usually associated with poor prognosis, including high WBC count and/or extramedullary involvement at diagnosis, higher frequencies of monocytic leukemia or megakaryocytic leukemia, and higher frequency of *MLL* gene rearrangement, and few infants have favorable cytogenetic characteristics, such as t(8;21)(q22;q22)/*RUNX1-RUNX1T1* and t(15;17)(q22;q12)/*PML-RARA* [1–3]. Thus, infants with AML are usually classified in the IR or HR groups and receive treatments appropriate for this level of risk.

Despite these 'unfavorable' characteristics, the outcomes of infants with AML are generally not much worse than those of older children with AML. For example, in the Japan Infant Leukemia Study Group, the 3-year pEFS and pOS were 72 and 76 %, respectively, for 35 infants with AML [12]. The AML Berlin–Frankfurt–Münster (BFM) Study Group reported that the 5-year pEFS and pOS were 44 and 61 %, respectively, in the AML-BFM 98 study ($n = 59$) and 51 and 75 %, respectively, in the AML-BFM 04 study ($n = 61$)[6]. The 5-year pEFS was 58 % in the United Kingdom (UK) Medical Research Council (MRC) AML-10 and AML-12 studies ($n = 151$) [7]. In other Japanese AML studies, similar results were reported for infant AML, with 5-year pEFS and pOS of 49.4 and 58.3 % in the Tokyo Children's Cancer Study Group (TCCSG) M91-13 and M96-14 studies ($n = 24$) [13] and 46.0 and 64.5 %, respectively, in the Japanese Childhood AML Cooperative Group AML99 study ($n = 27$) [10]. In the current study, pOS was lower in infants than in the other age groups. However, this disparity was due to the higher non-relapse mortality rate in infants and not the cumulative relapse rate, which was not significantly different among the three age groups.

The high early mortality rate (17.8 %, 5/28) observed in infants enrolled before the protocol amendments was somewhat unexpected, because the induction regimen, which consisted of etoposide (150 mg/m^2 per day on days 1–5), cytarabine (200 mg/m^2/day via 12 h continuous intravenous infusion on days 6–12), and mitoxantrone (5 mg/m^2/day on days 6–10), together with an age-adjusted dose of triple intrathecal chemotherapy on day 6, was identical to the induction regimen used in prior Japanese studies, although the dose modification method differed among these studies [10, 12, 13]. In the TCCSG M91-13 and M96-14 studies, the doses were adjusted for body surface area, with an additional dose reduction of 33 %, and the early mortality rate was 12.5 % (3/24). In the AML99 study, the doses were adjusted for body weight, as in our study before the protocol amendments, and the early mortality rate was 3.7 % (1/27).

It is notable that all non-leukemic deaths during the initial induction course were due to pulmonary complications following infectious complications. Of note, three patients had HPS. These findings suggest that infections in infants are likely to induce hypercytokinemia and result in severe conditions, such as ARDS and/or HPS. Consequently, we also amended the protocol to avoid the prophylactic administration of G-CSF, because this may aggravate inflammatory cytokine production, is often observed in engraftment syndrome after HSCT, and G-CSF was reported to have no benefits in the treatment of children with AML [14–16]. The BFM Study Group reported that, although hemorrhage and leukostasis were the main causes of death within the first 15 days of initial therapy, fatal infections were more common between days 15 and 42. They also reported that acute toxicities during the induction course, particularly severe infection and pulmonary toxicity, were more frequent in infants than in older children [17]. Therefore, the management of infectious complications is vital to prevent early death when treating infants with AML.

To prevent fatal infectious complications, we decided to modify the doses of chemotherapeutic drugs during the initial induction course, by reducing the doses of etoposide, cytarabine, and mitoxantrone by 33 %, but not of intrathecal chemotherapy. However, appropriate dose adjustments for infants are not well documented because very few pharmacokinetic studies have been performed in infants [18, 19]. Many factors may affect drug metabolism in infants, including higher total body water content and extracellular water content, higher unbound active fraction of drugs because of lower affinity of drugs to serum proteins, lower p450 enzyme activity, which could reduce or increase the cytotoxic effects, and lower renal clearance, which could increase systemic exposure of drugs because of reduced tubular and glomerular function until about 6 months of age. In addition, the ratio of body weight to body surface area is lower in infants than in older children. Therefore, if the doses are calculated based on the body surface area, the infants would be exposed to greater amounts of each drug. Thus, prior studies have used arbitrary methods to modify the drug doses for infants, including adjustment for body weight in the BFM [6], CCG-2891 [20], and AML99 [10] studies, while the doses were reduced by 25 % in the MRC AML-10 and AML-12 studies [7].

In addition to reducing the doses of chemotherapeutic drugs, we have also revised the guidelines for general supportive care and infection prevention, by recommending bacterial and fungal prophylaxis, intravenous immunoglobulin therapy to maintain IgG levels $\geq$500 mg/dl, and the use of rooms with positive air pressure with high-efficacy particulate air (HEPA) filtration. Although the incidence of RSV infection is low among patients with AML, it is associated with high mortality in children with AML [21], and it is well known that infants, both

immunocompetent and immunocompromised, are particularly vulnerable to RSV. Although RSV infection may be prevented with palivizumab, a humanized monoclonal antibody that targets the RSV epitope, it was not approved for use in malignant disease in Japan until August 2013, so its use was at the physician's discretion. An aerosol formulation of the anti-viral agent ribavirin is also effective against RSV infection, but only oral agents have been approved in Japan.

With these aforementioned amendments, most of the grade ≥ 3 non-hematological adverse events decreased, especially, clinically documented infections and pulmonary complications (Table 4). No more early death was observed and only 7.1 % cumulative incidence of non-relapse mortality was documented among the 17 infants enrolled after the protocol amendments. Although this reduction in early and non-relapse death led to achievement of >70 % pOS for the post-amendment cohort, improvements in pEFS and pOS were not statistically significant. Naturally, there are limitations for the exact explanations due to a relatively small number of patients included in the current study, but one must consider the possibility of increased relapse led by treatment reduction in the post-amendment cohort, although cumulative relapse rate itself was not statistically different.

In conclusion, appropriate dose reduction of chemotherapeutic drugs, particularly in the induction phase, together with enhanced supportive care is essential to prevent non-relapse death when treating AML in infants. As the conventional dose-intensifying approach is difficult to apply for this age group, less toxic agents targeting specific biological features are needed to improve the outcomes of infants with AML.

Acknowledgments This work was supported by a Grant for Clinical Cancer Research and a Grant-in-Aid for Cancer Research from the Ministry of Health, Labour, and Welfare of Japan.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Pui CH, Ribeiro RC, Campana D, Raimondi SC, Hancock ML, Behm FG, et al. Prognostic factors in the acute lymphoid and myeloid leukemias of infants. Leukemia. 1996;10(6):952–6.
2. Ishii E, Okamura J, Tsuchida M, Kobayashi M, Akiyama Y, Nakahata T, et al. Infant leukemia in Japan: clinical and biological analysis of 48 cases. Med Pediatr Oncol. 1991;19(1):28–32.
3. Sorensen PH, Chen CS, Smith FO, Arthur DC, Domer PH, Bernstein ID, et al. Molecular rearrangements of the MLL gene are present in most cases of infant acute myeloid leukemia and are strongly correlated with monocytic or myelomonocytic phenotypes. J Clin Investig. 1994;93(1):429–37.
4. Lion T, Haas OA, Harbott J, Bannier E, Ritterbach J, Jankovic M, et al. The translocation t(1;22)(p13;q13) is a nonrandom marker specifically associated with acute megakaryocytic leukemia in young children. Blood. 1992;79(12):3325–30.
5. Hauer J, Tosi S, Schuster FR, Harbott J, Kolb HJ, Borkhardt A. Graft versus leukemia effect after haploidentical HSCT in a MLL-negative infant AML with HLXB9/ETV6 rearrangement. Pediatr Blood Cancer. 2008;50(4):921–3.
6. Creutzig U, Zimmermann M, Bourquin JP, Dworzak MN, Kremens B, Lehrnbecher T, et al. Favorable outcome in infants with AML after intensive first- and second-line treatment: an AML-BFM study group report. Leukemia. 2012;26(4):654–61.
7. Gibson BE, Wheatley K, Hann IM, Stevens RF, Webb D, Hills RK, et al. Treatment strategy and long-term results in paediatric patients treated in consecutive UK AML trials. Leukemia. 2005;19(12):2130–8.
8. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. Blood. 2002;100(7):2292–302.
9. Ohta H, Iwamoto S, Kiyokawa N, Tsurusawa M, Deguchi T, Takase K, et al. Flow cytometric analysis of de novo acute myeloid leukemia in childhood: report from the Japanese Pediatric Leukemia/Lymphoma Study Group. Int J Hematol. 2011;93(1):135–7.
10. Tsukimoto I, Tawa A, Horibe K, Tabuchi K, Kigasawa H, Tsuchida M, et al. Risk-stratified therapy and the intensive use of cytarabine improves the outcome in childhood acute myeloid leukemia: the AML99 trial from the Japanese Childhood AML Cooperative Study Group. J Clin Oncol. 2009;27(24):4007–13.
11. Rubnitz JE, Inaba H, Dahl G, Ribeiro RC, Bowman WP, Taub J, et al. Minimal residual disease-directed therapy for childhood acute myeloid leukaemia: results of the AML02 multicentre trial. Lancet Oncol. 2010;11(6):543–52.
12. Kawasaki H, Isoyama K, Eguchi M, Hibi S, Kinukawa N, Kosaka Y, et al. Superior outcome of infant acute myeloid leukemia with intensive chemotherapy: results of the Japan Infant Leukemia Study Group. Blood. 2001;98(13):3589–94.
13. Tomizawa D, Tabuchi K, Kinoshita A, Hanada R, Kigasawa H, Tsukimoto I, et al. Repetitive cycles of high-dose cytarabine are effective for childhood acute myeloid leukemia: long-term outcome of the children with AML treated on two consecutive trials of Tokyo Children's Cancer Study Group. Pediatr Blood Cancer. 2007;49(2):127–32.
14. Schmid I, Stachel D, Pagel P, Albert MH. Incidence, predisposing factors, and outcome of engraftment syndrome in pediatric allogeneic stem cell transplant recipients. Biol Blood Marrow Transplant: J Am Soc Blood Marrow Transplant. 2008;14(4):438–44.
15. Lehrnbecher T, Zimmermann M, Reinhardt D, Dworzak M, Stary J, Creutzig U. Prophylactic human granulocyte colony-stimulating factor after induction therapy in pediatric acute myeloid leukemia. Blood. 2007;109(3):936–43.
16. Ehlers S, Herbst C, Zimmermann M, Scharn N, Germeshausen M, von Neuhoff N, et al. Granulocyte colony-stimulating factor (G-CSF) treatment of childhood acute myeloid leukemias that overexpress the differentiation-defective G-CSF receptor isoform IV is associated with a higher incidence of relapse. J Clin Oncol. 2010;28(15):2591–7.
17. Creutzig U, Zimmermann M, Reinhardt D, Dworzak M, Stary J, Lehrnbecher T. Early deaths and treatment-related mortality in children undergoing therapy for acute myeloid leukemia: analysis of the multicenter clinical trials AML-BFM 93 and AML-BFM 98. J Clin Oncol. 2004;22(21):4384–93.
18. Pieters R. Infant acute lymphoblastic leukemia: lessons learned and future directions. Curr Hematol Malig Rep. 2009;4(3):167–74.

19. Biondi A, Cimino G, Pieters R, Pui CH. Biological and therapeutic aspects of infant leukemia. Blood. 2000;96(1):24–33.
20. Smith FO, Alonzo TA, Gerbing RB, Woods WG, Arceci RJ. Long-term results of children with acute myeloid leukemia: a report of three consecutive phase III trials by the Children's Cancer Group: CCG 251, CCG 213 and CCG 2891. Leukemia. 2005;19(12):2054–62.
21. Sung L, Alonzo TA, Gerbing RB, Aplenc R, Lange BJ, Woods WG, et al. Respiratory syncytial virus infections in children with acute myeloid leukemia: a report from the Children's Oncology Group. Pediatr Blood Cancer. 2008;51(6):784–6.

Mutations of the *GATA2* and *CEBPA* genes in paediatric acute myeloid leukaemia

Hereditary *GATA2* mutations show predisposition to acute myeloid leukaemia (AML) and myelodysplastic syndrome (MDS) (Hahn *et al*, 2011). These mutations have also been reported in chronic myeloid leukaemia (Zhang *et al*, 2008) and monocytopenia and mycobacterial infection (MonoMAC) syndrome (Hsu *et al*, 2011). More recently, *GATA2* mutations have been identified in *de novo* AML, especially in adult patients with biallelic *CEBPA* mutations (Greif *et al*, 2012; Green *et al*, 2013). GATA2 and CEBPA are transcription factors that are crucial for haematopoietic development. These findings prompted us to identify possible *GATA2* and *CEBPA* mutations in patients with various paediatric leukaemias.

Direct Sequencing of *GATA2* was performed in 157 *de novo* AML patients, including 13 patients with acute promyelocytic leukaemia (APL; French–American–British type-M3) and 10 with Down syndrome (DS; Table S1), 22 secondary AML patients, 40 juvenile myelomonocytic leukaemia (JMML) patients, 50 acute lymphoblastic leukaemia (ALL) patients, 70 cell lines (25 B-cell precursor-ALL, 15 T-cell-ALL, 22 AML, and 8 neuroblastomas), and 60 healthy subjects. *GATA2* mutation analysis was performed by direct sequencing for all coding exons (exons 2–6) using an ABI PRISM 3130 Genetic Analyser (Applied Biosystems, Branchburg, NJ, USA) (Table S2). For AML patients, *CEBPA* and *NPM1* mutations were also examined. Mutational analyses of *FLT3*, *KIT*, *WT1* and *RAS* genes in our AML patients was performed as described previously (Shimada *et al*, 2006). Informed consent was obtained from the patients or the patients' parents according to guidelines based on the tenets of the revised Helsinki protocol. The institutional review boards of Gunma Children's Medical Centre approved this project.

GATA2 mutations were found in eight out of 157 AML patients (5·1%), including three APL patients (Fig 1A,B), but were absent in 18 patients with acute megakaryocytic leukaemia (FAB-M7; Table S3). Furthermore, there were no *GATA2* mutations in patients with other leukaemias, in the cell lines, or in the 60 healthy subjects, suggesting that *GATA2* mutations were indeed associated with leukaemogenesis in a subset of patients with *de novo* AML.

Germline *GATA2* mutations were also examined in five AML patients whose complete remission (CR) samples were available, and a germline mutation was identified in one patient. Furthermore, we performed *GATA2* mutation analyses of the patient's parents and two siblings, and identified the same *GATA2* mutations in her father (II-4) and brother (III-1) but not in her mother (II-5) or sister (III-2) (Fig 1C). Her father and brother lacked abnormalities in their full blood cell counts, lymphocyte subsets, or episodes of opportunistic infections. The proband experienced severe mycotic pneumonia during induction chemotherapy. Remarkably, she has been in CR for more than 11 years, despite discontinuation of chemotherapy. Three patients, for whom CR samples were not available, had no history of MonoMAC syndrome.

In addition, 16 *CEBPA* mutations (10·2%) and three *NPM1* mutations (1·9%) were found in 157 paediatric AML patients. Thirteen (81·3%) of 16 patients with *CEBPA* mutations had been in CR for more than 4 years, suggesting that *CEBPA* mutations may be associated with favourable outcomes. Although most *GATA2* mutations were found in patients with biallelic *CEBPA* mutations in adult AML (Greif *et al*, 2012; Green *et al*, 2013), only two of eight *GATA2* mutation-positive patients had monoallelic *CEBPA* mutations in this study (Table I).

We compared the clinical and molecular features between patients with and without *GATA2* mutations. However, there were no significant differences in terms of age, initial white blood cell count, gender, and cytogenetics (Table S3). Of the eight patients with *GATA2* mutations, one had a *WT1* mutation, one had a *KIT* mutation, and two patients had *RAS* mutations (Table I). *FLT3*-internal tandem duplication, *MLL*-partial tandem duplication, and *NPM1* mutations were not found in any patients with *GATA2* mutations (Table S3). All of the *GATA2* mutations were found in the intermediate risk subgroup or APL patients with t(15;17), whereas none were found in those with core-binding factor AML [i.e. t(8;21) and inv(16)]. *GATA2* mutations were found in two patients with 11q23 translocations, including t(11;19) and t(7;11), and three patients with complex chromosomal abnormalities, whereas most *GATA2* mutations were found in cytogenetically normal AML patients in previous reports (Table I) (Greif *et al*, 2012; Luesink *et al*, 2012).

GATA2 mutations were previously reported in patients with M1, M2, and M4 subtypes of AML (Greif *et al*, 2012; Luesink *et al*, 2012), which is in accordance with our results. *GATA2* mutations have not been previously reported in APL, but our study found these mutations in three APL patients. Of note, promyelocytic leukaemia protein has been shown to interact with GATA2 and potentiate its transactivation capacity (Tsuzuki *et al*, 2000).

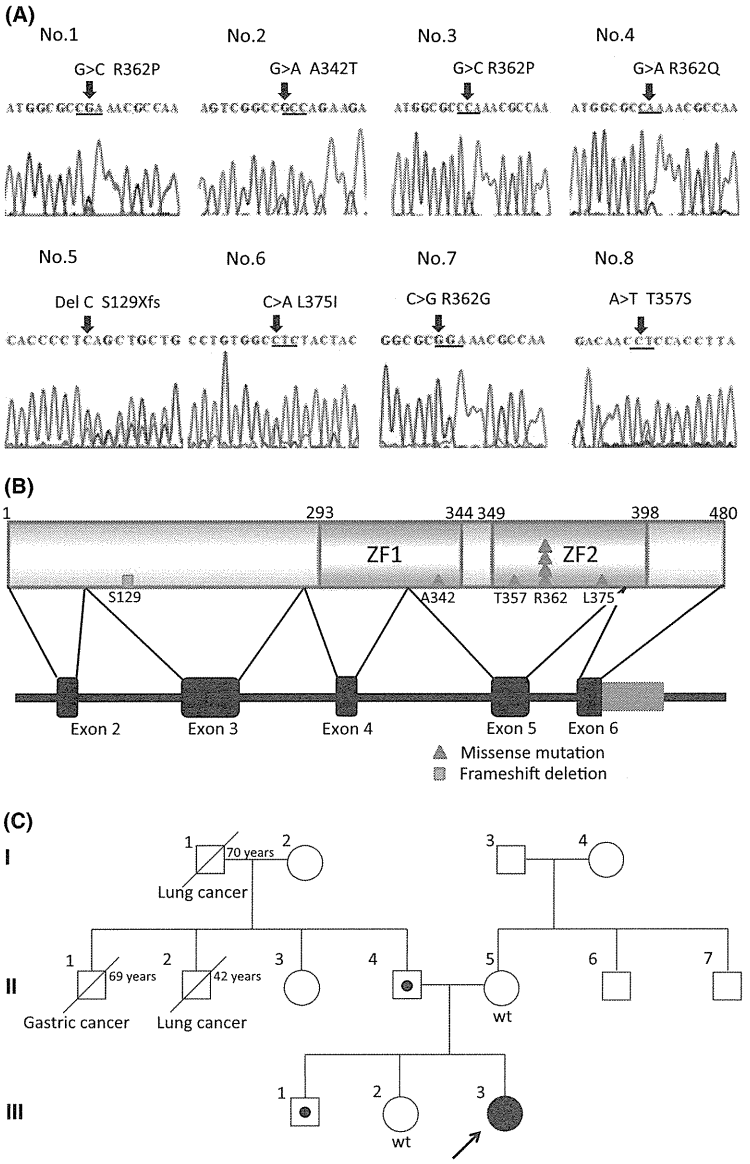


Fig 1. Identification of *GATA2* mutations by direct sequencing. (A) Eight *GATA2* mutations were identified in 157 Japanese paediatric *de novo* acute myeloid leukaemia (AML) patients (5·1%). Major missense mutations were R362 (R362P, R362Q, and R362G). Small vertical arrows indicate the mutated nucleotides. (B) Of the eight mutations, six mutations were identified in the ZF2 domain, one mutation was identified in the ZF1, and a mutation was identified on the outside of the ZF domain. (C) The family pedigree is shown. Squares indicate males and circles indicate females. The proband (III-3) is indicated by an arrow. The proband, her father (II-4), and her brother (III-1) harboured *GATA2* mutations (shown by squares containing dots). Her uncles and grandfather died of lung cancer (I-1 and II-2) and gastric cancer (II-1). wt, wild-type.

The outcomes of our patients with *GATA2* mutations was not poor (3-year overall survival and event free survival: 87·5%), which is in agreement with previous reports on *de novo* AML (Greif *et al*, 2012; Luesink *et al*, 2012): two of eight patients received autologous-stem cell transplantation (Auto-SCT), and one died of gastrointestinal haemorrhage after Auto-SCT. The remaining six patients who did not receive Auto-SCT were still alive (Table I).

In this study, one patient with a germline *GATA2* mutation developed AML. Her paternal grandfather (I-1) and second uncle (II-2) died of lung cancer at the age of 70 and 42 years, respectively, while her first uncle (II-1) died of gastric cancer at 69 years of age (Fig 1C). Increased GATA2 protein expression has been associated with biochemical recurrence and distant metastatic progression in prostate cancer (Böhm *et al*, 2009), as loss of GATA2 reduced the viability of Non-small cell lung cancer cells with RAS-pathway mutations, whereas wild-type cells were unaffected (Kumar *et al*, 2012). These facts indicate that GATA2 upregulation is strongly associated with maintenance of cancer cells. The association between *GATA2* mutations and solid tumours remains to be elucidated.

Our results indicate that *GATA2* mutations are associated with a favourable outcome in paediatric AML. Therefore, less aggressive treatment strategies without SCT may be

Table I. Clinical and molecular characteristics of patients with *GATA2* mutations.

Pt	Sex	Age (years)	FAB	WBC ($\times 10^9$/l)	Chromosome	Risk	Tx	Relapse	Prognosis (months)	*GATA2* mutation	Germline	Additional mutations
1	M	3	M4	23·8	46, XY, t(11;19) (q23; p13·1)	IR	Auto	Yes	16	R362P	N/A	–
2	F	7	M0	3·7	45,XX,add(3)(p13),del(6)(q?), der(8) t(3;8)(p21;q24), −13	IR	Chemo	No	+141	A342T	Yes	*NRAS*
3	F	8	M1	1·8	46, XX	IR	Chemo	No	+56	R362P	No	*KRAS*
4	M	14	M1	440·0	46XY [2/8], 46, XY, del(6) (q15 q21), −7, −9, −10, +3mar[1/8], 46, XY, ?de(3) (p25)[1/8], 47, XY, −5, −8, −10, add(12)(q24·1), −16, −18, +6mar [1/8], 46, XY, −2, −6, −8, +3mer [1/8], 46, XY, −8, +mar [1/8], 46, Y, ?add(X)(p11·2) [1/8]	IR	Auto	No	+51	R362Q	No	*WT1*, *CEBPA*-SM
5	M	11	M3	16·1	46,XX,inv(9)(p11q13),t(15;17)(q22;q11-21)	M3	Chemo	No	+50	S129X	N/A	–
6	M	3	M3	11·6	46,XY,t(15;17)(q22;q11?21)	M3	Chemo	No	+45	L375I	No	*CEBPA*-SM
7	M	10	M3	13·6	47,XY, +8, t(15;17)(q22;q11-21)	M3	Chemo	No	+41	R362G	N/A	*KIT*
8	F	2	M4	12·7	48, XX, +6, +10, t(11?; 7) (q23;q25)	IR	Chemo	No	+38	T357S	No	–

Pt, Patient; FAB, French–American–British classification; WBC, white blood cell count; Tx, Treatment; M, Male; F, Female; IR, Intermediate risk; Auto, Autologous stem cell transplantation; Chemo, Chemotherapy; N/A, not available; +, alive; SM, single mutation.

appropriate for paediatric AML patients with *GATA2* mutations, although most patients with *GATA2* mutations were classified into an intermediate risk group. Furthermore, the association between germline *GATA2* mutations and solid tumours remains to be elucidated.

Acknowledgements

We are grateful to all members of the Japanese Childhood AML Cooperative Study Group. Members of the Japanese Childhood AML Cooperative Study Group who contributed data to the study include Ryoji Hanada, Department of Haematology/Oncology, Saitama Children's Medical Centre; Masahiro Tsuchida, Department of Haematology/Oncology, Ibaraki Children's Medical Centre; Akira Morimoto, Department of Paediatrics, Kyoto Prefectural University of Medicine; Ryoji Kobayashi, Department of Paediatrics, Hokkaido University School of Medicine; Hiromasa Yabe, Department of Paediatrics, Tokai University School of Medicine; Kazuko Hamamoto, Department of Paediatrics, Hiroshima Red Cross Hospital; Shigeru Tsuchiya, Department of Paediatric Oncology, Institute of Development, Aging and Cancer, Tohoku University; Yuichi Akiyama, Department of Paediatrics, National Hospital Organization Kyoto Medical Centre; Hisato Kigasawa, Department of Haematology, Kanagawa Children's Medical Centre; Akira Ohara, Department of First Paediatrics, Toho University School of Medicine; Hideki Nakayama, Department of Paediatrics, Hamanomachi Hospital; Kazuko Kudo, Department of Paediatrics, Nagoya University Graduate School of Medicine; and Masue Imaizumi, Department of Haematology/Oncology, Miyagi Prefectural Children's Hospital. The authors would like to thank Enago (www.enago.jp) for the English language review. This work was supported by a grant for Cancer Research, a grant for Research on Children and Families, and Research on Intractable Diseases, Health and Labour Sciences Research Grants from the Ministry of Health, Labor and Welfare of Japan, by Grants-in-Aid for Scientific Research (B, C) and Exploratory Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan, by the Programme for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio) of Japan, and by a Research grant for Gunma Prefectural Hospitals.

Author contributions

Y.H. designed the study; M.F., S.A., M.K., A.K., M.S., A.T., K.H. and I.T. collected patient samples and clinical data; N.S., K.O., M.-J.P., Y.M. and S.M. performed the laboratory research; N.S., M.-J.P. and Y.H. analysed and interpreted the data; N.S. performed the statistical analysis; N.S. and Y.H. wrote the manuscript; H.A. and Y.H. supervised the work; and all authors critically reviewed the manuscript and gave their final approval.

Conflicts of interest

The authors declare no competing financial interests.

Norio Shiba[1,2]
Michinori Funato[3]
Kentaro Ohki[1]
Myoung-ja Park[1]
Yasuhiro Mizushima[4]
Souichi Adachi[5]
Masao Kobayashi[6]
Akitoshi Kinoshita[7]
Manabu Sotomatsu[1]
Hirokazu Arakawa[2]
Akio Tawa[8]
Keizo Horibe[9]
Ichiro Tsukimoto[10]
Yasuhide Hayashi[1]

[1]Department of Haematology/Oncology, Gunma Children's Medical Centre, Shibukawa, [2]Department of Paediatrics, Gunma University Graduate School of Medicine, Maebashi, [3]Department of Paediatrics, Graduate School of Medicine, Gifu University, Gifu, [4]Department of Paediatrics, Kyoto-Katsura Hospital, [5]Department of Human Health Sciences, Kyoto University Graduate School of Medicine, Kyoto, [6]Department of Paediatrics, Hiroshima University Graduate School of Biomedical and Health Sciences, Hiroshima, [7]Department of Paediatrics, St. Marianna University School of Medicine, Kawasaki, [8]Department of Paediatrics, National Hospital Organization Osaka National Hospital, Osaka, [9]Clinical Research Centre, National Hospital Organization Nagoya Medical Centre, Nagoya, and [10]First Department of Paediatrics, Toho University School of Medicine, Tokyo, Japan
E-mail: hayashiy-tky@umin.ac.jp

Keywords: *GATA2*, *CEBPA*, paediatric acute myeloid leukaemia, acute promyelocytic leukaemia, germline mutation

First published online 14 September 2013
doi: 10.1111/bjh.12559

Supporting Information

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

Table S1. Clinical and cytogenetically characteristics of 157 AML patients.

Table S2. PCR primers used for mutation screening.

Table S3. Clinical and molecular characteristics of *GATA2* mutation positive patients.

References

Böhm, M., Locke, W.J., Sutherland, R.L., Kench, J.G. & Henshall, S.M. (2009) A role for GATA-2 in transition to an aggressive phenotype in prostate cancer through modulation of key androgen-regulated genes. *Oncogene*, **28**, 3847–3856.

Green, C.L., Tawana, K., Hills, R.K., Bödör, C., Fitzgibbon, J., Inglott, S., Ancliff, P., Burnett, A.K., Linch, D.C. & Gale, R.E. (2013) GATA2 mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations. *British Journal of Haematology*, **161**, 701–705.

Greif, P.A., Dufour, A., Konstandin, N.P., Ksienzyk, B., Zellmeier, E., Tizazu, B., Sturm, J., Benthaus, T., Herold, T., Yaghmaie, M., Dörge, P., Hopfner, K.P., Hauser, A., Graf, A., Krebs, S., Blum, H., Kakadia, P.M., Schneider, S., Hoster, E., Schneider, F., Stanulla, M., Braess, J., Sauerland, M.C., Berdel, W.E., Büchner, T., Woermann, B.J., Hiddemann, W., Spiekermann, K. & Bohlander, S.K. (2012) GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia. *Blood*, **120**, 395–403.

Hahn, C.N., Chong, C.E., Carmichael, C.L., Wilkins, E.J., Brautigan, P.J., Li, X.C., Babic, M., Lin, M., Carmagnac, A., Lee, Y.K., Kok, C.H., Gagliardi, L., Friend, K.L., Ekert, P.G., Butcher, C.M., Brown, A.L., Lewis, I.D., To, L.B., Timms, A.E., Storek, J., Moore, S., Altree, M., Escher, R., Bardy, P.G., Suthers, G.K., D'Andrea, R.J., Horwitz, M.S. & Scott, H.S. (2011) Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nature Genetics*, **43**, 1012–1017.

Hsu, A.P., Sampaio, E.P., Khan, J., Calvo, K.R., Lemieux, J.E., Patel, S.Y., Frucht, D.M., Vinh, D.C., Auth, R.D., Freeman, A.F., Olivier, K.N., Uzel, G., Zerbe, C.S., Spalding, C., Pittaluga, S., Raffeld, M., Kuhns, D.B., Ding, L., Paulson, M.L., Marciano, B.E., Gea-Banacloche, J.C., Orange, J.S., Cuellar-Rodriguez, J., Hickstein, D.D. & Holland, S.M. (2011) Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*, **118**, 2653–2655.

Kumar, M.S., Hancock, D.C., Molina-Arcas, M., Steckel, M., East, P., Diefenbacher, M., Armenteros-Monterroso, E., Lassailly, F., Matthews, N., Nye, E., Stamp, G., Behrens, A. & Downward, J. (2012) The GATA2 transcriptional network is requisite for RAS oncogene-driven non-small cell lung cancer. *Cell*, **149**, 642–655.

Luesink, M., Hollink, I.H., van der Velden, V.H., Knops, R.H., Boezeman, J.B., de Haas, V., Trka, J., Baruchel, A., Reinhardt, D., van der Reijden, B.A., van den Heuvel-Eibrink, M.M., Zwaan, C.M. & Jansen, J.H. (2012) High GATA2 expression is a poor prognostic marker in pediatric acute myeloid leukemia. *Blood*, **120**, 2064–2075.

Shimada, A., Taki, T., Tabuchi, K., Tawa, A., Horibe, K., Tsuchida, M., Hanada, R., Tsukimoto, I. & Hayashi, Y. (2006) KIT mutations, and not FLT3 internal tandem duplication, are strongly associated with a poor prognosis in pediatric acute myeloid leukemia with t(8;21): a study of the Japanese Childhood AML Cooperative Study Group. *Blood*, **107**, 1806–1809.

Tsuzuki, S., Towatari, M., Saito, H. & Enver, T. (2000) Potentiation of GATA-2 activity through interactions with the promyelocytic leukemia protein (PML) and the t(15;17)-generated PML-retinoic acid receptor alpha oncoprotein. *Molecular and Cellular Biology*, **20**, 6276–6286.

Zhang, S.J., Ma, L.Y., Huang, Q.H., Li, G., Gu, B.W., Gao, X.D., Shi, J.Y., Wang, Y.Y., Gao, L., Cai, X., Ren, R.B., Zhu, J., Chen, Z. & Chen, S.J. (2008) Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia. *Proceedings of the National Academy of Sciences*, **105**, 2076–2081.

nature genetics

The landscape of somatic mutations in Down syndrome–related myeloid disorders

Kenichi Yoshida[1,2,17], Tsutomu Toki[3,17], Yusuke Okuno[1,17], Rika Kanezaki[3], Yuichi Shiraishi[4], Aiko Sato-Otsubo[1,2], Masashi Sanada[1,2], Myoung-ja Park[5], Kiminori Terui[3], Hiromichi Suzuki[1,2], Ayana Kon[1,2], Yasunobu Nagata[1,2], Yusuke Sato[1,2], RuNan Wang[3], Norio Shiba[5], Kenichi Chiba[4], Hiroko Tanaka[6], Asahito Hama[7], Hideki Muramatsu[7], Daisuke Hasegawa[8], Kazuhiro Nakamura[9], Hirokazu Kanegane[10], Keiko Tsukamoto[11], Souichi Adachi[12], Kiyoshi Kawakami[13], Koji Kato[14], Ryosei Nishimura[15], Shai Izraeli[16], Yasuhide Hayashi[5], Satoru Miyano[4,6], Seiji Kojima[7], Etsuro Ito[3,18] & Seishi Ogawa[1,2,18]

Transient abnormal myelopoiesis (TAM) is a myeloid proliferation resembling acute megakaryoblastic leukemia (AMKL), mostly affecting perinatal infants with Down syndrome. Although self-limiting in a majority of cases, TAM may evolve as non-self-limiting AMKL after spontaneous remission (DS-AMKL). Pathogenesis of these Down syndrome–related myeloid disorders is poorly understood, except for *GATA1* mutations found in most cases. Here we report genomic profiling of 41 TAM, 49 DS-AMKL and 19 non-DS-AMKL samples, including whole-genome and/or whole-exome sequencing of 15 TAM and 14 DS-AMKL samples. TAM appears to be caused by a single *GATA1* mutation and constitutive trisomy 21. Subsequent AMKL evolves from a pre-existing TAM clone through the acquisition of additional mutations, with major mutational targets including multiple cohesin components (53%), *CTCF* (20%), and *EZH2*, *KANSL1* and other epigenetic regulators (45%), as well as common signaling pathways, such as the JAK family kinases, *MPL*, *SH2B3* (*LNK*) and multiple RAS pathway genes (47%).

TAM represents a transient proliferation of immature megakaryoblasts that occurs in 5–10% of perinatal infants with Down syndrome[1,2]. Although morphologically indistinguishable from AMKL, TAM is self-limiting in the majority of cases and usually terminates spontaneously within 3–4 months of birth[1]. Hepatic infiltration of myeloid cells is a common finding and can be severe enough to be fatal, owing to hepatic failure, with liver fibrosis occurring in 5–16% of cases[2–4]. Moreover, even when spontaneous remission is achieved, approximately 20–30% of surviving infants develop DS-AMKL years after remission, although some DS-AMKL cases have no documented history of TAM[4]. In contrast to non–Down syndrome–related AMKL (non-DS-AMKL), which generally shows poor prognosis, individuals with DS-AMKL typically have a favorable prognosis. In molecular pathogenesis of these Down syndrome–related myeloid disorders, *GATA1* mutations are detected in virtually all affected infants, suggesting their central role in Down syndrome–related myeloid proliferation[5,6]. However, it is still open to question whether a *GATA1* mutation is sufficient for the development of TAM in individuals with Down syndrome, what is the cellular origin of the subsequent AMKL, whether additional gene mutations are required for progression to AMKL, and, if so, what are their gene targets, although several genes have been reported to be mutated in occasional cases with DS-AMKL, including *JAK1*, *JAK2* and *JAK3* (refs. 7–10), *TP53* (refs. 10,11), *FLT3* (ref. 8) and *MPL*[12]. We reasoned that identifying a comprehensive registry of gene mutations and tracking them at a clonal level using massively parallel sequencing would provide vital information for addressing these questions.

RESULTS

Genomic landscape of Down syndrome–related myeloid neoplasms

We performed whole-genome sequencing of 4 trios consisting of samples from TAM, AMKL and complete remission phases (**Supplementary Figs. 1** and **2** and **Supplementary Table 1**). In total,

[1]Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [2]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [3]Department of Pediatrics, Hirosaki University Graduate School of Medicine, Hirosaki, Japan. [4]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [5]Department of Hematology/Oncology, Gunma Children's Medical Center, Shibukawa, Japan. [6]Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [7]Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan. [8]Department of Pediatrics, St. Luke's International Hospital, Tokyo, Japan. [9]Department of Pediatrics, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan. [10]Department of Pediatrics, Graduate School of Medicine, University of Toyama, Toyama, Japan. [11]Division of Neonatology, National Center for Child Health and Development, Tokyo, Japan. [12]Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [13]Department of Pediatrics, Kagoshima City Hospital, Kagoshima, Japan. [14]Department of Hematology and Oncology, Children's Medical Center, Japanese Red Cross Nagoya First Hospital, Nagoya, Japan. [15]Department of Pediatrics, School of Medicine, Institute of Medical, Pharmaceutical and Health Sciences, Kanazawa University, Kanazawa, Japan. [16]Functional Genomics, Cancer Research Center, Sheba Medical Center, Tel Hashomer and Tel Aviv University, Tel Aviv, Israel. [17]These authors contributed equally to this work. [18]These authors jointly directed this work. Correspondence should be addressed to S.O. (sogawa-tky@umin.ac.jp) or E.I. (eturou@cc.hirosaki-u.ac.jp).

Received 3 May; accepted 19 August; published online 22 September 2013; corrected after print 30 October 2013; doi:10.1038/ng.2759

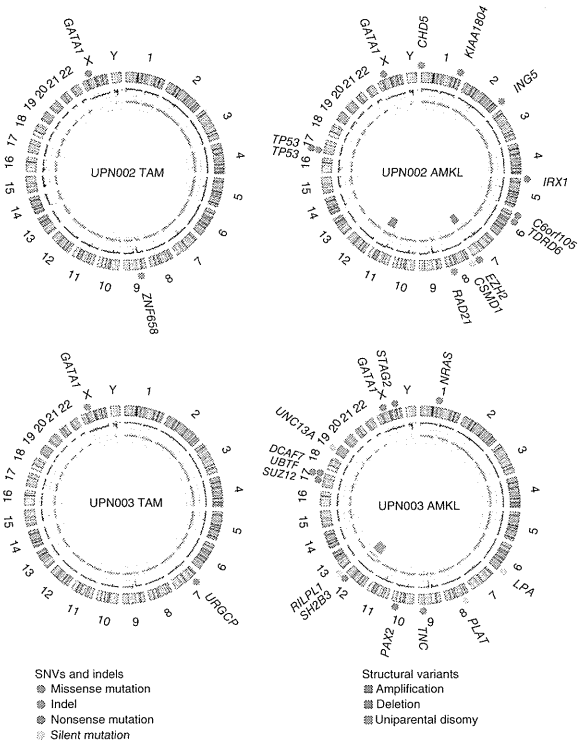


Figure 1 Representative Circos plots of paired TAM and DS-AMKL cases. Locations of somatic mutations, including of missense, frameshift, nonsense and silent mutations (colored circles), are indicated. Total (black) and allele-specific (red and green for alleles showing relatively larger and smaller copy numbers, respectively) genomic copy numbers, as well as somatic structural variants (colored bars), are indicated in the inner circle. Sample IDs are shown within each plot; plots were created with *Circus*[53].

we confirmed 411 single-nucleotide variants (SNVs) and 17 small nucleotide insertions and deletions (indels) by Sanger sequencing and/or deep resequencing (**Supplementary Fig. 1** and **Supplementary Table 2**). We detected only a few structural variants, including deletion, amplification and uniparental disomy, in the TAM and DS-AMKL genomes (**Fig. 1** and **Supplementary Fig. 3**). The mean number of validated somatic mutations in DS-AMKL samples (71 or 0.023 mutations/Mb) was twice the number observed in TAM samples (36 or 0.012 mutations/Mb) (**Supplementary Fig. 1a**). Mutation numbers in samples from both phases were substantially lower than in most other cancers (**Supplementary Fig. 4**), although differences in mutation rates could partly be affected by different definitions and algorithms for mutation calling. The spectrum of mutations was over-represented by C-to-T and G-to-A transitions in both TAM and DS-AMKL samples, resembling the mutational spectra in gastric and colorectal cancers[13] and in other blood cancers (**Supplementary Fig. 1b**)[14,15]. We unmasked the details of clonal evolution and expansion leading to AMKL through the use of deep sequencing of individual mutations detected by combined whole-genome and whole-exome sequencing (**Fig. 2** and **Supplementary Table 2**). Intratumoral heterogeneity was evident at initial diagnosis with TAM and in the AMKL phase in all cases (**Supplementary Fig. 5**). In UPN001, UPN002 and UPN004, AMKL evolved from one of the major subclones in the TAM phase with a shared *GATA1* mutation, as reported previously in relapsed acute myeloid leukemia (AML) in adults (**Fig. 2a,b,d**)[15]. In contrast, UPN003 showed a unique pattern of clonal evolution, in which AMKL originated from a minor subclone in the TAM phase that was totally unrelated to the predominant clone in terms of somatic mutations, with no mutation shared by both phases, and carried an independent *GATA1* mutation (**Fig. 2c**). In both scenarios, progression to AMKL seemed to be accompanied by many additional mutations, including common driver mutations that were absent in the original TAM population, indicating a multistep process of leukemogenesis.

Exome sequencing

We further investigated non-silent mutations by whole-exome sequencing of additional samples to generate a full registry of driver mutations that are relevant to the development of TAM and subsequent progression to AMKL (**Supplementary Fig. 6** and **Supplementary Table 1**). We detected *GATA1* mutations in all TAM and DS-AMKL cases, indicating sufficient sensitivity in our whole-exome analysis. In total, we confirmed 26 and 81 non-silent somatic mutations identified in the exome analysis of 15 TAM and 14 DS-AMKL samples, respectively, with 3 *GATA1* mutations common to both phases (**Supplementary Table 3**). The mean number of non-silent mutations was significantly higher in DS-AMKL samples (5.8; range of 1–11) than in TAM samples (1.7; range of 1–5) ($P = 0.0002$) (**Fig. 3a**). Of the 107 mutations, 84 were single-nucleotide substitutions that were mostly within coding sequences, except for 4 splice-site mutations. We also observed predominantly C-to-T and G-to-A transitions for non-silent substitutions (**Supplementary Fig. 7**). The remaining mutations were frameshift ($n = 21$) or non-frameshift ($n = 2$) indels, most frequently involving *GATA1* ($n = 13$). One individual with DS-AMKL (UPN004) had no SNVs or indels (**Fig. 3a**), but copy number analysis identified a large deletion at 16q involving the *CTCF* locus (**Supplementary Fig. 3**), suggesting that the alteration of *CTCF* could be a driver event in this case. Therefore, at least one additional genetic lesion other than *GATA1* mutation was detected in our whole-exome sequencing, despite the low frequency of leukemic cells appearing to show the morphology of immature megakaryoblasts (blast percentage) in many cases, which is a known characteristic of DS-AMKL samples[16,17]. Whole-exome sequencing results suggested the presence of intratumoral heterogeneity in the majority of DS-AMKL cases (**Fig. 3b**).

Spectrum of recurrent mutations in DS-AMKL

Recurrently affected genes are of primary interest in identifying driver mutations. Whereas *GATA1* was the only recurrent mutational target in TAM samples, an additional eight genes were recurrently mutated in the DS-AMKL samples, including *RAD21*, *STAG2*, *NRAS*, *CTCF*, *DCAF7*, *EZH2*, *KANSL1* and *TP53* (**Table 1**). These genes are expressed in a wide variety of hematopoietic compartments, including in both myeloid and lymphoid cells, except for *EZH2*, whose expression is largely confined to $CD34^+$ cells[18] (**Supplementary Fig. 8**). We also found that these genes were expressed in DS-AMKL cells at similar levels to common hematopoietic genes[19], although we did not observe significant difference in their expression levels in DS-AMKL and non-DS-AMKL cells (**Supplementary Fig. 9**).

We then performed targeted deep sequencing of these 8 genes in an extended set of 109 samples (including 29 samples in 25 discovery cases) consisting of 41 TAM, 49 DS-AMKL and 19 non-DS-AMKL samples (**Supplementary Tables 1** and **4**). We also included additional genes in targeted sequencing that were either functionally related to the above eight genes or were mutated only in single cases but had been previously reported to be mutated in DS-AMKL (*JAK3*) or other myeloid neoplasms (*SH2B3*, *SUZ12*, *SRSF2* and *WT1*), together with other common mutational targets in adult myeloid malignancies

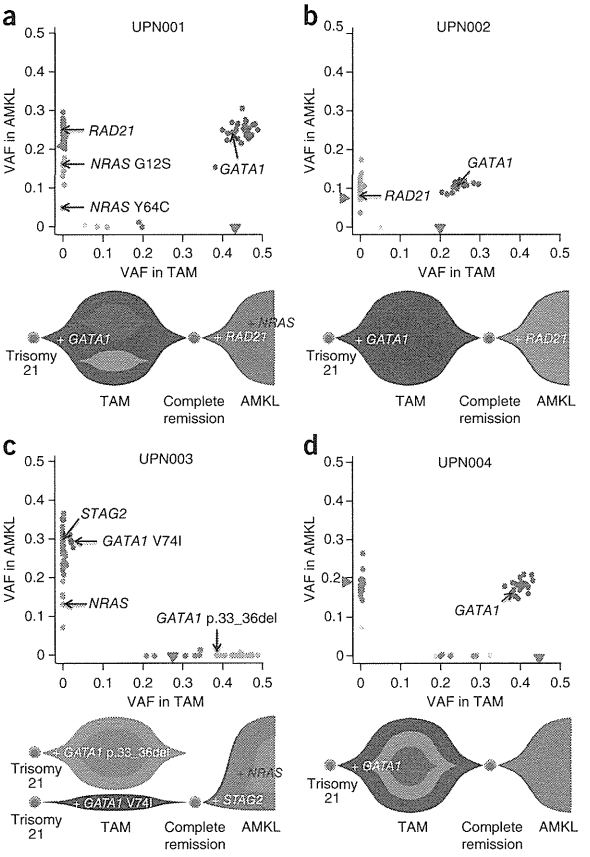


Figure 2 Clonal evolution of Down syndrome–related myeloid disorders. (**a**–**d**) Observed VAFs of validated mutations listed in **Supplementary Table 2** in both TAM and AMKL phases are shown in diagonal plots (top) for UPN001 (**a**), UPN002 (**b**), UPN003 (**c**) and UPN004 (**d**), where VAFs of genes on the X chromosome in male cases or in regions of uniparental disomy were halved. Half the value of the blast percentage, which corresponds to the allele frequency of a heterozygous mutation distributed in all tumor cells, is also shown by a red arrowhead, except for UPN003 AMKL, for which clinical data were not available. Driver mutations including in *GATA1*, *STAG2*, *RAD21* and *NRAS* are indicated by black arrows. Predicted chronological behaviors of different leukemia subclones are depicted below each diagonal plot. Distinct mutation clusters are indicated by color. In UPN001, UPN002 and UPN004, founding clones of TAM shown in blue became dominant in the AMKL samples, in which some subsequent subclones evolved through the serial acquisition of SNVs. In contrast, in UPN003, a subclone in the TAM phase (blue) and not the founding clone of TAM (aqua) became dominant in the AMKL sample. VAFs of some mutations were higher than for *GATA1* but seem to be actually equivalent to it given the error range of PCR-based deep sequencing.

(**Supplementary Fig. 10** and **Supplementary Tables 5** and **6**). We also analyzed by RT-PCR two recurrent fusion genes previously reported in non-DS-AMKL cases, *RBM15-MKL1* (*OTT-MAL*)[20,21] and *CBFA2T3-GLIS2* (refs. 22,23).

Mutations of cohesin and associated molecules

Major components of the cohesin complex, including *RAD21* and *STAG2*, were frequent targets of gene mutations in DS-AMKL (**Table 1**). Including an additional mutation in *NIPBL*, 8 of the 14 discovery DS-AMKL cases (57%) had a mutated cohesin or associated component (**Supplementary Table 3**). Cohesin is a multiprotein complex consisting of 4 core components, including the SMC1, SMC3, RAD21 and STAG proteins[24,25]. In concert with several functionally associated proteins, such as the NIPBL and ESCO proteins, cohesin is engaged in the cohesion of newly replicated sister chromatids by forming a ring-like structure[25], preventing their premature separation before late anaphase. Cohesin has also been implicated in post-replicative DNA repair and long-range regulation of gene expression[26–30]. Targeted deep sequencing confirmed recurrent mutations and deletions in all core cohesin components (*STAG2*, *RAD21*, *SMC3* and *SMC1A*) and in *NIPBL* in 26 of 49 DS-AMKL cases (53%) but in none of the 41 TAM cases, although 2 non-DS-AMKL cases (11%) had *STAG2* mutations (**Fig. 4a,b** and **Supplementary Tables 7** and **8**). Strikingly, all mutations and deletions in different cohesin components were completely mutually exclusive, suggesting that cohesin function was the common target of these mutations. All but one *STAG2* mutation (encoding a p.Arg370Gln substitution) was either a nonsense, frameshift or splice-site change (**Fig. 4a,b**, **Supplementary Figs. 11** and **12a**, and **Supplementary Table 7**). Similarly, 6 of 9 *RAD21* mutations were heterozygous nonsense or frameshift alterations. Four of the five mutations in *NIPBL*, *SMC1A* and *SMC3* were also nonsense or splice-site changes causing abnormal exon skipping (**Fig. 4a** and **Supplementary Table 7**). Thus, most of these mutations were thought to result in premature truncation, leading to loss of cohesin function. The leukemogenic mechanism of mutated cohesin components is still elusive; some studies have implicated aneuploidy caused by cohesin dysfunction in oncogenic actions[31]. However, DS-AMKL cases have been characterized by a largely normal karyotype[32]. We found no significant difference in the frequency of aneuploidy between cases with mutated and wild-type cohesin in the current DS-AMKL cohort. Many cases with mutated cohesin had completely normal karyotypes, except for constitutive trisomy 21, arguing against the hypothesis that aneuploidy has a major role in the pathogenesis of cohesin-mutated DS-AMKL (**Fig. 5a**).

CTCF mutations

Given the high frequency of cohesin mutations, new recurrent *CTCF* mutations were of particular interest because the functional interaction of cohesin and CTCF proteins has been of emerging interest in the long-range regulation of gene expression[26,30,33,34]. CTCF is a zinc-finger protein implicated in diverse regulatory functions, including transcriptional activation and/or repression, insulation, formation of chromatin barrier, imprinting and X-chromosome inactivation[35]. CTCF binds to target sequence elements and blocks the interaction of enhancers and promoters through DNA loop formation (insulator activity)[36], and several lines of evidence suggest that cohesin occupies CTCF-binding sites to contribute to the long-range regulation of gene expression by participating in the formation and stabilization of a repressive loop[26,37]. *CTCF* was mutated or deleted in ten DS-AMKL cases (20%), one TAM case (2%) and four non-DS-AMKL cases (21%), with seven mutations representing nonsense, frameshift or splice-site changes and an additional six alterations representing deletions resulting in the loss of protein function (**Fig. 4a,b**, **Supplementary Figs. 11** and **12b**, and **Supplementary Tables 7** and **8**). To our knowledge, this is the first report of frequent recurrent *CTCF* mutations in cancer, although rare mutations (occurring in approximately 2% of cases) have recently been reported in breast cancer sequencing[38].

Mutations in epigenetic regulators

EZH2, which encodes a catalytic subunit of the Polycomb repressive complex 2 (PRC2) that is responsible for di- and trimethylation of histone H3 lysine 27 (H3K27)[39], is another recurrent mutational target in DS-AMKL (**Table 1**). Inactivating mutations in *EZH2* have

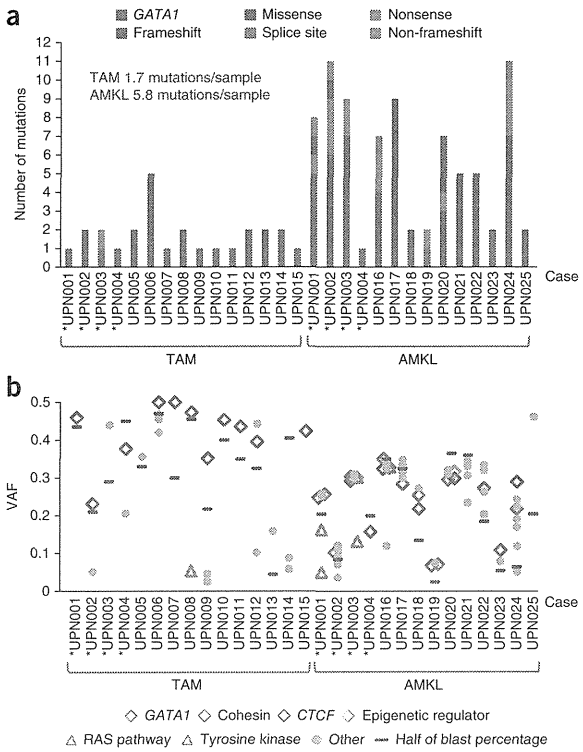


Figure 3 Somatic mutations detected by whole-exome sequencing of Down syndrome–related myeloid disorders. (**a**) Number of validated somatic mutations in 25 individuals with TAM and DS-AMKL identified by whole-exome sequencing. Paired samples are indicated by asterisks. The mutation rates per phase are given. (**b**) VAFs of individual mutations determined by deep sequencing, with VAFs adjusted for genomic copy numbers. Long indels of >3 bp were excluded from the analysis because their VAFs were difficult to accurately estimate. The VAF for each sample estimated on the basis of blast percentage is indicated by a purple horizontal bar.

been reported in up to 13% of myelodysplastic syndromes and related chronic myeloid neoplasms[40]. Although rarely mutated in adult AML[41], *EZH2* represents one of the most frequently mutated and deleted genes in childhood AMKL, as we identified mutations or deletions in 16 of 49 DS-AMKL cases (33%) and in 3 of 19 non-DS-AMKL cases (16%) (**Fig. 4a,b**, **Supplementary Fig. 12c** and **Supplementary Tables 7** and **8**). No other PRC2 components were mutated, except for *SUZ12*, which was mutated in a single DS-AMKL case (**Fig. 4a** and **Supplementary Table 7**). Although frequent mutations in other epigenetic regulators, including in *TET2*, *IDH1* or *IDH2*, *DNMT3A* and *ASXL1*, are cardinal features of myeloid neoplasms in adults, we rarely found these mutations in DS-AMKL and non-DS-AMKL cases, only identifying occasional *DNMT3A* ($n = 1$), *ASXL1* ($n = 1$) and *BCOR* ($n = 2$) mutations in DS-AMKL (**Fig. 4a**).

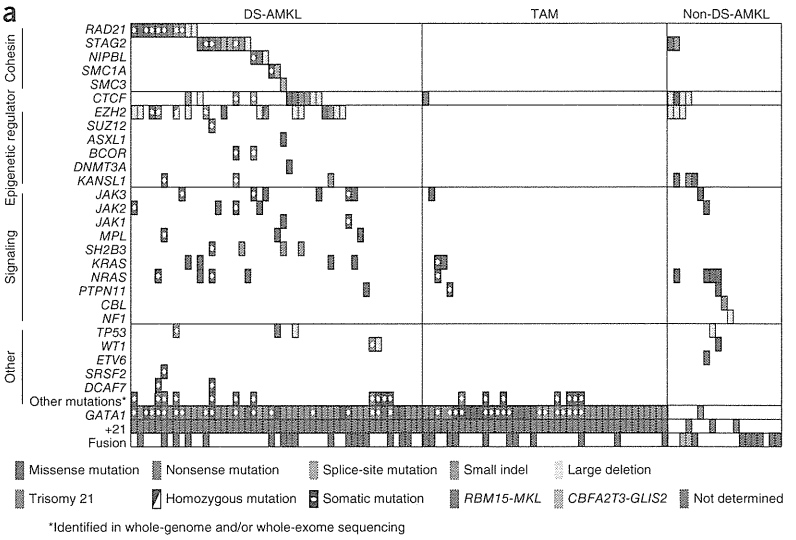
KANSL1 (encoding KAT8 regulatory NSL complex subunit 1; also known as MSL1V1 or NSL1) represents a new recurrent mutational target in human cancer (**Table 1**), although haploinsufficiency of *KANSL1* through germline deletions or mutations has been implicated in a congenital disease known as 17q21.31 microdeletion syndrome (MIM 610443)[42,43]. We found heterozygous mutations in *KANSL1* in three DS-AMKL and three non-DS-AMKL cases, and most of these mutations were nonsense or frameshifts, leading to loss of protein function (**Fig. 4a** and **Supplementary Table 7**). KANSL1 protein is necessary and sufficient for the activity of the KAT8 (MOF) histone acetyltransferase complex, which is engaged in the acetylation of histone H4 lysine 16 (H4K16), leading to transcriptional activation. Loss of acetylation of H4K16 has been reported to be a common hallmark of human cancer, and other histone acetyltransferases for H4K16 have been reported to form recurrent fusion partners in leukemia, including MOZ and MORF[44], suggesting a role for compromised H4K16 acetylation by *KANSL1* mutations in leukemogenesis. Of interest, KANSL1 is also responsible for the acetylation of the TP53 tumor suppressor that is important for TP53-dependent transcriptional activation[45]. KAT8 also interacts with a histone H3 lysine 4 (H3K4) methyltransferase, MLL, and the interaction of MLL and KAT8 complexes facilitates the cooperative recruitment of both complexes to gene promoters and enhances transcription initiation at target genes[45]. Thus, impaired TP53 function and/or deregulated expression of MLL gene targets could also contribute to leukemogenesis by *KANSL1* mutations.

Table 1 Recurrently mutated genes other than *GATA1* in DS-AMKL samples in whole-exome sequencing

Gene	Mutation type	RefSeq	Amino acid change	Nucleotide change	Sample (UPN) number
CTCF	Splice site	NM_006565	p.Gly318_splice	c.953-2A>G	016
CTCF	Frameshift	NM_006565	p.Asn314fs	c.940_941insAC	020
DCAF7	Missense	NM_005828	p.Leu340Phe	c.1018C>T	001
DCAF7	Missense	NM_005828	p.Leu340Phe	c.1018C>T	003
EZH2	Frameshift	NM_004456	p.710_716del	c.2129_2148delATCACAGGA TAGGTATTTTT	001
EZH2	Missense	NM_004456	p.Arg25Gln	c.74G>A	002
KANSL1	Frameshift	NM_001193466	p.Arg720fs	c.2159_2160insCG	020
KANSL1	Nonsense	NM_001193466	p.Arg462*	c.1384C>T	024
NRAS	Missense	NM_002524	p.Gly12Ser	c.34G>A	001
NRAS	Missense	NM_002524	p.Tyr64Cys	c.191A>G	001
NRAS	Missense	NM_002524	p.Gly12Ala	c.35G>C	003
RAD21	Nonsense	NM_006265	p.Arg139*	c.415A>T	001
RAD21	Frameshift	NM_006265	p.374_375del	c.1120_1124delTCTTT	002
RAD21	Missense	NM_006265	p.Leu611Arg	c.1832T>G	018
RAD21	Nonsense	NM_006265	p.Arg65*	c.193C>T	024
STAG2	Nonsense	NM_001042750	p.Arg604*	c.1810C>T	003
STAG2	Nonsense	NM_001042750	p.Arg216*	c.646C>T	019
STAG2	Frameshift	NM_001042750	p.Asn863fs	c.2588_2589insT	020
TP53	Nonsense	NM_000546	p.Glu68*	c.202G>T	002
TP53	Non-frameshift	NM_000546	p.157_162del	c.469_486delGTCCGCGCCA TGGCCATC	002

Other mutations in DS-AMKL

RAS pathway mutations are common in hematopoietic malignancies and other human cancers but have not to our knowledge been described in DS-AMKL. In the current cohort, we identified RAS pathway



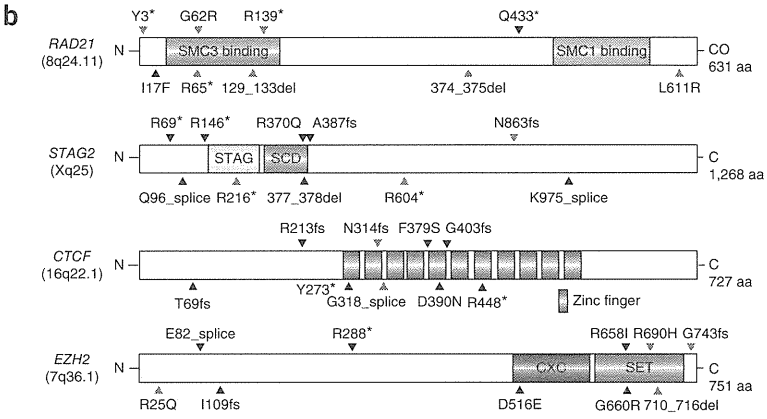


Figure 4 Driver mutations in Down syndrome–related myeloid disorders and non-DS-AMKL. (**a**) Driver mutations in 109 samples of 49 DS-AMKL, 41 TAM and 19 non-DS-AMKL cases. Types of mutations are distinguished by color. Each sample is also described in **Supplementary Table 12**. (**b**) Distribution of RAD21, STAG2, CTCF and EZH2 alterations. Alterations encoded by confirmed somatic mutations are indicated by red arrowheads.

mutations in the *NRAS*, *KRAS*, *PTPN11*, *NF1* and *CBL* genes in 8 DS-AMKL cases (16%) and 6 non-DS-AMKL cases (32%), but these mutations were rarely found in TAM cases (n = 3; 7%) (**Fig. 4a**). Tyrosine kinase and cytokine receptor mutations were also common in DS-AMKL. We found mutations in *JAK1*, *JAK2*, *JAK3*, *MPL* or *SH2B3* (*LNK*) in 17 DS-AMKL cases (35%) but rarely in TAM (n = 1) and non-DS-AMKL (n = 2) cases. We found no *FLT3* mutations in our cohort. The identified mutations were largely mutually exclusive. We found *JAK2* mutations in 4 DS-AMKL cases and 1 non-DS-AMKL case, including mutations encoding p.Val617Phe (n = 2), p.Leu611Ser (n = 1), p.Arg683Ser (n = 1) and p.Arg867Gln (n = 1); of these, *JAK2* mutations encoding p.Arg683Ser and p.Arg867Gln substitutions have been reported in acute lymphoblastic leukemia (ALL)[46,47] but not in myeloid malignancies[8,46]. Thus, we re-evaluated the diagnosis of AMKL in both UPN097 (p.Arg683Ser) and UPN023 (p.Arg867Gln), in whom the initial diagnosis of AMKL was strongly supported by typical surface marker expression of CD41, CD41b, CD117, CD13, CD33, CD34 and CD36 in UPN097 and of CD7, CD13, CD34, CD41a and CD42b in UPN023, together with characteristic cytomorphology. Similarly, the mutation encoding p.Leu611Ser was reported in both ALL[48] and polycythemia vera[49]. Thus, it seems that some *JAK2* mutations are involved in both myeloid and lymphoid leukemogenesis. As reported previously[10,11], *TP53* mutations were found in approximately 10% of DS-AMKL cases. Two identical somatic mutations found in the *DCAF7* gene (encoding p.Leu340Phe) might be interesting because the DCAF7 protein interacts with the DYRK1a kinase encoded within the Down syndrome critical region on chromosome 21 (ref. 50). DCAF7 has been shown to interact with DYRK1a through its N-terminal or C-terminal region, and the p.Leu340Phe substitution identified in our study was also located in the C-terminal domain. However, no additional mutation was detected in the extended cohort; therefore, the relevance of *DCAF7* remains to be determined.

Allelic burden of major recurrent mutations relative to *GATA1* mutations

We assessed intratumoral heterogeneity and the clonal origin of mutations by calculating the variant allele frequency (VAF) of each mutation relative to that of the *GATA1* mutation using deep sequencing. Mutations in cohesin components, *CTCF* and *EZH2* showed comparable VAFs to *GATA1* mutations (**Fig. 5b**), suggesting their role in the early stage of DS-AMKL development. In contrast, RAS pathway and other tyrosine kinases and cytokine receptor mutations showed significantly lower VAFs than corresponding *GATA1* mutations (P = 0.0001) (**Fig. 5b**), indicating that they are more likely to represent subclonal mutations, which were typically preceded by mutations in cohesin components, *CTCF* and *EZH2* and were involved in the evolution of multiple DS-AMKL subclones. Although RAS and JAK pathways activated by gene mutations represent potentially druggable targets and several promising compounds are currently available, this observation may largely preclude the efficient use of such compounds in eradicating founding DS-AMKL clones.

Distinct genetic features of Down syndrome– and non–Down syndrome–related AMKL

Despite their morphological similarities, both forms of AMKL in childhood are characterized by distinctive genetic features. According to the current study and a recent report of integrated analysis of non-DS-AMKL[22], *GATA1* mutations and trisomy 21 are less common in non-DS-AMKL than in DS-AMKL cases (**Fig. 4a** and **Supplementary Table 9**). In our series, DS-AMKL was characterized by high frequencies of mutations in the cohesin complex, *EZH2* and other epigenetic regulators, as well as in JAK family kinases, which were less

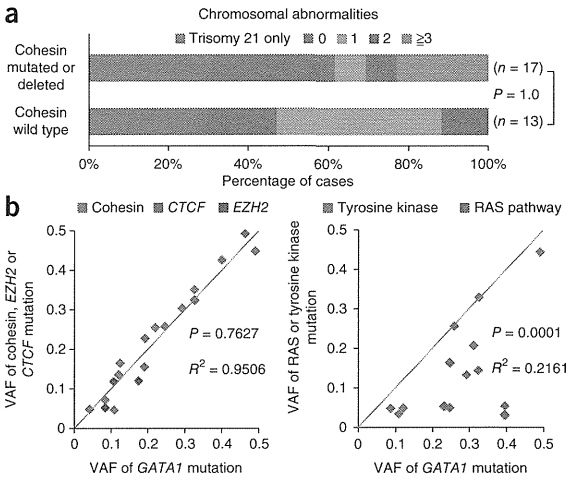


Figure 5 Relationship of cohesin mutations with karyotypes and comparison of mutation loads between major gene targets in DS-AMKL and *GATA1*. (**a**) The number of chromosomal abnormalities is compared between cases with and without cohesin mutations or deletions for DS-AMKL cases. Zero signifies chromosomal abnormalities without change in chromosome count, such as partial amplification or deletion of the chromosomal region or balanced translocation. (**b**) Diagonal plots of copy number–adjusted VAFs comparing coexisting *GATA1* and other pathway mutations, including cohesin, *CTCF*, *EZH2*, tyrosine kinase and the RAS pathway mutations, as indicated by color.

common mutational targets in non-DS-AMKL. Previous studies identified recurrent *CBFA2T3-GLIS2* and *RBM15-MKL* gene fusions in non-DS-AMKL, which were found in 27% and 15.2% of non-DS-AMKL cases, respectively[22,51], whereas these fusions were not detected in DS-AMKL cases in another report (n = 10 cases)[23]. Similarly, in the current cohort, RT-PCR analysis identified 2 *CBFA2T3-GLIS2* and 3 *RBM15-MKL* fusion genes in 19 non-DS-AMKL cases but not in TAM and DS-AMKL cases (**Fig. 4a** and **Supplementary Table 10**), illustrating the genetic differences between DS-AMKL and non-DS-AMKL. In addition, our RNA sequencing of the current cases (n = 17) (**Supplementary Table 11**) also showed no *CBFA2T3-GLIS2* and *RBM15-MKL* fusions.

DISCUSSION

Whole-genome and/or whole-exome analyses and follow-up targeted sequencing identified several new aspects of the pathogenesis of Down syndrome–related myeloid proliferation. First, the initial TAM phase was characterized by a paucity of somatic mutations. The mean number of non-silent mutations per sample (1.7; range of 1–5) was surprisingly small compared with that reported in other human cancers (**Supplementary Fig. 13**), in line with a recent report that identified 1.2 (range of 1–2) mutations per sample by whole-exome sequencing in 5 TAM samples[52]. In addition to reporting a low somatic mutation frequency in their initial TAM phase, Nikolaev *et al.*[52] also reported accumulation of somatic mutations (including single cases of *SMC3* and *EZH2* mutation) during progression from TAM to DS-AMKL. Excluding common *GATA1* mutations, we identified no other recurrent mutations, with only 0.7 non-silent mutations per case, indicating that TAM could be caused by a single acquired *GATA1* mutation in addition to constitutive trisomy 21.

Intratumoral heterogeneity was evident not only in the DS-AMKL phase but also at the initial diagnosis of TAM, and subsequent DS-AMKL originated from one of the multiple subclones present in the TAM phase, usually representing the progeny of the largest subpopulation. In most cases, the DS-AMKL clone was accompanied by newly acquired driver mutations not shared by the original TAM population, generating a unique landscape of gene mutations in DS-AMKL, which was characterized by high mutational frequencies in cohesin or *CTCF* (65%), other epigenetic regulators (45%), and RAS or signal-transducing molecules (47%) (**Fig. 4a**). Tumor recurrence or evolution has not to our knowledge been characterized by the distinct gene mutations in greater detail than in the present study. In total, 44 of the 49 DS-AMKL cases had additional mutations beyond those in *GATA1* (**Fig. 4a**), even though there was a clear limitation on capturing mutations using the targeted sequencing approach.

The very high frequency of cohesin (53%) and *EZH2* (33%) mutations and deletions in DS-AMKL but not in TAM or non-DS-AMKL cases was noteworthy because the reported mutation rates of cohesin and *EZH2* in adult AML and other human cancers remain approximately 10% (refs. 14,40,41), underscoring a major role for these mutations in the pathogenesis of DS-AMKL. The leukemogenic mechanism of mutated cohesin remains elusive, and frequent *CTCF* mutations also need further evaluation to characterize their possible cooperative role with cohesin mutations[26,30,33,34]. To our knowledge, *KANSL1* mutations have not been reported previously and represent a new recurrent mutational target in human cancer, although their functional impact on AMKL development remains unknown. Evaluation of the allelic burden of these mutations by deep sequencing disclosed a clonal hierarchy among different driver mutations in which clonal mutations in cohesin, *CTCF* and epigenetic regulators frequently preceded subclonal mutations in RAS and signal transduction molecules.

In conclusion, Down syndrome–related myeloid proliferation is shaped by multiple rounds of acquisition of new mutations and clonal selection, which are initiated by a *GATA1* mutation in the TAM phase and further driven by mutation in cohesin or *CTCF*, *EZH2* or other epigenetic regulators, and RAS or signal-transducing molecules, leading to AMKL. DS-AMKL and non-DS-AMKL showed similar phenotypes but had distinct genetic features, which may underlie their different clinical characteristics.

URLs. European Genome-phenome Archive (EGA), https://www.ebi.ac.uk/ega/; EBCall, https://github.com/friend1ws/EBCall; Catalogue of Somatic Mutations in Cancer (COSMIC), http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/; PubMed, http://www.ncbi.nlm.nih.gov/pubmed/; UCSC Genome Browser, http://genome.ucsc.edu/; Integrative Genomics Viewer, http://www.broadinstitute.org/igv/; DNACopy, http://biostatistics.oxfordjournals.org/content/5/4/557.full.pdf; Genomon-fusion (in Japanese), http://genomon.hgc.jp/rna/.

METHODS

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

Accession codes. Sequencing data have been deposited in the European Genome-phenome Archive (EGA) under accession EGAS00001000546.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank Y. Mori, M. Nakamura, O. Hagiwara and N. Mizota for their technical assistance. This work was supported by the Research on Measures for Intractable

Diseases Project and Health and Labor Sciences Research grants (Research on Intractable Diseases) from the Ministry of Health, Labour and Welfare, by Grants-in-Aid from the Ministry of Health, Labor and Welfare of Japan and KAKENHI (22134006, 23249052, 23118501, 23390266 and 25461579) and by the Japan Society for the Promotion of Science (JSPS) through the Funding Program for World-Leading Innovative Research and Development on Science and Technology (FIRST Program), initiated by the Council for Science and Technology Policy (CSTP) and research grants from the Japan Science and Technology Agency CREST.

AUTHOR CONTRIBUTIONS

Y.O., Y. Shiraishi, A.S.-O., K.C., H.T. and S.M. performed bioinformatics analyses of the resequencing data. M.S., A.S.-O., Y. Sato, A.H. and H.M. performed microarray experiments and analyses. R.K. and A.H. performed RT-PCR analyses. M.P., K. Terui, R.W., D.H., K.N., H.K., K. Tsukamoto, S.A., K. Kawakami, K. Kato, R.N., S.I., Y.H., S.K. and E.I. collected specimens and were involved in planning the project. K.Y., T.T., H.S., Y.N. and N.S. processed and analyzed genetic materials, prepared the library and performed sequencing. K.Y., T.T., Y.O., A.K. and S.O. generated figures and tables. E.I. and S.O. led the entire project. K.Y. and S.O. wrote the manuscript. All authors participated in discussions and interpretation of the data and results.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Khan, I., Malinge, S. & Crispino, J. Myeloid leukemia in Down syndrome. *Crit. Rev. Oncog.* **16**, 25–36 (2011).
2. Massey, G.V. *et al.* A prospective study of the natural history of transient leukemia (TL) in neonates with Down syndrome (DS): Children's Oncology Group (COG) study POG-9481. *Blood* **107**, 4606–4613 (2006).
3. Muramatsu, H. *et al.* Risk factors for early death in neonates with Down syndrome and transient leukaemia. *Br. J. Haematol.* **142**, 610–615 (2008).
4. Klusmann, J.H. *et al.* Treatment and prognostic impact of transient leukemia in neonates with Down syndrome. *Blood* **111**, 2991–2998 (2008).
5. Xu, G. *et al.* Frequent mutations in the *GATA-1* gene in the transient myeloproliferative disorder of Down syndrome. *Blood* **102**, 2960–2968 (2003).
6. Wechsler, J. *et al.* Acquired mutations in *GATA1* in the megakaryoblastic leukemia of Down syndrome. *Nat. Genet.* **32**, 148–152 (2002).
7. Walters, D.K. *et al.* Activating alleles of *JAK3* in acute megakaryoblastic leukemia. *Cancer Cell* **10**, 65–75 (2006).
8. Malinge, S. *et al.* Activating mutations in human acute megakaryoblastic leukemia. *Blood* **112**, 4220–4226 (2008).
9. Blink, M. *et al.* Frequency and prognostic implications of *JAK 1–3* aberrations in Down syndrome acute lymphoblastic and myeloid leukemia. *Leukemia* **25**, 1365–1368 (2011).
10. Hama, A. *et al.* Molecular lesions in childhood and adult acute megakaryoblastic leukaemia. *Br. J. Haematol.* **156**, 316–325 (2012).
11. Malkin, D., Brown, E.J. & Zipursky, A. The role of p53 in megakaryocyte differentiation and the megakaryocytic leukemias of Down syndrome. *Cancer Genet. Cytogenet.* **116**, 1–5 (2000).
12. Hussein, K. *et al.* MPL^{W515L} mutation in acute megakaryoblastic leukaemia. *Leukemia* **23**, 852–855 (2009).
13. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
14. Welch, J.S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
15. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
16. Creutzig, U. *et al.* Diagnosis and management of acute myeloid leukemia in children and adolescents: recommendations from an international expert panel. *Blood* **120**, 3187–3205 (2012).
17. Swerdlow, S.H., Jaffe, E.S. & International Agency for Research on Cancer & World Health Organization *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues* (International Agency for Research on Cancer, Lyon, France, 2008).
18. Wu, C. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* **10**, R130 (2009).
19. Bourquin, J.P. *et al.* Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling. *Proc. Natl. Acad. Sci. USA* **103**, 3339–3344 (2006).
20. Mercher, T. *et al.* Involvement of a human gene related to the *Drosophila spen* gene in the recurrent t(1;22) translocation of acute megakaryocytic leukemia. *Proc. Natl. Acad. Sci. USA* **98**, 5776–5779 (2001).
21. Ma, Z. *et al.* Fusion of two novel genes, *RBM15* and *MKL1*, in the t(1;22)(p13;q13) of acute megakaryoblastic leukemia. *Nat. Genet.* **28**, 220–221 (2001).
22. Gruber, T.A. *et al.* An *inv*(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. *Cancer Cell* **22**, 683–697 (2012).
23. Thiollier, C. *et al.* Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models. *J. Exp. Med.* **209**, 2017–2031 (2012).
24. Gruber, S., Haering, C.H. & Nasmyth, K. Chromosomal cohesin forms a ring. *Cell* **112**, 765–777 (2003).
25. Nasmyth, K. & Haering, C.H. Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* **43**, 525–558 (2009).
26. Wendt, K.S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
27. Ström, L. *et al.* Postreplicative formation of cohesion is required for repair and induced by a single DNA break. *Science* **317**, 242–245 (2007).
28. Watrin, E. & Peters, J.M. The cohesin complex is required for the DNA damage–induced G2/M checkpoint in mammalian cells. *EMBO J.* **28**, 2625–2635 (2009).
29. Dorsett, D. *et al.* Effects of sister chromatid cohesion proteins on *cut* gene expression during wing development in *Drosophila*. *Development* **132**, 4743–4753 (2005).
30. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422–433 (2008).
31. Solomon, D.A. *et al.* Mutational inactivation of *STAG2* causes aneuploidy in human cancer. *Science* **333**, 1039–1043 (2011).
32. Forestier, E. *et al.* Cytogenetic features of acute lymphoblastic and myeloid leukemias in pediatric patients with Down syndrome: an iBFM-SG study. *Blood* **111**, 1575–1583 (2008).
33. Rubio, E.D. *et al.* CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. USA* **105**, 8309–8314 (2008).
34. Stedman, W. *et al.* Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J.* **27**, 654–666 (2008).
35. Ohlsson, R., Bartkuhn, M. & Renkawitz, R. CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. *Chromosoma* **119**, 351–360 (2010).
36. Phillips, J.E. & Corces, V.G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
37. Wendt, K.S. & Peters, J.M. How cohesin and CTCF cooperate in regulating gene expression. *Chromosome Res.* **17**, 201–214 (2009).
38. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
39. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
40. Ernst, T. *et al.* Inactivating mutations of the histone methyltransferase gene *EZH2* in myeloid disorders. *Nat. Genet.* **42**, 722–726 (2010).
41. Patel, J.P. *et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N. Engl. J. Med.* **366**, 1079–1089 (2012).
42. Koolen, D.A. *et al.* Mutations in the chromatin modifier gene *KANSL1* cause the 17q21.31 microdeletion syndrome. *Nat. Genet.* **44**, 639–641 (2012).
43. Zollino, M. *et al.* Mutations in *KANSL1* cause the 17q21.31 microdeletion syndrome phenotype. *Nat. Genet.* **44**, 636–638 (2012).
44. Yang, X.J. The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Res.* **32**, 959–976 (2004).
45. Li, X., Wu, L., Corsa, C.A., Kunkel, S. & Dou, Y. Two mammalian MOF complexes regulate transcription activation by distinct mechanisms. *Mol. Cell* **36**, 290–301 (2009).
46. Bercovich, D. *et al.* Mutations of *JAK2* in acute lymphoblastic leukaemias associated with Down's syndrome. *Lancet* **372**, 1484–1492 (2008).
47. Mullighan, C.G. *et al.* JAK mutations in high-risk childhood acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. USA* **106**, 9414–9418 (2009).
48. Kratz, C.P. *et al.* Mutational screen reveals a novel JAK2 mutation, L611S, in a child with acute lymphoblastic leukemia. *Leukemia* **20**, 381–383 (2006).
49. Nussenzveig, R.H. *et al.* Detection of *JAK2* mutations in paraffin marrow biopsies by high resolution melting analysis: identification of L611S alone and in *cis* with V617F in polycythemia vera. *Leuk. Lymphoma* **53**, 2479–2486 (2012).
50. Miyata, Y. & Nishida, E. DYRK1A binds to an evolutionarily conserved WD40-repeat protein WDR68 and induces its nuclear translocation. *Biochim. Biophys. Acta* **1813**, 1728–1739 (2011).
51. de Rooij, J.D. *et al.* *NUP98/JARID1A* is a novel recurrent abnormality in pediatric acute megakaryoblastic leukemia with a distinct *HOX* gene expression pattern. *Leukemia* doi:10.1038/leu.2013.87 (27 March 2013).
52. Nikolaev, S.I. *et al.* Exome sequencing identifies putative drivers of progression of transient myeloproliferative disorder to AMKL in infants with Down Syndrome. *Blood* **122**, 554–561 (2013).
53. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

ONLINE METHODS

Subjects and samples. Genomic DNA from 84 individuals with Down syndrome–related myeloid disorders (41 samples from the TAM phase and 49 from the AMKL phase) and 19 with non-DS-AMKL were analyzed by whole-genome and/or whole-exome and/or targeted deep sequencing. In six cases with Down syndrome–related myeloid disorders, samples were collected from both the TAM and AMKL phases. RNA sequencing was also performed for 12 of the 49 DS-AMKL cases and for 5 additional DS-AMKL cases. RNA samples were also available for RT-PCR analysis from 30 cases with TAM, 32 cases with DS-AMKL and 15 cases with non-DS-AMKL. Written informed consent was obtained from each subject's parents before sample collection (**Supplementary Note**). This study was approved by the Ethics Committees of the University of Tokyo according to the Helsinki convention. *GATA1* mutations were detected by Sanger sequencing of all TAM and DS-AMKL samples according to the previously described procedure[5]. Detailed information on subjects and samples is provided in **Supplementary Tables 1, 4, 11** and **12**. Tumor DNA was extracted from bone marrow– or peripheral blood–derived mononuclear cells at diagnosis. Genomic DNA samples from peripheral blood from subjects in remission or from nail tissues at diagnosis were used as germline controls. Genomic DNA was extracted using a QIAamp DNA Blood Mini kit and a QIAamp DNA Investigator kit (Qiagen). Total RNA was extracted using the RNeasy kit (Qiagen) with RNase-free DNase (Qiagen).

Whole-genome sequencing. DNA samples were processed for whole-exome sequencing using NEBNext DNA sample Prep Reagent (New England Biolabs) according to the modified Illumina protocol. Sequence data were generated on the Illumina HiSeq 2000 platform in 100-bp paired-end reads. Data processing and variant calling were performed as described previously[54]. All candidate variants were validated by deep sequencing.

Validation and quantitative measurements of the frequencies of mutant alleles by deep sequencing. Individual mutation sites were amplified by genomic PCR using primers tagged with NotI cleavage sites and subjected to high-throughput sequencing as described previously[55], except that target DNA was not pooled. Deep sequencing was performed using the MiSeq or HiSeq 2000 platform. Data processing was performed according to the previously described method with minor modifications[55]. Briefly, each read was aligned to a set of PCR-amplified target sequences using BLAT[56], and dichotomic variant alleles were differentially enumerated. For indels, individual reads were first aligned to each of the wild-type and indel sequences and then assigned to the one to which better alignment was obtained in terms of the number of matched bases. Each SNV and indel whose VAF in the tumor sample was equal to or greater than 2.0% and significantly higher than the frequency in the germline sample was adopted as a somatic mutation. The error size for estimated VAFs was evaluated by assuming binomial distributions in deep sequencing, which were confirmed by observed allele frequencies at heterozygous SNPs in normal DNA samples (**Supplementary Fig. 14a**), in which the variance (σ^2) ranged from $4.0–11.0 \times 10^{-4}$ (**Supplementary Fig. 14b**).

Clustering analysis of mutations. To identify the chronological behavior of the structure of the tumor subpopulation for the TAM and AMKL phases, somatic mutations detected in both phases by whole-genome sequencing were clustered according to their VAFs as measured by deep sequencing. Copy number–adjusted deep sequencing data, in which the VAFs of genes on the X chromosome in male cases or in regions of uniparental disomy were halved, were subjected to unsupervised clustering. Six mutations located in amplified or deleted genomic regions were excluded from the analysis. Long indels of >3 bp, except for those affecting key genes such as *GATA1* and *RAD21*, and mutations in repetitive regions were excluded from the analysis because their VAFs could tend to be underestimated.

All validated mutations were grouped into three categories according to the following criteria: (i) mutations found only in TAM (VAF in AMKL < 0.02), (ii) mutations found only in AMKL (VAF in TAM < 0.02) and (iii) mutations found in both TAM and AMKL (VAF in TAM > 0.02 and VAF in AMKL > 0.02). Clustering of mutations in each category was performed using Mclust, provided as an R package, on the basis of the VAFs of the mutations in the TAM and AMKL phases, where one-dimensional clustering of mutations in categories (i) and (ii) was performed on the basis of the homoscedastic model and two-dimensional clustering was performed for mutations in category (iii) on the basis of the ellipsoidal model. The most appropriate number of clusters was determined by using the Bayesian information criterion (BIC) score. Singleton points identified by this algorithm were regarded as outliers. Clonal subpopulations within tumors were also evaluated by kernel density analysis (**Supplementary Fig. 5**), where we drew kernel density estimate plots for the VAFs of validated variants using the density function in R.

Whole-exome sequencing and detection of somatic mutations. Exome capture was performed using SureSelect Human All Exon V3 or V4 (Agilent Technologies) or the TruSeq Exome Enrichment kit (Illumina). Enriched exome fragments were then subjected to massively parallel sequencing using the Genome Analyzer IIx or HiSeq 2000 platform (Illumina). Candidate somatic mutations were detected using our in-house pipeline EBCall (Empirical Bayesian mutation Calling; see URLs)[57]. All candidates were validated by Sanger sequencing or independent deep sequencing.

PCR-based targeted deep sequencing. Deep sequencing of *DCAF7*, *EED*, *JAK1*, *JAK3*, *KANSL1*, *SH2B3*, and *SUZ12* was performed using the primers tagged with NotI cleavage sites whose sequences are listed in **Supplementary Table 6**. Data processing and variant calling were performed as described previously[58]. All candidate variants were validated by Sanger sequencing or independent deep sequencing using non-amplified DNA.

Targeted deep sequencing. In total, 39 gene targets were exhaustively examined for mutations in all 109 cases using deep sequencing (**Supplementary Table 5**). Genomic DNA (1–1.5 μg) from bone marrow–derived mononuclear cells or peripheral blood was enriched for target exons using a SureSelect custom kit (Agilent Technologies) designed to capture all of the coding exons from the 39 target genes, and high-throughput sequencing was performed on the enriched targets using the HiSeq 2000 platform with a standard 100-bp paired-end read protocol. Sequencing reads were aligned to hg19 using Burrows-Wheeler Aligner (BWA) version 0.5.8 with default parameters. The allele frequencies of SNVs and indels were calculated at each genomic position by enumerating the relevant reads with SAMtools[59]. Initially, all variants showing VAF > 0.02 were extracted and annotated using ANNOVAR[60] for further consideration if they were found in >6 reads out of >10 total reads and appeared in both plus- and minus-strand reads. For the cases for which no germline DNA was available, relevant somatic mutations were called by eliminating the following entries, unless they were registered in the Catalogue of Somatic Mutations in Cancer (COSMIC) v60 (ref. 61) or reported as somatic mutations in PubMed: (i) synonymous variants and those having ambiguous (unknown) annotations, (ii) known SNPs in public and private databases, including dbSNP131, the 1000 Genomes Project as of 23 November 2010 and our in-house database, (iii) sequencing or mapping errors, (iv) all missense SNVs with allele frequencies of 0.45–0.55 and (v) variants localized to duplicated regions found in SegDups of the UCSC Genome Browser. To eliminate sequencing errors in category (iii), we excluded all variants found in 31 normal Japanese samples at, on average, allele frequency > 0.25. Mapping errors were removed by visual inspection with the Integrative Genomics Viewer browser[62]. All candidate variants were validated by Sanger sequencing or independent deep sequencing.

Calculation of copy numbers for target exons. Letting $d_j^{i,s}$ be the sequencing depth at the ith nucleotide of the jth exon in sample s, the standardized depth of the jth exon is calculated as

$$D_j^s = k_s \sum_i d_i^{j,s}$$

where k_s is determined to satisfy

$$k_0 = \sum_j D_j^s$$

for a fixed constant k_0 (for example, $k_0 = 1$). The correlation coefficient ($R = R^{s,t}$) between two vectors D_i^s and D_i^t was calculated, where D_i^s and D_i^t represent the depth for a given sample (sample s) and each of the 443

samples (sample t), analyzed for other projects, with completely normal copy numbers in array-comparative genomic hybridization (aCGH; $t = 1, 2, 3, \ldots, 443$), respectively, through which a total of m_0 (= 12) control samples showing the largest R values were selected (T_m; $m = 1, 2, 3, \ldots, m_0$) and used for copy number calculation. The copy number of the ith target exon of sample s (Cn_i^s) was calculated as

$$Cn_i^s = D_i^s / \hat{D}_i^s$$

where $\hat{D}_i^s$ was calculated by averaging m_0 samples by

$$\hat{D}_i^s = \sum_{m=1}^{m_0} D_i^{T_m} / m_0$$

Copy numbers were calculated for exons with mean depth of >500. Circular binary segmentation was also used to identify discrete copy number segments using DNACopy (see URLs); segmented copy number ($\widehat{Cn_i^s}$) was defined for the ith exon of sample s. The distribution of $\widehat{Cn_i^s}$ was calculated for all samples, and exons showing $|\widehat{Cn_i^s} - E(\widehat{Cn_i^s})| > 4$ s.d. were considered to have copy number losses or gains.

Screening for *CBFA2T3-GLIS2* and *RBM15-MKL1* fusion genes. *CBFA2T3-GLIS2* and *RBM15-MKL1* fusion genes were screened by RT-PCR[22,63]. Primer sequences are given in **Supplementary Table 13**. PCR amplification was performed by 40 cycles at 94 °C for 2 min, 60 °C for 30 s and 68 °C for 1 min, followed by denaturation at 94 °C for 2 min and extension at 68 °C for 7 min.

SNP array analyses. All tumor samples subjected to whole-exome sequencing were also analyzed for copy number alterations using SNP arrays (Affymetrix GeneChip Human Mapping 250K NspI Array or Genome-Wide Human SNP Array 6.0) as described previously[10,64,65].

RT-PCR analysis of *STAG2* and *CTCF* transcripts. To confirm abnormal splicing of *CTCF* in UPN016 and UPN071 and that of *STAG2* in UPN067, RT-PCR were performed using cDNA derived from each subject, with cDNA from CMK11-5 (DS-AMKL-derived cell line with no known mutations in both genes) used as a control (**Supplementary Fig. 11**). Primer sequences are given in **Supplementary Table 14**. Total RNA (1 μg) was subjected to reverse transcription using M-MLV reverse transcriptase (Invitrogen) according to the manufacturer's instructions. Electrophoresis was performed using Experion (Bio-Rad).

RNA sequencing. Detailed information on samples is provided in **Supplementary Table 11**. Library preparation and sequencing were performed as described previously[54]. Fusion transcripts were detected using Genomon-fusion.

Gene expression analysis of recurrently mutated genes. Expression data for the recurrently mutated genes in whole-exome sequencing were retrieved from the BioGPS database[18] for normal hematopoietic cells, including whole bone marrow, $CD33^+$ myeloid cells, $CD34^+$ cells, $CD19^+$ B cells and $CD4^+$ T cells, and from published data[19] and our RNA sequencing data for DS-AMKL samples.

Statistical analysis. The number of non-silent mutations identified by whole-exome sequencing in TAM and DS-AMKL samples (**Fig. 2a**) and the number of chromosome abnormalities in DS-AMKL cases with and without cohesin mutations or deletions (**Fig. 5a**) were compared using the Mann-Whitney U test. The difference in VAF between two mutations (**Fig. 5b**) was tested by Wilcoxon signed-rank test.

54. Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* **45**, 860–867 (2013).
55. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
56. Kent, W.J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
57. Shiraishi, Y. *et al.* An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. *Nucleic Acids Res.* **41**, e89 (2013).
58. Sakaguchi, H. *et al.* Exome sequencing identifies secondary mutations of *SETBP1* and *JAK3* in juvenile myelomonocytic leukemia. *Nat. Genet.* **45**, 937–941 (2013).
59. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
60. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
61. Forbes, S.A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
62. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
63. Torres, L. *et al.* Acute megakaryoblastic leukemia with a four-way variant translocation originating the *RBM15-MKL1* fusion gene. *Pediatr. Blood Cancer* **56**, 846–849 (2011).
64. 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).
65. Yamamoto, G. *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of Affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am. J. Hum. Genet.* **81**, 114–126 (2007).

Corrigendum: The landscape of somatic mutations in Down syndrome–related myeloid disorders

Kenichi Yoshida, Tsutomu Toki, Yusuke Okuno, Rika Kanezaki, Yuichi Shiraishi, Aiko Sato-Otsubo, Masashi Sanada, Myoung-ja Park, Kiminori Terui, Hiromichi Suzuki, Ayana Kon, Yasunobu Nagata, Yusuke Sato, RuNan Wang, Norio Shiba, Kenichi Chiba, Hiroko Tanaka, Asahito Hama, Hideki Muramatsu, Daisuke Hasegawa, Kazuhiro Nakamura, Hirokazu Kanegane, Keiko Tsukamoto, Souichi Adachi, Kiyoshi Kawakami, Koji Kato, Ryosei Nishimura, Shai Izraeli, Yasuhide Hayashi, Satoru Miyano, Seiji Kojima, Etsuro Ito & Seishi Ogawa

Nat. Genet. **45, 1293–1299 (2013); published online 22 September 2013; corrected after print 30 October 2013**

In the version of this article initially published, the discussion of cited reference 52 should also have noted that the work "reported accumulation of additional somatic mutations (including single cases of *SMC3* and *EZH2* mutation) during progression from TAM to DS-AMKL." The error has been corrected in the HTML and PDF versions of the article.

GENES, CHROMOSOMES & CANCER 52:683–693 (2013)

NUP98-NSD1 Gene Fusion and Its Related Gene Expression Signature Are Strongly Associated with a Poor Prognosis in Pediatric Acute Myeloid Leukemia

Norio Shiba,[1,2] Hitoshi Ichikawa,[3] Tomohiko Taki,[4] Myoung-Ja Park,[1] Aoi Jo,[3] Sachiyo Mitani,[3] Tohru Kobayashi,[2] Akira Shimada,[5] Manabu Sotomatsu,[1] Hirokazu Arakawa,[2] Souichi Adachi,[6] Akio Tawa,[7] Keizo Horibe,[8] Masahiro Tsuchida,[9] Ryoji Hanada,[10] Ichiro Tsukimoto,[11] and Yasuhide Hayashi[1]*

[1]Department of Hematology/Oncology, Gunma Children's Medical Center, Shibukawa, Japan
[2]Department of Pediatrics, Gunma University Graduate School of Medicine, Maebashi, Japan
[3]Division of Genetics, National Cancer Center Research Institute, Tokyo, Japan
[4]Department of Molecular Diagnostics and Therapeutics, Kyoto Prefectural University of Medicine Graduate School of Medical Science, Kyoto, Japan
[5]Department of Pediatrics, Okayama University Graduate School of Medicine, Okayama, Japan
[6]Department of Human Health Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan
[7]Department of Pediatrics, National Hospital Organization Osaka National Hospital, Osaka, Japan
[8]Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Japan
[9]Department of Pediatrics, Ibaraki Children's Hospital, Ibaraki, Japan
[10]Division of Hematology/Oncology, Saitama Children's Medical Center, Saitama, Japan
[11]Department of First Pediatrics, Toho University School of Medicine, Tokyo, Japan

The cryptic t(5;11)(q35;p15.5) creates a fusion gene between the *NUP98* and *NSD1* genes. To ascertain the significance of this gene fusion, we explored its frequency, clinical impact, and gene expression pattern using DNA microarray in pediatric acute myeloid leukemia (AML) patients. *NUP98-NSD1* fusion transcripts were detected in 6 (4.8%) of 124 pediatric AML patients. Supervised hierarchical clustering analyses using probe sets that were differentially expressed in these patients detected a characteristic gene expression pattern, including 18 *NUP98-NSD1*-negative patients (*NUP98-NSD1*-like patients). In total, a *NUP98-NSD1*-related gene expression signature (*NUP98-NSD1* signature) was found in 19% (24/124) and in 58% (15/26) of cytogenetically normal cases. Their 4-year overall survival (OS) and event-free survival (EFS) were poor (33.3% in *NUP98-NSD1*-positive and 38.9% in *NUP98-NSD1*-like patients) compared with 100 *NUP98-NSD1* signature-negative patients (4-year OS: 86.0%, 4-year EFS: 72.0%). Interestingly, t(7;11)(p15;p15)/*NUP98-HOXA13*, t(6;11)(q27;q23)/MLL-*MLLT4* and t(6;9)(p22;q34)/*DEK-NUP214*, which are known as poor prognostic markers, were found in *NUP98-NSD1*-like patients. Furthermore, another type of *NUP98-NSD1* fusion transcript was identified by additional RT-PCR analyses using other primers in a *NUP98-NSD1*-like patient, revealing the significance of this signature to detect *NUP98-NSD1* gene fusions and to identify a new poor prognostic subgroup in AML. © 2013 Wiley Periodicals, Inc.

INTRODUCTION

Acute myeloid leukemia (AML) is a complex disease caused by mutations, epigenetic modifications, and deregulated expression of genes, leading to increased proliferation and decreased differentiation of hematopoietic progenitor cells (Frohling et al., 2005; Marcucci et al., 2011; Pui et al., 2011). Several important molecular markers have been discovered in AML that have not only helped to characterize better patients, but also to improve risk stratification (Marcucci et al., 2011; Pui et al., 2011). However, in a subset of AML patients, no prognosis-associated cytogenetic aberrations or mutations are known (Frohling et al., 2005; Marcucci et al., 2011; Pui et al., 2011). In

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

Supported by: a grant for Cancer Research, a grant for Research on Children and Families, and Research on Intractable Diseases, Health and Labor Sciences Research Grants from the Ministry of Health, Labor, and Welfare of Japan, Grants-in-Aid for Scientific Research (B, C) and Exploratory Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan, the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio) of Japan, and a Research grant for Gunma Prefectural Hospitals.

*Correspondence to: Yasuhide Hayashi, Department of Hematology/Oncology, Gunma Children's Medical Center, 779, Shimohakoda, Hokkitsu, Shibukawa, 377-8577 Gunma, Japan. E-mail: hayashiy-tky@umin.ac.jp

Received 21 December 2012; Accepted 15 March 2013

DOI 10.1002/gcc.22064

Published online 30 April 2013 in Wiley Online Library (wileyonlinelibrary.com).

hematological malignancies, 11p15 translocations involving the nucleoporin 98-kDa (NUP98) protein gene are relatively rare. This notwithstanding, more than 20 different chromosomal rearrangements have been identified (Romana et al., 2006). These translocations fuse *NUP98* with respective partner genes, including many homeobox genes and nuclear non-homeobox genes (Arai et al., 1997; Taketani et al., 2002a,b,c). Although the *NUP98* fusion genes are rare, they have provided valuable information regarding the role of homeobox proteins in leukemogenesis (Nakamura et al., 1996; Nakamura, 2005).

The cryptic t(5;11)(q35;p15.5), which is frequently accompanied by deletion of the long arm of chromosome 5, del(5q), creates a fusion gene between *NUP98* and the nuclear receptor-binding SET-domain protein 1 (*NSD1*) gene (Jaju et al., 2001). This fusion gene has mainly been identified in pediatric AML, by the use of fluorescence in situ hybridization (FISH) with subtelomeric probes (Jaju et al., 2001; Brown et al., 2002; Panarello et al., 2002; Cerveira et al., 2003), it is rare in adult AML (Brown et al., 2002; Casas et al., 2003; Nebral et al., 2005; Walter et al., 2009). Recently, the *NUP98-NSD1* gene fusion, identified by high-resolution genome-wide copy number analysis and reverse transcription (RT)-PCR in pediatric and adult AML patients, was associated with poor prognosis (Hollink et al., 2011). To increase our understanding of this gene fusion, we explored the frequency, clinical significance, and gene expression pattern of *NUP98-NSD1* using DNA microarray in pediatric AML patients.

MATERIALS AND METHODS

Patients and Samples

From January 2000 to December 2002, 318 patients were diagnosed with de novo AML. The diagnosis of AML was based on the French-American-British (FAB) classification, and cytogenetic analysis was performed using conventional G-banding. Of these patients, samples from 124 patients with known mutation status and gene expression profiling data were available, including 10 patients with FAB-M3 and six patients with Down syndrome who were treated on different treatment protocols (Kudo et al., 2007; Tsukimoto et al., 2009; Imaizumi et al., 2011). Age and initial white blood cell (WBC) count were higher, patients with t(8;21) were more frequent, and M7 patients with Down syndrome were fewer in the present study cohort than in the non-analyzed patients (Supporting Information Table S1). There were no significant differences in survival between these two groups [4-year overall survival (OS): 76.6% versus 79.9% and 4-year event-free survival (EFS): 65.3% versus 71.1%]. Informed consent was obtained from the patients or the patients' parents, according to guidelines based on the tenets of the revised Helsinki protocol. The institutional review boards of Gunma Children's Medical Center and National Cancer Center approved this project.

Reverse Transcription-PCR and Sequence Analysis

Total RNA extracted from leukemic cells at diagnosis was reverse transcribed to cDNA with a cDNA Synthesis Kit (GE Healthcare, Tokyo, Japan). PCR was performed with AmpliTaq Gold DNA polymerase (Applied Biosystems, Branchburg, NJ), using a DNA thermal cycler (Applied Biosystems). For the detection of *NUP98-NSD1* and the reciprocal fusion transcript, *NUP98*−5F and *NSD1*−1R, and *NSD1*−2F and *NUP98*−6R were used, respectively (Supporting Information Table S2) (Jaju et al., 2001; Brown et al., 2002). PCR conditions were as follows: initial denaturation at 94°C (9 min), 40 cycles of 96°C (45 sec), 58°C (45 sec), and 72°C (1 min), followed by final elongation at 72°C (7 min). For sequencing, PCR products were amplified using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) under the following conditions: 95°C (2 min) followed by 25 cycles of 95°C (10 sec), 50°C (5 sec), and 60°C (4 min). Direct sequencing was performed using an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). To confirm and identify the gene fusions precisely, we used various primer sets for RT-PCR followed by direct sequencing (Supporting Information Table S2). Mutations of *NPM1* were also examined as previously reported (Döhner et al., 2005). Mutation analyses of the *DNMT3A*, *FLT3*, *MLL*, *KIT*, *NRAS*, *KRAS*, and *WT1* genes have been reported previously (Shimada et al., 2006, 2008; Sano et al., 2012; Shiba et al., 2012).

Microarray Analysis

Gene expression profiling data for the 124 patients (Gene Expression Omnibus accession number, GSE35784) were obtained and analyzed as follows. Total RNA was re-purified using the RNeasy MinElute cleanup kit (Qiagen, Hilden,

Germany), and the integrity of the purified RNA was confirmed using a 2100 Bioanalyzer and an RNA 6000 Nano LabChip kit (Agilent Technologies, Santa Clara, CA). The DNA microarray used was a Human Genome U133 plus 2.0 array (Affymetrix, Santa Clara, CA). Target cRNA was prepared from 20 ng purified RNA with a two-cycle cDNA synthesis kit and 3′-amplification reagents for IVT labeling (Affymetrix). Hybridization to the microarrays, washing, and staining with the antibody amplification procedure and scanning were performed according to the manufacturer's instructions. Using the GeneChip Operating Software version 1.4 (Affymetrix), the scanned image data were processed and the expression value (signal) of each probe set was calculated. The signal values were normalized so that the mean in each experiment was set at 100 to adjust for minor differences between the experiments. Statistical analyses and fold change calculations were performed using expression values that were log-transformed after addition of 10 to reduce adverse effects caused by noise at low expression levels (Ichikawa et al., 2006; Jo et al., 2009). To identify differentially expressed probe sets between *NUP98-NSD1*-positive and -negative patients, *P*-values in Student's *t*-test and fold change values were used. Hierarchical clustering analysis was performed using the Cluster and Tree View software (Eisen et al., 1998). For this analysis, log-transformed expression values were normalized for each probe set by subtracting the mean, and uncentered correlation metric and complete linkage clustering methods were used. Unsupervised clustering analyses was performed and visualized as previously described (Ichikawa et al., 2006).

Statistical Analysis

All analyses were carried out using the SPSS statistical package program (version 18.0J; SPSS Tokyo, Japan). Survival distributions were assessed using the Kaplan-Meier method and the differences were compared using the log-rank test. EFS and OS were defined as the times from diagnosis to event (relapse or death of any cause) and from diagnosis to death from any cause, respectively. Statistical analyses were performed using Fisher's exact test for categorical variables and Mann-Whitney's U test for continuous variables. For all analyses, the *P*-values were two-tailed, and $P < 0.05$ was considered significant. To identify independent predictors of poor prognosis, multivariate Cox regression analysis was performed using factors that were significant in frequencies between *NUP98-NSD1* signature positive and negative cases.

RESULTS

Detection of *NUP98-NSD1*-Positive Patients

We identified the *NUP98-NSD1* fusion transcript in 6 (4.8%) out of 124 Japanese pediatric AML patients. The reciprocal fusion transcript *NSD1-NUP98* was detected in five of the patients (Supporting Information Fig. S1A). Sequence analysis of the PCR products confirmed that *NUP98* and *NSD1* were fused in-frame, joining *NUP98* exon 12 with *NSD1* exon 6 (Supporting Information Fig. S1B) (Jaju et al., 2001; Brown et al., 2002; Panarello et al., 2002; Casas et al., 2003; Cerveira et al., 2003; Hollink et al., 2011). The reciprocal fusion products were also fused in-frame, joining *NSD1* exon 5 with *NUP98* exon 13 (Supporting Information Fig. S1C).

Detection of *NUP98-NSD1*-Like Patients by Gene Expression Analysis

To investigate the gene expression pattern characteristic of *NUP98-NSD1*-positive patients, we used gene expression profiling data on the 124 patients. In unsupervised hierarchical clustering analyses (Supporting Information Fig. S2), *NUP98-NSD1*-positive patients did not form a cluster, whereas patients with t(8;21)(q22;q22), t(15;17)(q22;q21), and inv(16)(p13;q22) formed unique clusters. Comparing the 6 *NUP98-NSD1*-positive patients with the other 118 patients resulted in the identification of 87 differentially expressed probe sets ($P < 0.001$, fold change ≥ 2) (Supporting Information Table S3). Fifty-one probe sets, including several *HOX* genes, were overexpressed, and 36 probe sets were underexpressed. Interestingly, in supervised hierarchical clustering analysis using those 87 probe sets, a relatively large cluster including 18 *NUP98-NSD1*-negative patients was found, not only a cluster of the 6 *NUP98-NSD1*-positive patients (Fig. 1). These 24 patients also formed a single cluster in similar supervised clustering analyses in which probe sets were selected by slightly different criteria (Supporting Information Fig. S3). The robustness in clustering suggests that these 24 patients have certain common features and that they constitute a distinct subtype. Thus, we termed this characteristic gene expression pattern the *NUP98-NSD1* signature. We also designated the 18 patients lacking the *NUP98-NSD1* gene fusion but displaying the *NUP98-NSD1* signature as *NUP98-NSD1*-like AML patients. Of the 24 patients displaying the *NUP98-NSD1* signature, a common overexpression of 3 *HOXB* genes (*HOXB3*,