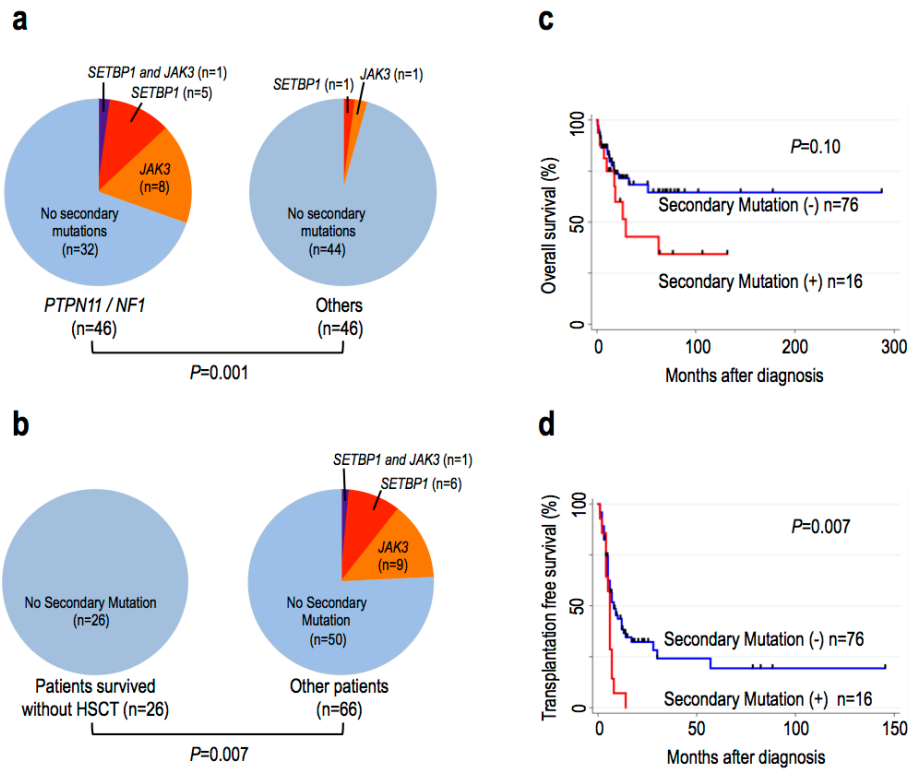

a
PTPN11
NF1
N/KRAS
c-CBL
SETBP1
JAK3
RAS pathway mutations
confirmed somatic
confirmed germline
not confirmed
compound
homozygous
other mutations
NS/MPD
b
PTPN11
592
SH2
SH2
PTPc
N58Y
G60R
D61V
D61Y
E69K
A72T
E139D
T468M
S502A
G503A
G503V
A72V
T73I
E76K
E76G
E76A
NRAS
189
RAS
G12A
G12D
G12S
G13D
G13R
Y64D
KRAS
188
RAS
G12D
G12V
G13D
c-CBL
906
TKB
RING
UBA
E366_F468del
E369_D390del
Y371H
C401S
T406fs
W408R
E410 splice
A677V
NF1
2818
RasGAP
SEC14
T59P
R304fs
R461X
Y628X
T676fs
F710fs
R997fs
R1513X
Y1657C
W1976fs
C2202R
V2338fs
H2436fs
SETBP1
1596
SKI
SETBD
D868N
S869R
G870S
JAK3
1224
SH2
PTK
PTK
R657Q
V765D
L857P
E958K
Heterozygous mutation in JMML
Homozygous mutation in JMML
Heterozygous mutation in NS/MPD

Figure 2


a
SETBP1 and JAK3 (n=1)
SETBP1 (n=5)
JAK3 (n=8)
No secondary mutations (n=32)
PTPN11 / NF1 (n=46)
SETBP1 (n=1)
JAK3 (n=1)
No secondary mutations (n=44)
Others (n=46)
P=0.001
b
No Secondary Mutation (n=26)
Patients survived without HSCT (n=26)
SETBP1 and JAK3 (n=1)
SETBP1 (n=6)
JAK3 (n=9)
No Secondary Mutation (n=50)
Other patients (n=66)
P=0.007
c
Overall survival (%)
0
50
100
P=0.10
Secondary Mutation (-) n=76
Secondary Mutation (+) n=16
0
100
200
300
Months after diagnosis
d
Transplantation free survival (%)
0
50
100
P=0.007
Secondary Mutation (-) n=76
Secondary Mutation (+) n=16
0
50
100
150
Months after diagnosis

References

1. Niemeyer, C. *et al.* Differentiating juvenile myelomonocytic leukemia from infectious disease. *Blood* **91**, 365-7 (1998).
2. Loh, M.L. *et al.* Mutations in CBL occur frequently in juvenile myelomonocytic leukemia. *Blood* **114**, 1859-63 (2009).
3. Muramatsu, H. *et al.* Mutations of an E3 ubiquitin ligase c-Cbl but not TET2 mutations are pathogenic in juvenile myelomonocytic leukemia. *Blood* **115**, 1969-75 (2010).
4. Perez, B. *et al.* Genetic typing of CBL, ASXL1, RUNX1, TET2 and JAK2 in juvenile myelomonocytic leukaemia reveals a genetic profile distinct from chronic myelomonocytic leukaemia. *Br J Haematol* **151**, 460-8 (2010).
5. Ng, S.B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790-3 (2010).
6. Minakuchi, M. *et al.* Identification and characterization of SEB, a novel protein that binds to the acute undifferentiated leukemia-associated protein SET. *Eur J Biochem* **268**, 1340-51 (2001).
7. Damm, F. *et al.* SETBP1 mutations in 658 patients with myelodysplastic syndromes, chronic myelomonocytic leukemia and secondary acute myeloid leukemias. *Leukemia* (2013).
8. Laborde, R.R. *et al.* SETBP1 mutations in 415 patients with primary myelofibrosis or chronic myelomonocytic leukemia: independent prognostic impact in CMML. *Leukemia* (2013).
9. 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 i(17)(q10), ASXL1 and CBL mutations. *Leukemia* (2013).
10. Piazza, R. *et al.* Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. *Nat Genet* **45**, 18-24 (2013).
11. Thol, F. *et al.* SETBP1 mutation analysis in 944 patients with MDS and AML. *Leukemia* (2013).
12. Panagopoulos, I. *et al.* Fusion of NUP98 and the SET binding protein 1 (SETBP1) gene in a paediatric acute T cell lymphoblastic leukaemia with t(11;18)(p15;q12). *Br J Haematol* **136**, 294-6 (2007).
13. Cristobal, I. *et al.* SETBP1 overexpression is a novel leukemogenic mechanism that predicts adverse outcome in elderly patients with acute myeloid leukemia. *Blood* **115**, 615-25 (2010).
14. Goyama, S. *et al.* Evi-1 is a critical regulator for hematopoietic stem cells and transformed leukemic cells. *Cell Stem Cell* **3**, 207-20 (2008).
15. 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**, 401-9 (2006).
16. Hoischen, A. *et al.* De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* **42**, 483-5 (2010).
17. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64-9 (2011).
18. Flotho, C. *et al.* Genome-wide single-nucleotide polymorphism analysis in juvenile myelomonocytic leukemia identifies uniparental disomy surrounding the NF1 locus in cases associated with neurofibromatosis but not in cases with mutant RAS or PTPN11. *Oncogene* **26**, 5816-21 (2007).
19. Walters, D.K. *et al.* Activating alleles of JAK3 in acute megakaryoblastic leukemia. *Cancer Cell* **10**, 65-75 (2006).
20. Sato, T. *et al.* Functional analysis of JAK3 mutations in transient myeloproliferative disorder and acute megakaryoblastic leukaemia accompanying Down syndrome. *Br J Haematol* **141**, 681-8 (2008).
21. De Vita, S. *et al.* Loss-of-function JAK3 mutations in TMD and AMKL of Down syndrome. *Br J Haematol* **137**, 337-41 (2007).

22. Norton, A. *et al.* Analysis of JAK3, JAK2, and C-MPL mutations in transient myeloproliferative disorder and myeloid leukemia of Down syndrome blasts in children with Down syndrome. *Blood* **110**, 1077-9 (2007).
23. Kiyoi, H., Yamaji, S., Kojima, S. & Naoe, T. JAK3 mutations occur in acute megakaryoblastic leukemia both in Down syndrome children and non-Down syndrome adults. *Leukemia* **21**, 574-6 (2007).
24. Elliott, N.E. *et al.* FERM domain mutations induce gain of function in JAK3 in adult T-cell leukemia/lymphoma. *Blood* **118**, 3911-21 (2011).
25. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157-63 (2012).
26. Koo, G.C. *et al.* Janus Kinase 3-Activating Mutations Identified in Natural Killer/T-cell Lymphoma. *Cancer Discov* (2012).
27. Zhang, J. *et al.* A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* **481**, 329-34 (2012).
28. Piazza, R. *et al.* Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. *Nat Genet* **45**, 18-24 (2012).
29. Makishima, H. *et al.* Somatic SETBP1 mutations in myeloid malignancies. *Nat Genet* (2013).
30. Crozatier, M. & Meister, M. Drosophila haematopoiesis. *Cell Microbiol* **9**, 1117-26 (2007).
31. Changelian, P.S. *et al.* Prevention of organ allograft rejection by a specific Janus kinase 3 inhibitor. *Science* **302**, 875-8 (2003).

Online Methods

Patients

We studied 92 children (61 boys and 31 girls) with JMML, including 7 patients with Noonan syndrome-associated myeloproliferative disorder (NS/MPD), who were diagnosed as having JMML in institutions throughout Japan. Written informed consent was obtained from patients' parents before sample collection. This study was approved by the Ethics Committees of Nagoya University Graduate School of Medicine and of Graduate School of Medicine, the University of Tokyo in accordance with the Declaration of Helsinki. Diagnosis of JMML was made based on the internationally accepted criteria.[1] Characteristics of the 92 JMML cases are summarized in Table 2. The median age at diagnosis was 16 months (1-160 months). Karyotypic abnormalities were detected in 16 patients, including 8 with monosomy 7. Fifty-six of the 92 (61 %) patients received allogeneic hematopoietic stem cell transplantation (HSCT).

Sample preparation

Genomic DNA was extracted using QIAamp DNA Blood Mini kit and a QIAamp DNA Investigator kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. T Cell Activation/Expansion kits, human (Miltenyi Biotec, Bergisch Gladbach, Germany) were used for the expansion of CD3-positive T cells from patients' PB- or BM-MNCs.

Whole-exome sequencing

Exome capture from paired tumor-reference DNA was performed using SureSelect® Human All Exon V3 (Agilent Technologies, Palo Alto, CA) covering 50 Mb of coding exons according to the manufacturers' protocol. Enriched exome fragments were subjected to massively parallel sequencing using HiSeq 2000 (Illumina). Candidate somatic mutations were detected through our in-house pipeline (Genomon: http://genomon.hgc.jp/exome/) as previously described (also see below).[2]

Detection of the mutations from whole-exome sequencing data

Detection of candidate somatic mutations was performed according to the previously described algorithms with minor modifications.[2] Briefly, the number of the reads containing single nucleotide variations (SNVs) and indels in both tumor and reference samples was enumerated using samtools[3] and the null hypothesis of equal allele frequencies between tumor and reference was tested using the two-tailed Fisher's exact test. A variant was adopted as a candidate somatic mutation, if it had $p < 0.01$, observed in bi-directional reads (i.e. in both + and – strands of the reference sequence), and its allele frequency was less than

0.25 in the corresponding reference sample. For the detection of germline mutations of the RAS pathway genes, the SNVs and indels having the allele frequency of more than 0.25 (SNVs) and 0.10 (indels) were interrogated among the 46 genes, which consisted of the known JMML-related RAS pathway genes and the genes registered in the pathway databases ('Ras Signaling Pathway' in BioCarta (http://www.biocarta.com) and 'Signaling to RAS' in Reactome[4]). For variant calls of the tumor samples, for which the paired normal reference was not available, the candidate variants in the RAS pathway were detected at the allele frequency > 0.10. Finally, the list of candidate somatic/germline mutations was generated by excluding synonymous SNVs and other variants registered in either dbSNP131 (http://www.ncbi.nlm.nih.gov/projects/SNP/) or in-house SNP database constructed from 180 individual samples. All candidates were validated by Sanger sequencing as previously described.

Estimation of tumor contents

The tumor content in the bone marrow specimen was estimated from the allele frequency of the somatic mutations obtained by deep sequencing. For homozygous mutations, as indicated by the allele frequency > 0.75, the tumor content (F^{tum}) was calculated from the observed frequency (F^{obs}) of the mutation according to the following formula; $F^{tum} = 2 \times F^{obs} - 1$. For heterozygous mutations, the tumor content was calculated by doubling the allele frequency.

Power analysis of whole-exome sequencing

The power of detecting somatic mutations at each nucleotide position in whole-exome sequencing was estimated by Monte-Carlo simulation (N = 1000) based on the observed mean depth for each exon in germline and tumor samples and the observed tumor content for that sample, which were estimated through the allele frequencies of the observed mutations. For the samples with no observed somatic mutations, the average tumor content of the informative samples was employed. Simulations were performed across a total of 192,424 exons.

Copy number analysis in whole-exome sequencing data

To detect the copy number lesions at a single exon level, the mean coverage of each exon normalized by the mean depth of the entire sample, was compared with that of 12 unrelated normal DNA samples. The exons showing the normalized coverage beyond the 3 standard deviations from the mean coverage of the reference samples were called as candidates

undergoing copy number alterations. All candidate exons of the RAS pathway genes were visually inspected using Integrative Genomics Viewer,[5] and validated by Sanger sequencing of corresponding putative breakpoint-containing fragments.

Target deep sequencing

Deep sequencing of the targeted genes was performed essentially as described in the section of 'deep sequencing of pooled target exons' in the reference,[2] except that the target DNA was not pooled. Briefly, all exons of *PTPN11*, *NF1*, *KRAS*, *NRAS*, *c-CBL*, *SETBP1*, *JAK3, and SH3BP1* were PCR-amplified by Quick Taq HS DyeMix (TOYOBO, Osaka, Japan) and PrimeSTAR GXL DNA Polymerase Kit (Takara Bio, Otsu, Japan) using NotI-tagged primers (Supplementary Table 3). The PCR products from an individual sample were combined and purified by QIAquick PCR Purification Kit (QIAGEN) for the subsequent digestion with NotI (Fermentas, Glen Burnie, MD, US). The digested PCR product was again purified, concatenated with T4 DNA ligase (Takara Bio), and sonicated to 150bp fragments on average using Covaris®. The fragments were processed for sequencing according to a modified Illumina pair-end library protocol, and sequences were read by Hiseq 2000 using a 100 bp pair-end reads protocol.

Variant calls in target deep sequencing

The data processing and variant calling was performed with a set of modifications to the previous publication.[2] Each read was aligned to the set of target sequences of PCR amplification, for which BLAT,[6] instead of BWA,[7] was used with -fine option. The mapping information in the .psl format was converted to the .sam format with pair read information. Out of the successfully mapped reads, the following reads were excluded from further analysis, which were mapped to multiple sites, mapped with more than 4 mismatched bases or had more than 10 soft-clipped bases. Next, the 'Estimation_CRME' script was run to eliminate strand-specific errors and exclude cycle-dependent errors. A strand-specific mismatch ratio was calculated for each nucleotide variant for both strands using those bases between 11 to 50 cycles. By excluding the top 5 cycles showing the highest mismatch rates, strand-specific mismatch rates were recalculated and the smaller value between both strands was adopted as a nominal mismatch ratio for that variant. After excluding those variants found dbSNP131 or in-house SNP database, the non-silent variants having more than 0.05 mismatch ratio were called as candidates, unless the median value of the mismatch ratio at the relevant nucleotide positions in the 92 samples exceeded 0.01, since it was likely to be caused by systematic PCR problems. Finally, the candidates with mismatch ratios > 0.15 were further validated by Sanger sequencing.

Annotation of the detected mutations

The detected mutations were annotated using ANNOVAR.[8] The positions of the mutations were based on the following RefSeq transcript sequences: NM_002834.3 for *PTPN11*, NM_000267.3 for *NF1*, NM_002524.4 for *NRAS*, NM_004985.3 for *KRAS*, NM_005188.3 for *c-CBL*, NM_015559.2 for *SETBP1*, and NM_000215.3 for *JAK3*. The effect of the mutations to the protein function was assessed by SIFT,[9] PolyPhen2,[10] and MutationTaster.[11]

Whole-genome sequencing

The paired tumor-reference DNA was sequenced with Hiseq 2000 according to manufacturer's instructions to obtain 30x read coverage for reference and 40x for tumor. Obtained FASTQ sequences were aligned to human reference genome (hg19) using Burrows-Wheeler Aligner[12] 0.5.8 with default parameters. The pairs of the sequences, at least one of which were not mapped or considered having possible mapping problems (mapping quality less than 40, insertions or deletions, soft clipped sequence of more than 10% length of the original sequence, irregular pair read orientation, or mate distance longer than 2000 bp), were tried to align by BLAT[13] with default parameters except stepSize=5 and repMatch=2253. Statistics of the mapping were calculated by counting the bases of each genomic position by SAMtools.[14] For variant calling, the numbers of the variant and reference bases with base quality > 30 were counted in both germline and tumor samples and Fisher's exact test was applied. The variants with $p < 0.01$ were called. The variants having > 0.25 allele frequency in germline sample were excluded. The variants found in 12 unrelated germline samples at > 0.01 allele frequency on average were also excluded due to the high possibility of false positive calls. Copy number estimation was performed by calculating the averaged ratio of read depths in germline and tumor samples in 10k bases bin. Allele-specific copy number plot was generated by measuring the allele frequency of the tumor sample in the positions, in which more than 25% of the allele mismatch was observed in germline samples. For the detection of the chromosomal structural variations (SVs), the soft clipped sequences which could be mapped on a unique genomic position were harvested. The SV candidates, which had more than 4 supporting read pairs in total and at least 1 read pair from each side of the breakpoint, were called. The contig sequences were generated by assembling the reads around 200 bp of the breakpoint by CAP3,[15] and SVs having the contig sequence which could be aligned to the alternate assembly of the hg19 genome with more than 93% identity were excluded as false positives. The SVs with the read depth of more than 150 at least on side of the breakpoint were considered to be on the repeat element and also excluded. For the detection of the viruses, the unmapped sequences were aligned to the collection of all viral

genomes in RefSeq database (http://www.ncbi.nlm.nih.gov/RefSeq/) using BLAT. The virus were considered to be detected if its genome were covered with >1 mean read coverage.

cDNA Sequencing

Total RNA was extracted using RNeasy Mini Kit (QIAGEN) and was reverse transcribed by ThermoScript RT-PCR system (Life technologies, Carlsbad, CA). Target sequences were PCR-amplified by PrimeSTAR GXL DNA Plymerase Kit (Takara Bio) using primers listed in Supplementary Table 3 and sequenced.

Statistical analysis

For comparison of the frequency of mutations or other clinical features between disease groups, categorical variables were analyzed using the Fisher's exact test and continuous variables were tested using Mann-Whitney U test. Overall survival (OS) and transplantation free survival (TFS) were estimated by the Kaplan-Meier method. Hazard ratios for survivals with 95%CIs were estimated according to the Cox-proportional hazard model and the difference in survivals was tested by log-rank test. STATA version 12.0 (StataCorp, College Station, TX) was used for all statistical calculations.

References

1. Niemeyer, C. *et al.* Differentiating juvenile myelomonocytic leukemia from infectious disease. *Blood* **91**, 365-7 (1998).
2. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64-9 (2011).
3. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
4. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res* **37**, D619-22 (2009).
5. Thorvaldsdóttir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* (2012).
6. Kent, W.J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-64 (2002).
7. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
8. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
9. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-81 (2009).
10. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-9 (2010).
11. Schwarz, J.M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575-6 (2010).
12. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
13. Kent, W.J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-64 (2002).
14. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
15. Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res* **9**, 868-77 (1999).

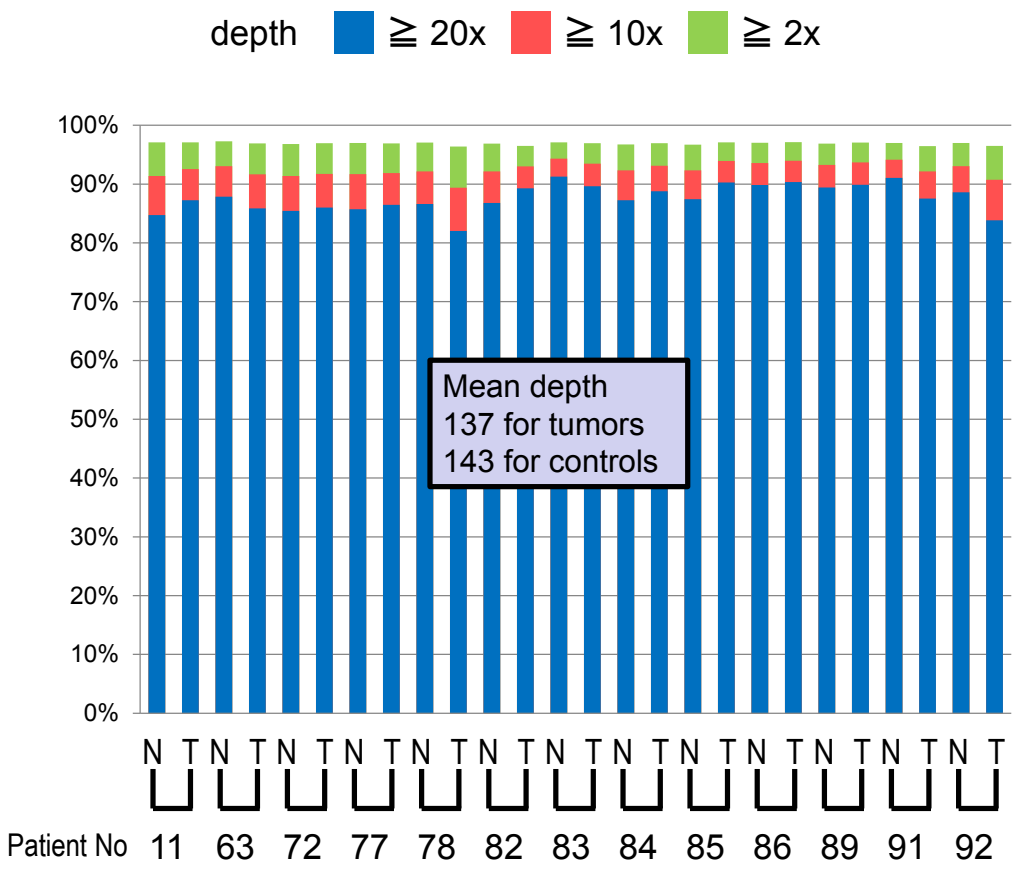
Manuscript number: NG-LE34408R
Corresponding author: Seiji Kojima

NATURE GENETICS SUPPLEMENTARY INFORMATION LIST

Supplementary item	Descriptive title
Supplementary Figure 1.	Mean coverage of whole-exome sequencing
Supplementary Figure 2	Power analysis of whole-exome sequencing
Supplementary Figure 3	Sanger sequencing of the mutations found in whole-exome sequencing
Supplementary Figure 4	Molecular pathways of SETBP1 and JAK3 in leukemogenesis
Supplementary Figure 5	Mean coverage of target deep sequencing
Supplementary Figure 6	The summary of the additional analyses in 2 samples without any RAS pathway alterations in deep sequencing and whole-exome sequencing
Supplementary Figure 7	The summary of the additional analyses in 8 samples without any RAS pathway mutations
Supplementary Figure 8	Confirmation of the bi-allelic expression of the mutated genes
Supplementary Figure 9	The allele frequencies of the SNPs in deep sequencing
Supplementary Figure 10	Secondary mutations in JMML
Supplementary Figure 11	The number of the somatic mutations across cancer types
Supplementary Table 1	Patients analyzed by whole-exome sequencing
Supplementary Table 2.	List of the detected mutations
Supplementary Table 3	List of primers

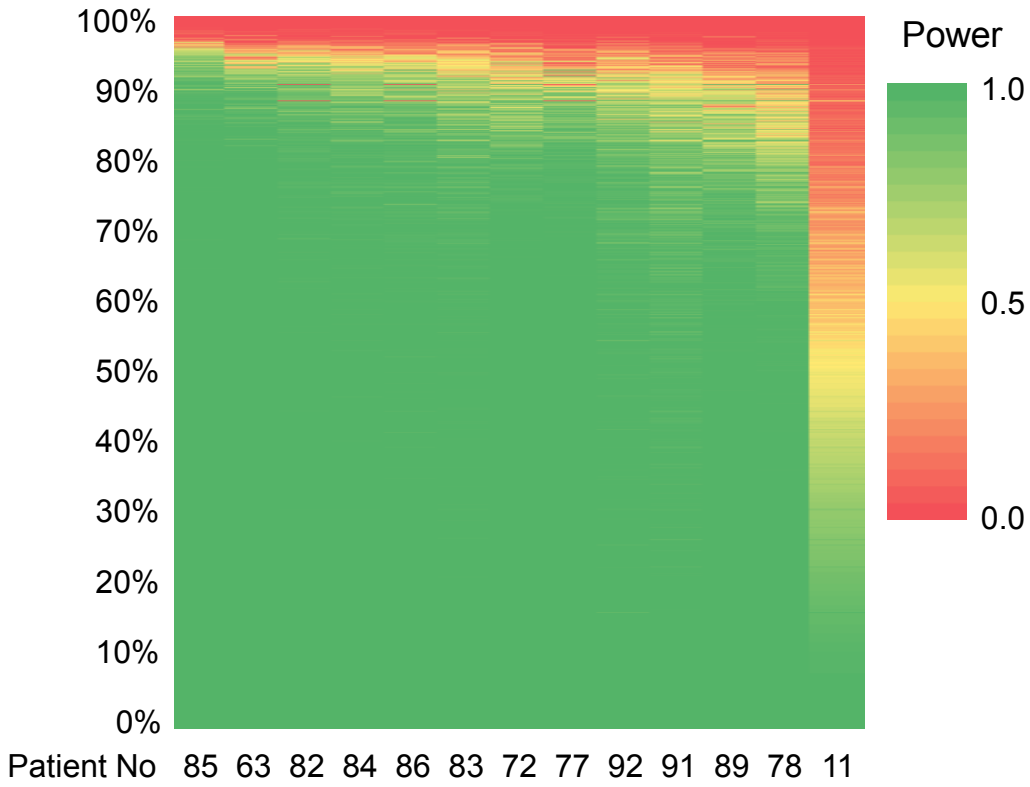
Supplementary Information

Supplementary Figure 1. Mean coverage of whole-exome sequencing



The coverage of the target coding regions analyzed at indicated depth is plotted for each pair of tumor and normal specimens. The percentage indicated by blue, red and green bars means the percentage of the region covered with 20x, 10x and 2x read depth, respectively. The mean depth for all tumor and normal samples is also indicated.

Supplementary Figure 2. Power analysis of whole-exome sequencing

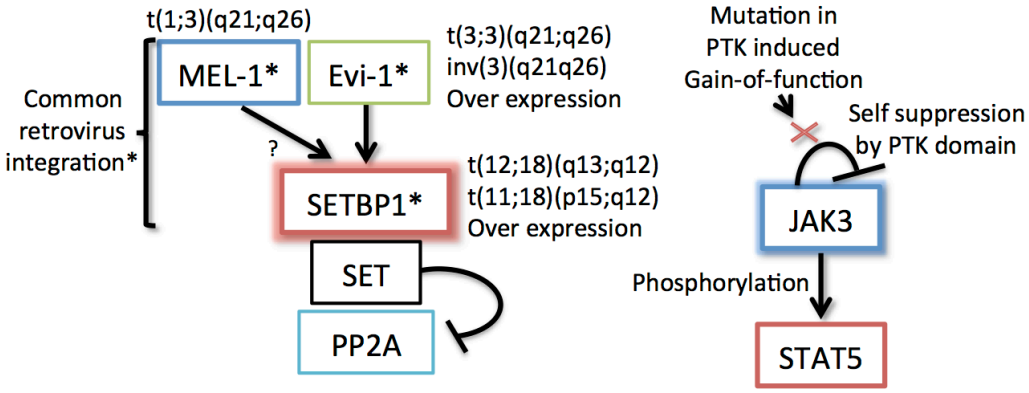


The heatmap shows the estimated power for the mutation detection for the entire coding region, calculated from the mean depth for each exon and the estimated tumor content of the tumor and germline samples. The power is indicated by color scale (0=red, 1=green). Y axis shows the percentage of the coding region. The mean power across the samples is 0.88. 79% of the coding region is covered with the detection power > 0.9.

Supplementary Figure 3. Sanger sequencing of the mutations found in whole-exome sequencing

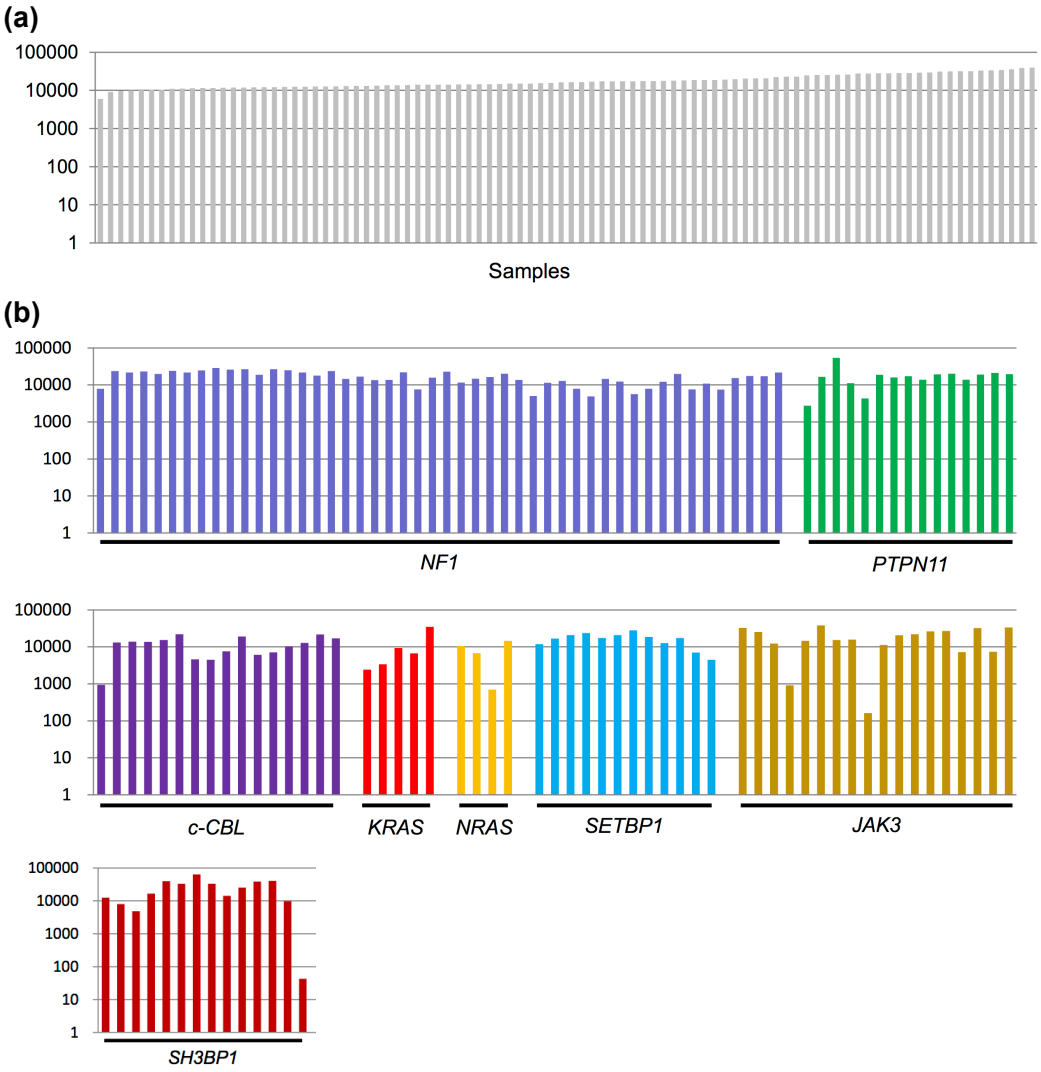
The result of Sanger sequencing for the mutations found in whole-exome sequencing is shown. Upper panel of each presentation shows the result of the whole bone marrow, while lower panel shows that of germline control samples. The variant alleles are indicated below the reference sequence. The *c-CBL* c.1096-110del643 mutation starts at the outside of the coding region and therefore no reading frame information is annotated.

Supplementary Figure 4. Molecular pathways of *SETBP1* and *JAK3* in leukemogenesis



Activation of *SETBP1*, *Evi1* and *MEL1*, through retrovirus integration was implicated in recovery of granulopoiesis after gene therapies for chronic glanulomatous disease. *SETBP1*, of which overexpression and associated translocations are observed in leukemia patients, causes PP2A inhibition through stabilization of SET. *JAK3* mutations in PTK domain induce loss of self-suppression and STAT5 phosphorylation associated with GM-CSF hypersensitivity.

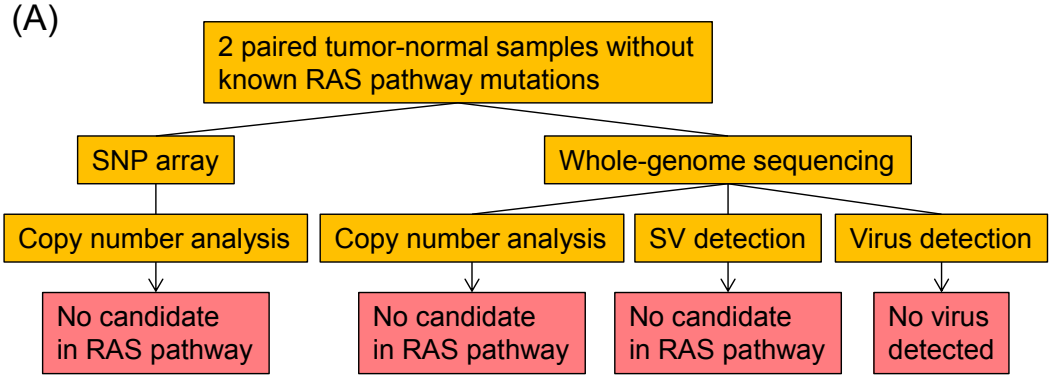
Supplementary Figure 5. Mean coverage of target deep sequencing



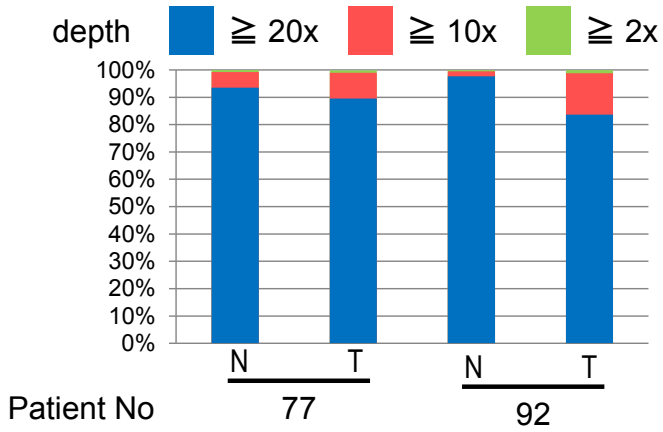
(a) The mean coverage of deep sequencing for 92 individual samples, sorted in ascending order, is shown.
(b) The mean coverage of the PCR amplicons for the indicated genes, sorted by the order which corresponds to that of the primers listed in Supplemetary Table 3, is shown.

Supplementary Figure 6. The summary of the additional analyses in 2 samples without any RAS pathway alterations in deep sequencing and whole-exome sequencing

(A)



(B)



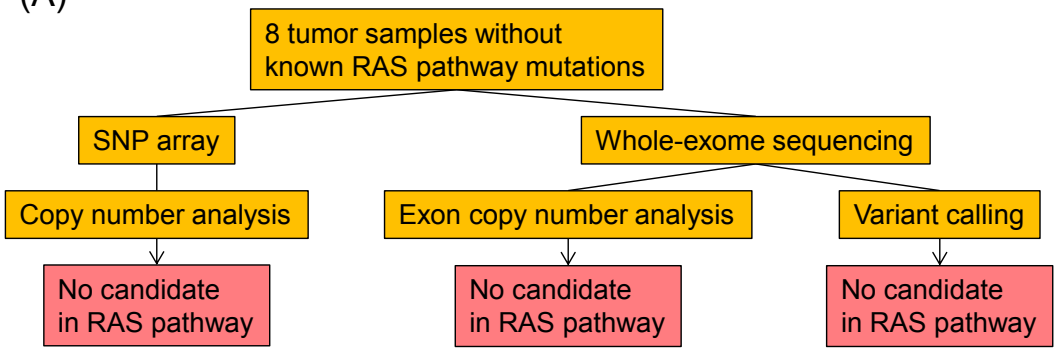
(C)

Patient No	position	type	affected genes
92	chr2:3424531-34502125	tandem duplication	-

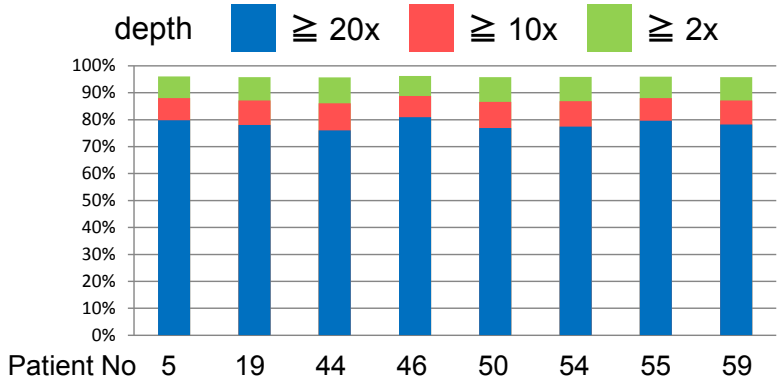
(A) The flowchart of the analysis of the 2 samples, in which no RAS pathway alterations were found in either deep sequencing or whole-exome sequencing. The SNP array analysis and the whole-genome sequencing were applied to the 2 samples. In whole-genome sequencing, the copy number analysis, the chromosomal structural variation (SV) detection, and the virus detection were performed. However, no candidate alterations were found. (B) The mean coverage for the entire coding region of the whole-genome sequencing. (C) The summary of the detected SVs. A 260 kb tandem duplication was observed in patient #92. No gene was annotated in the region in RefSeq database.

Supplementary Figure 7. The summary of the additional analyses in 8 samples without any RAS pathway mutations

(A)



(B)

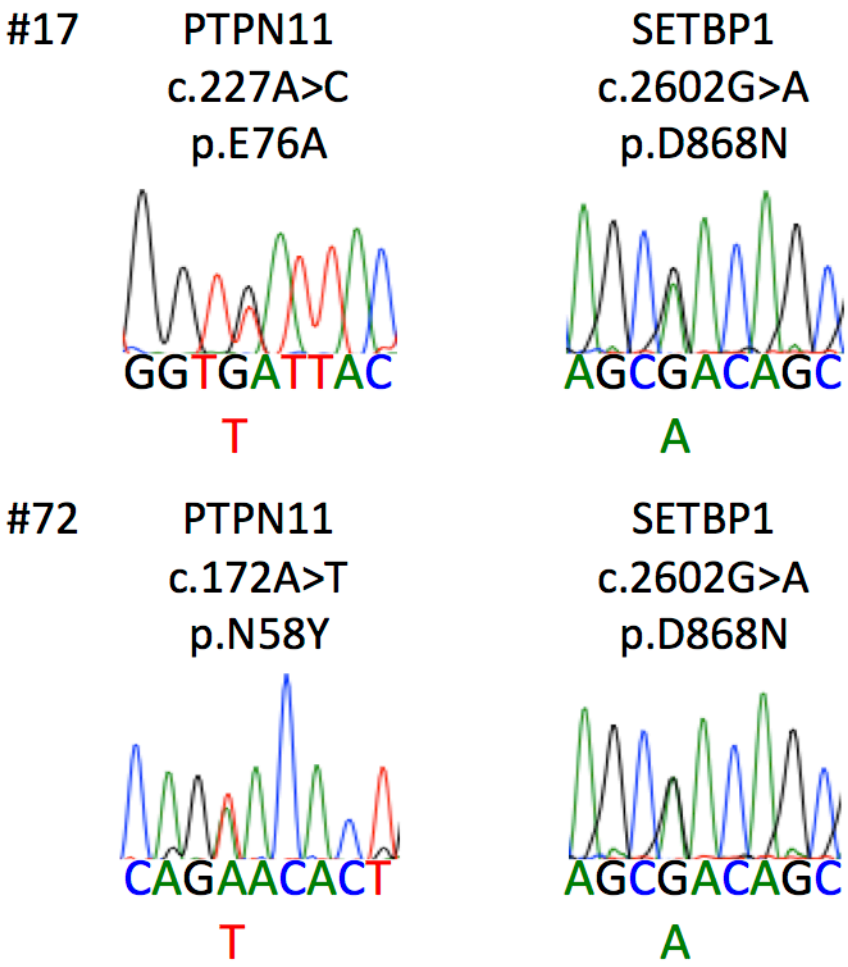


(C)

Patient No	Number of variants
5	234
19	234
44	254
46	264
50	245
54	213
55	255
59	228

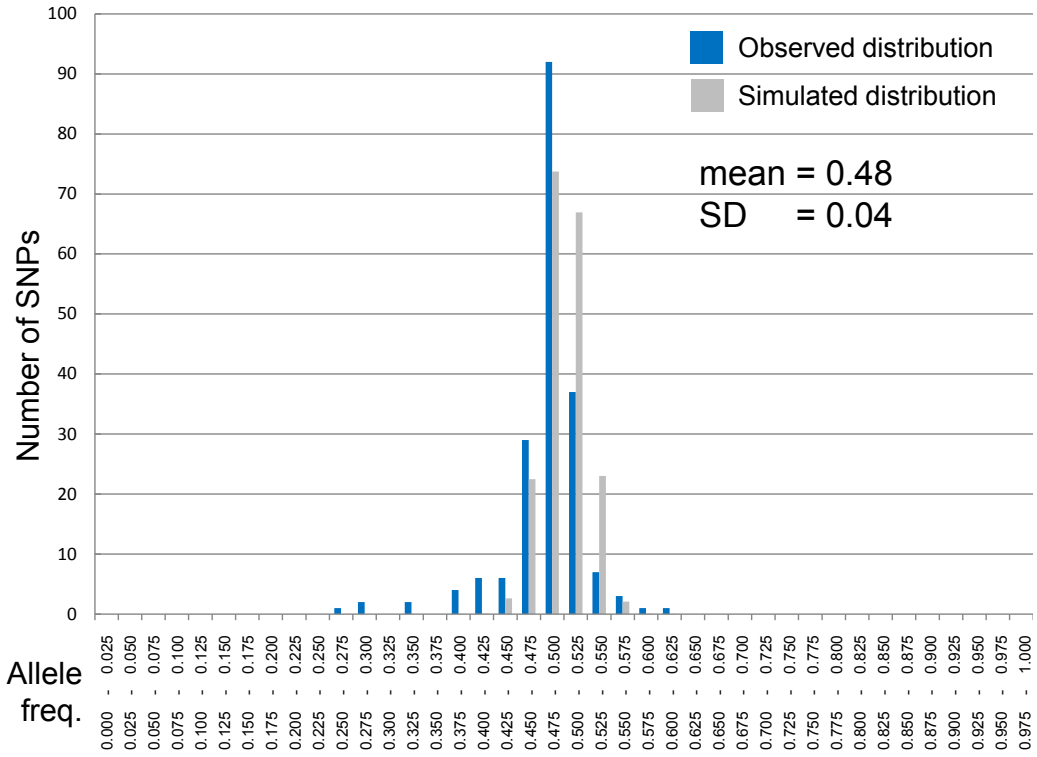
(A) The flowchart of the analysis of the 8 samples, in which no RAS pathway mutations were found in deep sequencing. No control DNA was available for these samples and only DNA from tumor specimen was analyzed. The SNP array analysis and the whole-exome sequencing were applied to all samples. However, no candidate alterations were found in RAS pathway. (B) The mean coverage of the whole-exome sequencing. (C) The summary of variant calling in these samples. The number of the non-silent variants, including germline and somatic ones, are indicated.

Supplementary Figure 8. Confirmation of the bi-allelic expression of the mutated genes



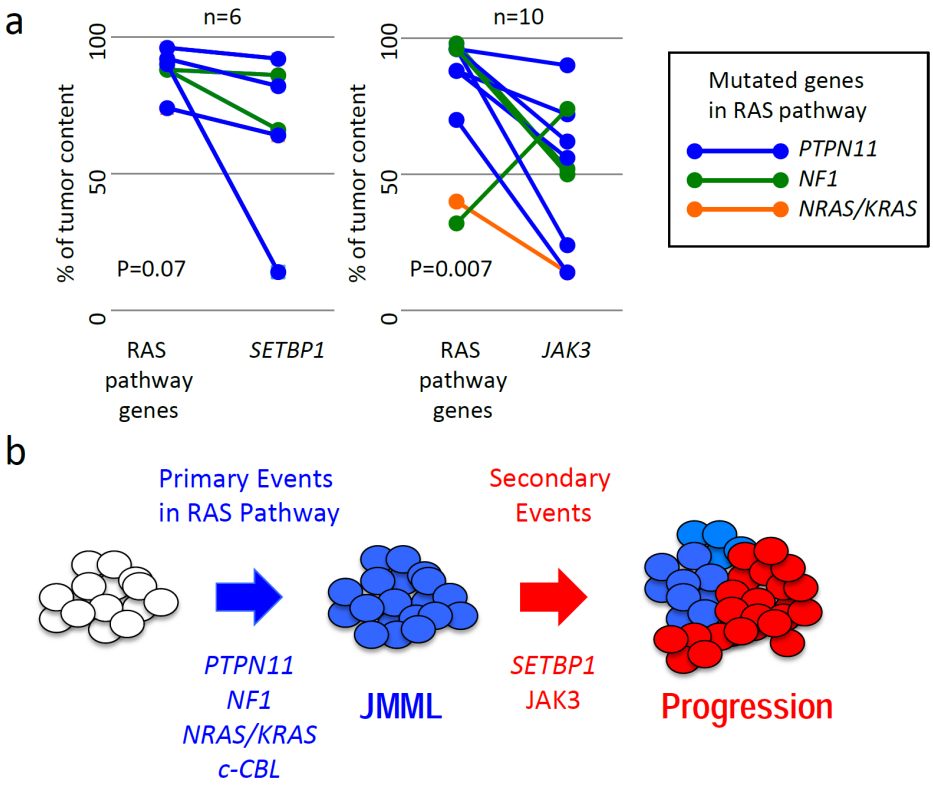
cDNAs from 2 patients, which had both *PTPN11* and *SETBP1* mutations, were analyzed for their expression level by Sanger sequencing. The bi-allelic expression was confirmed in both patients.

Supplementary Figure 9. The allele frequencies of the SNPs in deep sequencing



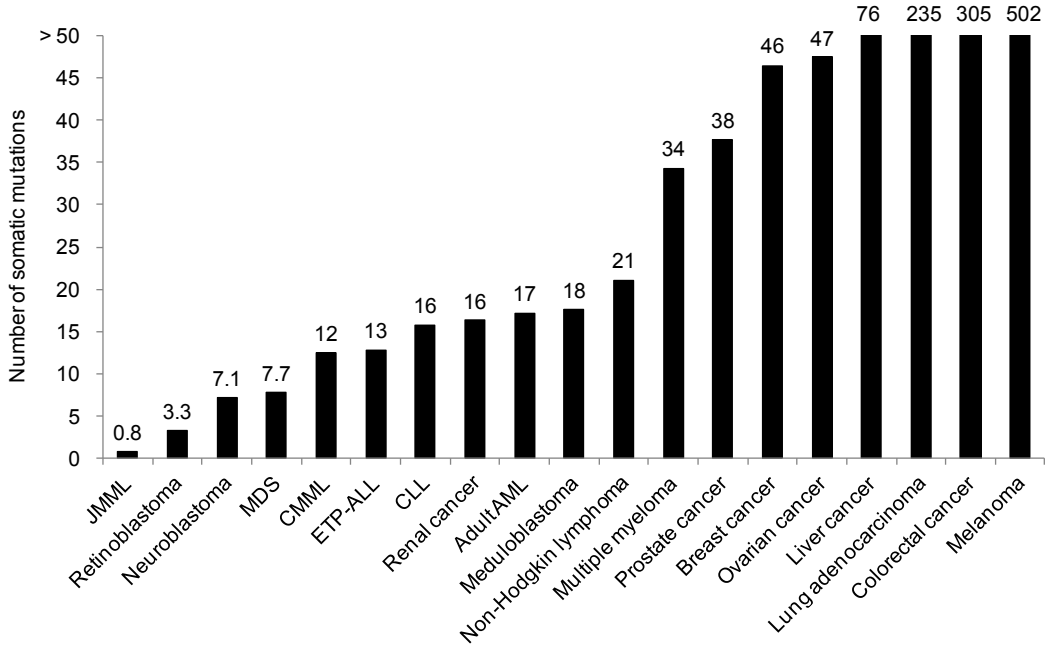
The blue bars of the histogram show the distribution of the allele frequency of 191 heterozygous SNPs observed in deep sequencing. The gray bars show the simulated distribution based on binomial distribution (trials = 500).

Supplementary Figure 10. Secondary mutations in JMML



(a) Comparison of allele frequencies between RAS pathway mutations and either of the *SETBP1* and *JAK3* mutations with *p*-values calculated by two-sided paired-T test. **(b)** Clonal history of gene mutations in JMML.

Supplementary Figure 11. The number of the somatic mutations across cancer types



The numbers of the somatic mutations in the coding region for the indicated cancer types are shown. The numbers are referenced from the literatures.[1-18] Note the exceptionally small number of the somatic mutations in JMML among the cancer types listed. Abbreviations: MDS, myelodysplastic syndrome; CMML, chronic myelomonocytic leukemia; ETP-ALL, early T-cell precursor acute lymphoblastic leukemia; CLL, chronic lymphocytic leukemia; AML, Acute myeloblastic leukemia.

Supplementary Table 1. Patients analyzed by whole-exome sequencing

Patient No.	Gender	Age at diagnosis (m)	WBC ($x10^9$/L)	Monocyte ($x10^9$/L)	Hb (g/dL)	HbF (%)	PLT ($x10^9$/L)	Karyotype	Tumor Sample	Reference Sample
11	M	160	15.1	0.9	8.5	24.7	29.4	46,XY	BM-MNC	CD3(+) cells
63	F	105	24.0	9.4	5.4	3	36	48,XX,+8,+19	BM-MNC	Umbilical cord
72	M	28	27.3	3.6	9.6	53	19	46,XY	BM-MNC	CD3(+) cells
77	F	21	17.6	1.9	9.7	9.2	53	47,XX,+8	BM-MNC	CD3(+) cells
78	M	13	49.9	9.0	9.2	1	251	46,XY	BM-MNC	CD3(+) cells
82	F	7	24.3	6.6	11	5.9	217	46,XX	BM-MNC	CD3(+) cells
83*	M	3	13.1	1.7	11.3	14	201	46,XY	BM-MNC	CD3(+) cells
84	M	1	131.3	18.4	9.8	68	98	46,XY	BM-MNC	CD3(+) cells
85	M	24	19.8	5.6	8.2	62.6	73	46,XY	BM-MNC	CD3(+) cells
86	F	4	34.1	12.3	8.1	15.6	28	46,XX	BM-MNC	CD3(+) cells
89*	M	1	52.1	6.8	13	NA	48	46,XY	BM-MNC	CD3(+) cells
91*	M	2	20.5	3.1	10.5	NA	191	46,XY	BM-MNC	CD3(+) cells
92*	M	1	40.2	3.0	12.7	NA	483	NA	BM-MNC	CD3(+) cells

*Noonan syndrome-associated myeloproliferative disease; WBC, white blood cell; Hb, hemoglobin; HbF, hemoglobin F; PLT, platelets; BM-MNC, bone marrow-derived mononuclear cells; NA, not assessed.

Supplementary Table 2. List of the detected mutations

Patient No.	Gene	Chr	Nucleotide change	Amino acid change	Allele Freq.	SIFT	PolyPhen2	MutationTaster
1	*PTPN11*	chr12	182A>T	D61V	0.4	deleterious	probably damaging	disease causing
2	*c-CBL*	chr11	1111T>C	Y371H	0.19	deleterious	probably damaging	disease causing
2	*NF1*	chr17	2990delG	R997fs	0.12	N/A	N/A	disease causing
2	*PTPN11*	chr12	181G>T	D61Y	0.46	deleterious	probably damaging	disease causing
3	*KRAS*	chr12	35G>A	G12D	0.38	deleterious	possibly damaging	disease causing
4	*c-CBL*	chr11	1111T>C	Y371H	0.44	deleterious	probably damaging	disease causing
6	*NF1*	chr17	2027delC	T676fs	0.47	N/A	N/A	disease causing
6	*NF1*	chr17	1884C>A	Y628X	0.41	tolerated	possibly damaging	disease causing
7	*NRAS*	chr1	190T>G	Y64D	0.46	deleterious	probably damaging	disease causing
7	*PTPN11*	chr12	215C>T	A72V	0.42	deleterious	probably damaging	disease causing
8	*PTPN11*	chr12	182A>T	D61V	0.48	deleterious	probably damaging	disease causing
9	*PTPN11*	chr12	214G>A	A72T	0.38	deleterious	probably damaging	disease causing
10	*NRAS*	chr1	34A>G	G12S	0.05	deleterious	probably damaging	disease causing
10	*PTPN11*	chr12	226G>A	E76K	0.47	deleterious	probably damaging	disease causing
11	*NF1*	chr17	5927delG	W1976fs	0.44	N/A	N/A	disease causing
11	*NF1*	chr17	4537C>T	R1513X	0.42	tolerated	possibly damaging	disease causing
11	*SETBP1*	chr18	2602G>A	D868N	0.33	deleterious	probably damaging	polymorphism
12	*JAK3*	chr19	1970G>A	R657Q	0.07	deleterious	probably damaging	disease causing
12	*KRAS*	chr12	35G>T	G12V	0.2	deleterious	probably damaging	disease causing
13	*JAK3*	chr19	2872G>A	E958K	0.28	deleterious	probably damaging	disease causing
13	*JAK3*	chr19	1970G>A	R657Q	0.07	deleterious	probably damaging	disease causing
13	*PTPN11*	chr12	227A>G	E76G	0.44	deleterious	probably damaging	disease causing
14	*NRAS*	chr1	35G>C	G12A	0.08	deleterious	probably damaging	disease causing
14	*PTPN11*	chr12	181G>T	D61Y	0.45	deleterious	probably damaging	disease causing
14	*SETBP1*	chr18	2608G>A	G870S	0.07	deleterious	probably damaging	disease causing
15	*JAK3*	chr19	1970G>A	R657Q	0.25	deleterious	probably damaging	disease causing
15	*NF1*	chr17	911delG	R304fs	0.96	N/A	N/A	disease causing
16	*c-CBL*	chr11	1202G>C	C401S	0.95	deleterious	probably damaging	disease causing
17	*PTPN11*	chr12	181G>T	D61Y	0.37	deleterious	probably damaging	disease causing
17	*SETBP1*	chr18	2602G>A	D868N	0.32	deleterious	probably damaging	polymorphism
18	*PTPN11*	chr12	226G>A	E76K	0.46	deleterious	probably damaging	disease causing
20	*KRAS*	chr12	35G>A	G12D	0.47	deleterious	possibly damaging	disease causing
21	*JAK3*	chr19	1970G>A	R657Q	0.45	deleterious	probably damaging	disease causing
21	*PTPN11*	chr12	215C>T	A72V	0.48	deleterious	probably damaging	disease causing
22	*NF1*	chr17	1381C>T	R461X	0.12	tolerated	possibly damaging	disease causing

22	*NF1*	chr17	2130_2131insCG	F710fs	0.1	N/A	N/A	disease causing
22	*PTPN11*	chr12	215C>T	A72V	0.45	deleterious	probably damaging	disease causing
23	*c-CBL*	chr11	1103del66	E369_D390del	0.39	N/A	N/A	N/A
24	*PTPN11*	chr12	215C>T	A72V	0.48	deleterious	probably damaging	disease causing
25	*KRAS*	chr12	35G>T	G12V	0.47	deleterious	probably damaging	disease causing
26	*NRAS*	chr1	34G>A	G12S	0.5	deleterious	benign	disease causing
27	*c-CBL*	chr11	1111T>C	Y371H	0.97	deleterious	probably damaging	disease causing
28	*PTPN11*	chr12	227A>C	E76A	0.46	deleterious	probably damaging	disease causing
28	*SETBP1*	chr18	2602G>A	D868N	0.41	deleterious	probably damaging	polymorphism
29	*PTPN11*	chr12	227A>G	E76G	0.47	deleterious	probably damaging	disease causing
30	*NRAS*	chr1	35G>A	G12D	0.48	deleterious	possibly damaging	disease causing
31	*PTPN11*	chr12	1508G>C	G503A	0.47	deleterious	probably damaging	disease causing
32	*PTPN11*	chr12	227A>G	E76G	0.34	deleterious	probably damaging	disease causing
33	*JAK3*	chr19	1970G>A	R657Q	0.37	deleterious	probably damaging	disease causing
33	*NF1*	chr17	7012_7013insT	V2338fs	0.16	N/A	N/A	disease causing
34	*KRAS*	chr12	38G>A	G13D	0.27	deleterious	probably damaging	disease causing
35	*c-CBL*	chr11	2030C>T	A677V	0.07	tolerated	possibly damaging	disease causing
35	*KRAS*	chr12	38G>A	G13D	0.42	deleterious	probably damaging	disease causing
36	*PTPN11*	chr12	226G>A	E76K	0.43	deleterious	probably damaging	disease causing
36	*PTPN11*	chr12	1403C>T	T468M	0.37	deleterious	probably damaging	disease causing
37	*PTPN11*	chr12	226G>A	E76K	0.46	deleterious	probably damaging	disease causing
38	*c-CBL*	chr11	1111T>C	Y371H	0.99	deleterious	probably damaging	disease causing
39	*PTPN11*	chr12	205G>A	E69K	0.44	deleterious	probably damaging	disease causing
40	*PTPN11*	chr12	227A>C	E76A	0.47	deleterious	probably damaging	disease causing
41	*PTPN11*	chr12	226G>A	E76K	0.4	deleterious	probably damaging	disease causing
42	*JAK3*	chr19	2294T>A	V765D	0.07	tolerated	benign	polymorphism
42	*PTPN11*	chr12	181G>T	D61Y	0.35	deleterious	probably damaging	disease causing
42	*PTPN11*	chr12	226G>A	E76K	0.07	deleterious	probably damaging	disease causing
43	*c-CBL*	chr11	1228-2A>G	splice site	0.96	N/A	N/A	N/A
45	*JAK3*	chr19	2570T>C	L857P	0.36	deleterious	probably damaging	disease causing
45	*PTPN11*	chr12	226G>A	E76K	0.44	deleterious	probably damaging	disease causing
47	*PTPN11*	chr12	172A>T	N58Y	0.42	deleterious	probably damaging	disease causing
48	*JAK3*	chr19	1970G>A	R657Q	0.26	deleterious	probably damaging	disease causing
48	*NF1*	chr17	6604T>C	C2202R	0.98	deleterious	probably damaging	disease causing
48	*NF1*	chr17	175A>C	T59P	0.95	deleterious	possibly damaging	disease causing
49	*JAK3*	chr19	1970G>A	R657Q	0.12	deleterious	probably damaging	disease causing
49	*PTPN11*	chr12	1508G>T	G503V	0.48	deleterious	probably damaging	disease causing

51	*NRAS*	chr1	35G>A	G12D	0.42	deleterious	possibly damaging	disease causing
52	*PTPN11*	chr12	182A>T	D61V	0.47	deleterious	probably damaging	disease causing
53	*KRAS*	chr12	38G>A	G13D	0.45	deleterious	probably damaging	disease causing
56	*NRAS*	chr1	38G>A	G13D	0.49	deleterious	probably damaging	disease causing
57	*KRAS*	chr12	38G>A	G13D	0.45	deleterious	probably damaging	disease causing
58	*KRAS*	chr12	38G>A	G13D	0.45	deleterious	probably damaging	disease causing
60	*NRAS*	chr1	35G>A	G12D	0.48	deleterious	possibly damaging	disease causing
61	*NRAS*	chr1	35G>A	G12D	0.49	deleterious	possibly damaging	disease causing
62	*PTPN11*	chr12	178G>C	G60R	0.58	deleterious	probably damaging	disease causing
63	*KRAS*	chr12	38G>A	G13D	0.44	deleterious	probably damaging	disease causing
64	*NF1*	chr17	7307_7308delAT	H2436fs	0.44	N/A	N/A	disease causing
64	*SETBP1*	chr18	2607C>A	S869R	0.43	deleterious	probably damaging	disease causing
65	*PTPN11*	chr12	227A>C	E76A	0.42	deleterious	probably damaging	disease causing
66	*c-CBL*	chr11	1111T>C	Y371H	0.86	deleterious	probably damaging	disease causing
67	*c-CBL*	chr11	1111T>C	Y371H	0.88	deleterious	probably damaging	disease causing
67	*NRAS*	chr1	38G>A	G13D	0.53	deleterious	probably damaging	disease causing
68	*PTPN11*	chr12	226G>A	E76K	0.47	deleterious	probably damaging	disease causing
69	*PTPN11*	chr12	226G>A	E76K	0.42	deleterious	probably damaging	disease causing
70	*NRAS*	chr1	37G>C	G13R	0.48	deleterious	probably damaging	disease causing
71	*NRAS*	chr1	38G>A	G13D	0.47	deleterious	probably damaging	disease causing
72	*JAK3*	chr19	1970G>A	R657Q	0.31	deleterious	probably damaging	disease causing
72	*PTPN11*	chr12	172A>T	N58Y	0.48	deleterious	probably damaging	disease causing
72	*SETBP1*	chr18	2602G>A	D868N	0.46	deleterious	probably damaging	polymorphism
72	*SH3BP1*	chr22	830C>T	S277L	0.48	deleterious	possibly damaging	disease causing
73	*NRAS*	chr1	38G>A	G13D	0.47	deleterious	probably damaging	disease causing
74	*KRAS*	chr12	38G>A	G13D	0.44	deleterious	probably damaging	disease causing
75	*c-CBL*	chr11	1222T>C	W408R	0.93	deleterious	probably damaging	disease causing
76	*c-CBL*	chr11	1111T>C	Y371H	0.95	deleterious	probably damaging	disease causing
77	*SETBP1*	chr18	2602G>A	D868N	0.33	deleterious	probably damaging	polymorphism
78	*NRAS*	chr1	35G>C	G12A	0.46	deleterious	possibly damaging	disease causing
79	*KRAS*	chr12	35G>A	G12D	0.44	deleterious	possibly damaging	disease causing
80	*NRAS*	chr1	37G>C	G13R	0.48	deleterious	probably damaging	disease causing
81	*PTPN11*	chr12	182A>T	D61V	0.49	deleterious	probably damaging	disease causing
82	*c-CBL*	chr11	1217del22	T406fs	0.35	N/A	N/A	N/A
84	*c-CBL*	chr11	1096-110del643	E366_F468del	N/A	N/A	N/A	N/A
85	*PTPN11*	chr12	226G>A	E76K	0.47	deleterious	probably damaging	disease causing
86	*KRAS*	chr12	38G>A	G13D	0.39	deleterious	probably damaging	disease causing

83*	*NF1*	chr17	4970A>G	Y1657C	0.5	deleterious	probably damaging	disease causing
87*	*PTPN11*	chr12	218C>T	T73I	0.49	deleterious	probably damaging	disease causing
88*	*PTPN11*	chr12	417G>C	E139D	0.5	deleterious	probably damaging	disease causing
89*	*PTPN11*	chr12	1504T>G	S502A	0.5	tolerated	probably damaging	disease causing
90*	*PTPN11*	chr12	218C>T	T73I	0.5	deleterious	probably damaging	disease causing
91*	*PTPN11*	chr12	218C>T	T73I	0.49	deleterious	probably damaging	disease causing

Asterisks indicate Noonan syndrome-associated myeloproliferative disease.

Underlines at the gene indicate the confirmed somatic mutations. N/A, not available.

Supplementary Table 3. List of primers

Primer name	Forward primer (5'->3')	Reverse primer (5'->3')
NF1_exon1	AAGCGGCCGCGAGCCTGCACTCCACAGACCC	AAGCGGCCGCTCCTCCAGAGCCTGAGGCAGC
NF1_exon2	AAGCGGCCGCATGGCAAGTAAGTTATTTATGGTC	AAGCGGCCGCTATCCAAAGTCCACAGAAAATCA
NF1_exon3	AAGCGGCCGCGAGGTAAAATGGAAGACTATTGTTG	AAGCGGCCGCAAAATTGTCTGTCACCAGGTCAG
NF1_exon4	AAGCGGCCGCTGGTAAGGATGGCAGGGGGATTG	AAGCGGCCGCTCCCATGTGGATTTACACACTAACC
NF1_exon5	AAGCGGCCGCGATGTCTTGCTATGTTGCCCAGG	AAGCGGCCGCATTGCCAAGATTTAAAATGCTCA
NF1_exon6–7	AAGCGGCCGCAATCATTGATGTCCAAGGCATAT	AAGCGGCCGCTTTCCTATTTGACACCAGTTGAC
NF1_exon8	AAGCGGCCGCCCAGGGATTTTGTTCCTATCTAA	AAGCGGCCGCCAGAGGTATCGTTTAGTCTTCTCAGA
NF1_exon9	AAGCGGCCGCTTCTTTATAGTATGAGTTTTAGAGGC	AAGCGGCCGCGAACTTATCAACGAAGAGTCAGA
NF1_exon10–11	AAGCGGCCGCGGGTGCTTTGTGCTTCTTCTGGC	AAGCGGCCGCAAGAAATACGCAAAGAAAAGAAAGA
NF1_exon12	AAGCGGCCGCTGGAAATCATGGTGTGTGTTTGC	AAGCGGCCGCAAAATAACCAAAGCAGCAGGAAT
NF1_exon13	AAGCGGCCGCAATACTGACCTTATGCTTACTATTGA	AAGCGGCCGCTATCCTCAAGGTCTTGGCGTTTC
NF1_exon14	AAGCGGCCGCTTGAAGTTTCCTTTTTTTCCTTGCA	AAGCGGCCGCAAACCACACACCAAAGGAACATCAT
NF1_exon15	AAGCGGCCGCAATGAAAGAGCTCAATTTCTTAGCA	AAGCGGCCGCAAACCATAAAACCTTTGGAAGTGTA
NF1_exon16	AAGCGGCCGCCTCTAAACTTGTATTCATTATGGGAGA	AAGCGGCCGCATCTCTCACCATTACCATTCCAA
NF1_exon17	AAGCGGCCGCATTCCTCTTGGTTGTCAGTGCTT	AAGCGGCCGCAACAAACAGAGCACATAAAATGATACA
NF1_exon18	AAGCGGCCGCTTATACATAAAATTACCCAAGTTGCA	AAGCGGCCGCACTGAGTAAAAAAAACCACTATTCACA
NF1_exon19–20	AAGCGGCCGCTGGCAGGCAGGGCTCTAAGTG	AAGCGGCCGCTAACAGACAAAAGTCAACTTTACAGA
NF1_exon21	AAGCGGCCGCTAAGGTTTAATTCATGCTTTGCA	AAGCGGCCGCATAGAGAAAGGTGAAAAATAAGAGAAC
NF1_exon22–23	AAGCGGCCGCCTCTAGGGGGTCTGTCTTCTGGG	AAGCGGCCGCTGGACCATATTCCCAAGCACACG
NF1_exon24	AAGCGGCCGCTGTCACTTAGGTTATCTGGCAAA	AAGCGGCCGCCTGTAGACTATTCTTCATAAACTGACAA C
NF1_exon25	AAGCGGCCGCTTTAAGGTAGCCATTTTGCCAAGAT	AAGCGGCCGCTTGCTTTATGTTTTTTGGTGACT
NF1_exon26–27	AAGCGGCCGCTTGTTTTGGAATGTCTGGTTAGC	AAGCGGCCGCACTTTTCTTCCCCGCTTACTCTA
NF1_exon28–29	AAGCGGCCGCTTCCTACCTAAGAATAAAAATGGGA	AAGCGGCCGCAACAGCGGTTCTATGTGAAAAGAT
NF1_exon30	AAGCGGCCGCCGTAAGCCATCCAGCCCTGTCAA	AAGCGGCCGCCAATTCTCAATGTATTATTCATCCAAAC
NF1_exon31	AAGCGGCCGCTGTTGCTGTATGTAGTCGGTGCT	AAGCGGCCGCTTTACAGTGAAGGTCAAATAGGC
NF1_exon32	AAGCGGCCGCGGACTGATTGATTCAGAGTTTTTATG	AAGCGGCCGCGCACATAACTGAAAACCATAGGG
NF1_exon33	AAGCGGCCGCTATTTGGGAAGGTTAGAAACACT	AAGCGGCCGCGAGCAACTGAGTAAGTGGCAAGA
NF1_exon34	AAGCGGCCGCATAAGTCTGGGTGTATCTGGTGT	AAGCGGCCGCAAAGAGCAAATCTGTGATTTCTT
NF1_exon35	AAGCGGCCGCGGGAAAAGTAGTGGACTGTGAAGC	AAGCGGCCGCGTGGCAAACTCTCCTTCTCAACC
NF1_exon36	AAGCGGCCGCTACCCTTTAGAATGCCTGTTGCT	AAGCGGCCGCAACTTGCCATCTCTCTATATTTGCTA
NF1_exon37	AAGCGGCCGCATTCCCACTGTTTTCTTCCTTTC	AAGCGGCCGCGAGGCCAGGATATAGTCTAGTTAGTCA
NF1_exon38	AAGCGGCCGCATTCTTCTCCACTTCACCCCGTC	AAGCGGCCGCACCCCAAATCAAACTGAAGAGAA
NF1_exon39	AAGCGGCCGCAAGGGGTATTTTGGTTTTACTGTAG	AAGCGGCCGCTAAAGTACCAAACTCTTGCCGCT

NF1_exon40	AAGCGGCCGCCAGGCCTGATTCTAGGTAATAGTCTT	AAGCGGCCGCACTGTGTTTTTTACAACTCTATCCCTA
NF1_exon41–42	AAGCGGCCGCATCTCTTAATCTCTGAAGGAGTCAA	AAGCGGCCGCTCAGGTGAAGTAAAATGGAGAAA
NF1_exon43	AAGCGGCCGCAATACTCAGTGCCAGTTGACCAT	AAGCGGCCGCATGCCTAAAAAAGGGGATACTCT
NF1_exon44–45	AAGCGGCCGCCAATATGTATTCAGAGTATCCCCTTT	AAGCGGCCGCATATTTGGGAGAAGTGAGGGCGG
NF1_exon46	AAGCGGCCGCTCCGAGATTCAGTTTAGGAGTTA	AAGCGGCCGCCTAATATGACTACTTCAAACAACTAAA
NF1_exon47	AAGCGGCCGCGAAGAATCAACAAACCTTGGTGA	AAGCGGCCGCGCAACAAGAAAAGATGGAAGAGTAC
NF1_exon48	AAGCGGCCGCAAGAAAGCTACTGTGTGAACCTCAT	AAGCGGCCGCGCTCAAGCAATCCTTCCATCTAT
NF1_exon49	AAGCGGCCGCGTCAGGGAAGAAGACCTCAGCAG	AAGCGGCCGCACTGTGAACTTTCTGCTCTGCCA
NF1_exon50	AAGCGGCCGCGTTTCTCTACTCAGCAACACTTAGC	AAGCGGCCGCAGCACAATCAGACTGGAAGAATAA
NF1_exon51	AAGCGGCCGCACTTGGAAGGAGCAAACGATGGT	AAGCGGCCGCAGCAAAGGCAAAACAAAATAAGG
NF1_exon52	AAGCGGCCGCTTAAACACTTTATGTCCAAACATTT	AAGCGGCCGCTGGCTACCTATTTACAATGCTGT
NF1_exon53–54	AAGCGGCCGCGGTGAAGTGATTATCCAGGTGTT	AAGCGGCCGCTTAACTTAAAGACAGGCACGAAG
NF1_exon55–56	AAGCGGCCGCAATGAAGAAATGCCCCAGAAAG	AAGCGGCCGCCATTGTGTGTTCTTAAAGCAGGC
NF1_exon57	AAGCGGCCGCTATTTTTGGCTTCAGATGGGGAT	AAGCGGCCGCAAAGTCAAGTCAGTTACAAGGTA
NF1_exon58	AAGCGGCCGCAATGTGTCCCCGTTGTTAAGCGA	AAGCGGCCGCGGGCAAGGACAGGGAAGGGG
PTPN11_exon1	AAGCGGCCGCAGCGGAGCCTGAGCAAG	AAGCGGCCGCGGGGACGAGGAGGGAAC
PTPN11_exon2	AAGCGGCCGCACTGAATCCCAGGTCTCTACCAAG	AAGCGGCCGCCAGCAAGCTATCCAAGCATGGT
PTPN11_exon3	AAGCGGCCGCTTCTTTCAACACTTAGGTAAAATCC	AAGCGGCCGCTCTGACACTCAGGGCACAAG
PTPN11_exon4	AAGCGGCCGCAGGAGAGCTGACTGTATACAGTAG	AAGCGGCCGCCATCTGTAGGTGATAGAGCAAGA
PTPN11_Exon5	AAGCGGCCGCATGCTGCAGTGAACATGAGAGTGC	AAGCGGCCGCAGGAGAATCGCTTGAACCCAGGAA
PTPN11_exon6	AAGCGGCCGCTTTGCATTAACACCGTTTTCTG	AAGCGGCCGCTCAGTTTCAAGTCTCTCAGGTCC
PTPN11_exon7	AAGCGGCCGCGAACATTTCCTAGGATGAATTCC	AAGCGGCCGCGGTACAGAGGTGCTAGGAATCA
PTPN11_exon8	AAGCGGCCGCAAACTTGGACTAGGCTGGGG	AAGCGGCCGCCTTTCAGGACATGAGGAAGG
PTPN11_exon9	AAGCGGCCGCTCCTTCCTCATGTCCTGAAAG	AAGCGGCCGCTCCTAAACATGGCCAATCTG
PTPN11_exon10	AAGCGGCCGCTGCTTGTTTAGGCTTTTATTTCAG	AAGCGGCCGCGGCAAGACCCTGAATTCCTAC
PTPN11_exon11	AAGCGGCCGCCGGGTGATTCCTCAACCTC	AAGCGGCCGCGGGTAGGTAAAAGCAAGCCC
PTPN11_exon12	AAGCGGCCGCGCTCCAAAGAGTAGACATTGTTTC	AAGCGGCCGCGACTGTTTTCGTGAGCACTTTC
PTPN11_exon13	AAGCGGCCGCCAACTTCATCCTGGCTCTGC	AAGCGGCCGCCCTGTCCTCCTGCTCAAAAG
PTPN11_exon14	AAGCGGCCGCTCAGTATTCTCAACCCGTCTATC	AAGCGGCCGCAGTAGGGCAACAGGAACGTG
PTPN11_exon15	AAGCGGCCGCACCATATCTGGTGCCCAAAG	AAGCGGCCGCACATTCCCAAATTGCTTGCC
c-CBL_exon1	AAGCGGCCGCCTCCCTCGCTCGCAGTC	AAGCGGCCGCGTCCGCTCGTTCCCCTC
c-CBL_exon2	AAGCGGCCGCGCAATGGGGTTATGGATCTG	AAGCGGCCGCTGTGTTACCCATTCAGGCAG
c-CBL_exon3	AAGCGGCCGCTCTTGTATGGTGAATTTGGTGC	AACCAAAGCCAGGAAATACATAC
c-CBL_exon4	AAGCGGCCGCTCCTTTCCTTGATTATGGCG	AAGCGGCCGCGCTTTCTCTTCACCGAAGTAGC
c-CBL_exon5	AAGCGGCCGCCTCATTGCCCTCTGAGTTGG	AAGCGGCCGCGGCTATTGCGAAACTAATTAAGG

c-CBL_exon6	AAGCGGCCGCTGTATCTTGCCTTGCCTTCC	AAGCGGCCGCGAGGTTGGACAGCCCCTAAG
c-CBL_exon7	AAGCGGCCGCAACTCCCAGATTCCATTTGTC	AAGCGGCCGCGGTTCCAATGAATTCAATTATGC
c-CBL_exon8	AAGCGGCCGCAAATTTGTATAGGAAACAAGTCTTCAC	AAGCGGCCGCAGGCCACCCCTTGTATCAG
c-CBL_exon9	AAGCGGCCGCCCTGGCTTTTGGGGTTAGG	AAGCGGCCGCGGATATCGTTAAGTGTTTTACGGC
c-CBL_exon10	AAGCGGCCGCTTGAAAGATGCCATTTCCC	AAGCGGCCGCAAGCAGGGTGAAAGCAAATC
c-CBL_exon11	AAGCGGCCGCGTTCCCTCAAGTGCTTCTGC	AAGCGGCCGCCAACCCATAAACTTCTATCCACAC
c-CBL_exon12	AAGCGGCCGCCAGAGGCTCAGCTGTGGTAAG	AAGCGGCCGCTCTGGGTTTTCTCAATTTTCTG
c-CBL_exon13	AAGCGGCCGCCCCTAGGTGACATGTATTTTGC	AAGCGGCCGCAGTATGTGCAGGTAGCAAGCC
c-CBL_exon14	AAGCGGCCGCATTGGCAAAACGAGAAGATG	AAGCGGCCGCTCCCAGCTACAGAAGAATTTTG
c-CBL_exon15	AAGCGGCCGCCTGTTGAATGAGATAAATGATTGC	AAGCGGCCGCAACATACAGGCCACACTGCC
c-CBL_exon16	AAGCGGCCGCTGTTGTTAAATGAGGATTTCCC	AAGCGGCCGCACTCCTGCCCTTCTAGGTGC
KRAS_Exon2	AAGCGGCCGCTTTGTATTAAAAGGTACTGGTGGAG	AAGCGGCCGCCCTTTATCTGTATCAAAGAATGGTC
KRAS_Exon3	AAGCGGCCGCTGGGTATGTGGTAGCATCTC	AAGCGGCCGCTATGCATGGCATTAGCAAAG
KRAS_Exon4	AAAGCGGCCGCGAAGGAAGGAAAATTTGGTG	AAGCGGCCGCAGAAGCAATGCCCTCTCAAG
KRAS_Exon5a	AAGCGGCCGCACCTGTACACATGAAGCCATCG	AAGCGGCCGCCTAACAGTCTGCATGGAGCAGGAA
KRAS_Exon5b	AAGCGGCCGCAACTTCTTGCACATGGCTTTC	AAGCGGCCGCGTGGTTGCCACCTTGTTACC
NRAS_Exon2	AAGCGGCCGCGATGTGGCTCGCCAATTAAC	AAGCGGCCGCGAATATGGGTAAAGATGATCCGAC
NRAS_Exon3	AAGCGGCCGCGTTAGATGCTTATTTAACCTTGGC	AAGCGGCCGCTGTGGTAACCTCATTTCCCC
NRAS_Exon4	AAGCGGCCGCCCAAAGTGCTGAGATTGCAGGCAT	AAGCGGCCGCAGGCAGGAGAATCACTTGAACCCA
NRAS_Exon5	AAGCGGCCGCCAAGAGAGCTTATAATTTGGATTGTG	AAGCGGCCGCCAATAACACCAGCACTCCTCC
SETBP1_exon2	AAGCGGCCGCTCACCTTTCCCTTTTCCC	AAGCGGCCGCCCCTAAGAAGTTGAGGATAAAATG
SETBP1_exon3	AAGCGGCCGCTGGTAAGTCCATTGCTGGTC	AAGCGGCCGCATCCAAGGTTCGGGTTTCTG
SETBP1_exon4a	AAGCGGCCGCCATGCTCATCTTTGTTTCTCTCTC	AAGCGGCCGCATGCTTTCTGGGCATTCTTG
SETBP1_exon4b	AAGCGGCCGCCCCCAGAACCACCTACGG	AAGCGGCCGCTTTGGATGCTGGATTTCTGG
SETBP1_exon4c	AAGCGGCCGCCCCAGGAGGTGTGTCTAAGC	AAGCGGCCGCCTTTTGCCTTCAGAGCAACG
SETBP1_exon4d	AAGCGGCCGCAGAAAGTTGGAAAGCTCGGC	AAGCGGCCGCGGACAGCGTGATTTCCTTTAG
SETBP1_exon4e	AAGCGGCCGCACACAGTGGAACCTGGAAGC	AAGCGGCCGCAGGCCTACCACGCTTCTTC
SETBP1_exon4f	AAGCGGCCGCACCACGAGAATCCATATCCC	AAGCGGCCGCCTTTGTCTGCGCTACTCAGC
SETBP1_exon4g	AAGCGGCCGCCACAAGCATAAGCACAAGGAAG	AAGCGGCCGCATGTGTCTGAGGTGCAAAGC
SETBP1_exon5	AAGCGGCCGCTGTTGTCTATCTTCCTGTTCCC	AAGCGGCCGCTCAACAGGCCATTCTCAGTG
SETBP1_exon6a	AAGCGGCCGCAGAGGCTGAGTTGAAGGCAC	AAGCGGCCGCGGTGGCCATGATGGTGTC
SETBP1_exon6b	AAGCGGCCGCGACCTGCCCAGCAAAAGAG	AAGCGGCCGCCGCTTCCCACGTGTCAG
JAK3_exon2	AAGCGGCCGCCAGAAGTCCAATCCCCTCTG	AAGCGGCCGCAGCTTCCAATCTTGGCCC
JAK3_exon3	AAGCGGCCGCGATCTGGACGGTTGGGTATG	AAGCGGCCGCCAAGTGACCCACCCTCTCTG
JAK3_exon4-5	AAGCGGCCGCCCCCACCATAATGTCACTCC	AAGCGGCCGCAGCCCACGTTGCTCACTC

JAK3_exon6	AAGCGGCCGCTTTGTGTGTGTCCCCGC	AAGCGGCCGCTACCACTCTCCGGCCCC
JAK3_exon7-8	AAGCGGCCGCATAGGGAGTGGATGGTGTGG	AAGCGGCCGCTGAGGCATAGAGAAGGGGAG
JAK3_exon9	AAGCGGCCGCTACCTGAATTTGAGCCCAGG	AAGCGGCCGCATTCAAACGTCACCTCCTCC
JAK3_exon10	AAGCGGCCGCCTGGAATGAGTTCATGGTGC	AAGCGGCCGCGGCTTTAATGAGCAAGTGGC
JAK3_exon11-12	AAGCGGCCGCAGAGGGTACCTCAAACTAAGGC	AAGCGGCCGCCTCTGCATCCACGACCC
JAK3_exon13	AAGCGGCCGCTTGGGATTATTGGAGTGGAAG	AAGCGGCCGCAGAGGTGGGAAGAACAGCC
JAK3_exon14	AAGCGGCCGCGGTGTTGGCAGAACCTCCTC	AAGCGGCCGCAATTCCTCTTCCACCCAGAG
JAK3_exon15-16	AAGCGGCCGCGCCAAACAGACTCTTCATTCATC	AAGCGGCCGCGACTGGATGTCAGTCTGCCC
JAK3_exon17	AAGCGGCCGCAGATTGGGGTGGGTCTATTG	AAGCGGCCGCACCACGCTCCTTCCACTG
JAK3_exon18-19	AAGCGGCCGCGTGGGGCCCAGGATGAG	AAGCGGCCGCGCAGGAGGGTAAGAATGTGC
JAK3_exon20	AAGCGGCCGCGCAAAACTGAGGTCGAGAGG	AAGCGGCCGCACCCCAAACCACTCCTCAG
JAK3_exon21	AAGCGGCCGCGTCACGCTTGGGGTACCTG	AAGCGGCCGCCTGGGGAGCAAAGCAGC
JAK3_exon22	AAGCGGCCGCCCCCTCTTCCTGTCCTTTC	AAGCGGCCGCCAGGCGCAGACAGGTTG
JAK3_exon23	AAGCGGCCGCATCACAGATGGCCCCTACC	AAGCGGCCGCACCCCGGCCCAATCTAC
JAK3_exon24	AAGCGGCCGCATGGGAGTTGTGTCCTTTGG	AAGCGGCCGCTGTTCCTGAAGTAGAGGACCC
SH3BP1_exon1	AAGCGGCCGCGAGAGGCAGGCTGGACC	AAGCGGCCGCACGACCCCTCAAGTCCTG
SH3BP1_exon2-3	AAGCGGCCGCTAACCCAGGCAGGCTCAG	AAGCGGCCGCCTTCCCCATTGGATGGAAG
SH3BP1_exon4-5	AAGCGGCCGCCTCAGGCAGGGTGGAGG	AAGCGGCCGCGGGTCAGTGTGCTGGGG
SH3BP1_exon6	AAGCGGCCGCCCTGTGCCCTGCTTTCAG	AAGCGGCCGCCCATCTCTGTGGGCTGACC
SH3BP1_exon7	AAGCGGCCGCACACTTGGCTCTGGACATGC	AAGCGGCCGCCTTCCTTGCCTGTGAGATGG
SH3BP1_exon8-9	AAGCGGCCGCTCCCCTCTAGACCCCACAG	AAGCGGCCGCTTCAGGATGAAGTGCGAGC
SH3BP1_exon10	AAGCGGCCGCCCCCTGCCTGACTGCTC	AAGCGGCCGCGGTTTGGGACAGAATGAGGG
SH3BP1_exon11	AAGCGGCCGCAGGCTACAAGTGCGCCAG	AAGCGGCCGCGATCGTTTCAAGCCACTGC
SH3BP1_exon12-13	AAGCGGCCGCGAGTGGGAGAGCGGGTG	AAGCGGCCGCAGGCTCAAGCCCAGGGTC
SH3BP1_exon14	AAGCGGCCGCAGAATTCTCAGGGCTGGGTG	AAGCGGCCGCTTGGGATGGGATCTGCC
SH3BP1_exon15	AAGCGGCCGCAGCCAGAGAGTTGGGGTCAC	AAGCGGCCGCTGGTGAGTTTGTGGGGCTAC
SH3BP1_exon16	AAGCGGCCGCAGTGGAGTGAGGAGCCTGG	AAGCGGCCGCGCAGGCTGATTGTGGGAG
SH3BP1_exon17	AAGCGGCCGCGGCAGCTGGAGACTCCTC	AAGCGGCCGCTGCCTGTCCCCTCTCCC
SH3BP1_exon18	AAGCGGCCGCGGATTCACCACAGGCCAC	AAGCGGCCGCGAGAGCTGTCCTGGGAGG
MLL2_5007delC	TCTCAGTGGCAGGATGACAAGGTT	TGTCGCACCATGAAACCACCAATG
SETBP1_exon4e_cDNA	ACACAGTGGAACCTGGAAGC	AGGCCTACCACGCTTCTTC
PTPN11_exon3_cDNA	GAGACTTCACACTTTCCGTTAGA	CAGAGAGATGTC CATGAAACCA

References

1. Quesada, V. *et al.* Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet* 44, 47-52 (2012).
2. Wang, L. *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med* 365, 2497-506 (2011).
3. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* 481, 157-63 (2012).
4. Jones, D.T. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* 488, 100-5 (2012).
5. Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* 488, 43-8 (2012).
6. Molenaar, J.J. *et al.* Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* 483, 589-93 (2012).
7. Network, C.G.A.R. Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609-15 (2011).
8. Morin, R.D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 476, 298-303 (2011).
9. Pasqualucci, L. *et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet* 43, 830-7 (2011).
10. Fujimoto, A. *et al.* Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* 44, 760-4 (2012).
11. Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 465, 473-7 (2010).
12. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* 469, 539-42 (2011).
13. Chapman, M.A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* 471, 467-72 (2011).
14. Banerji, S. *et al.* Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 486, 405-9 (2012).
15. Stephens, P.J. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 462, 1005-10 (2009).
16. Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* 150, 251-63 (2012).
17. Network, C.G.A. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330-7 (2012).
18. Barbieri, C.E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* 44, 685-9 (2012).