Table IV. Previous reports on the incidence of secondary cancers among survivors of childhood ALL.

Authors	Group	Total ALL patients (n)	Treatment Year	Follow-up, years Total person-years (P-Y)	*Patients with* secondary cancer (n)	Type of secondary cancer	Cumulative incidence	SIR (95%CI)
Neglia *et al* (1991)	CCG	9720	1972–88	4·7 (0·2–16) 43 446 P-Y	43	10 leukeamia/lymphoma, 24 brain tumours, 9 other tumours	0·3% (0·2–0·5) at 5 years 1·5% (1·1–2·1) at 10 years 2·5% (1·7–3·4) at 15 years	42/6·1 = 6·85
Bhatia *et al* (2002)	CCG	8831	1979–95	5·5 (0–16·1) 54 883 P-Y	70	14 AML/MDS, 6 NHL, 2 HL, 19 brain tumours, 4 sarcoma, 4 thyroid cancers, 4 parotid tumours, 4 other tumours	1·3% (0·8–1·5)at 10 years 2·1% (1·4–2·8) at 15 years	7·2 (5·5–9·1)
Nygaard *et al* (1991)	Norway (NOPHO)	895	1958–85	10·5 7·2 (4·3–26·5) 6295 P–Y	8 (6)	3 brain tumours, 2 basal cell carcinoma, 1 thyroid cancer, 2 sarcoma	2·9% (SE 1·4) at 20 years	5·9 (2·2–12·9)
Kimball Dalton *et al* (1998)	DFCI	1597	1972–95	7·6 (0–24·0)	13	3 leukaemia/lymphoma, 5 brain tumours, 5 other solid tumours	2·7% (0·7–4·7)	N/A
Loning *et al* (2000)	BFM	5006	1979–95	5·7 (1·5–18) 28 605 P-Y	52	16 AML, 1 CML, 6 lymphoma, 13 brain tumours, 3 thyroid cancers, 13 other solid tumours	0·5% (0·4–0·6) at 5 years 1·5% (1·3–1·9) at 10 years 3·3% (1·6–5·1) at 15 years	14·1 (11–18)
Hijiya *et al* (2007)	St. Jude	2169	1962–98	18·7 (2·4–41·3) 29 179 P-Y	123	45 AML, 2 CML, 10 MDS, 6 lymphoma, 48 brain tumours, 9 sarcoma, 48 other solid tumours	4·2% (SE 0·5) at 15 years 10·9% (SE 1·3) at 30 years	13·5 (11–17)
Schmiegelow *et al* (2009)	NOPHO	1614	1992–01	10·4 (50% range: 8·0–12·6)	20	8 AML, 8 MDS, 1 brain tumour, 1 oral cancer, 1 LPD after SCT, 1 thyroid cancer	1·6% (SE 0·4) at 12 years	N/A
Mody *et al* (2008)	CCSS	5760	1970–86	21·2 (5–35)	185 (199)	4 AML, 7 NHL, 106 brain tumours, 11 breast cancer, 16 thyroid cancers, 13 sarcomas, 9 skin cancer, 26 others	5·2% (4·3–6·1) at 25 years	5·0 (4·1–6·0)
Reulen *et al* (2011)	BCCSS	No. of leukaemia patients was not available Total 17 981	1940–91	24·3 (50% range: 17·9–32·4) 80 028 P-Y	115 all leukaemias (not limited ALL)	7 leukaemia, 3 lymphoma, 27 brain tumours, 17 thyroid cancers, 61 other solid tumours	N/A	4·3 (3·6–5·2)
This study	TCCSG	2807	1984–05	9·5 (0·2–27) 27 658 P-Y	37	11 AML,5 MDS, 2 lymphoma, 13 brain tumours, 6 other solid tumour	1·0% (0·7–1·4) at 10 years 1·4% (0·9–2·0) at 15 years 2·4% (1·5–3·7) at 20 years	9·3 (6·5–12·8)

CCG, Children's Cancer Group; NOPHO, Nordic Society for Paediatric Haematology and Oncology; DFCI, Dana-Farber Cancer Institute; BFM, Berlin-Frankfürt-Münster; St. Jude, St. Jude Children's Research Hospital, St.Jude; CCSS, Childhood Cancer Survivor Study; BCCSS, British Childhood Cancer Survivor Study; TCCSG, Tokyo Children's Cancer Study Group; ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia; MDS, myelodysplastic syndrome; NHL, Non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CML, chronic myeloid leukaemia; LPD, lymphoproliferative disease; SCT, stem cell transplantation; SE, standard error; SIR, standardized incidence ratio; 95%CI, 95% confidence interval; N/A, not available.

2008; Schmiegelow *et al*, 2009) Another important finding in the present study is that no marked difference was observed in cumulative incidence between the five TCCSG protocols, but the multivariate analysis adjusting for confounders including CRT resulted in a statistically significant increased risk of secondary cancers in patients treated with the recent protocol (Table III). Because of the relatively short follow-up duration for patients included in the recent protocol and potentially complex interplay with treatment, at this point we consider this finding preliminary and it will be followed-up in future studies. Despite considerable reduction in the use of CRT over time, particularly for the more recent treatment protocols, we observed no evidence of a reduction in incidence of secondary cancers among children with ALL (Fig 2D). One explanation may be that CRT is probably linked mainly to secondary brain tumours, which only comprise approximately a third of all the secondary cancers. Furthermore, examination of the data showed that secondary brain tumours were diagnosed in patients enrolled in the earlier treatment protocols, a time when CRT was still commonly administered. Secondary cancers diagnosed in patients enrolled in the more recent treatment protocols were predominantly haematological. Given that the latency for secondary brain tumours is generally longer than that of haematological cancers, it is possible that a longer follow-up period may be needed to observe the effects of the reduction in CRT use. Finally, considering the poor prognosis of secondary AML/MDS (Fig 3D), it is important to identify the associated factors and minimize the development of secondary haematological cancers.

Therapy-related secondary cancers have been identified in patients receiving radiotherapy, chemotherapy, or combined modality therapy for ALL. Our study identified CRT as a strong risk factor, which was also found in the BFM study. (Loning *et al*, 2000; Borgmann *et al*, 2008) The cumulative incidence of secondary cancer for the irradiated group continued to increase with time even after more than 15 years following ALL diagnosis, possibly suggesting a long-term effect of irradiation on the rates of secondary cancers (Fig 2C). Even among the patients treated with more recent non-irradiated protocol, it is currently unknown whether the cumulative risk will remain constant, or whether secondary cancers might arise after a longer latency period. Multi-agent chemotherapy as part of multimodality therapy for cancers has increased the difficulty of assessing which agents might play a causative role in the development of secondary cancers. Alkylating agents, and more recently DNA-topoisomerase II inhibitors, have been linked to the development of secondary AML and MDS.(Hawkins *et al*, 1992; Le Deley *et al*, 2003) In contrast to previous reports, we were not able to demonstrate a clear relationship between the anthracyclines, etoposide or methotrexate and the occurrence of secondary cancers or specific types of secondary cancers.(Relling *et al*, 1999) The crude HR of CPM showed an increased risk of secondary cancers, but adjustment for confounders in multivariate analyses resulted in an attenuated and non-statistically significant finding.

Lastly, we found that cumulative incidence of secondary cancers in patients remaining in first CR was significantly higher than the patients who experienced a relapse of their primary ALL, changed treatment regimen, were lost to follow-up or died during first CR unexpectedly (Fig 2B). This finding was unexpected as it could be hypothesized that, because relapsed patients usually receive additional therapeutic exposures, they may potentially be at a higher risk of developing of secondary cancers. Nevertheless, a few studies provide some supportive data for our observations, including Borgmann *et al* (2008) who reported that the cumulative incidence of secondary cancers was unexpectedly low (1·3% at 15 years) despite repeated exposure to intense frontline and relapse treatment using BFM ALL-REZ Study data. In the St. Jude study (Hijiya *et al*, 2007), secondary neoplasms were observed in 123 out of 2,169 (5·7%) patients with continued first CR and in 45 out of 879 patients (5·1%) with relapse. In contrast, however, Bhatia *et al* (2002) demonstrated that the 10-year cumulative incidence of second malignancy was 0·91% in the patients with continued first CR compared with 1·2% in the entire cohort. The interpretation of these inconsistent results is difficult. It could be partially influenced by differences in OS among the patients with continued first CR and patients with relapse across the various studies.

One strength of the current study is that treatment of patients according to TCCSG therapeutic protocols ensured uniform access to standard therapy, giving us the opportunity to explore risk factors associated with secondary cancers in this cohort. Secondly, the follow-up duration was relatively long compared to previous prospective clinical studies and allowed us to describe the incidence of secondary cancers among patients treated on contemporary therapeutic protocols.

The results of this study should be interpreted in the context of acknowledged limitations. One major limitation is that this study was smaller than some previous studies, such as the CCG, CCSS and BCCSS, which may have affected our statistical power for certain analyses. Although all the patients in our cohort were treated according to therapeutic protocols, we do not have detailed information regarding actual cumulative exposures doses after relapse, which potentially could have influenced the development of second cancers. To address this concern, we conducted a sensitivity analysis (per protocol analysis) that included only patients who had completed all planned treatment leading to first CR. These results were largely consistent with the primary analyses (Table S1). Also, we were unable to compare the clonal phenotypes and genotypes between the primary ALL in L84-11/L89-12 and certain secondary ALL candidates. The difficulty in distinguishing between the primary and secondary type of recurrence using current standard techniques is well-recognized. In both of these events, some clonal markers

are maintained between the original diagnosis and recurrence but others can be altered.(Szczepanski *et al*, 2001; Zuna *et al*, 2007) Thus, they were not included in the analysis.

In conclusion, we showed that cumulative incidence of secondary cancer after TCCSG-ALL therapy is relatively low (1·0% at 10 years and 2·4% at 20 years) compared to the previous reports, although it is still 9 times higher than in the general population. We confirm that CRT is a strong risk factor of secondary cancer, but we did not observe evidence for a decrease in incidence despite the marked reduction in CRT treatment in the more recent protocols. In view of the long latency periods and long life expectancy of ALL patients treated in childhood, long and careful follow-up of these patients is warranted. Efforts to identify the causative carcinogenic factors should continue, and future treatment protocols should take these factors into account to maximize the chances of a long and healthy life, while preserving the efficacy of ALL treatment.

Acknowledgements

We sincerely thank Kaori Itagaki for preparing the TCCSG-ALL data, and Ryota Hosoya for his careful supervision. The authors thank the members of the ALL Committee of the TCCSG: Atsushi Manabe, Kei-ichi Isoyama, Akitoshi Kinoshita, Takehiko Kamijo, Masa-aki Kumagai, Hiromasa Yabe, Yasuhide Hayashi, Tsuyoshi Morimoto, Miho Maeda, Ken-ichi Sugita, Yasushi Noguchi, Takashi Kaneko, Kanji Sugita, Manabu Sotomatsu, Michiko Kajiwara, Takeyuki Sato, Yuri Okimoto, Setsuo Ohta, Masahiro Saito, Hiroyuki Takahashi, and Koichiro Ikuta.

Authorship and Disclosures

Conception and design: Yasushi Ishida (YI) and Miho Maeda (MM); Financial support: YI and Akira Ohara (AO); Administrative support: MM, Masahiro Tuchida(MT) and AO; Provision of study patients: YI, MM, Chikako Kiyotani, Yuki Aoki, Yoko Kato, Shoko Goto, Sachi Sakaguchi, Kenichi Sugita, Mika Tokuyama (MT), Naoya Nakadate, Eizaburo Ishii, MT and AO; Collection and assembly of data: YI, MM and MT; Data analysis and YI, MM, and Kevin Urayama (KU); Data interpretation: YI, MM, MT and AO; Manuscript writing: YI and KU, Final approval of manuscript: All authors.

Financial Disclosure

This study was supported in part by a grant from the Children's Cancer Association of Japan and a research grant from the St. Luke's Life Science Institute. The authors indicated no potential conflicts of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Schemas of the TCCSG (Tokyo Children's Cancer Study Group) protocols.

Fig. S2. Cumulative incidence of secondary cancers according to therapy of primary ALL.

Table S1. Cox- regression analysis limited to per protocol group evaluating the association between select characteristics of the primary ALL diagnosis and risk of developing a secondary cancer.

References

Bhatia, S., Sather, H.N., Pabustan, O.B., Trigg, M.E., Gaynon, P.S. & Robison, L.L. (2002) Low incidence of second neoplasms among children diagnosed with acute lymphoblastic leukaemia after 1983. *Blood*, **99**, 4257–4264.

Borgmann, A., Zinn, C., Hartmann, R., Herold, R., Kaatsch, P., Escherich, G., Moricke, A., Henze, G. & von Stackelberg, A. (2008) Secondary malignant neoplasms after intensive treatment of relapsed acute lymphoblastic leukaemia in childhood. *European Journal of Cancer*, **44**, 257–268.

Friedman, D.L., Whitton, J., Leisenring, W., Mertens, A.C., Hammond, S., Stovall, M., Donaldson, S.S., Meadows, A.T., Robison, L.L. & Neglia, J.P. (2010) Subsequent neoplasms in 5-year survivors of childhood cancer: the Childhood Cancer Survivor Study. *Journal of the National Cancer Institute*, **102**, 1083–1095.

Gooley, T.A., Leisenring, W., Crowley, J. & Storer, B.E. (1999) Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*, **18**, 695–706.

Hasegawa, D., Manabe, A., Ohara, A., Kikuchi, A., Koh, K., Kiyokawa, N., Fukushima, T., Ishida, Y., Saito, T., Hanada, R. & Tsuchida, M. & Tokyo Children's Cancer Study, G. (2012) The utility of performing the initial lumbar puncture on day 8 in remission induction therapy for childhood acute lymphoblastic leukaemia: TCCSG L99-15 study. *Pediatric Blood & Cancer*, **58**, 23–30.

Hawkins, M.M. & Robison, L.L. (2006) Importance of clinical and epidemiological research in defining the long-term clinical care of pediatric cancer survivors. *Pediatric Blood & Cancer*, **46**, 174–178.

Hawkins, M.M., Wilson, L.M., Stovall, M.A., Marsden, H.B., Potok, M.H., Kingston, J.E. & Chessells, J.M. (1992) Epipodophyllotoxins, alkylating agents, and radiation and risk of secondary leukaemia after childhood cancer. *BMJ*, **304**, 951–958.

Hijiya, N., Hudson, M.M., Lensing, S., Zacher, M., Onciu, M., Behm, F.G., Razzouk, B.I., Ribeiro, R.C., Rubnitz, J.E., Sandlund, J.T., Rivera, G.K., Evans, W.E., Relling, M.V. & Pui, C.H. (2007) Cumulative incidence of secondary neoplasms as a first event after childhood acute lymphoblastic leukaemia. *JAMA*, **297**, 1207–1215.

Igarashi, S., Manabe, A., Ohara, A., Kumagai, M., Saito, T., Okimoto, Y., Kamijo, T., Isoyama, K., Kajiwara, M., Sotomatsu, M., Sugita, K., Sugita, K., Maeda, M., Yabe, H., Kinoshita, A., Kaneko, T., Hayashi, Y., Ikuta, K., Hanada, R. & Tsuchida, M. (2005) No advantage of dexamethasone over prednisolone for the outcome of standard- and intermediate-risk childhood acute lymphoblastic leukaemia in the Tokyo Children's Cancer Study Group L95-14 protocol. *Journal of Clinical Oncology*, **23**, 6489–6498.

Japanese National Cancer Centre Hospital (2013) The regional cancer registry, Available at: http://ganjoho.jp/professional/statistics/monita.html. [Accessed July 13, 2013].

Kanda, Y. (2013) Free statistical software: EZR (Easy R) on R commander. Available at: http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmedEN.html. [Accessed April 1, 2013].

Kimball Dalton, V.M., Gelber, R.D., Li, F., Donnelly, M.J., Tarbell, N.J. & Sallan, S.E. (1998) Second malignancies in patients treated for childhood acute lymphoblastic leukaemia. *Journal of Clinical Oncology*, **16**, 2848–2853.

Le Deley, M.C., Leblanc, T., Shamsaldin, A., Raquin, M.A., Lacour, B., Sommelet, D., Chompret, A., Cayuela, J.M., Bayle, C., Bernheim, A., de Vathaire, F., Vassal, G., Hill, C., Societe Francaise d'Oncologie, P. (2003) Risk of secondary leukaemia after a solid tumor in childhood according to the dose of epipodophyllotoxins and anthracyclines: a case-control study by the Societe Francaise d'Oncologie Pediatrique. *Journal of Clinical Oncology*, **21**, 1074–1081.

Loning, L., Zimmermann, M., Reiter, A., Kaatsch, P., Henze, G., Riehm, H. & Schrappe, M. (2000) Secondary neoplasms subsequent to Berlin-Frankfurt-Munster therapy of acute lymphoblastic leukaemia in childhood: significantly lower risk without cranial radiotherapy. *Blood*, **95**, 2770–2775.

Manabe, A., Tsuchida, M., Hanada, R., Ikuta, K., Toyoda, Y., Okimoto, Y., Ishimoto, K., Okawa, H., Ohara, A., Kaneko, T., Koike, K., Sato, T., Sugita, K., Bessho, F., Hoshi, Y., Maeda, M., Kinoshita, A., Saito, T., Tsunematsu, Y. & Nakazawa, S. (2001) Delay of the diagnostic lumbar puncture and intrathecal chemotherapy in children with acute lymphoblastic leukaemia who undergo routine corticosteroid testing: Tokyo Children's Cancer Study Group study L89-12. *Journal of Clinical Oncology*, **19**, 3182–3187.

Meadows, A.T., Friedman, D.L., Neglia, J.P., Mertens, A.C., Donaldson, S.S., Stovall, M., Hammond, S., Yasui, Y. & Inskip, P.D. (2009) Second neoplasms in survivors of childhood cancer: findings from the Childhood Cancer Survivor Study cohort. *Journal of Clinical Oncology*, **27**, 2356–2362.

Mody, R., Li, S., Dover, D.C., Sallan, S., Leisenring, W., Oeffinger, K.C., Yasui, Y., Robison, L.L. & Neglia, J.P. (2008) Twenty-five-year follow-up among survivors of childhood acute lymphoblastic leukaemia: a report from the Childhood Cancer Survivor Study. *Blood*, **111**, 5515–5523.

Neglia, J.P., Meadows, A.T., Robison, L.L., Kim, T.H., Newton, W.A., Ruymann, F.B., Sather, H.N. & Hammond, G.D. (1991) Second neoplasms after acute lymphoblastic leukaemia in childhood. *New England Journal of Medicine*, **325**, 1330–1336.

Neglia, J.P., Friedman, D.L., Yasui, Y., Mertens, A.C., Hammond, S., Stovall, M., Donaldson, S.S., Meadows, A.T. & Robison, L.L. (2001) Second malignant neoplasms in five-year survivors of childhood cancer: childhood cancer survivor study. *Journal of the National Cancer Institute*, **93**, 618–629.

Nygaard, R., Garwicz, S., Haldorsen, T., Hertz, H., Jonmundsson, G.K., Lanning, M. & Moe, P.J. (1991) Second malignant neoplasms in patients treated for childhood leukaemia. A population-based cohort study from the Nordic countries. The Nordic Society of Pediatric Oncology and Hematology (NOPHO). *Acta Paediatrica Scandinavica*, **80**, 1220–1228.

Pui, C.H., Campana, D., Pei, D., Bowman, W.P., Sandlund, J.T., Kaste, S.C., Ribeiro, R.C., Rubnitz, J.E., Raimondi, S.C., Onciu, M., Coustan-Smith, E., Kun, L.E., Jeha, S., Cheng, C., Howard, S.C., Simmons, V., Bayles, A., Metzger, M.L., Boyett, J.M., Leung, W., Handgretinger, R., Downing, J.R., Evans, W.E. & Relling, M.V. (2009) Treating childhood acute lymphoblastic leukaemia without cranial irradiation. *New England Journal of Medicine*, **360**, 2730–2741.

Relling, M.V., Rubnitz, J.E., Rivera, G.K., Boyett, J.M., Hancock, M.L., Felix, C.A., Kun, L.E., Walter, A.W., Evans, W.E. & Pui, C.H. (1999) High incidence of secondary brain tumours after radiotherapy and antimetabolites. *Lancet*, **354**, 34–39.

Reulen, R.C., Frobisher, C., Winter, D.L., Kelly, J., Lancashire, E.R., Stiller, C.A., Pritchard-Jones, K., Jenkinson, H.C. & Hawkins, M.M. & British Childhood Cancer Survivor Study Steering, G. (2011) Long-term risks of subsequent primary neoplasms among survivors of childhood cancer. *JAMA*, **305**, 2311–2319.

Schmiegelow, K., Al-Modhwahi, I., Andersen, M.K., Behrendtz, M., Forestier, E., Hasle, H., Heyman, M., Kristinsson, J., Nersting, J., Nygaard, R., Svendsen, A.L., Vettenranta, K., Weinshilboum, R., Nordic Society for Paediatric, H. & Oncology (2009) Methotrexate/6-mercaptopurine maintenance therapy influences the risk of a second malignant neoplasm after childhood acute lymphoblastic leukaemia: results from the NOPHO ALL-92 study. *Blood*, **113**, 6077–6084.

Szczepanski, T., Willemse, M.J., Kamps, W.A., van Wering, E.R., Langerak, A.W. & van Dongen, J.J. (2001) Molecular discrimination between relapsed and secondary acute lymphoblastic leukaemia: proposal for an easy strategy. *Medical and Pediatric Oncology*, **36**, 352–358.

Toyoda, Y., Manabe, A., Tsuchida, M., Hanada, R., Ikuta, K., Okimoto, Y., Ohara, A., Ohkawa, Y., Mori, T., Ishimoto, K., Sato, T., Kaneko, T., Maeda, M., Koike, K., Shitara, T., Hoshi, Y., Hosoya, R., Tsunematsu, Y., Bessho, F., Nakazawa, S. & Saito, T. (2000) Six months of maintenance chemotherapy after intensified treatment for acute lymphoblastic leukaemia of childhood. *Journal of Clinical Oncology*, **18**, 1508–1516.

Tsuchida, M., Ohara, A., Manabe, A., Kumagai, M., Shimada, H., Kikuchi, A., Mori, T., Saito, M., Akiyama, M., Fukushima, T., Koike, K., Shiobara, M., Ogawa, C., Kanazawa, T., Noguchi, Y., Oota, S., Okimoto, Y., Yabe, H., Kajiwara, M., Tomizawa, D., Ko, K., Sugita, K., Kaneko, T., Maeda, M., Inukai, T., Goto, H., Takahashi, H., Isoyama, K., Hayashi, Y., Hosoya, R. & Hanada, R. & Tokyo Children's Cancer Study, G. (2010) Long-term results of Tokyo Children's Cancer Study Group trials for childhood acute lymphoblastic leukaemia, 1984-1999. *Leukaemia*, **24**, 383–396.

Tsunematsu, Y., Kikuchi, K. & Koide, R. (1974) Severe hyponatremia in two cases of childhood acute myeloid leukaemia–syndrome of inappropriate secretion of antidiuretic hormone secondary to vincristine therapy (author's transl). *Rinsho Ketsueki*, **15**, 1149–1155.

Zuna, J., Cave, H., Eckert, C., Szczepanski, T., Meyer, C., Mejstrikova, E., Fronkova, E., Muzikova, K., Clappier, E., Mendelova, D., Boutard, P., Schrauder, A., Sterba, J., Marschalek, R., van Dongen, J.J., Hrusak, O., Stary, J. & Trka, J. (2007) Childhood secondary ALL after ALL treatment. *Leukaemia*, **21**, 1431–1435.

nature genetics

Somatic *SETBP1* mutations in myeloid malignancies

Hideki Makishima[1], Kenichi Yoshida[2], Nhu Nguyen[3], Bartlomiej Przychodzen[1], Masashi Sanada[2,4], Yusuke Okuno[2,5], Kwok Peng Ng[1], Kristbjorn O Gudmundsson[3], Bandana A Vishwakarma[3], Andres Jerez[1], Ines Gomez-Segui[1], Mariko Takahashi[2], Yuichi Shiraishi[6], Yasunobu Nagata[2], Kathryn Guinta[1], Hiraku Mori[7], Mikkael A Sekeres[8], Kenichi Chiba[6], Hiroko Tanaka[9], Hideki Muramatsu[5], Hirotoshi Sakaguchi[5], Ronald L Paquette[10], Michael A McDevitt[11], Seiji Kojima[5], Yogen Saunthararajah[1], Satoru Miyano[6,9], Lee-Yung Shih[12], Yang Du[3,13], Seishi Ogawa[2,4,13] & Jaroslaw P Maciejewski[1,13]

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)[1]. Closely positioned somatic *SETBP1* mutations encoding changes in Asp868, Ser869, Gly870, Ile871 and Asp880, which match germline mutations in Schinzel-Giedion syndrome (SGS)[2], 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[3–5]. 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*[8,9], SNP array karyotyping led to the discovery of mutations in *CBL*[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*[13], *IDH1* (ref. 14) and spliceosomal genes[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[18,19].

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 SET[20,21] and is activated to support the recovery of granulopoiesis in chronic granulomatous disease[22]. Mutations in *SETBP1* are causative in SGS, a congenital disease characterized by a higher than normal prevalence of tumors, typically neuroepithelial neoplasia[23,24]. 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[1,3–5,25,26], 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

[1]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio, USA. [2]Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [3]Department of Pediatrics, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA. [4]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [5]Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan. [6]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [7]Department of Hematology, Showa University, Tokyo, Japan. [8]Department of Hematologic Oncology and Blood Disorders, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio, USA. [9]Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [10]Department of Medicine, Hematology/Oncology, University of California, Los Angeles, Los Angeles, California, USA. [11]Department of Medicine and Oncology, Division of Hematology and Hematological Malignancy, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [12]Department of Internal Medicine, Division of Hematology-Oncology, Chang Gung Memorial Hospital, Chang Gung University, Taipei, Taiwan. [13]These authors contributed equally to this work. Correspondence should be addressed to J.P.M. (maciejj@ccf.org), S.O. (sogawa-tky@umin.ac.jp) or Y.D. (yang.du@usuhs.edu).

Received 14 November 2012; accepted 13 June 2013; published online 7 July 2013; doi:10.1038/ng.2696

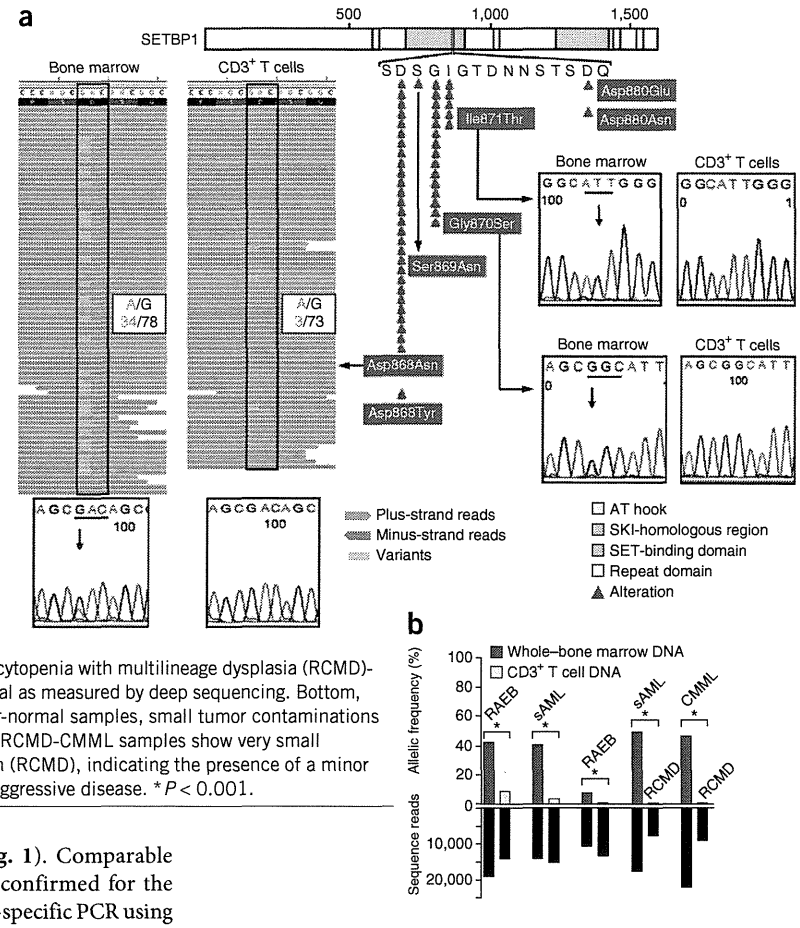

Figure 1 Somatic *SETBP1* mutations as detected by next-generation whole-exome sequencing and Sanger sequencing. (**a**) Distribution of *SETBP1* mutations detected in 52 of 727 myeloid neoplasms, all of which were located within the portion of the gene encoding the SKI-homologous domain. Bottom and right, representative mutations confirmed by Sanger sequencing (horizontal lines and vertical arrows indicate affected codons and nucleotides, respectively). Left, a somatic *SETBP1* mutation (encoding a p.Asp868Asn alteration) detected by whole-exome sequencing of paired tumor (bone marrow) 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. $^*P < 0.001$.

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 DNA from $CD3^+$ T cells and/or serial samples ($n = 21$).

Of the cases with *SETBP1* mutations, 12 had clinical material available to successfully analyze samples collected serially at multiple clinical time points. None of the 12 cases had *SETBP1* mutations at the time of initial presentation, indicating that the mutations were acquired only upon or during leukemic evolution (**Figs. 1** and **2**). Most of the *SETBP1* mutations (17 of 19) showed comparable or higher allele frequencies relative to other secondary events, suggesting a potential permissive role of *SETBP1* mutations (**Supplementary Fig. 5**). Such a secondary nature for *SETBP1* mutations was confirmed by mutational analysis of colonies derived from individual progenitor cells grown in methylcellulose culture (**Supplementary Fig. 6**).

To test potential associations with additional genetic defects, the frequencies of mutations in 13 common genes relevant to myeloid 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 cells[10,27]. *SETBP1* mutations were also found in aCML[1] and juvenile chronic myelomonocytic leukemia[28], 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*

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

Characteristic	Wild-type *SETBP1*	Mutant *SETBP1*	*P*[a]
Number	675	52	
Age at study entry (years), mean ± s.d.	61 ± 15	67 ± 12	**0.01**[b]
Age range (years)	16–91	26–83	
Ancestry, number			0.27
Caucasian	222	29	
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]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. [b]A Wilcoxon test was used to calculate the *P* value.

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 CMML 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

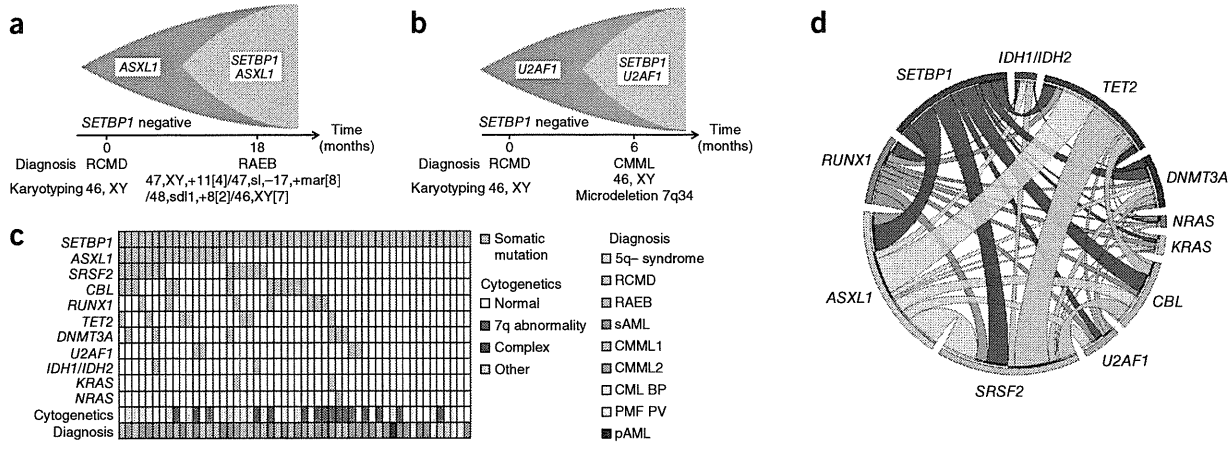

Figure 2 The relationship of *SETBP1* mutations with other common mutations. (**a**,**b**) Clonological profiles of gene mutations in two representative cases with MDS that transformed to RAEB (**a**) and CMML (**b**). Initially, hypocellular MDS (RCMD) was diagnosed on the basis of hypocellular bone 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**. CMML1 and CMML2 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.

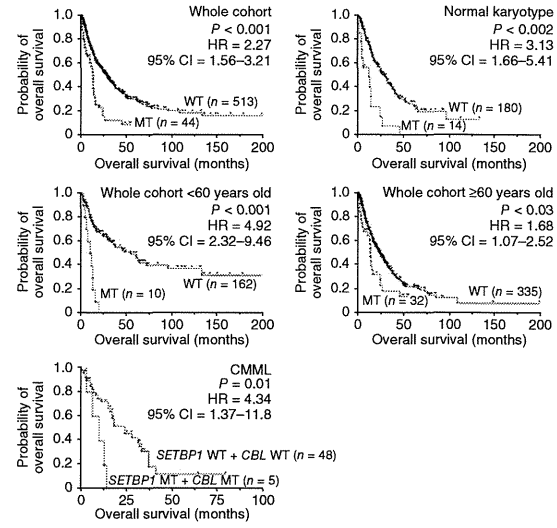

Figure 3 Impact of *SETBP1* mutations on clinical outcome. In the whole cohort, cases with *SETBP1* mutations (MT) had worse overall survival compared with those with wild-type *SETBP1* (WT). *SETBP1* mutations were poor prognostic factors for individuals with normal karyotype. *SETBP1* mutations were poor prognostic factors for individuals regardless of age (>60 and ≤60 years). In the CMML cohort, individuals with double mutations of *SETBP1* and *CBL* had worse prognosis than those with both wild-type genes. Data points indicate events and censors. The Kaplan-Meier method was used to analyze survival outcomes by the log-rank test.

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 precursors[31]. 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 observation[1]. 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 transformation[20,32]. *SETBP1* was shown to regulate PP2A activity via binding to SET[20], and decreased PP2A activity has been described in AML[21,33]. 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.

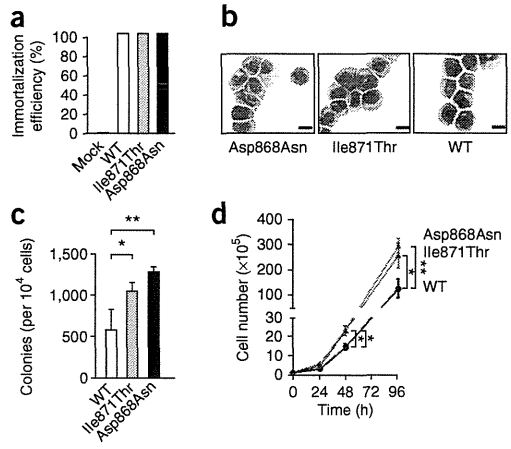

Figure 4 Immortalization of mouse myeloid progenitors by *SETBP1* mutations. (**a**) Efficiencies of empty pMYs retroviral vector (mock) or of pMYs constructs expressing wild-type Setbp1 (WT) or Setbp1 mutants (Asp868Asn and Ile871Thr) in immortalizing mouse myeloid progenitor cells in three independent experiments. (**b**) Wright-Giemsa staining of cells immortalized by transduction with retroviruses expressing the indicated wild-type or mutant Setbp1 proteins. Scale bars, 50 μm. (**c**) Mean and s.d. values for the colony-forming potentials of mouse myeloid progenitors immortalized by the expression of wild-type or mutant Setbp1 on methylcellulose medium in the presence of stem cell factor (SCF) (100 ng/ml) and interleukin (IL)-3 (20 ng/ml). Combined results from three independent myeloid progenitor populations immortalized by each retroviral construct are shown. (**d**) Expansion of myeloid progenitors immortalized by expression of wild-type or mutant Setbp1 in liquid medium with SCF and IL-3 over a 96-h period. Results from three independent populations immortalized by each retroviral construct are presented. Cell numbers were counted by trypan blue staining. Error bars, s.d. $*P < 0.05$, $**P < 0.005$; *t* tests were used for comparisons between strains.

URLs. The February 2009 human reference sequence (GRCh37) produced by the Genome Reference Consortium was used as the reference genome (UCSC Genome Browser; http://genome.ucsc.edu/cgi-bin/hgGateway). Basewise conservation scores were calculated using PhyloP in the UCSC Genome Browser. Expression array and methylation array data were extracted from Oncomine (https://www.oncomine.org/), BioGPS (http://biogps.org/) and The Cancer Genome Atlas (TCGA; http://cancergenome.nih.gov/) and were analyzed by Matlab software (http://www.mathworks.com/). Somatic mutation data were searched in the Catalogue of Somatic Mutations in Cancer (COSMIC) database on the Welcome Trust Sanger Institute website (http://www.sanger.ac.uk/genetics/CGP/cosmic/). Each potential mutation was compared against databases of known SNPs, including Entrez Gene (http://www.ncbi.nlm.nih.gov/gene) and the Ensembl Genome Browser (http://useast.ensembl.org/index.html). SAMtools (http://samtools.sourceforge.net/) and Integrative Genomics Viewer (http://www.broadinstitute.org/igv/) software were used. The Database of Genomic Variants is a publically available database of copy number variations (http://projects.tcag.ca/variation).

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. 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 RO1 HL-082983 (J.P.M.), U54 RR019391 (J.P.M.), K24 HL-077522 (J.P.M.) and RO1 CA-143193 (Y.D.), by a grant from the AA & MDS International Foundation, by the Robert Duggan Charitable Fund (J.P.M.), by a Scott Hamilton CARES grant (H. Makishima) and by Grants-in-Aid from the Ministry of Health, Labor and Welfare of Japan and KAKENHI (23249052, 22134006 and 21790907; 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-EX100-10003NI 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.N., B.P., K.O.G., B.A.V., A.J., I.G.-S., Y. Shiraishi, Y.N., M.S., M.T., K.C., H.T., H. Muramatsu, H.S., S.M. and L.-Y.S. performed research and analyzed data. K.G. and H. Mori collected data. M.A.S., R.L.P., M.A.M., S.K. and Y. Sauntharajah designed research, 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.

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 (CMML): independent prognostic impact in CMML. *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/PEBP2αB gene associated with myeloblastic leukemias. *Blood* **93**, 1817–1824 (1999).
7. Levine, R.L. *et al.* Activating mutation in the tyrosine kinase *JAK2* in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* **7**, 387–397 (2005).
8. 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 and oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* **85**, 1629–1633 (1988).
9. Lyons, J., Janssen, J.W., Bartram, C., Layton, M. & Mufti, G.J. Mutation of Ki-*ras* and N-*ras* oncogenes in myelodysplastic syndromes. *Blood* **71**, 1707–1712 (1988).
10. Sanada, M. *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**, 1090–1098 (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 binds to 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. Rodríguez, J.I., 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 (22 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 i(17)(q10), *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. & Stavnezer, 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[30]. 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 CNAG (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 cRNA 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[15]. 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[39].

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[15]. 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.Asp868Asn 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.Asp868Asn 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 (5×10^5) were infected with retrovirus (1×10^5 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.Asp868Asn 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 (Wako Junyaku) 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 α 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).

nature genetics

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

Hirotoshi Sakaguchi[1,8], Yusuke Okuno[2,8], Hideki Muramatsu[1,8], Kenichi Yoshida[2,8], Yuichi Shiraishi[3], Mariko Takahashi[2], Ayana Kon[2], Masashi Sanada[2,4], Kenichi Chiba[3], 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[3,5], Seishi Ogawa[2,4,9] & Seiji Kojima[1,9]

Juvenile myelomonocytic leukemia (JMML) is an intractable pediatric leukemia with poor prognosis[1] whose molecular pathogenesis is poorly understood, except for somatic or germline mutations of RAS pathway genes, including *PTPN11*, *NF1*, *NRAS*, *KRAS* and *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 individuals with JMML (cases), which was followed by deep sequencing of 8 target genes in 92 tumor samples. JMML was characterized by a paucity of gene mutations (0.85 non-silent mutations per sample) with somatic or germline RAS pathway involvement in 82 cases (89%). The *SETBP1* and *JAK3* genes were among common targets for secondary mutations. Mutations in the latter were often subclonal and may be involved in the progression rather than the initiation of leukemia, and these mutations associated with poor clinical outcome. Our findings provide 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*, *NRAS*, *KRAS*, *PTPN11* and *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 if secondary mutations have a role in the development and progression of this cancer. To address these issues and to better define the molecular pathogenesis of JMML, we performed whole-exome sequencing of paired tumor-normal DNA from 13 cases (**Supplementary Table 1**). We obtained mean coverage in exome sequencing of 137× for tumor samples and 143× for normal samples (**Supplementary Fig. 1**). A Monte-Carlo simulation indicated that the study detected 88% of the existing somatic mutations (Online Methods and **Supplementary Fig. 2**).

Sanger sequencing of 25 candidate non-silent somatic nucleotide alterations confirmed 1 nonsense and 10 missense mutations (**Table 1** and **Supplementary Fig. 3**), with the low true positive rate consistent with the very low numbers of somatic mutations in JMML. Of the 11 somatic mutations, 6 involved known RAS pathway genes. In addition, non-overlapping RAS pathway mutations (6 somatic and 6 germline) were confirmed in 11 of the 13 discovery cases (86%; **Table 1**). For the remaining two 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 subjects 77 and 92, respectively, but these mutations did not affect known RAS pathway genes or other cancer-related genes, including the ones registered in the pathway databases (Online Methods). A frameshift deletion in *KMT2D* (also known as *MLL2*; encoding p.Val1670fs) was found in subject 92, who had been diagnosed as having Noonan syndrome on the basis of typical features such as hypertelorism, webbed neck and congenital heart disease (**Supplementary Fig. 3**) but lacked the distinctive facial appearance of Kabuki syndrome, which was shown to be caused by germline *KMT2D* mutations[5].

Five of the 11 somatic mutations were non–RAS pathway mutations, involving *SETBP1* (3 p.Asp868Asn alterations), *JAK3* (1 p.Arg657Gln alteration) and *SH3BP1* (1 p.Ser277Leu alteration), which had not been reported in JMML cases. SETBP1 was originally isolated as a 170-kDa nuclear protein that interacts with SET, a small protein inhibitor of the putative tumor suppressors PP2A and NM23-H1 (ref. 6). Several lines of recent evidence suggest that SETBP1 has a role in leukemogenesis (**Supplementary Fig. 4**)[7–11]. *SETBP1* participates in

[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]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [4]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [5]Laboratory of Sequence 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. [8]These authors contributed equally to this work. [9]These authors jointly directed this work. Correspondence should be addressed to S.O. (sogawa-tky@umin.ac.jp) or S.K. (kojimas@med.nagoya-u.ac.jp).

Received 6 November 2012; accepted 17 June 2013; published online 7 July 2013; doi:10.1038/ng.2698

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

	RAS pathway mutations								Other somatic mutations			
	Somatic				Germline							
Subject number	Gene	Change at DNA level	Change at protein level	VAF[a]	Gene	Change at DNA level	Change at protein level	VAF[a]	Gene	Change at DNA level	Change at protein level	VAF[a]
11[b]	*NF1*	c.4537C>T	p.Arg1513*	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*	c.2602G>A	p.Asp868Asn	45.9/2.5
									JAK3	c.1970G>A	p.Arg657Gln	30.5/2.2
									SH3BP1	c.830C>T	p.Ser277Leu	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	–	–	–	–	*CBL*	c.1217del22	p.Thr406fs	34.7/38.9	–	–	–	–
83	–	–	–	–	*NF1*	c.4970A>G	p.Tyr1657Cys	50.0/51.0	–	–	–	–
84	–	–	–	–	*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[c]	–	–	–	–	*PTPN11*	c.1502T>G	p.Ser502Ala	50.0/49.9	–	–	–	–
91[c]	–	–	–	–	*PTPN11*	c.218C>T	p.Thr73Ile	49.0/48.0	–	–	–	–
92[c]	–	–	–	–	–	–	–	–	–	–	–	–

NA, not available.
[a]Variant allele frequency (VAF) in tumor/reference samples, where the reference was CD3+ T cells, except for subject 63, for whom umbilical cord was used as the reference. [b]Substantial contamination of tumor cell components in the CD3+ T cell reference. [c]Noonan syndrome–associated myeloproliferative disorder.

translocations that result in an aberrant fusion gene (*NUP98-SETBP1*) and overexpression of *SETBP1* in T cell acute lymphoblastic leukemia (T-ALL) and acute myeloid leukemia (AML), respectively[12,13]. *SETBP1* is one of the downstream targets induced by the Evi-1 oncoprotein[14] and, together with *EVI1* and its homolog *PRDM16* (also known as *MEL1*), was reported to be activated through retrovirus integration. SETBP1 is also known to augment the recovery of granulopoiesis after gene therapies for chronic granulomatous disease[15]. *SETBP1* overexpression is found in more than 27% of adult AML cases and is associated with poor survival[13]. The discovery of recurrent hotspot mutations of *SETBP1* provides unequivocal evidence for the leukemogenic role of deregulated SETBP1 function. Notably, the *SETBP1* mutation encoding p.Asp868Asn was identical to one of the *de novo* mutations reported to be causative in Schinzel-Giedion syndrome (SGS; MIM 269150), which is a highly recognizable congenital disease characterized by severe mental retardation, distinctive facial features and

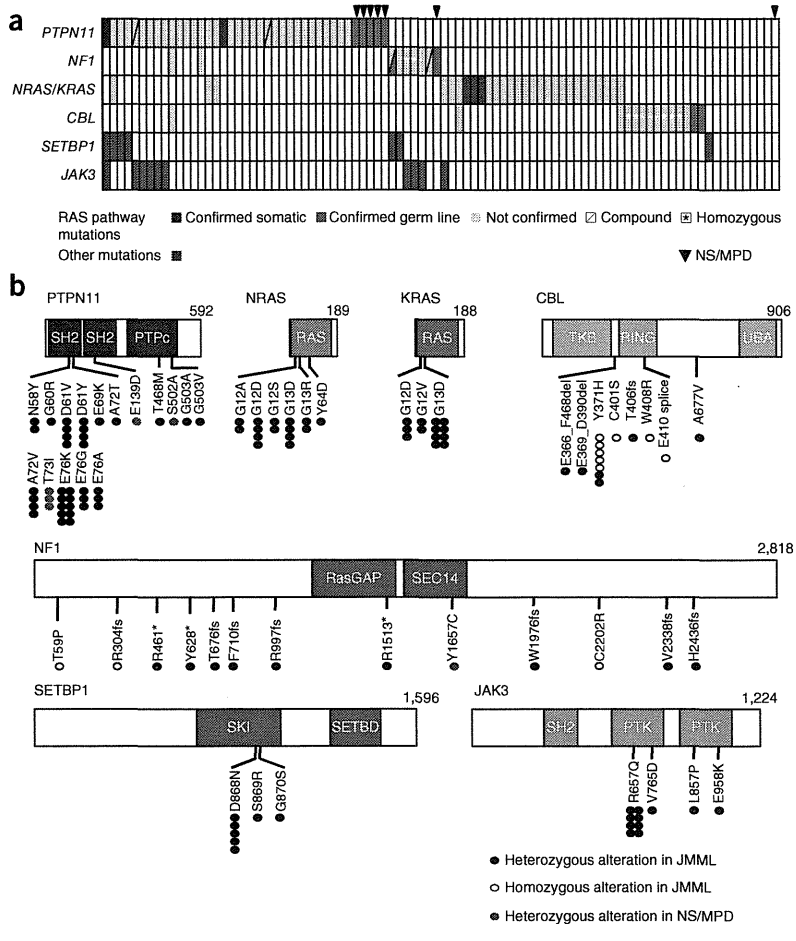

Figure 1 Mutation profiles of 92 JMML cases. (**a**) The mutation status of RAS pathway genes and 2 newly identified gene targets in a cohort of 92 JMML cases is summarized. NS/MPD, Noonan syndrome–associated myeloproliferative disorder. (**b**) The distribution of alterations is shown for each protein. SH2, Src homology 2 domain; PTPc, protein tyrosine phosphatase, catalytic domain; RAS, Ras GTPase family domain; TKB, tyrosine kinase–binding domain; RING, RING-finger domain; UBA, ubiquitin-associated domain; RasGAP, a region of similarity with the catalytic domain of the mammalian p120RasGAP protein in neurofibromin; SEC14, Sec14p-like lipid-binding domain; SKI, v-ski sarcoma viral oncogene homolog domain; SETBD, SET-binding domain; PTK, pseudokinase domain of the protein tyrosine kinases.

Table 2 Subject characteristics

Characteristic	Total cohort (n = 92)	Secondary mutations		P value
		Yes (n = 16)	No (n = 76)	
Sex (male/female)	61/31	12/4	49/27	NS
Median age at diagnosis in months (range)	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* or *KRAS*)	28 (15/13)	2 (1/1)	26 (14/12)	0.08
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 count at diagnosis $\times 10^9$/l, median (range)	30.0 (1.0–563)	29.6 (5.6–563)	30.0 (1.0–131)	NS
Monocyte count at diagnosis $\times 10^9$/l, median (range)	4.6 (0.2–31.6)	3.1 (0.5–15.2)	4.9 (0.2–31.6)	NS
Percent HbF at diagnosis, median (range)	21 (0–68)	26 (9–55)	16 (0–68)	NS
PLT at diagnosis $\times 10^9$/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/deceased	62/30	7/9	55/21	
Percent probability of 5-year overall survival (95% CI)	60 (46–71)	33 (10–59)	65 (49–77)	0.10
Percent probability of 5-year transplantation-free survival (95% CI)	15 (6–27)	0 (0–0)	18 (8–33)	0.007

JMML, juvenile myelomonocytic leukemia; NS/MPD, Noonan syndrome–associated myeloproliferative disorder; WBC, white blood cell; HbF, hemoglobin F; HSCT, hematopoietic stem cell transplantation; NS, not significant. We compared the difference between the subjects with and without secondary mutation, and P values were calculated by two-sided Fisher's exact test or Mann-Whitney U test.

multiple congenital malformations. Individuals with SGS with this mutation have a higher than normal prevalence of tumors, including of 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 3 genes together with known RAS pathway targets using deep sequencing[17] (**Supplementary Fig. 5**).

RAS pathway mutations were found in 82 of 92 cases (89%) in a mutually exclusive manner, with *PTPN11* mutations predominant, followed by *NRAS*, *KRAS*, *CBL* and *NF1* mutations (**Fig. 1a** and **Table 2**). In accordance with previous reports, most of the *CBL* (8/14) and *NF1* (4/9) mutations were biallelic (**Fig. 1a,b** and **Supplementary Table 2**)[2,3,18], whereas the majority of mutations in *PTPN11*, *NRAS* and *KRAS* were heterozygous[4]. The individuals 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**), with these mutations more frequent in cases with mutated *PTPN11* (and possibly *NF1*) than in cases with mutated *NRAS*, *KRAS*

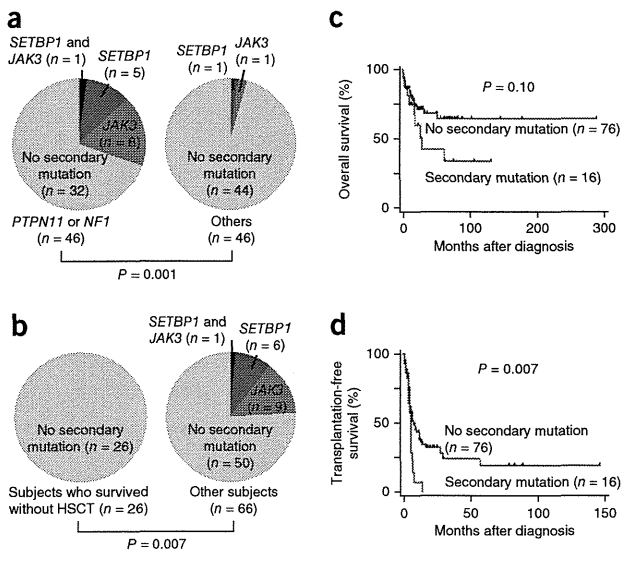

Figure 2 Clinical features of JMML cases with or without secondary mutations. (**a**,**b**) Frequency of secondary mutations in individuals with JMML depending on the type of RAS pathway mutations (left, *PTPN11* or *NF1*; right, other or no mutations) (**a**) and the status of HSCT (**b**). P values were calculated by two-sided Fisher's exact test. (**c**,**d**) The impact of secondary mutations on overall (**c**) and transplantation-free (**d**) survival is shown in Kaplan-Meier survival curves, where statistical significance was tested by log-rank test.

or *CBL* (**Fig. 2a**). Mutations in *SH3BP1*, encoding SH3 domain–binding protein 1, were not recurrent. All *SETBP1* mutations were heterozygous and occurred within the portion of the gene encoding the SKI domain, with six identical to the *de novo* recurrent mutations reported in SGS and five identical to the mutation encoding the p.Asp868Asn alteration (**Fig. 1b**). RT-PCR analysis showed that the wild-type and mutant alleles of *SETBP1* were equally expressed (**Supplementary Fig. 8**). Similarly, 8 of the 11 *JAK3* mutations in 10 cases were the well-described activating mutation (encoding a p.Arg657Gln alteration) found in various hematological malignancies, including Down syndrome–associated acute megakaryoblastic leukemia[19–23], ALL[24,25] and natural killer (NK)/T cell lymphoma[26], and the remaining 3 were also within the portions of the gene encoding the pseudokinase or kinase domain, suggestive of gain of function.

Deep sequencing of the relevant mutant alleles enabled an accurate estimation of allele frequencies for individual mutations (**Supplementary Fig. 9**). *SETBP1* and *JAK3* mutations showed lower allele frequencies (but not with statistical significance for *SETBP1*) than did the corresponding RAS pathway mutations (**Supplementary Fig. 10a**), indicating that the former mutations represent secondary genetic hits that contributed to clonal evolution after the main tumor population was established (**Supplementary Fig. 10b**). Individuals with secondary mutations had shorter lengths of survival compared to those without mutations: 5-year overall survival (hazards ratio (HR) = 1.90, 95% CI = 0.87–4.19). In addition, none of the individuals with JMML who survived without hematopoietic stem cell transplantation (HSCT; $n = 26$) harbored any of the secondary mutations, and individuals with secondary mutations showed significantly inferior 5-year transplant-free survival (HR = 2.18, 95% CI = 1.18–4.02) (**Fig. 2b–d** and **Table 2**).

JMML is characterized by a paucity of gene mutations. The average number of mutations per sample (0.85; range of 0–4) was unexpectedly low compared to those reported in other human cancers (**Supplementary Fig. 11**); excluding common RAS pathway mutations, only 5 mutations were detected in 3 of the 13 discovery cases. This small number of mutations is only comparable to the figure reported for retinoblastoma (mean of 3.3 per case; range of 0–5) (ref. 27) and is in 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 represented known driver changes (ref. 17 and K.Y., M.S., Y.S., D. Nowak, Y. Nagata *et al.*, unpublished data), underscoring the distinct pathogenesis in these two neoplasms that show indistinguishable morphology. The impact of germline events is underscored by the fact that 6 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 cases had no documented RAS pathway mutations, even after whole-exome analysis in the two RAS pathway mutation–negative cases, raising the possibility that the latter cases represent a genetically distinct myeloproliferative neoplasm in childhood.

Another key 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 are thought to be involved in the progression rather than the establishment of JMML and were associated with poor clinical outcome. *SETBP1* is a newly identified proto-oncogene, and identical mutations in this gene have recently been reported in 15–25% of adult cases with atypical chronic myeloid leukemia (CML)[10], CMML and secondary AML[28]. Affecting one of three highly conserved 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 amounts[10,28]. Although the precise leukemogenic mechanisms of *SETBP1* mutations are still unclear, we have shown that mutant *SETBP1* alleles confer self-renewal capability to myeloid progenitors *in vitro*, and *SETBP1* mutations in adult leukemia were associated with increases in *HOXA9* and *HOXA10* expression[28]. Recurrent *JAK3* mutations in JMML are also noteworthy. The JAK-STAT pathway is a key component of normal hematopoiesis[29]. As in other hematopoietic malignancies[20], the p.Arg657Gln alteration represents the most frequent change in JMML. This alteration confers interleukin (IL)-3 independence to Ba/F3 cells and induces STAT5 phosphorylation[20]. Targeting the JAK-STAT pathway with a pan-JAK inhibitor such as CP-690550 (ref. 30) could be a promising therapeutic possibility for patients with *JAK3*-mutated JMML.

In conclusion, our whole-exome sequencing analysis identified the spectrum of gene mutations in JMML. Together with the 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 were associated with poor clinical outcomes. Our findings provide an important clue to understanding the pathogenesis of JMML that may help in the development of novel diagnostics and therapeutics for this leukemia.

URLs. Genomon, http://genomon.hgc.jp/exome/en/; BioCarta, http://www.biocarta.com/; dbSNP131, http://www.ncbi.nlm.nih.gov/projects/SNP/; RefSeq database, http://www.ncbi.nlm.nih.gov/RefSeq/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession code. We deposited whole-genome and whole-exome sequence data in the European Genome-phenome Archive under accession EGAS00001000521.

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

ACKNOWLEDGMENTS

We thank the subjects and their parents for participating in this study. This work was supported by the Research on Measures for Intractable Diseases Project from the Ministry of Health, Labor and Welfare, by Grants-in-Aid from the Ministry of Health, Labor and Welfare of Japan and KAKENHI (23249052, 22134006 and 21790907), by the Project for the Development of Innovative Research on Cancer Therapeutics (P-DIRECT) and by the Japan Society for the Promotion of Science through the Funding Program for World-Leading Innovative R&D on Science and Technology.

AUTHOR CONTRIBUTIONS

H.S., Y.O., H. Muramatsu, K.Y., M.T., A.K. and M.S. designed and performed the research, analyzed the data and wrote the manuscript. 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 manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Pinkel, D. *et al.* Differentiating juvenile myelomonocytic leukemia from infectious disease. *Blood* **91**, 365–367 (1998).
2. Loh, M.L. *et al.* Mutations in *CBL* occur frequently in juvenile myelomonocytic leukemia. *Blood* **114**, 1859–1863 (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**, 1969–1975 (2010).
4. Pérez, 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**, 460–468 (2010).
5. Ng, S.B. *et al.* Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. *Nat. Genet.* **42**, 790–793 (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**, 1340–1351 (2001).
7. Damm, F. *et al. SETBP1* mutations in 658 patients with myelodysplastic syndromes, chronic myelomonocytic leukemia and secondary acute myeloid leukemias. *Leukemia* **27**, 401–403 (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* published online; doi:10.1038/leu.2013.97 (5 April 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* published online; doi:10.1038/leu.2013.133 (30 April 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* published online; doi:10.1038/leu.2013.145 (7 May 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**, 294–296 (2007).
13. 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).
14. 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).
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**, 401–409 (2006).
16. Hoischen, A. *et al. De novo* mutations of *SETBP1* cause Schinzel-Giedion syndrome. *Nat. Genet.* **42**, 483–485 (2010).
17. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (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**, 5816–5821 (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**, 681–688 (2008).
21. De Vita, S. *et al.* Loss-of-function *JAK3* mutations in TMD and AMKL of Down syndrome. *Br. J. Haematol.* **137**, 337–341 (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**, 1077–1079 (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**, 574–576 (2007).
24. Elliott, N.E. *et al.* FERM domain mutations induce gain of function in JAK3 in adult T-cell leukemia/lymphoma. *Blood* **118**, 3911–3921 (2011).
25. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163 (2012).
26. Koo, G.C. *et al.* Janus kinase 3–activating mutations identified in natural killer/T-cell Lymphoma. *Cancer Discov.* **2**, 591–597 (2012).
27. Zhang, J. *et al.* A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* **481**, 329–334 (2012).
28. Makishima, H. *et al.* Somatic *SETBP1* mutations in myeloid malignancies. *Nat. Genet.* published online; doi:10.1038/ng.2696 (7 July 2013).
29. Crozatier, M. & Meister, M. *Drosophila* haematopoiesis. *Cell. Microbiol.* **9**, 1117–1126 (2007).
30. Changelian, P.S. *et al.* Prevention of organ allograft rejection by a specific Janus kinase 3 inhibitor. *Science* **302**, 875–878 (2003).

ONLINE METHODS

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

Sample preparation. Genomic DNA was extracted using the QIAamp DNA Blood Mini kit and the QIAamp DNA Investigator kit (Qiagen) according to the manufacturer's instructions. The T Cell Activation/Expansion kit, human (Miltenyi Biotec) was used for the expansion of $CD3^+$ T cells from subjects' peripheral blood or bone marrow mononuclear cells[3].

Whole-exome sequencing. Exome capture from paired tumor-reference DNA was performed using SureSelect Human All Exon V3 (Agilent Technologies), covering 50 Mb of coding exons, according to the manufacturer's protocol. Enriched exome fragments were subjected to massively parallel sequencing using the HiSeq 2000 platform (Illumina). Candidate somatic mutations were detected through our in-house pipeline (Genomon) as previously described[17].

Detection of mutations from whole-exome sequencing data. Detection of candidate somatic mutations was performed according to previously described algorithms with minor modifications[17]. Briefly, the number of reads containing single-nucleotide variations (SNVs) and indels in both tumor and reference samples was determined using SAMtools[31], and the null hypothesis of equal allele frequencies in tumor and reference samples was tested using the two-tailed Fisher's exact test. A variant was adopted as a candidate somatic mutation if it had $P < 0.01$, if it was observed in bidirectional reads (in both plus and minus strands of the reference sequence) and if its allele frequency was less than 0.25 in the corresponding reference sample. For the detection of germline mutations in RAS pathway genes, SNVs and indels having allele frequencies of more than 0.25 (SNVs) and 0.10 (indels) were interrogated for 46 genes, which consisted of known JMML-related RAS pathway genes and genes registered in the pathway databases ('Ras signaling pathway' in BioCarta and 'signaling to RAS' in Reactome[32]). For variant calls in tumor samples for which the paired normal reference was not available, candidate variants in the RAS pathway were detected at an allele frequency of >0.10. Finally, the list of candidate somatic and/or germline mutations was generated by excluding synonymous SNVs and other variants registered in either dbSNP131 or an in-house SNP database constructed from 180 individual samples. All candidates were validated by Sanger sequencing as previously described.

Estimation of tumor content. The tumor content of bone marrow specimens was estimated from the allele frequency of the somatic mutations identified by deep sequencing. For homozygous mutations, as indicated by an allele frequency of >0.75, the tumor content (F_{tumor}) was calculated from the observed frequency ($F_{observed}$) of the mutation according to the following equation: $F_{tumor} = 2 \times F_{observed} - 1$. 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 ($n = 1{,}000$) on the basis of the observed mean depth of coverage for each exon in germline and tumor samples and the observed tumor content for each sample, which were estimated using 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 copy number lesions at a single-exon level, the mean coverage of each exon normalized by the mean depth of coverage of the entire sample was compared with that of 12 unrelated normal DNA samples. Exons showing normalized coverage greater than 3 s.d. from the mean coverage of the reference samples were called as candidates for copy number alterations. All candidate exons of RAS pathway genes were visually inspected using the Integrative Genomics Viewer[33] and were validated by Sanger sequencing of corresponding putative breakpoint-containing fragments.

Targeted deep sequencing. Deep sequencing of the targeted genes was performed essentially as described in the 'deep sequencing of pooled target exons' section in ref. 17, except that target DNA was not pooled. Briefly, all exons of *PTPN11*, *NF1*, *KRAS*, *NRAS*, *CBL*, *SETBP1*, *JAK3* and *SH3BP1* were PCR amplified with Quick Taq HS DyeMix (TOYOBO) and the PrimeSTAR GXL DNA Polymerase kit (Takara Bio) using primers including the NotI restriction site (**Supplementary Table 3**). The PCR products from an individual sample were combined and purified with the QIAquick PCR Purification kit (Qiagen) for subsequent digestion with NotI (Fermentas). Digested PCR product was purified, concatenated with T4 DNA ligase (Takara Bio) and sonicated to generate fragments with an average size of 150 bp using Covaris. Fragments were processed for sequencing according to a modified Illumina paired-end library protocol, and sequences were read by a HiSeq 2000 instrument using a 100-bp paired-end read protocol.

Variant calls in targeted deep sequencing. Data processing and variant calling were performed with modifications to the protocol described in a previous publication[17]. Each read was aligned to the set of targeted sequences from PCR amplification, with BLAT[34] instead of Burrows-Wheeler Aligner (BWA)[35] used with the -fine option. Mapping information in the .psl format was converted to the .sam format with paired-read information. Of the successfully mapped reads, reads were excluded from further analysis if they mapped to multiple sites, mapped with more than four mismatched bases or had more than ten soft-clipped bases. Next, the Estimation_CRME script was run to eliminate strand-specific errors and exclude PCR-derived errors. A strand-specific mismatch ratio was calculated for each nucleotide variant for both strands using the bases from read cycles 11 to 50 on the next-generation sequencer. By excluding the top five 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 variants found in dbSNP131 or the in-house SNP database, non-silent variants having a mismatch ratio of greater than 0.05 were called as candidates, unless they had median values of the mismatch ratio at the relevant nucleotide positions in the 92 samples of greater than 0.01, as such variants were likely to be caused by systematic PCR problems. Finally, candidates with mismatch ratios of >0.15 were further validated by Sanger sequencing.

Annotation of the detected mutations. Detected mutations were annotated using ANNOVAR[36]. 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 *CBL*, NM_015559.2 for *SETBP1* and NM_000215.3 for *JAK3*. The effect of the mutations on protein function was assessed by SIFT[37], PolyPhen-2 (ref. 38) and MutationTaster[39].

Whole-genome sequencing. Paired tumor-reference DNA samples were sequenced with the HiSeq 2000 platform according to the manufacturer's instructions to obtain 30× read coverage for reference samples and 40× coverage for tumor samples. Obtained FASTQ sequences were aligned to the human reference genome (hg19) using BWA[35] 0.5.8 with default parameters. Alignment of pairs of sequences, at least one of which was not mapped or was considered to have possible mapping problems (with mapping quality of less than 40, insertions or deletions, soft-clipped sequence of more than 10% of the length of the original sequence, irregular paired-read orientation or mate distance of greater than 2,000 bp), was attempted with BLAT[34] using default parameters, except for stepSize = 5 and repMatch = 2,253. Mapping statistics were calculated by counting the bases at each genomic position with SAMtools[31]. For variant calling, variant and reference bases with base quality of >30 were counted in both germline and tumor samples, and the Fisher's

doi:10.1038/ng.2698

exact test was applied. Variants with P of <0.01 were called. Variants having allele frequency of >0.25 in the germline sample were excluded. Variants found in 12 unrelated germline samples with an allele frequency of >0.01 on average were also excluded owing to the high probability that they represented false positive calls. Copy number estimation was performed by calculating the averaged ratio of read depths in germline and tumor samples in 10,000-base bins. An allele-specific copy number plot was generated by measuring the allele frequency of the tumor sample at the positions in which more than 25% of the allele mismatch was observed in germline samples. For the detection of chromosomal structural variations, soft-clipped sequences that could be mapped to a unique genomic position were selected. Structural variation candidates that had more than four supporting read pairs in total and at least one read pair from each side of the breakpoint were called. Contig sequences were generated by assembling the reads within 200 bp of the breakpoint with CAP3 (ref. 40), and structural variations having the contig sequence that could be aligned to the alternate assembly of the hg19 genome with more than 93% identity were excluded as false positives. Structural variations with read depth of greater than 150 on at least one side of the breakpoint were considered to be mapped to a repeat element and were also excluded. For detection of viruses, unmapped sequences were aligned to the collection of all viral genomes in the RefSeq database using BLAT. A virus was considered to be detected if its genome was covered by mean read coverage of >1.

cDNA sequencing. Total RNA was extracted using the RNeasy Mini kit (Qiagen) and was reverse transcribed with the ThermoScript RT-PCR system (Life Technologies). Target sequences were PCR amplified with the PrimeSTAR GXL DNA Polymerase kit using the primers listed in **Supplementary Table 3** and were 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 the Mann-Whitney U test. Overall survival and transplantation-free survival were estimated by the Kaplan-Meier method. Hazard ratios for survival with 95% CIs were estimated according to the Cox proportional hazards model, and difference in survival was tested by log-rank test. STATA version 12.0 (StataCorp) was used for all statistical calculations.

31. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
32. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–D622 (2009).
33. Thorvaldsdóttir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
34. Kent, W.J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
35. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
36. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
37. 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**, 1073–1081 (2009).
38. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
39. Schwarz, J.M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
40. Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877 (1999).

blood

2013 121: 862-863
doi:10.1182/blood-2012-11-465633

Rabbit antithymocyte globulin and cyclosporine as first-line therapy for children with acquired aplastic anemia

Yoshiyuki Takahashi, Hideki Muramatsu, Naoki Sakata, Nobuyuki Hyakuna, Kazuko Hamamoto, Ryoji Kobayashi, Etsuro Ito, Hiroshi Yagasaki, Akira Ohara, Akira Kikuchi, Akira Morimoto, Hiromasa Yabe, Kazuko Kudo, Ken-ichiro Watanabe, Shouichi Ohga and Seiji Kojima

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.

To the editor:

Rabbit antithymocyte globulin and cyclosporine as first-line therapy for children with acquired aplastic anemia

Horse antithymocyte globulin (hATG) and cyclosporine have been used as standard therapy for children with acquired aplastic anemia (AA) for whom an HLA-matched family donor is unavailable. However, in 2009, hATG (lymphoglobulin; Genzyme) was withdrawn and replaced by rabbit ATG (rATG; thymoglobulin; Genzyme) in Japan. Many other countries in Europe and Asia are facing the same situation.[1] Marsh et al recently reported outcomes for 35 adult patients with AA who were treated with rATG and cyclosporine as a first-line therapy.[2] Although the hematologic response rate was 40% at 6 months, several patients subsequently achieved late responses. The best response rate was 60% compared with 67% in a matched-pair control group of 105 patients treated with hATG. The overall and transplantation-free survival rates appeared to be significantly inferior with rATG compared with hATG at 68% versus 86% ($P = .009$) and 52% versus 76% ($P = .002$), respectively. These results are comparable to those from a prospective randomized study reported by Scheinberg et al comparing hATG and rATG.[3] Both studies showed the superiority of hATG over rATG.[2,3]

We recently analyzed outcomes for 40 Japanese children (median age, 9 years; range, 1-15) with AA treated using rATG and cyclosporine. The median interval from diagnosis to treatment was 22 days (range, 1-203). The numbers of patients with very severe, severe, and nonsevere disease were 14, 10, and 16, respectively. The ATG dose was 3.5 mg/kg/day for 5 days. The median follow-up time for all patients was 22 months (range, 6-38). At 3 months, no patients had achieved a complete response (CR) and partial response (PR) was seen in only 8 patients (20.0%). At 6 months, the numbers of patients with CR and PR were 2 (5.0%) and 17 (42.5%), respectively. After 6 months, 5 patients with PR at 6 months had achieved CR and 4 patients with no response at 6 months had achieved PR, offering a total best response rate of 57.5%. Two patients relapsed at 16 and 19 months without receiving any second-line treatments. Two patients with no response received a second course of rATG at 13 and 17 months, but neither responded. Sixteen patients underwent hematopoietic stem cell transplantation (HSCT) from alternative donors (HLA-matched unrelated donors, n = 13; HLA-mismatched family donors, n = 3). Two deaths occurred after rATG therapy, but no patients died after HSCT. Causes of death were intracranial hemorrhage at 6 months and acute respiratory distress syndrome at 17 months. The overall 2-year survival rate was 93.8% and the 2-year transplantation-free survival rate was 50.3% (Figure 1).

In our previous prospective studies with hATG, the response rates after 6 months were 68% and 70%, respectively, with no increases in response rates observed after 6 months.[4,5] Our results support the notion that rATG is inferior to hATG for the treatment of AA in children. First-line HSCT from an alternative donor may be justified, considering the excellent outcomes in children who received salvage therapies using alternative donor HSCT.

Yoshiyuki Takahashi
Department of Pediatrics, Nagoya Graduate School of Medicine, Nagoya, Japan

Hideki Muramatsu
Department of Pediatrics, Nagoya Graduate School of Medicine, Nagoya, Japan

Naoki Sakata
Department of Pediatrics, Kinki University School of Medicine, Osaka, Japan

Nobuyuki Hyakuna
Center of Bone Marrow Transplantation, Ryukyu University Hospital, Okinawa, Japan

Kazuko Hamamoto
Department of Pediatrics, Hiroshima Red Cross Hospital, Hiroshima, Japan

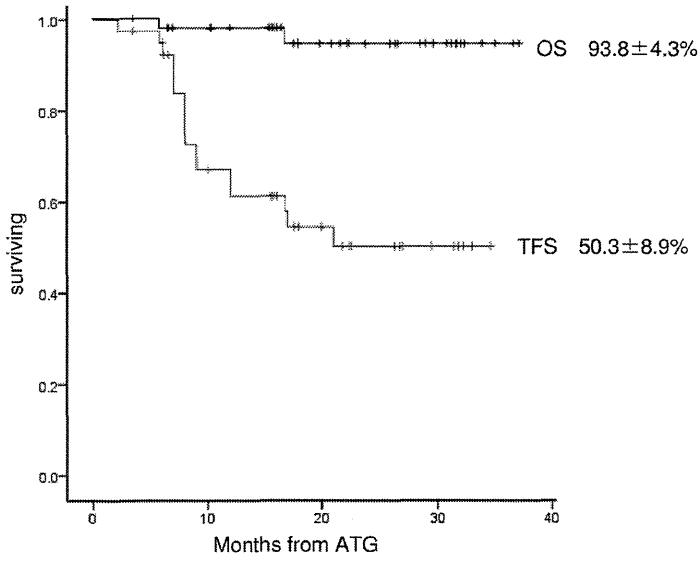

Figure 1. Kaplan-Meier estimates of overall survival (OS) and transplantation-free survival (TFS) in 40 Japanese children with AA. Survival was investigated using Kaplan-Meier methods. OS for all patients with AA after rATG and cyclosporine as first-line therapy included patients who later received HSCT for nonresponse to rATG. In the analysis of TFS for all patients treated with rATG and CSA, transplantation was considered an event.