Brief Report

MYELOID NEOPLASIA

Naturally occurring oncogenic GATA1 mutants with internal deletions in transient abnormal myelopoiesis in Down syndrome

Tsutomu Toki,[1] Rika Kanezaki,[1] Eri Kobayashi,[2,3] Hiroshi Kaneko,[3] Mikiko Suzuki,[3] RuNan Wang,[1] Kiminori Terui,[1] Hirokazu Kanegane,[4] Miho Maeda,[5] Mikiya Endo,[6] Tatsuki Mizuochi,[7] Souichi Adachi,[8] Yasuhide Hayashi,[9] Masayuki Yamamoto,[2] Ritsuko Shimizu,[3] and Etsuro Ito[1]

[1]Department of Pediatrics, Hirosaki University Graduate School of Medicine, Hirosaki, Japan; [2]Department of Medical Biochemistry and [3]Department of Molecular Hematology, Tohoku University Graduate School of Medicine, Sendai, Japan; [4]Department of Pediatrics, Graduate School of Medicine, University of Toyama, Toyama, Japan; [5]Department of Pediatrics, Nippon Medical School, Tokyo, Japan; [6]Department of Pediatrics, Iwate Medical University, Morioka, Japan; [7]Department of Pediatrics and Child Health, Kurume University School of Medicine, Kurume, Japan; [8]Human Health Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan; and [9]Department of Hematology/Oncology, Gunma Children's Medical Center, Gunma, Japan

Key Points

- Naturally occurring oncogenic GATA1 mutants with internal deletions contribute to transient abnormal myelopoiesis in Down syndrome.

Children with Down syndrome have an increased incidence of transient abnormal myelopoiesis (TAM) and acute megakaryoblastic leukemia. The majority of these cases harbor somatic mutations in the *GATA1* gene, which results in the loss of full-length GATA1. Only a truncated isoform of GATA1 that lacks the *N*-terminal 83 amino acids (GATA1-S) remains. We found through genetic studies of 106 patients with TAM that internally deleted GATA1 proteins (GATA1-IDs) lacking amino acid residues 77-119 or 74-88 (created by splicing mutations) contributed to the genesis of TAM in 6 patients. Analyses of GATA1-deficient embryonic megakaryocytic progenitors revealed that the GATA1 function in growth restriction was disrupted in GATA1-IDs. In contrast, GATA1-S promoted megakaryocyte proliferation more profoundly than that induced by GATA1 deficiency. These results indicate that the internally deleted regions play important roles in megakaryocyte proliferation and that perturbation of this mechanism is involved in the pathogenesis of TAM. (*Blood.* 2013;121(16):3181-3184)

Introduction

Children with Down syndrome (DS) are known to have a high risk of developing transient abnormal myelopoiesis (TAM) and subsequent acute megakaryoblastic leukemia (DS-AMKL).[1-4] Blast cells in the majority of patients with TAM and DS-AMKL have mutations in the second exon of the *GATA1* gene.[5,6] The mutations turn off the production of full-length GATA1. Instead, *N*-terminally truncated GATA1 protein (GATA1-S) was translated from the second methionine at codon 84, which is identical to the truncated GATA1 isoform found in the healthy human.[7] In contrast, only a few patients with AMKL have been reported to harbor 21-disomy blasts with the GATA1 mutation.[8,9] Therefore, GATA1-S is believed to be a prerequisite for the pathogenesis of TAM and DS-AMKL in children with DS, and unrestricted proliferation of megakaryocytic progenitors in DS-AMKL is thought to be provoked by a mechanism involving GATA1-S. However, the molecular mechanism of how GATA1-S contributes to the genesis of TAM and DS-AMKL remains elusive.

GATA1 regulates the proliferation of immature megakaryocytic progenitors. Indeed, active proliferation of immature megakaryocytic progenitors derived from GATA1-deficient mouse embryos is restricted by introduction of wild-type GATA1, but not by GATA-S.[10] GATA1-deficient mice rescued with transgenic expression of GATA1-S (or GATA1-ΔNT) are found to exhibit hyper-megakaryopoiesis in a limited embryonic and postnatal period, resembling the phenotype in human TAM cases.[11] In contrast, another report indicates that targeting mice expressing GATA1 protein with a deletion of 64 *N*-terminal amino acids, but retaining the 65th to 83rd amino acid residues intact, has demonstrated that the embryos display a transient megakaryocytic phenotype only during the early embryonic stage, not in the late-embryonic and postnatal stages.[12] We surmise that this difference simply may be a result of missing the region corresponding to the 65th to 83rd amino acids.

Here, we have identified novel GATA1 mutants with internal deletions (IDs) of either amino acid residues 77-119 or 74-88 (GATA1-IDs) in 6 patients. We found that the GATA1-IDs lost their activity in the regulation of megakaryocyte growth. These results demonstrate that disruption of ID regions is implicated in the pathogenesis of TAM.

Study design

This study was approved by the Ethics Committee of the Hirosaki University Graduate School of Medicine. All animal experiments

Submitted January 20, 2012; accepted February 7, 2013. Prepublished online as *Blood* First Edition paper, February 25, 2013; DOI 10.1182/blood-2012-01-405746.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

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.

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

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

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).

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

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).

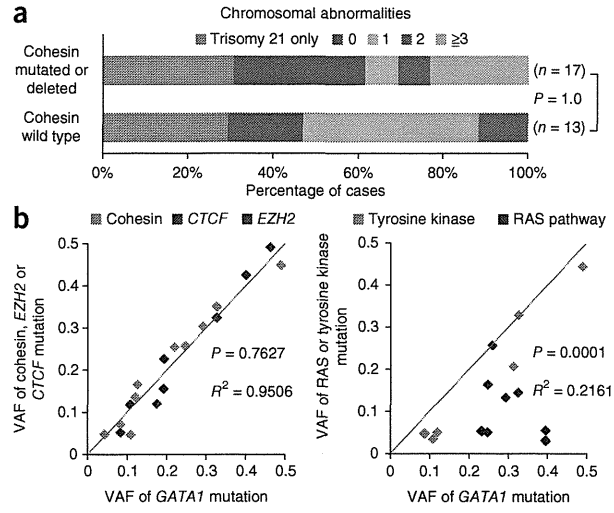

Figure 5 Relationship of cohesin mutations with karyotypes and comparison of mutation loads between major gene targets in DS-AMKL and *GATA1*. (**a**) The number of chromosomal abnormalities is compared between cases with and without cohesin mutations or deletions for DS-AMKL cases. Zero signifies chromosomal 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

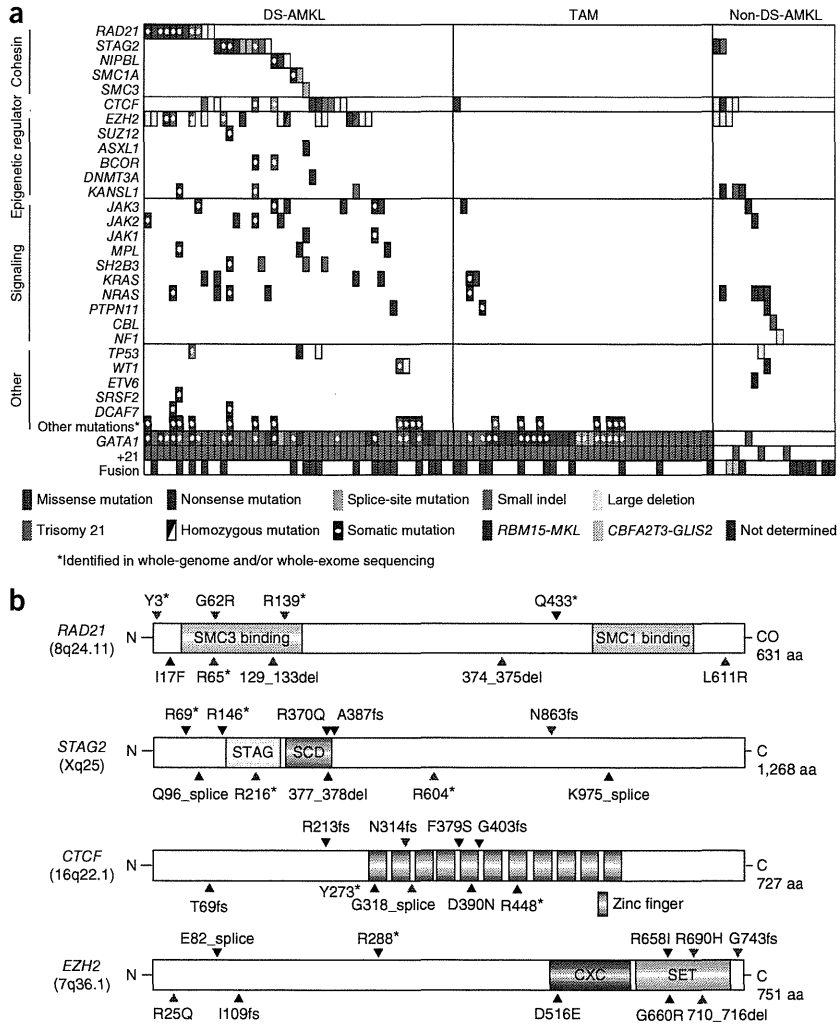

Figure 4 Driver mutations in Down syndrome–related myeloid disorders and non-DS-AMKL. (**a**) Driver mutations in 109 samples of 49 DS-AMKL, 41 TAM and 19 non-DS-AMKL cases. Types of mutations are distinguished by color. Each sample is also described in **Supplementary Table 12**. (**b**) Distribution 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

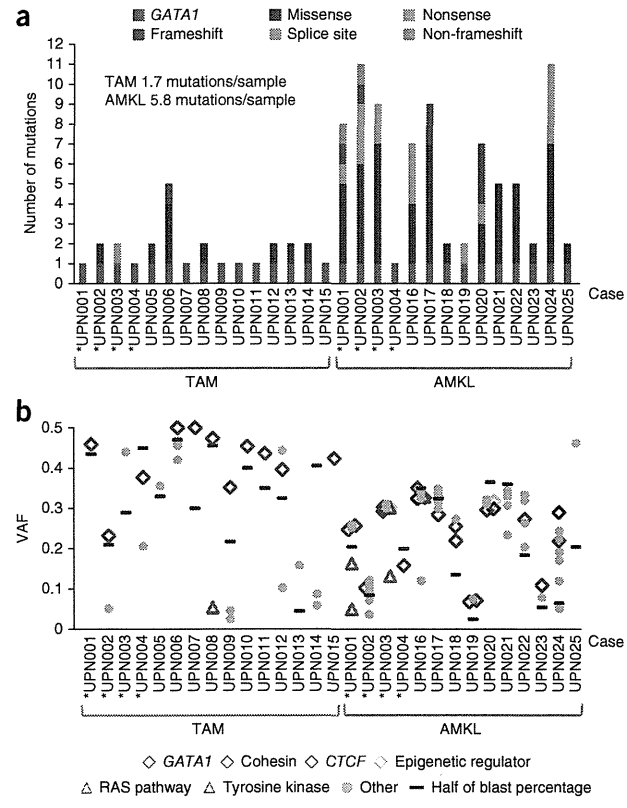

Figure 3 Somatic mutations detected by whole-exome sequencing of Down syndrome–related myeloid disorders. (**a**) Number of validated somatic mutations in 25 individuals with TAM and DS-AMKL identified by whole-exome sequencing. Paired samples are indicated by asterisks. The mutation rates per phase 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_2148delATCACAGGATAGGTATTTTT	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_486delGTCCGCGCCATGGCCATC	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

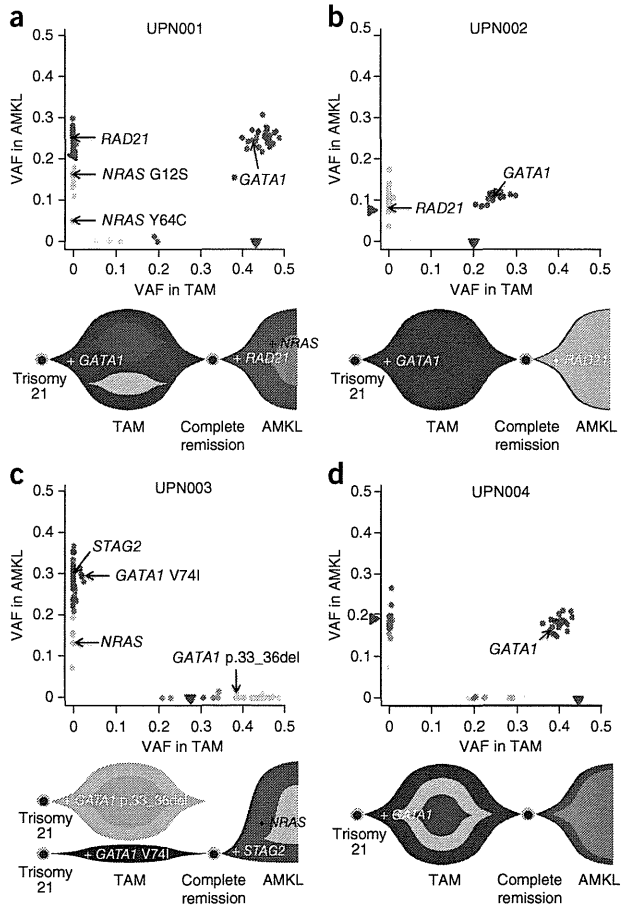

Figure 2 Clonal evolution of Down syndrome–related myeloid disorders. (**a**–**d**) Observed VAFs of validated mutations listed in **Supplementary Table 2** in both TAM and AMKL phases are shown in diagonal plots (top) for UPN001 (**a**), UPN002 (**b**), UPN003 (**c**) and UPN004 (**d**), where VAFs 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

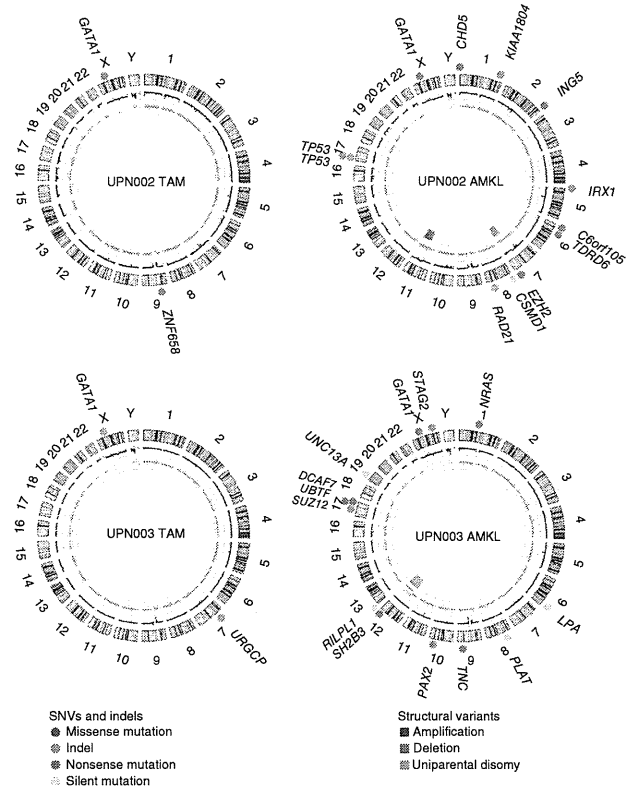

Figure 1 Representative Circos plots of paired TAM and DS-AMKL cases. Locations of somatic mutations, including of missense, frameshift, nonsense and silent mutations (colored circles), are indicated. Total (black) and allele-specific (red and green for alleles showing relatively larger and smaller copy 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

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

7. Groet J, McElwaine S, Spinelli M, et al. Acquired mutations in GATA1 in neonates with Down's syndrome with transient myeloid disorder. *Lancet.* 2003;361(9369):1617-1620.
8. Xu G, Nagano M, Kanezaki R, et al. Frequent mutations in the GATA-1 gene in the transient myeloproliferative disorder of Down syndrome. *Blood.* 2003;102(8):2960-2968.
9. Weiss MJ, Orkin SH. Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc Natl Acad Sci USA.* 1995;92(21):9623-9627.
10. Ferreira R, Ohneda K, Yamamoto M, Philipsen S. GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol Cell Biol.* 2005; 25(4):1215-1227.
11. Hayashi Y, Eguchi M, Sugita K, et al. Cytogenetic findings and clinical features in acute leukemia and transient myeloproliferative disorder in Down's syndrome. *Blood.* 1988;72(1):15-23.
12. Forestier E, Izraeli S, Beverloo B, et al. Cytogenetic features of acute lymphoblastic and myeloid leukemias in pediatric patients with Down syndrome: an iBFM-SG study. *Blood.* 2008; 111(3):1575-1583.
13. Xu G, Kato K, Toki T, Takahashi Y, Terui K, Ito E. Development of acute megakaryoblastic leukemia from a minor clone in a Down syndrome patient with clinically overt transient myeloproliferative disorder. *J Pediatr Hematol Oncol.* 2006;28(10): 696-698.
14. Vyas P, Crispino JD. Molecular insights into Down syndrome-associated leukemia. *Curr Opin Pediatr.* 2007;19(1):9-14.
15. Kirsammer G, Jilani S, Liu H, Davis E, Gurbuxani S, Le Beau MM, Crispino JD. Highly penetrant myeloproliferative disease in the Ts65Dn mouse model of Down syndrome. *Blood.* 2008;111(2): 767-775.
16. Carmichael CL, Majewski IJ, Alexander WS, Metcalf D, Hilton DJ, Hewitt CA, Scott HS. Hematopoietic defects in the Ts1Cje mouse model of Down syndrome. *Blood.* 2009;113(9): 1929-1937.
17. Alford KA, Slender A, Vanes L, et al. Perturbed hematopoiesis in the Tc1 mouse model of Down syndrome. *Blood.* 2010;115(14):2928-2937.
18. Malinge S, Bliss-Moreau M, Kirsammer G, Diebold L, Chlon T, Gurbuxani S, Crispino JD. Increased dosage of the chromosome 21 ortholog Dyrk1a promotes megakaryoblastic leukemia in a murine model of Down syndrome. *J Clin Invest.* 2012;122(3):948-962.
19. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med.* 1997;3(7):730-737.
20. Clappier E, Gerby B, Sigaux F, et al. Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. *J Exp Med.* 2011;208(4): 653-661.
21. Notta F, Mullighan CG, Wang JCY, et al. Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. *Nature.* 2011; 469(7330):362-367.
22. Anderson K, Lutz C, van Delft FW, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature.* 2011;469(7330):356-361.
23. Hiramatsu H, Nishikomori R, Heike T, Ito M, Kobayashi K, Katamura K, Nakahata T. Complete reconstitution of human lymphocytes from cord blood CD34+ cells using the NOD/SCID/gammacnull mice model. *Blood.* 2003;102(3):873-880.
24. Fujino H, Hiramatsu H, Tsuchiya A, et al. Human cord blood CD34+ cells develop into hepatocytes in the livers of NOD/SCID/gamma(c)null mice through cell fusion. *FASEB J.* 2007;21(13):3499-3510.
25. Kato M, Sanada M, Kato I, et al. Frequent inactivation of A20 in B-cell lymphomas. *Nature.* 2009;459(7247):712-716.
26. Kato I, Niwa A, Heike T, et al. Identification of hepatic niche harboring human acute lymphoblastic leukemic cells via the SDF-1/ CXCR4 axis. *PLoS ONE.* 2011;6(11):e27042.
27. Chen J, Li Y, Doedens M, Wang P, Shago M, Dick JE, Hitzler JK. Functional differences between myeloid leukemia-initiating and transient leukemia cells in Down's syndrome. *Leukemia.* 2010;24(5):1012-1017.
28. Ito M, Hiramatsu H, Kobayashi K, et al. NOD/ SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells. *Blood.* 2002;100(9):3175-3182.
29. Lacombe F, Durrieu F, Briais A, et al. Flow cytometry CD45 gating for immunophenotyping of acute myeloid leukemia. *Leukemia.* 1997;11(11): 1878-1886.
30. Suda J, Eguchi M, Akiyama Y, et al. Differentiation of blast cells from a Down's syndrome patient with transient myeloproliferative disorder. *Blood.* 1987;69(2):508-512.
31. Silva ML, do Socorro Pombo-de-Oliveira M, Raimondi SC, et al. Unbalanced chromosome 1 abnormalities leading to partial trisomy 1q in four infants with Down syndrome and acute megakaryocytic leukemia. *Mol Cytogenet.* 2009;2:7.
32. Welch JS, Ley TJ, Link DC, et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell.* 2012;150(2):264-278.
33. Ahmed M, Sternberg A, Hall G, et al. Natural history of GATA1 mutations in Down syndrome. *Blood.* 2004;103(7):2480-2489.
34. Greaves M, Maley CC. Clonal evolution in cancer. *Nature.* 2012;481(7381):306-313.
35. Shimada A, Xu G, Toki T, Kimura H, Hayashi Y, Ito E. Fetal origin of the GATA1 mutation in identical twins with transient myeloproliferative disorder and acute megakaryoblastic leukemia accompanying Down syndrome. *Blood.* 2004; 103(1):366.
36. Hitzler JK, Zipursky A. Origins of leukaemia in children with Down syndrome. *Nat Rev Cancer.* 2005;5(1):11-20.
37. Khan I, Malinge S, Crispino J. Myeloid leukemia in Down syndrome. *Crit Rev Oncog.* 2011;16(1-2): 25-36.
38. Nowell PC. The clonal evolution of tumor cell populations. *Science.* 1976;194(4260):23-28.
39. Blink M, van den Heuvel-Eibrink MM, Aalbers AM, et al. High frequency of copy number alterations in myeloid leukaemia of Down syndrome. *Br J Haematol.* 2012;158(6):800-803.
40. Malkin D, Brown EJ, Zipursky A. The role of p53 in megakaryocyte differentiation and the megakaryocytic leukemias of Down syndrome. *Cancer Genet Cytogenet.* 2000;116(1):1-5.
41. Norton A, Fisher C, Liu H, et al. Analysis of JAK3, JAK2, and C-MPL mutations in transient myeloproliferative disorder and myeloid leukemia of Down syndrome blasts in children with Down syndrome. *Blood.* 2007;110(3):1077-1079.
42. Malinge S, Ragu C, Della-Valle V, et al. Activating mutations in human acute megakaryoblastic leukemia. *Blood.* 2008;112(10):4220-4226.

However, there was no clear association between the percentage of TAM blast cells in transplanted samples and successful engraftment (supplemental Table 1). Likewise, frozen samples from 3 patient samples (patients 1, 2, and 9) were as efficient for engraftment as fresh samples from these patients. Therefore, although the number of injected TAM cells and technical issues may affect engraftment, we speculate that engraftment efficiency is an intrinsic property of each TAM-derived cell population.

In addition to trisomy 21, somatic *GATA1* mutation is considered an early essential event of TAM and ML-DS occurring in utero.[35,36] Interestingly, our TAM-NOG mice model enhanced the emergence of a minor clone with a distinct *GATA1* mutation that was not detectable in the original patient sample by conventional sequencing methods. In our model, a minor *GATA1* mutant clone expanded predominantly in some recipients and acquired CNAs independently of clones with the original *GATA1* mutation, raising the possibility that leukemic evolution occurred from this minor clone, similar to the clinical observation in our previous report.[13] In this scenario, a common founder clone of TAM/ML-DS may be established before the acquisition of the *GATA1* mutation, or TAM clones with distinct *GATA1* mutations may arise independently in the fetal period.

It has long been considered that the linear sequential acquisition of genetic alterations induces disease progression in TAM/ML-DS.[37] By contrast, recent studies using high-throughput genomic technology indicate that evolutionary trajectories are more complex and branching in other cancers and leukemias, as previously proposed by Nowell.[38] In this theory, genomic instability in founder cells gives rise to heterogeneous mutant subclones, and under selective pressure, some subclones expand to result in disease progression, whereas others become extinct or remain dormant. Thus, leukemic clones may evolve and emerge through the complex interaction of selectively advantageous "driver" mutations, additional advantageous "cooperating" mutations, neutral "passenger" mutations, and deleterious mutations.[32,38] It is clinically true that genomic alterations are more frequently observed in ML-DS than in TAM.[2,11,39] In this paper, we showed that diverse subclones with various CNAs can be generated in TAM, and these events occurred preferentially in a patient who later developed ML-DS. These findings suggest the presence of leukemic driver mutations in the early phase of TAM in this patient, which may have induced genomic instability. We were unable to find any candidate tumor-associated genes on the deletion sites (3q24, 9q22, 12p12, 16q22, and 16q24) of TAM-derived cells using the The National Center for Biotechnology Information database, suggesting that other genetic mutations and epigenetic events may contribute to the progression to ML-DS, including a few candidate mutations identified previously.[40-42] It is also noteworthy that subclones in each recipient mouse showed different repopulating capacities in this study. The dominant clones in each recipient were not always identical in the 1° generations, and the dominant clone in a certain recipient did not always propagate dominantly in the next generation recipients (Figure 4A). Differences between the recipient mice or technical problems may have caused variations in engraftment outcome, which is a potential weakness of this xenograft model; however, it is more likely that cooperating genetic event(s) important for leukemogenesis led to the cells of a specific TAM clone becoming the dominant population in each recipient. Such cooperating event(s) could have a considerable impact on a preleukemic TAM clone, and clonal selection might occur in a somewhat random manner. Thus, leukemic evolution may depend on random chance to an extent. Our TAM xenograft model may help demonstrate the branching architecture of clonal evolution in a preleukemic phase, which contrasts with a linear and deterministic pattern of evolution.[34,38] Further genomewide analysis is needed to elucidate the true driver or cooperating mutation(s) and unravel the evolutionary process of leukemia.

In conclusion, we established a xenograft model of TAM using highly immunodeficient NOG mice. Our model enabled the observation of clonal selection and expansion of minor mutant TAM clones and is likely to mimic the early phase of the leukemic evolutionary process, demonstrating the striking genetic heterogeneity and the propagating potential of minor clones in a preleukemic phase. Our xenograft model could be valuable tool for gaining insight into the leukemogenesis of ML-DS and for evaluating the prognosis of TAM patients.

Acknowledgments

The authors thank all the TAM patients and their families for their participation. The authors thank Drs Akira Niwa, Masashi Sanada, Hironao Numabe, Tomoki Kawai, Takahiro Yasumi, and Ryuta Nishikomori for technical advice.

This work was supported by grants from the Japanese Ministry of Education, Culture, Sports, Science, and Technology, and from the Japanese Ministry of Health, Labor and Welfare.

Authorship

Contribution: S.S., I.K., T.M., H.F., and K.U. performed sample collection and processing; S.S., K.T., K.Y., and R.W. performed experiments; A.S.-O. performed microarray analysis (accession number GSE44739); Y.S. and S.M. provided expert statistical analysis; S.S., Y.O., and T.T. analyzed results and made the figures; M.I. and T.N. generated NOG mice; and S.S., K.W., H.H., S.A., E.I., S.O., and T.H. designed the research and wrote the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Toshio Heike, Department of Pediatrics, Graduate School of Medicine, Kyoto University, 54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto, 606-8507, Japan; e-mail: heike@kuhp.kyoto-u.ac.jp.

References

1. Pine SR, Guo Q, Yin C, Jayabose S, Druschel CM, Sandoval C. Incidence and clinical implications of GATA1 mutations in newborns with Down syndrome. *Blood.* 2007;110(6):2128-2131.
2. Massey GV, Zipursky A, Chang MN, et al; Children's Oncology Group (COG). 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.* 2006;107(12):4606-4613.
3. Zipursky A, Poon A, Doyle J. Leukemia in Down syndrome: a review. *Pediatr Hematol Oncol.* 1992;9(2):139-149.
4. Hitzler JK. Acute megakaryoblastic leukemia in Down syndrome. *Pediatr Blood Cancer.* 2007; 49(7 Suppl):1066-1069.
5. Hitzler JK, Cheung J, Li Y, Scherer SW, Zipursky A. GATA1 mutations in transient leukemia and acute megakaryoblastic leukemia of Down syndrome. *Blood.* 2003;101(11):4301-4304.
6. Wechsler J, Greene M, McDevitt MA, Anastasi J, Karp JE, Le Beau MM, Crispino JD. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat Genet.* 2002; 32(1):148-152.

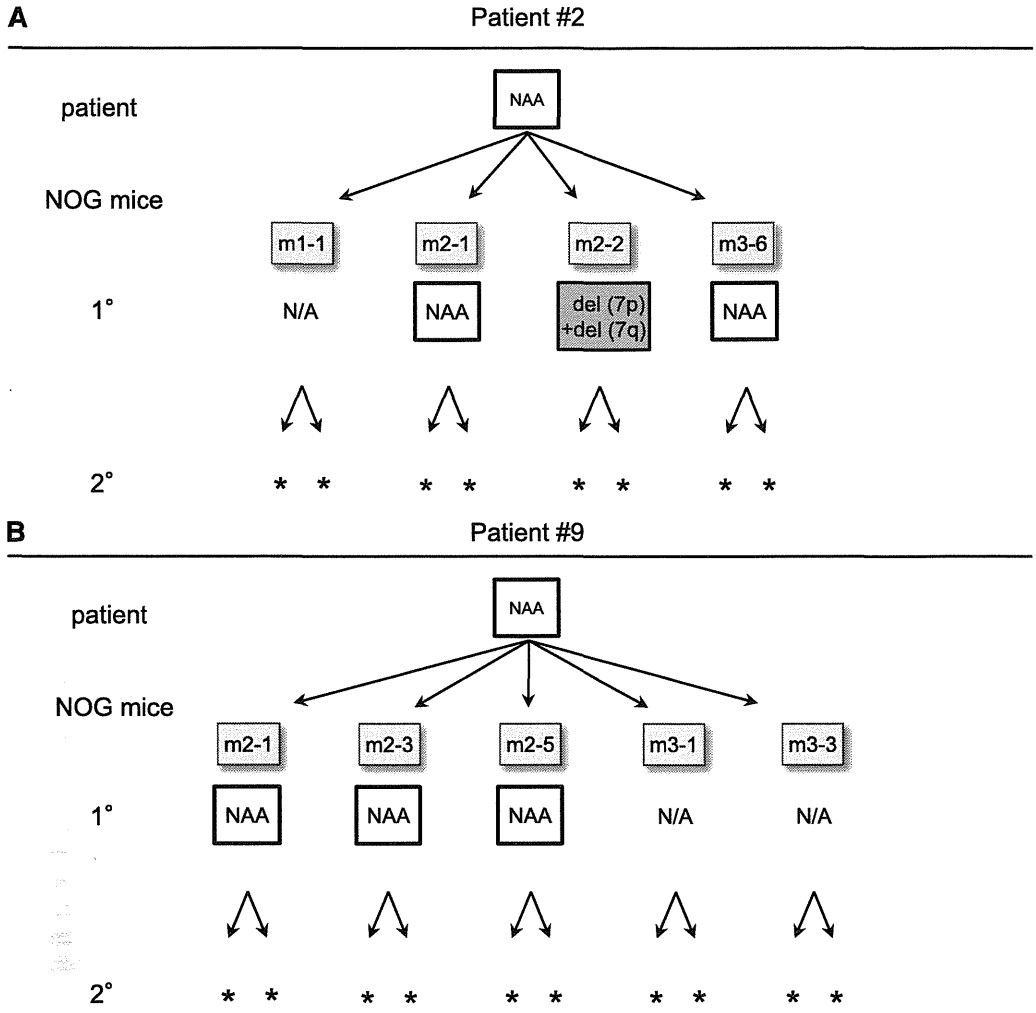

Figure 6. Serial transplantation and CNA profiling of TAM-derived cells from patients that did not develop ML-DS. (A) Serial transplantation assay using TAM cells from patient 2. Four attempts resulted in successful analysis in 1° recipients. CNAs with del(7p) and del(7q) were observed in 1 recipient (m2-2). No additional CNAs were observed in any other recipients. No engraftment was observed in 2° recipients. (B) Serial transplantation assay of TAM cells from patient 9. Five attempts resulted in successful analysis in 1° recipients. No additional CNAs were observed in any analyzed recipients, and no engraftment was observed in 2° recipients. NAA, no additional alteration; N/A, not assessed because of low cell count. *No engraftment.

leukemia,[19-22] we hypothesized that xenograft models of TAM cells would be an attractive method to investigate leukemogenesis. In this report, we demonstrated the long-term engraftment of primary TAM cells in NOG mice and showed that TAM cells from a patient that subsequently developed ML-DS had the potential to gain diverse additional genomic alterations and self-renewal capacity. Although we were unable to determine whether the clonal evolution of TAM cells observed in our model reflected the clinical phenotype of the original patient because of insufficient sample from the ML-DS phase, our model is likely to enable the prospective evaluation of leukemic evolution and can be a powerful tool to study the pathophysiology of leukemogenesis. Our model using NOG mice contrasts somewhat with the study by Chen et al,[27] who reported that TAM cells resided only in the BM after intra-BM infusion into NOD/SCID mice. We speculate that a severe and unique immunodeficient microenvironment may have contributed to the successful engraftment of TAM cells in NOG mice.

In the present study, primary TAM cell samples from 3 of 11 patients engrafted in NOG mice (Figure 1), but serial engraftment was successful only with cells obtained from the patient who developed ML-DS at the age of 1 year (Figures 2, 3, and 6). The results of extensive serial transplantation revealed the emergence of subclones with various additional CNAs characteristic of ML-DS (Figures 3 and 4). Furthermore, we showed that minor subclones with various CNAs and a distinct *GATA1* mutation were already present in the ML-DS patient during the early TAM phase (Figures 4 and 5), as previously described for polyclonality of TAM.[2,33] These findings suggest that several preleukemic clones with high leukemia-initiating potential may already reside as minor clones in TAM cells of patients fated to develop ML-DS and show high repopulating capacity in the special microenvironment of NOG mice. Our findings support the hypothesis that ML-DS develops from a pool of heterogeneous minor clones through clonal selection, illustrating the early evolutionary process of leukemia.[34]

Long-term engraftment of TAM-derived cells was observed for only a minority of TAM patients. This finding suggests that factors other than the properties of the TAM-derived cells, such as technical issues, affected engraftment efficiency. In this regard, increasing the number of transplanted cells resulted in a higher rate of engraftment in recipients of samples from patient 9 (supplemental Table 1).

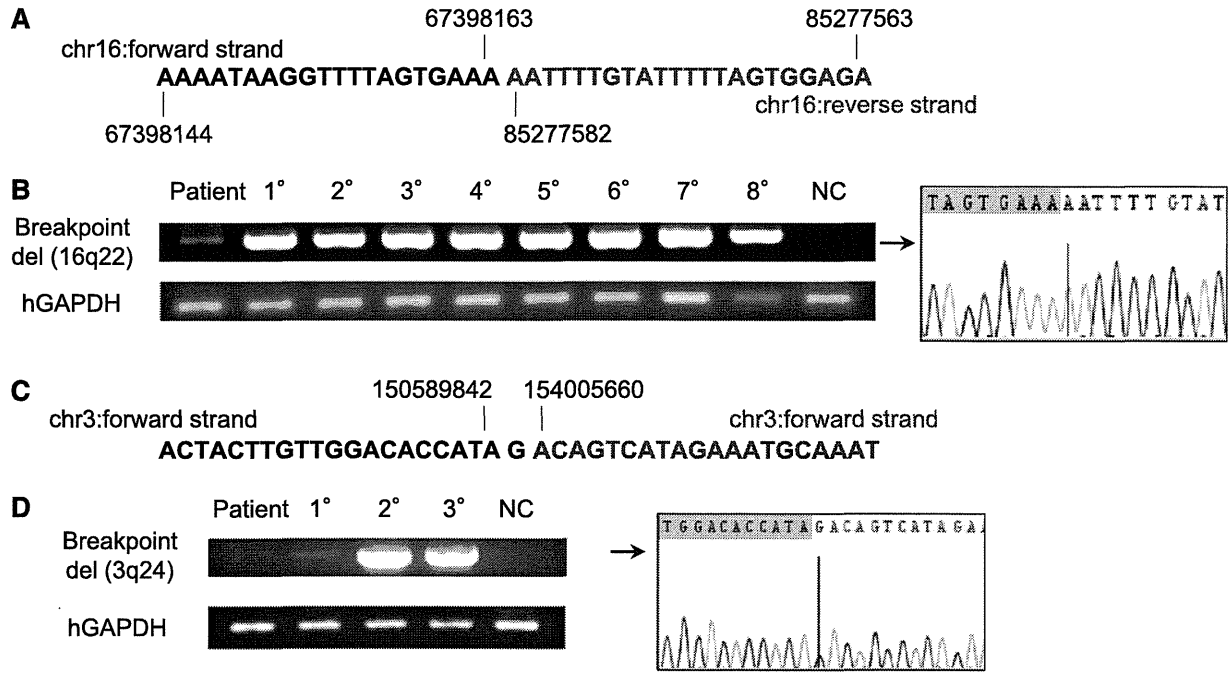

Figure 5. A minor subclone with additional CNAs was present in the primary TAM patient sample, whereas a new clone emerged in a 1° recipient. (A) Contig of del (16q22) breakpoint deduced by whole genome sequencing of the clone containing del(16q22) and del(16q24) in patient 1. Details are shown in supplemental Figure 6. (B) Breakpoint-specific PCR for the del(16q22) clone using genomic DNA from the original patient sample (1; PB in TAM phase), 1° to 8° xenografts (hCD45$^+$ BM cells; Figure 3), and NC (negative control; PBMCs from a healthy adult). Cells from 1° to 8° recipients showed a bright band. The original patient sample showed a faint band, and direct sequencing revealed the presence of the deduced breakpoint for del(16q22). Human glyceraldehyde-3-phosphate dehydrogenase (hGAPDH) was used as an internal control. (C) Contig of del(3q24) breakpoint deduced by whole-genome sequencing. Details are shown in supplemental Figure 9. (D) Breakpoint-specific PCR for the del(3q24) clone using genomic DNA from the original patient sample (1; PB in TAM phase), 1° to 3° xenografts (hCD45$^+$ BM cells, m2–5; Figure 4A), and NC. Cells from 2° and 3° recipients showed a bright band. No band was detected in the original patient sample, but a faint band was detected in the 1° recipient sample. hGAPDH was used as an internal control. Direct sequencing confirmed the presence of cells with the deduced breakpoint for del(3q24) in the 1° recipient.

However, because the sensitivity of the specific PCR targeting of the 3q24 deletion was ~0.1% as determined by the dilution assay (data not shown), it is also possible that this minor clone already existed in the primary patient sample at a frequency below the sensitivity limit. Collectively, our results provide evidence that subclones with additional genetic alterations already exist in the TAM phase and suggest that clonal selection occurs continuously in this xenograft model.

TAM cells derived from patients who did not develop ML-DS had limited self-renewal capacity and fewer additional CNAs than those from the patient who developed ML-DS

To assess whether TAM cells derived from the patients who did not develop ML-DS had similar self-renewal capacity and genetic instability to those from patient 1, CNA analysis of TAM-derived cells was performed by transplanting the preserved PBMC samples of patients 2 and 9. In patient 2, 4 1° transplantation attempts resulted in successful engraftment. The primary sample of patient 2 had no CNAs (Figure 6A). However, TAM-derived cells in 1 of the 1° recipients (m2-2) showed 7p and 7q deletions, suggesting that a subclone with these CNAs may exist in the primary patient sample. The other 2 1° recipients had no additional CNAs (m2-1 and m3-6). In patient 9, engraftment succeeded in 5 1° recipients, and no additional CNAs were detected in either primary patient sample or engrafted TAM-derived cells (Figure 6B). The engrafted cells in all of the recipient mice harbored the same *GATA1* mutation as that of the primary samples of patients 2 and 9. In these 2 cases, our xenograft assay did not detect potent TAM clones with self-renewal capacity in serial transplantation assays (Figure 6A-B; supplemental Table 1), which may reflect the favorable clinical outcome of these patients.

Taken together, the results show that only the TAM cells derived from patients who subsequently developed ML-DS had long-term self-renewal capacity with additional CNAs in our serial transplantation assay.

Discussion

New genomic technologies have led to a better understanding of the complex clonal architecture of leukemia and have shown that disease progression occurs through clonal evolution.[20-22,32] However, most studies have been based on the retrospective analysis of frank leukemia samples, and data on the evolutionary process occurring in the preleukemic phase are limited because primary preleukemia samples are rarely available and are difficult to maintain in vitro or in vivo. TAM is a unique hematologic condition associated with DS that is mostly self-limited but leads to ML-DS in 20% of cases after spontaneous remission. Therefore, TAM has been considered a preleukemic state and is a suitable pathological condition to analyze the evolutionary process of leukemia.

Because mice models in which primary human leukemic cells were transplanted into immunodeficient hosts provided significant clues to advance our understanding of the pathogenesis of human

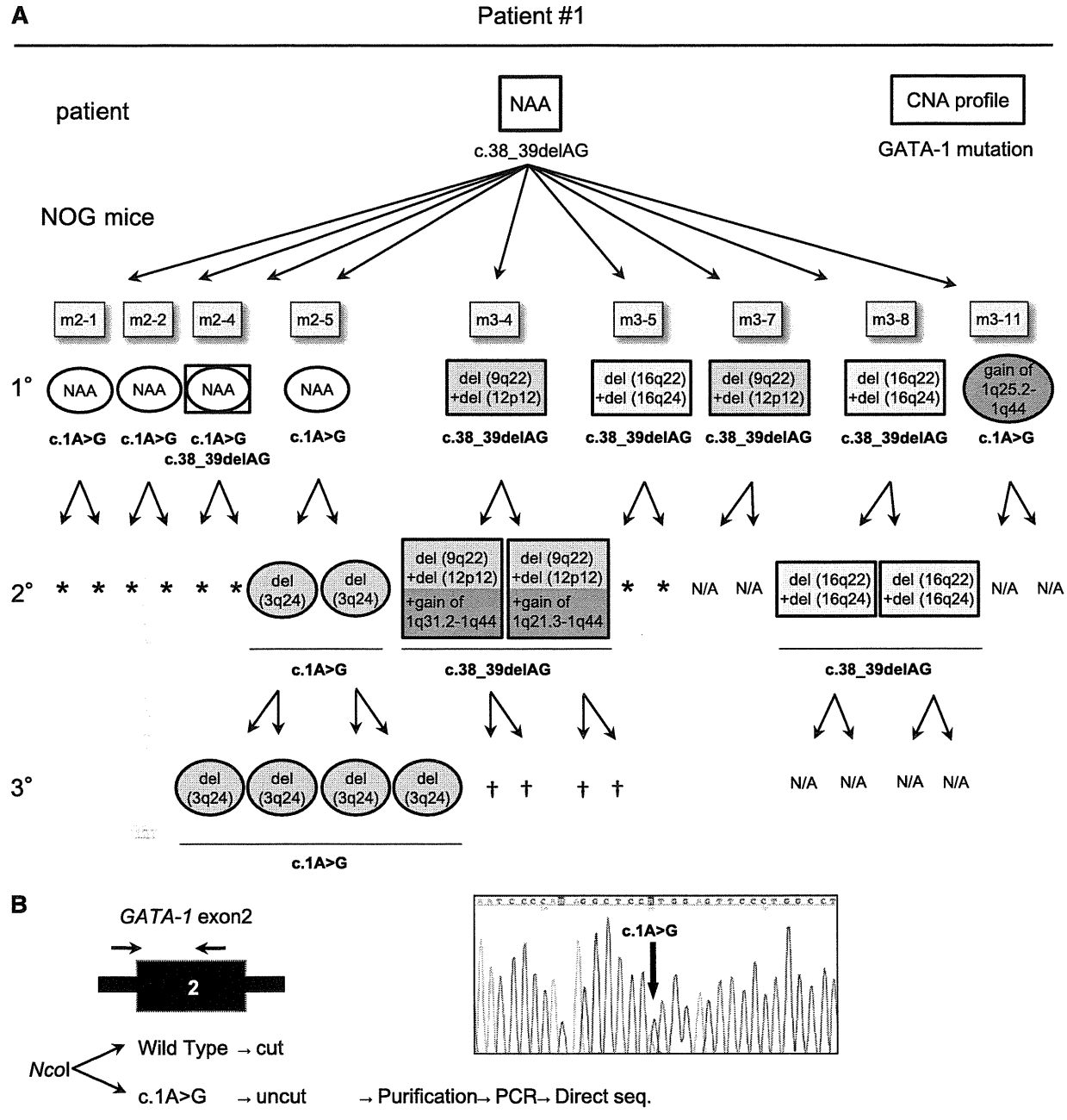

Figure 4. TAM-derived cells show genetic and functional diversity. (A) Summary of the serial transplantation of TAM cells of patient 1 and the results of CNA profiling and *GATA1* mutation analysis. The original patient sample had a single *GATA1* mutation, c.38G_39delAG, and no additional CNAs. Diverse subpopulations with or without additional CNAs expanded in each recipient. *GATA1* mutation analysis showed 2 distinct mutations in recipients: one identical to that of the original patient (c.38_39delAG) and a different mutation (c.1A>G). The mice harboring cells with the original mutation (c.38_39delAG) are shown in rectangles, and the mice with cells harboring the other mutation (c.1A>G) are shown in ovals, with a CNA profile note inside. The *GATA1* mutation is indicated below the symbol. NAA, no additional alteration; N/A, not assessed because of low blast cell count. †Death of recipient before analysis. *No engraftment. (B) Detection of a minor clone with the c.1A>G mutation in the original sample of patient 1. NcoI digestion of a DNA fragment obtained by PCR of *GATA1* exon 2 yielded 2 fragments in the wild type, whereas the mutant allele (c.1A>G) remained undigested. PCR of the undigested band and direct sequence analysis identified the same *GATA1* mutation (c.1A>G mutation) in the patient sample. Black arrow indicates the primer set.

The same method was used to detect a subclone with a 3q24 deletion in the primary patient sample (m2-5; Figure 4A; supplemental Figure 6). At the site of the deletion, genomic breakage was confirmed, and the ends were bound by insertion of a G-nucleotide (Figure 5C; supplemental Figure 7). Consistent with the results of CNA profiling, PCR using DNA from engrafted cells in the 2° and 3° mice (m2-5; Figure 4A) produced a bright DNA fragment, which was confirmed to contain the deletion breakpoint in 3q24 by Sanger sequencing (Figure 5D). Engrafted cells from the BM of the 1° recipient produced a faint DNA fragment, although CNAs were not detected in these cells by array-based methods. We could not detect the corresponding DNA fragment in the primary sample of patient 1. These results suggest the subclone with the 3q24 deletion arose in the 1° recipient mouse as a minor population, emerged as a major population in the 2° recipient, and subsequently engrafted into the 3° recipients.

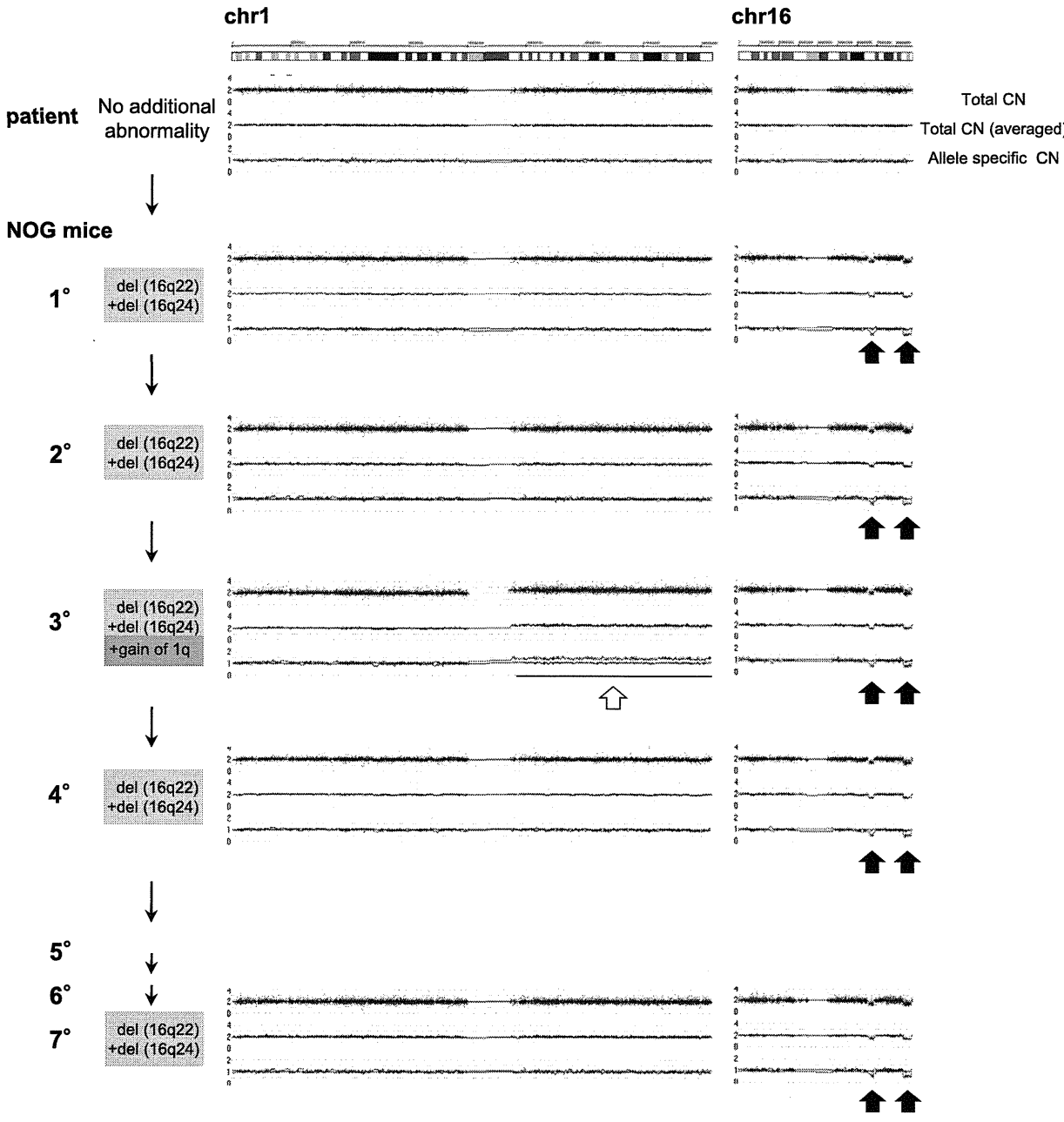

Figure 3. Sequential CNA analysis of TAM-derived cells in the recipients of patient 1. DNA obtained from the original patient sample and sorted hCD45$^+$ recipient BM cells were analyzed by Affymetrix GeneChip Mapping 250K arrays and compared with the PB sample of the original patient in complete remission phase. The primary sample of the patient in TAM phase (blast 92%) had no CNA. hCD45$^+$ BM cells of 1° to 7° recipients had a hemi-allelic deletion in regions 16q22 and 16q24 (black arrows). The 3° recipient had a gain of the entire arm of chromosome 1q (white arrow) in addition to deletion of 16q22 and 16q24. Arrowhead indicates abnormal CNA.

(Figure 4A). To determine whether these subclones were present at low levels in the primary sample of patient 1, specific PCR for the 16q22 deletion was performed using primer pairs designed to bookend the deletion site. CNA analysis and genome sequencing data in these deletion sites (16q22 and 16q24) revealed the presence of genomic breakage and inversion (Figure 5A; supplemental Figures 3 and 4; see supplemental Methods for details). A primer set was designed to detect the deduced breakpoint and used to perform PCR on TAM-derived cells from patient 1 in the recipients with 16q22 and 16q24 deletions. PCR using genomic DNA from TAM-derived cells in the 1° to 8° recipients of the first series of transplantations (Figure 3) produced a uniformly bright DNA fragment of the same size, consistent with the results of CNA profiling (Figure 5B). A faint fragment was detected by applying this PCR method to genomic DNA from the primary patient sample (patient 1), which was confirmed to contain the deletion breakpoint in 16q22 by Sanger sequencing. These results demonstrated that TAM cells with the 16q22 and 16q24 deletions already resided as a minor population in the original sample of patient 1. The frequency of the mutant cells was estimated to be ~1.0% to 0.2% of the patient's PBMCs by a serial dilution assay (supplemental Figure 5).

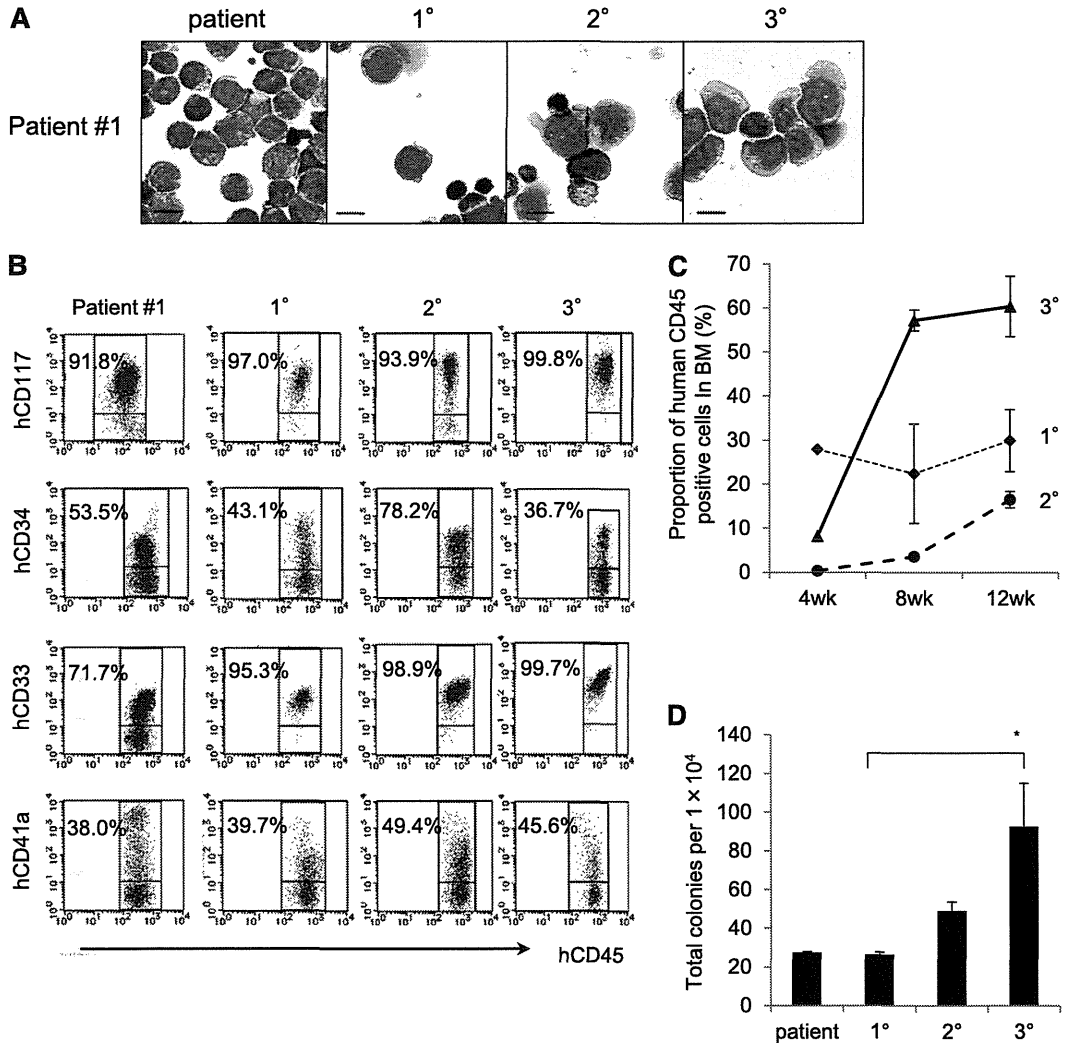

Figure 2. The NOG mouse model can support self-renewal of TAM-derived cells. (A) May-Giemsa staining of TAM-derived cells in recipients of patient 1. (B) Surface marker analysis of TAM-derived cells in recipients by flow cytometry. Viable cells were gated according to their forward scatter (FSC) and 4′,6-diamidino-2-phenylindole staining, blast cells were identified by CD45/SSC gating, and hCD45$^+$-gated cells were tested for the expression of hCD117, hCD34, hCD33, and hCD41a. (C) Proportion of hCD45$^+$ cells in BM of 1°, 2°, and 3° recipient mice at 4, 8, and 12 weeks after transplantation. (D) Colony assay of hCD45$^+$ cells in BM of 1°, 2°, and 3° recipient mice. hCD45$^+$ cells were seeded at 1.0×10^4 cells per 35-mm dish in triplicate, and the number of colonies in each dish was counted. Bars represent the standard deviation of the mean of 3 independent experiments. *Significant difference ($P < .05$).

acquired various CNAs and showed divergent repopulating capacity in our xenograft model.

TAM-NOG xenograft model revealed the presence of a minor clone with a distinct GATA1 mutation

ML-DS can arise from a minor TAM clone with a *GATA1* mutation that is distinct from that of the major TAM clone in a patient.[13] To determine whether the *GATA1* mutation in the primary patient's TAM cells was preserved in engrafted TAM-derived cells, *GATA1* mutation analysis was performed. TAM-derived cells in the series m3-4, m3-5, m3-7, and m3-8 had the same *GATA1* mutation (c.38_39delAG) as that of patient 1 (Figure 4A). Surprisingly, this mutation was not detected in TAM-derived cells in m2-1, m2-2, m2-5, and m3-11; instead, these samples showed a distinct *GATA1* mutation (c.1A>G) that was not detectable in the primary patient sample by direct sequencing. One of the 1° recipients (m2-4) showed both *GATA1* mutations. These results suggested that a minor clone with a distinct *GATA1* mutation (c.1A>G) was present in the primary patient sample and that this minor clone coexisted with, or predominated over, other clones in some 1° recipients. Therefore, a mutation-specific restriction enzyme digestion assay was performed using the primary sample from patient 1, which confirmed the presence of cells with the *GATA1* mutation (c.1A>G) as a minor clone (Figure 4B). Moreover, this minor clone propagated and acquired CNAs in NOG mice independently of the major clones (Figure 4A), further demonstrating the genetic heterogeneity of TAM cells. Interestingly, the major clone in the original patient 1 sample with a c.38_39delAG *GATA1* mutation and no CNAs did not become dominant in any of the recipients.

Minor subclone with additional CNAs was present in the primary TAM patient sample

TAM-derived cells in multiple 1° recipients derived from patient 1 had various CNAs including deletions of 16q22 and 16q24

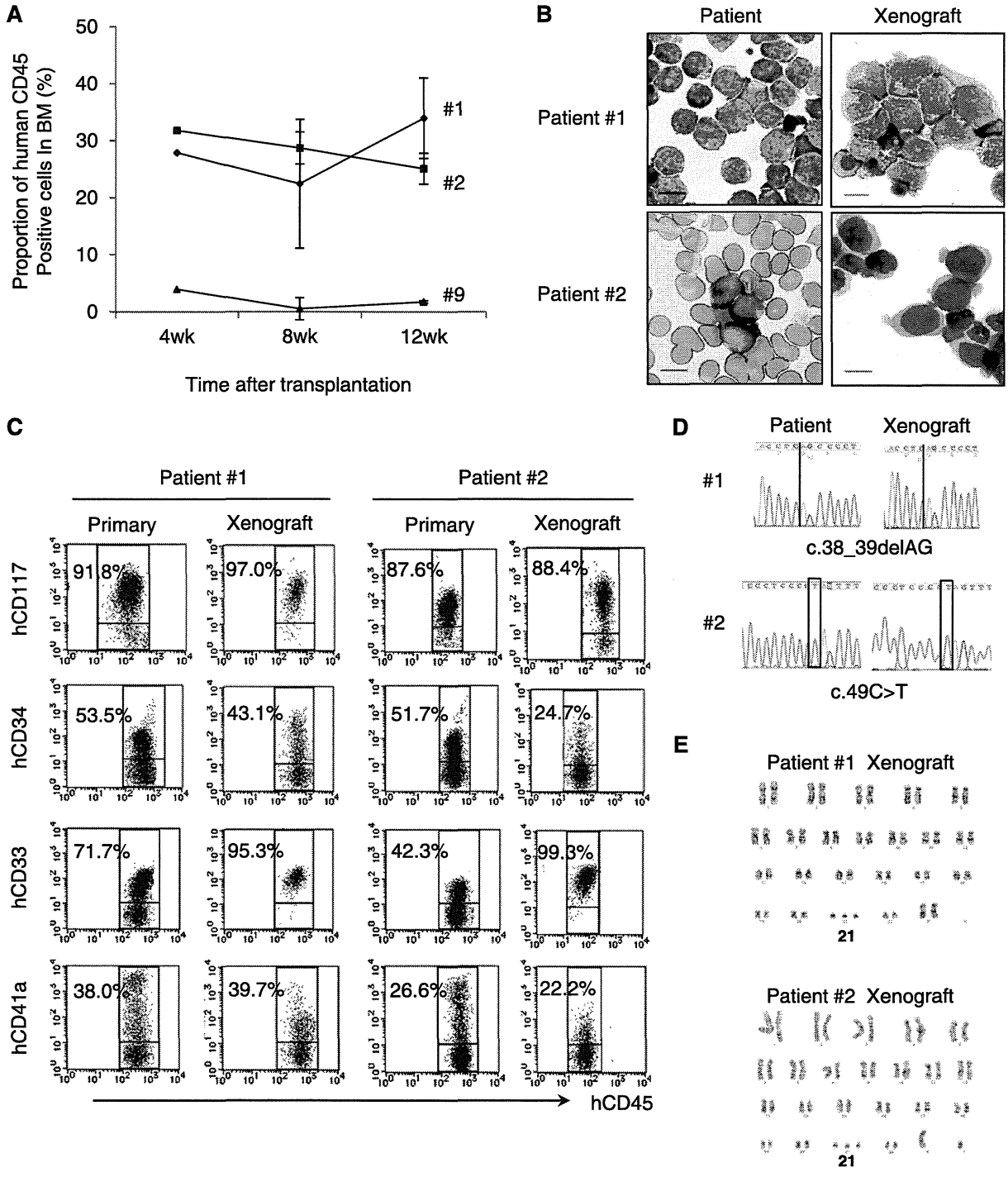

Figure 1. TAM cells engrafted in NOG mice. (A) Proportion of human CD45$^+$ cells in the BM of NOG mice at 4, 8, and 12 weeks after transplantation (n = 3–5 per group). (B) May-Giemsa staining of the BM smear of patients and cytospin preparation of human CD45$^+$ cells in the recipient NOG mice. Blast cells with cytoplasmic blebbing consistent with megakaryocytic differentiation were present in the BM of recipient mice. (C) Surface marker analysis of engrafted TAM cells. Human CD45$^+$ TAM-derived cells expressing hCD117, hCD34, hCD33, and hCD41a are detected in the recipient's BM. Blast cells were identified by CD45/SSC gating, and debris (low forward scatter) and dead cells (4′,6-diamidino-2-phenylindole positive) were excluded from the analysis. A representative result of >3 experiments is shown. (D) Genomic direct sequencing shows the presence of concordant GATA1 mutation in xenograft and original patients (1 and 2). (E) G-band karyotyping of TAM-derived cells in recipient murine BM shows no additional chromosome abnormality apart from constitutional trisomy 21, consistent with the findings in the original patients. The GATA1 mutation and the karyotype of engrafted cells from patient 9 were not assessed because of a low cell number.

gain of 1q was recurrently observed in this series, the duplicated regions were diverse: 1q25.2-1q44 (1°, m3-11), 1q21.3-1q44 (2°, m3-4), 1q31.2-1q44 (2°, m3-4), and the whole arm of chromosome 1q (3° in Figures 3 and 4A; supplemental Figure 2). In m2-5, a deletion of 3q24 appeared in the 2° and 3° recipients. These results demonstrated that TAM cells derived from patient 1

Table 1. Clinical characteristics of 11 TAM patients

Patient no.	Gender	Period of gestation, wk	Weight at birth, g	PB at diagnosis of TAM				Cytogenetics International System for Human Cytogenetic Nomenclature (2009)	*GATA1* mutation	Treatment		Clinical outcome	Onset of ML-DS (m of age)	Follow-up interval
				WBC, $\times 10^3$/ μL	Blast, %	Hb, g/dL	Plt, $\times 10^3$/ μL			Exchange transfusion	Low-dose Ara-C			
1	F	36	3050	159	91	9.1	247	47,XX,+21 [20]	c.38_39delAG	No	Yes	Alive	Yes (14)	27
2	M	37	1868	45.0	65	19.0	80	47,XY,+21 [20]	c.49C>T	No	Yes	Alive	No	24
3	F	39	3102	40.3	37	15.1	304	47,XX,+21 [20]	c.59_174del116	No	No	Alive	No	23
4	M	37	2780	15.6	24	17.0	50	47,XY,+21 [20]	c.163_169del	No	No	Alive	No	22
5	M	39	3052	60.9	48	20.7	258	47,XY,+21 [19]	c.37G>T	No	No	Alive	No	21
6	F	37	2050	13.6	12	20.9	291	47,XX,+21 [20]	N/A	No	No	Alive	No	19
7	M	38	2694	280	87	13.6	26	47,XY,+21 [20]	c.186C>G	Yes	Yes	DOD	No	1
8	M	35	2070	174	80	19.2	117	47,XY,+21 [20]	c.-19-1G>A, c.1A>G	Yes	No	Alive	No	17
9	M	39	3380	56.2	65	20.2	156	47,XY,+21 [20]	c.35C>G	No	Yes	Alive	No	14
10	M	36	2131	199	84	12.2	73	47,XY,+21 [20]	c.19_20insCCTGA	Yes	Yes	Alive	No	14
11	F	33	2032	254	90	14.3	178	47,XX,+21 [20]	c.-19-62_-5delinsA	Yes	Yes	Alive	No	10

Brackets under International System for Human Cytogenetic Nomenclature indicate the number of analyzed cells in metaphase.
Ara-C, cytosine arabinoside; DOD, died of disease; Hb, hemoglobin; N/A, not assessed; Plt, platelet; WBC, white blood cell.

expanded rapidly in the 3° recipients (Figure 2C). The colony-forming ability of the engrafted cells also increased in subsequent generations (Figure 2D). These cells could be grown by serial transplantation for >1 year and ⩾8° recipients, indicating that some TAM clones had long-term self-renewal capacity, a characteristic of leukemia. Indeed, patient 1 developed ML-DS at the age of 1 year, whereas the other patients did not (Table 1).

TAM-NOG xenograft model recapitulates leukemic evolution from TAM

Additional chromosomal alterations are frequently observed in ML-DS in comparison with TAM, suggesting that these alterations in genomic structure could be related to the evolution of ML-DS from TAM.[2,11,12] Therefore, we first investigated whether the serially engrafted TAM-derived cells (from patient 1) had DNA copy number alterations (CNAs) using Affymetrix GeneChip Mapping 250K arrays. Primary samples from patient 1 had no CNAs other than the gain of chromosome 21. However, the TAM-derived cells in the 1° recipients showed heterozygous deletion of 16q22 and 16q24 (Figure 3). To determine whether these deletions were present in the same cell, we calculated the signal intensities of each deletion using array data. Nearly 100% of TAM-derived cells harbored each deletion, indicating that these 2 deletions exist in a single TAM-derived cell. Although 2° recipients showed the same CNAs, 3° recipients showed additional CNAs, namely the gain of the entire chromosome 1q (Figure 3; supplemental Figure 1A). Interestingly, the 1q gain was not detected in the 4° to 7° recipients, whereas deletions of 16q22 and 16q24 were present (Figure 3). In this series of transplantations, the original *GATA1* mutation found in the primary patient sample (patient 1) remained unchanged (supplemental Figure 1B).

Gain of 1q and deletions in 16q are recurrent chromosomal abnormalities in ML-DS.[11,12,31] The result of G-band karyotyping of TAM-derived cells in 3° recipients was 47,XX,+1, der (1;15)(q10;q10),+21 in 20/20 metaphase cells (supplemental Figure 1C), confirming genomic structural change, which is a hallmark of ML-DS. These data suggest that leukemic evolution of TAM-derived cells was observed in our NOG mouse model.

Genetically heterogeneous subclones with varying repopulating capacity expanded in the TAM-NOG xenograft model

To examine the kinetics of the leukemic evolution of TAM cells, another 2 sets of serial transplantations were performed using the preserved patient 1 sample (Figure 4A). Four of 5 mice in the second group (m2-1–m2-5) and 5 of 11 mice in the third group (m3-1–m3-11) harbored TAM cells from the patient. Of the total of 9 engrafted mice, 2 had the same CNAs detected in the first series of serial transplantations: deletion of 16q22 and 16q24 (m3-5 and m3-8; Figure 3). Moreover, 2 combinations of new CNAs were detected in the 1° recipients: deletion of 9q22 +12p12 (m3-4 and m3-7) and gain of 1q25.2-1q44 (m3-11). No CNAs other than the gain of chromosome 21 were detected in the other recipients (m2-1, m2-2, m2-4, and m2-5).

Each 1° engrafted mouse was subjected to 2° transplantation, and 5 of 9 series (m2-5, m3-4, m3-7, m3-8, and m3-11) successfully gave rise to the xenografts in the 2° recipients. It is noteworthy that the TAM-derived cells of the 2° recipients in 2 of the 3 analyzed series (m2-5 and m3-4) acquired additional CNAs, whereas the CNAs in 2 descendent 2° recipients of m3-8 remained unchanged. The additional CNA of gain of 1q was detected in the 2° recipients of m3-4, similar to that observed in the 3° recipient in Figure 3. Although