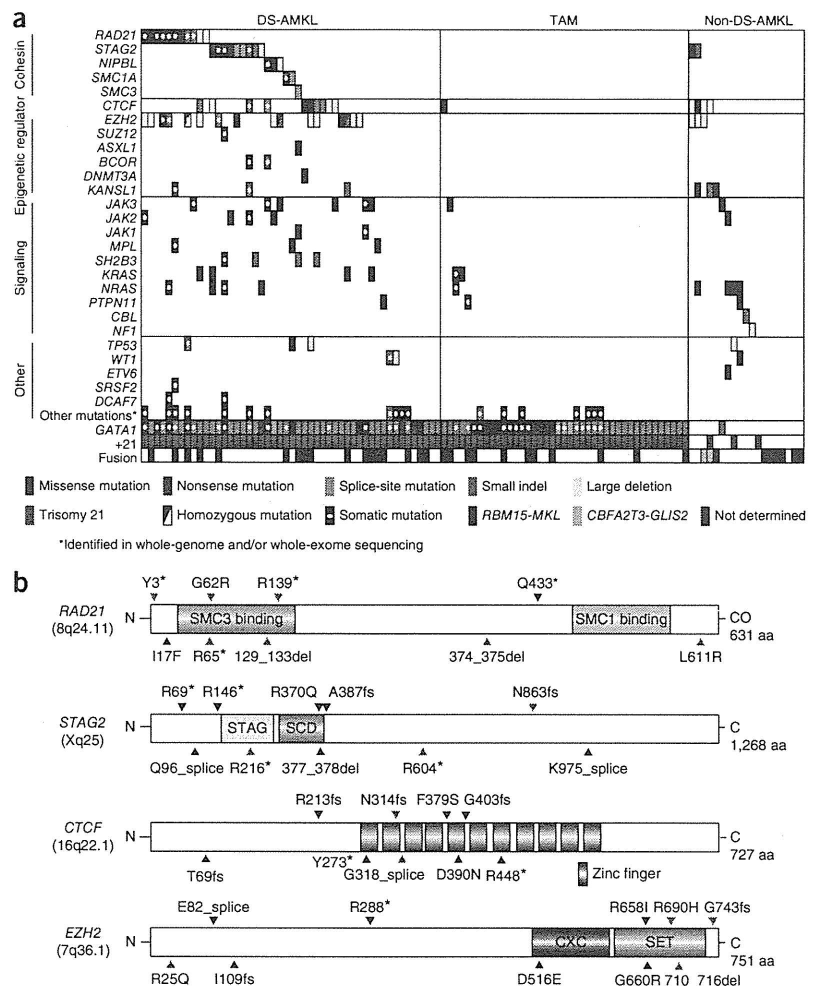


Figure 4 Driver mutations in Down syndrome–related myeloid disorders and non-DS-AMKL. (**a**) Driver mutations in 109 samples of 49 DS-AMKL, 41 TAM and 19 non-DS-AMKL cases. Types of mutations are distinguished by color. Each sample is also described in **Supplementary Table 12**. (**b**) Distribution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

© 2013 Nature America, Inc. All rights reserved.

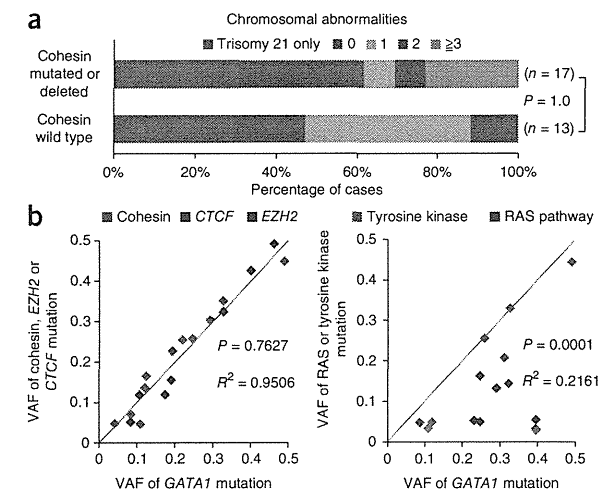


Figure 5 Relationship of cohesin mutations with karyotypes and comparison of mutation loads between major gene targets in DS-AMKL and *GATA1*. (**a**) The number of chromosomal abnormalities is compared between cases with and without cohesin mutations or deletions for DS-AMKL cases. Zero signifies chromosomal 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]. Excluding common *GATA1* mutations, we identified no other recurrent mutations, with only 0.7 non-silent mutations per case, indicating that TAM could be caused by a single acquired *GATA1* mutation in addition to constitutive trisomy 21.

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

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

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

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

METHODS

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

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

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

ACKNOWLEDGMENTS
We thank Y. Mori, M. Nakamura, O. Hagiwara and N. Mizota for their technical assistance. This work was supported by the Research on Measures for Intractable Diseases Project and Health and Labor Sciences Research grants (Research on Intractable Diseases) from the Ministry of Health, Labour and Welfare, by Grants-in-Aid from the Ministry of Health, Labor and Welfare of Japan and

© 2013 Nature America, Inc. All rights reserved.
npg

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.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

© 2013 Nature America, Inc. All rights reserved.

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

ONLINE METHODS

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

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

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

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

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

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

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

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

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

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

where k_s is determined to satisfy

$$k_0 = \sum_j D_j^s$$

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

© 2013 Nature America, Inc. All rights reserved.

doi:10.1038/ng.2759

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

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

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

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

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

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

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

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

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

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

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

54. Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* **45**, 860–867 (2013).
55. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
56. Kent, W.J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
57. Shiraishi, Y. *et al.* An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. *Nucleic Acids Res.* **41**, e89 (2013).
58. Sakaguchi, H. *et al.* Exome sequencing identifies secondary mutations of *SETBP1* and *JAK3* in juvenile myelomonocytic leukemia. *Nat. Genet.* **45**, 937–941 (2013).
59. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
60. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
61. Forbes, S.A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
62. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
63. Torres, L. *et al.* Acute megakaryoblastic leukemia with a four-way variant translocation originating the *RBM15-MKL1* fusion gene. *Pediatr. Blood Cancer* **56**, 846–849 (2011).
64. Nannya, Y. *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res.* **65**, 6071–6079 (2005).
65. Yamamoto, G. *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of Affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am. J. Hum. Genet.* **81**, 114–126 (2007).

© 2013 Nature America, Inc. All rights reserved.

Ann Hematol
DOI 10.1007/s00277-012-1564-5

ORIGINAL ARTICLE

Clinical and genetic characteristics of congenital sideroblastic anemia: comparison with myelodysplastic syndrome with ring sideroblast (MDS-RS)

Rie Ohba · Kazumichi Furuyama · Kenichi Yoshida · Tohru Fujiwara · Noriko Fukuhara · Yasushi Onishi · Atsushi Manabe · Etsuro Ito · Keiya Ozawa · Seiji Kojima · Seishi Ogawa · Hideo Harigae

Received: 30 May 2012 / Accepted: 21 August 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Sideroblastic anemia is characterized by anemia with the emergence of ring sideroblasts in the bone marrow. There are two forms of sideroblastic anemia, i.e., congenital sideroblastic anemia (CSA) and acquired sideroblastic anemia. In order to clarify the pathophysiology of sideroblastic anemia, a nationwide survey consisting of clinical and molecular genetic analysis was performed in Japan. As of January 31, 2012, data of 137 cases of sideroblastic anemia, including 72 cases of myelodysplastic syndrome (MDS)–refractory cytopenia with multilineage dysplasia (RCMD), 47 cases of MDS–refractory anemia with ring sideroblasts (RARS), and 18 cases of CSA, have been collected. Hemoglobin and MCV level in CSA are significantly lower than those of MDS, whereas serum iron level in CSA is significantly higher than those of MDS. Of 14 CSA for which DNA was available for genetic analysis, 10 cases were diagnosed as X-linked sideroblastic anemia due to *ALAS2* gene mutation. The mutation of *SF3B1* gene, which was frequently mutated in MDS-RS, was not detected in CSA patients. Together with the difference of clinical data, it is

Electronic supplementary material The online version of this article (doi:10.1007/s00277-012-1564-5) contains supplementary material, which is available to authorized users.

R. Ohba · T. Fujiwara · N. Fukuhara · Y. Onishi · H. Harigae (✉)
Department of Hematology and Rheumatology,
Tohoku University Graduate School of Medicine,
1-1 Seiryo-machi, Aoba-ku,
Sendai 980-8574, Japan
e-mail: harigae@med.tohoku.ac.jp

K. Furuyama
Department of Molecular Biology and Applied Physiology,
Tohoku University Graduate School of Medicine,
Sendai, Japan

K. Yoshida · S. Ogawa
Cancer Genomics Project, Graduate School of Medicine,
The University of Tokyo,
Tokyo, Japan

T. Fujiwara · H. Harigae
Department of Molecular Hematology/Oncology,
Tohoku University Graduate School of Medicine,
Sendai, Japan

A. Manabe
Department of Pediatrics, St Luke's International Hospital,
Hirosaki University Graduate School of Medicine,
Hirosaki, Japan

E. Ito
Department of Pediatrics,
Hirosaki University Graduate School of Medicine,
Hirosaki, Japan

K. Ozawa
Division of Hematology, Jichi Medical University,
Shimotsuke, Japan

S. Kojima
Department of Pediatrics,
Nagoya University Graduate School of Medicine,
Nagoya, Japan

Published online: 16 September 2012

suggested that genetic background, which is responsible for the development of CSA, is different from that of MDS-RS.

Keywords Congenital sideroblastic anemia · Myelodysplastic syndrome · ALAS2

Introduction

Sideroblastic anemia is characterized by anemia with the emergence of ring sideroblasts in the bone marrow. Ring sideroblasts are formed by the irregular accumulation of iron in mitochondria. There are two forms of sideroblastic anemia i.e., congenital sideroblastic anemia (CSA) and acquired sideroblastic anemia. Most of acquired sideroblastic anemia cases were included in myelodysplastic syndrome (MDS). To date, mutations of genes involved in heme biosynthesis, Fe–S cluster biogenesis, or the biology of mitochondria have been reported in CSA [1–5]. Impaired function of these genes is speculated to result in disutilization of iron, leading to accumulation of iron in mitochondria. Acquired sideroblastic anemia in MDS is categorized either as refractory cytopenia with multilineage dysplasia (RCMD) or refractory anemia with ring sideroblasts (RARS) depending on the level of dysplasia. In contrast CSA, mechanism of forming ring sideroblasts in MDS is not clarified, although it was recently suggested that the mutations of splicing pathway are involved in the pathogenesis of MDS [6]. It is possible that there is a common mechanism between CSA and MDS; however, mutations in genes, which are responsible for development of the CSA, have not been identified in MDS.

The most common form of CSA is X-linked sideroblastic anemia (XLSA), which is caused by mutation of erythroid-specific 5-aminolevulinate synthase (*ALAS2*), the first enzyme of heme synthesis in erythroid cells [7–10]. More than half of the patients with XLSA respond to the administration of pyridoxine [vitamin B6 (Vit.B6)], or pyridoxal 5-phosphate (PLP), which is the coenzyme of *ALAS2* [11]. In XLSA, adult onset cases have been reported [12, 13]; therefore, it is possible that some cases of CSA may be misdiagnosed as MDS, especially RARS. However, the clinical and pathological features of congenital and acquired sideroblastic anemia have not been fully clarified because there have been no comprehensive studies, including clinical and genetic analyses, focusing on sideroblastic anemia.

Here, we performed a nationwide survey of sideroblastic anemia in Japan to investigate the epidemiology and pathogenesis of this disease. The difference of clinical data and results of genetic analysis suggest that genetic background, which is responsible for the development of CSA, is distinct from that of MDS-RS.

Materials and methods

Data acquisition

This study consisted of three investigations. First, patients with sideroblastic anemia were searched by questionnaire sent to hospitals with hematology department (493 hospitals) and pediatric hematology department (593 hospitals) asking for information about patients diagnosed as sideroblastic anemia (first investigation) over the past 10 years. Next, detailed clinical data of sideroblastic anemia patients were collected from the hospital based on responses to the first investigation (second investigation). Survey items were age of onset, gender, family history, hematological and biochemical findings, treatment, and cause of death. Then, genetic analysis of patients, who were diagnosed as CSA and MDS without chromosomal anomaly, was performed in cases for which genome sample was available (third investigation).

This study was approved by the ethics committee of Tohoku University Graduate School of Medicine, the center responsible for clinical and genetic analysis. Informed consent for the genetic analysis was obtained in all cases.

Diagnostic procedure

Ring sideroblasts were defined following the 2001 World Health Organization (WHO) classification. Sideroblastic anemia patients were diagnosed in the respective institutions. In all cases, bone marrow smears were investigated, and at least 15 % ring sideroblasts were confirmed by iron staining. Furthermore, diagnosis for RARS was made when dysplasia restricted to erythroid lineage in bone marrow was recognized. Diagnosis for RCMD was made when there is multilineage dysplasia. Thereafter, in the present study, RCMD correspond to refractory cytopenia with multilineage dysplasia and ringed sideroblasts (RCMD-RS) of the 2001 WHO classification. Diagnosis for CSA was made when the patient had a family history or the disease onset during infancy, or fulfilled the characteristic features of XLSA, such as onset at a young age, microcytic anemia, and responsiveness to Vit.B6.

Genetic analysis of patients with sideroblastic anemia

In the genetic analysis, mutations in *ALAS2*, *SLC25A38*, *GLRX5*, *ABCB7*, *PUS1*, and *SLC19A2*, which are known to be responsible for CSA, were examined in 14 cases of CSA and 10 cases of MDS. In addition, *SF3B1*, which was very recently reported to be mutated in sideroblastic anemia in MDS at a high incidence, were analyzed as well. Mutation analysis for the *ALAS2* gene was performed first in all candidates, and then the analysis proceeded to the other

genes if no mutations in *ALAS2* were detected. For mutation analysis of *ALAS2*, genomic DNA was extracted from the proband's peripheral blood using QIAamp DNA blood midi kit (QIAGEN, Valencia, CA, USA). The proximal promoter region [14], erythroid enhancer in intron 8 [15], and all exons and exon–intron boundaries of the *ALAS2* gene were amplified using ExTaq DNA polymerase (Takara Bio, Shiga, Japan) [16]. Amplified products were purified using a QIAquick gel extraction kit (QIAGEN) after agarose gel electrophoresis. They were then subjected to direct sequencing analysis using BigDye Terminator Cycle sequencing kit v1.1 with an ABI3100 genetic analyzer (Life Technologies Corp., Carlsbad, CA, USA). Mutation of the gene was confirmed by repeated polymerase chain reaction (PCR) followed by direct sequencing analysis. Genes other than *ALAS2* were sequenced by Hiseq2000® [6]. Briefly, genomic DNA was amplified using REPLI-g mini kit® (QIAGEN Science). After adjusting the concentration of amplified DNA, DNA from consecutive 12 samples was combined into one DNA pool, and the entire coding sequences were amplified by primers to which *Not*I linker was attached. The products were digested with *Not*I, and ligated with T4 ligase. Then, DNA was sonicated into ~200-bp fragments, and sequencing libraries were generated. Libraries were subjected to deep sequencing on Hiseq2000®. Sequencing data was analyzes as described previously. Detected mutations were validated by direct sequence.

Analysis of enzymatic activity of recombinant ALAS2 protein

For preparing recombinant ALAS2 proteins, complementary DNA (cDNA) encoding mature human ALAS2 protein was amplified using a following primer set (5′-GGTGGTCATATGATCCACCTTAAGGCAACAAAGG-3′and 5′-GGCATAGGTGGTGACATACTG-3′). The amplified cDNA was then treated with *Nde*I restriction enzyme and was cloned between *Nde*I and blunt-ended *Sap*I site of pTXB1 plasmid (New England Biolabs, Ipswich, MA, USA), resulting in pTXB-AEm. From this plasmid, mature ALAS2 protein was expressed as an inducible fusion protein with Intein and chitin-binding domain in *E. coli*. To obtain the mutant protein, the identified mutation was introduced into pTXB-AEm using PrimeStar Max site-directed mutagenesis kit (Takara Bio, Shiga, Japan). For expression and purification of wild-type and mutant ALAS2 proteins, *E. coli* BL21 (DE3) was transformed with each plasmid. The induction and purification of the recombinant proteins were performed using Impact system (New England Biolabs) according to manufacturer's instruction. Briefly, each recombinant protein was induced in *E. coli* with 0.1 mM IPTG at 25 °C for overnight. Then, cells were resuspended with lysis buffer (20 mM Tris–HCl pH8.5, 500 mM NaCl, 1 mM EDTA, 0.1 % Triton X-100, 1 mM PMSF, 1 μg/ml of antipain, pepstatin, and leupeptin). After the sonication and centrifugation, cleared cell lysates were incubated with chitin beads for 1 h at 4 °C, then washed with wash buffer (20 mM Tris–HCl pH8.5, 500 mM NaCl, 1 mM EDTA, and 0.1 % Triton X-100). Tag-free recombinant mature ALAS2 protein was obtained by on-column cleavage with 50 mM DTT in wash buffer at room temperature for 16 h. After the elution from the column, protein concentration was determined using Bio-Rad Protein assay reagent (Bio-Rad Laboratories, Inc., Hercules, CA, USA). The ALAS activity of each recombinant protein was measured in vitro, as described previously [8].

Statistical analysis

Results are presented as means±SD with the exception of the age of onset, which is expressed as the median. Statistical analysis was performed using Student's *t* test, and $P<0.05$ was taken to indicate statistical significance.

Results

Epidemiology of sideroblastic anemia

As of 31 January 2012, detailed data for 148 sideroblastic anemia, including MDS and secondary sideroblastic anemia, patients have been collected. Excluding 10 cases of refractory anemia with excess blasts (RAEB) and one case of sideroblastic anemia due to alcohol, the remaining 137 cases were classified as 18 cases of CSA, 47 cases of RARS, and 72 cases of RCMD. Of 18 CSA cases, 7 were already confirmed to be XLSA due to mutation of *ALAS2* before registration in this study, and the others were diagnosed as CSA based on family history or clinical findings, including responsiveness to Vit.B6 treatment. Clinical findings and family history, which suggest the porphyria, were not observed in any CSA patients.

Analysis of the pathology of congenital sideroblastic anemia

Laboratory data of CSA, RARS, and RCMD are shown in Tables 1 and 2. Median age at onset of CSA was younger than those of RARS and RCMD (19, 72.5, and 71 years old, respectively). Hemoglobin and mean corpuscular volume (MCV) values of CSA were significantly lower than those of RARS and RCMD cases (7.1 g/dl and 69.0 fl, 8.7 g/dl and 106.8 fl; and 8.3 g/dl and 106.5 fl, respectively). Serum iron level in CSA was significantly higher than that in RARS or RCMD (210.7, 162.8, and 171.1 μg/dl, respectively). These data have possibilities of reflecting the states of the iron over-loaded of CSA; however, as serum iron concentration is very instable and depends from different factors, this finding should be carefully evaluated.

Table 1 Clinical data of CSA, RARS, and RCMD (1)

	CSA (n=18)	RARS (n=47)	RCMD (n=72)	p-value (between CSA and RARS)	p-value (between CSA and RCMD)
Gender					
Male	17	33	44		
Female	1	14	28		
Median age at onset (year)	19.0 (±20.2)	72.5 (±10.4)	71.0 (±13.0)	<0.01	<0.01
White blood cells (/μl)	5547 (±2022)	4741 (±2561)	4105 (±1847)	0.24	<0.01
Red blood cells (×10^4/μl)	383.4 (±100.0)	245.6 (±45.6)	239.4 (±56.4)	<0.01	<0.01
Hemoglobin (g/dl)	7.1 (±2.21)	8.7 (±1.7)	8.3 (±1.8)	<0.01	0.02
Mean corpuscular volume (fl)	69.0 (±11.6)	106.8 (±9.0)	106.5 (±9.2)	<0.01	<0.01
Platelet (×10^4/μl)	28.5 (±12.62)	25.9 (±15.5)	23.9 (±24.1)	0.53	0.44
Reticulocyte (‰)	12.1 (±10.9)	17.7 (±10.8)	21.5 (±20.1)	0.07	0.05

When iron-related laboratory data were examined in transfusion independent cases (CSA, 13; RARS, 26; RCMD, 34), Serum iron level in CSA was tended to be higher than that in RARS or RCMD (210.6, 180.3, and 166.6 μg/dl, respectively), although the difference was not significant (p=0.07, data not shown). Serum ferritin level in CSA, RARS and RCMD were elevated in these transfusion independent cases (1,087.9, 898.1, and 732.2 ng/ml, respectively), suggesting that most of sideroblastic anemia patients were iron-overloaded before transfusion. There were no significant differences in other biochemical data among the three groups.

Chromosomal abnormalities of MDS

Data regarding cytogenetic abnormalities were available for all RARS patients and for 68 of 72 RCMD patients. Figure 1 shows the cytogenetic findings of RARS and RCMD. In RARS cases, chromosomal abnormalities were found in 17 patients (36.2 %). Abnormalities consisted of abnormality including +8 (3 cases), complex abnormality with deletion 5 (2 cases), and complex abnormality with 20q− (3 cases). Chromosomal abnormalities in RCMD were found in bone marrow samples from 27 RCMD patients (39.7 %). Abnormality including +8 was detected in nine cases (33.3 %) and abnormality of idic (X) (q13), associated with the *ABCB7* gene [17], was found in one case. In addition, −7, which was not identified in RARS, was identified in four RCMD patients (14.8 %).

Treatment and outcome

Analysis of the available data regarding treatment indicated that 17 of 47 RARS cases and 26 of 72 RCMD cases were administered Vit.B6 (data not shown). The effectiveness was judged according to the criteria of IWG [18], and one RARS patient obtained a major response, and three RARS patients and one RCMD patient obtained minor responses. Thus, 4 of 17 RARS patients and 1 of 26 RCMD patients responded to Vit.B6 treatment. However, improvement of Hb was not sustained in two RARS patients; Hb level gradually returned to or dropped below the pretreatment level. Therefore, Vit.B6 treatment may not be effective for MDS, or the effect if any may be very limited. The clinical outcomes of patients are shown in Supplemental Table 1. The median follow-up from the time of diagnosis in CSA patients was 30.5 months, and two patients died due to sepsis (one case) and cardiac failure (one case). One patient

Table 2 Clinical data of CSA, RARS, and RCMD (2)

	CSA (n=18)	RARS (n=47)	RCMD (n=72)	p-value (between CSA and RARS)	p-value (between CSA and RCMD)
Total bilirubin (mg/dl)	1.1 (±0.8)	1.3 (±0.9)	1.1 (±0.7)	0.47	0.78
AST (GOT) (IU/l)	33.0 (±24.3)	24.9 (±11.7)	27.9 (±20.8)	0.08	0.38
LDH (IU/l)	218.3 (±98.9)	263.5 (±119.2)	246.1 (±97.7)	0.16	0.28
CRP (mg/dl)	0.13 (±0.15)	0.40 (±1.16)	1.17 (±3.81)	0.37	0.30
Serum iron (mg/dl)	210.7 (±77.6)	162.8 (±73.6)	171.1 (±66.2)	0.03	0.04
UIBC (mg/dl)	80.4 (±113.6)	102.4 (±82.7)	79.9 (±60.7)	0.48	0.93
Ferritin (ng/ml)	1239.8 (±1306.8)	743.4 (±815.3)	804.3 (±990.2)	0.08	0.13

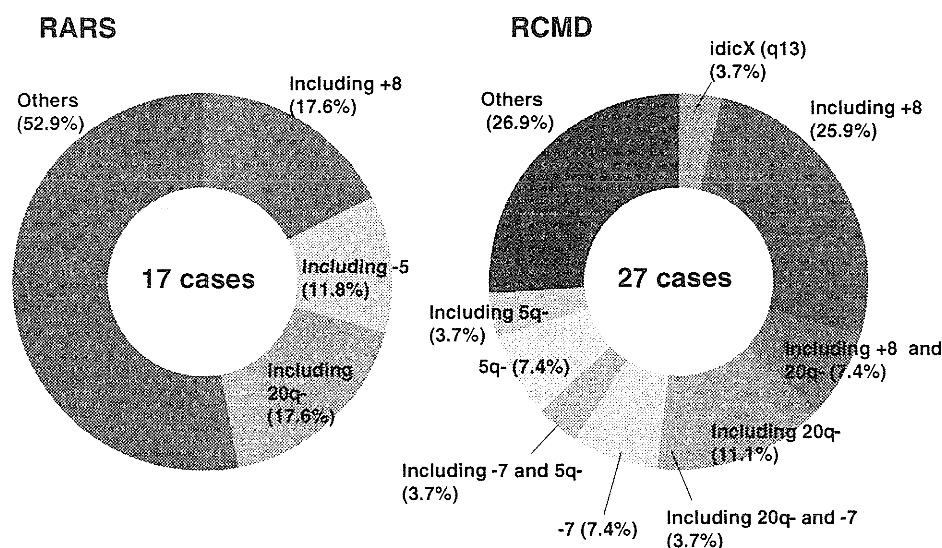


Fig. 1 Chromosomal abnormalities in RARS and RCMD. Data of chromosomal analysis in RARS and RCMD are shown. +8 was most common both in RARS and RCMD. -7 was only seen in RCMD

who died due to cardiac failure was heavily iron overloaded as defined by serum ferritin level, suggesting that cardiac complications may be caused by hemochromatosis. The median follow-up from the time of diagnosis in RARS patients was 23 months, and 6 patients (12.8 %) died due to pneumonia (two cases), evolution to leukemia (one case), and others (three cases). The median follow-up from the time of diagnosis in RCMD patients was 19.5 months, and 20 patients (27.8 %) died due to pneumonia (7 cases), cardiac failure (3 cases), evolution to leukemia (2 cases), sepsis (1 case), and others (7 cases). These results suggest that the prognosis of RCMD is worse than that of RARS.

Gene analysis of congenital sideroblastic anemia

Eighteen CSA patients were candidates for gene analysis; however, mutation analysis for genes responsible for CSA was not performed in four patients. One patient was diagnosed as having PMPS based on clinical findings, and DNA samples were not available for the remaining three patients. Therefore, gene analysis was performed in 14 of 18 CSA patients. Ten of these 14 patients were diagnosed as XLSA due to *ALAS2* mutation. Table 3 summarizes the results of gene analysis in XLSA. Case 2 (R411C), case 4 (D190V), case 6 (M567I), and case 7 (V562A) were reported previously [19–21]. Since amino acid substitution at Arg170, 411, and 452 were observed in plural patients, there are hot spots of mutation of *ALAS2* gene.

Patient with D190V (case 4), R170L (Case 10) and two patients with R452C (cases 3 and 5) did not respond to Vit.B6 treatment, whereas six patients responded to Vit.B6 treatment, although the increment of hemoglobin varied from 1.7 to 8.1 g. Interestingly, case 8 responded to Vit.B6 treatment, whereas case 10 did not, although both of them harbor the same mutation, R170L. Therefore, the activity of R170L mutant proteins was examined to determine the property, especially the Vit.B6 responsiveness. The enzymatic activities of wild type and R170L mutant protein were 7,193±138 nmol ALA/mg protein/h and 2,240±145 nmol ALA/mg protein/h, respectively, in the absence of PLP (Fig. 2). With an excess amount of PLP (100 μM) in the assay mixture, higher enzymatic activities were obtained with wild-type and mutant proteins (12,662±311 nmol ALA/mg protein/h and 7,700± 49 nmol ALA/mg protein/h, respectively) (Fig. 2). In addition, the enzymatic activity of R170C, which is another substitution at Arg170 found in this study, was also examined. As shown in Fig. 2, The enzymatic activity of mutant protein was significantly lower than wild-type protein without PLP (4,612± 87 nmol ALA/mg protein/h vs 7,193±138 nmol ALA/mg protein/h), and the activity was restored by addition of excess amount of PLP (100 μM) in the assay mixture. These in vitro data suggest that amino acid substitution at Arg 170, at least R170L and R170C, results in the decrease in enzymatic activity, but the decrease can be recovered by excess amount of PLP. The enzymatic activity of mutant proteins, which were identified in this study, is summarized in Table 3. The enzymatic activities of R411C, D190V, M567I, and V562A were referred from previous reports [19–21]. The levels of activity and PLP responsiveness in vitro were not correlated with clinical responsiveness to PLP in some cases. It is possible that the variety of mechanisms, such as the decrease in enzymatic activity of mutant ALAS2 protein, the changes of amount of *ALAS2* transcript, and physiological and environmental status of the patients, are responsible for the development of the disease.

Data for CSA patients other than XLSA are summarized in Table 4. Case 15 was diagnosed as PMPS. Gene analysis was not performed for cases 16 and 17; however, XLSA was strongly suspected because these patients were male and had microcytic anemia that was responsive to Vit.B6 treatment.

Table 3 Congenital sideroblastic anemia (XLSA)

Case number	Age at diagnosis (y.o.)	Gender	Position of *ALAS2* mutation	*SF3B1* mutation	Hb at onset (g/dl)	MCV at onset (fl)	Increment of Hb by Vit.B6 treatment (g/dl)	In vitro enzymatic activity of mutant protein[a]	
								Without PLP	With PLP
1	0	M	R170C	N/D	4.8	52.5	1.7	64.1 %	72.5 %[b]
2	20	M	R411C	N/D	4.8	52.5	5.2	11.9 %	25.1 % [19]
3	68	M	R452C	–	6.0	67.3	No effect	99.9 %	94.0 % [21]
4	17	M	D190V	N/D	8.9	66.9	No effect	98.6 %	98.5 % [20]
5	36	M	R452C	–	7.4	70.0	No effect	99.9 %	94.0 % [21]
6	36	M	M567I	N/D	6.5	64.4	3.4	38.1 %	25.2 % [21]
7	14	M	V562A	–	8.1	61.2	4.7	150.6 %	116.9 % [21]
8	31	M	R170L	–	4.1	50.8	8.1	31.1 %	60.8 %[b]
9	3	M	R411C	–	5.4	54.4	2.9	11.9 %	25.1 % [19]
10	62	M	R170L	N/D	8.0	73.9	No effect	31.1 %	60.8 %[b]

[a] % of WT

[b] Present study

ALAS2 mutations were not identified in cases 11, 12, 13, and 14. Therefore, mutations of *SLC25A38*, *GLRX5*, *ABCB7*, *PUS1*, *SLC19A2*, and *SF3B1* were examined in these cases; however, no mutations were identified in these cases. In contrast to other cases, case 18 was female and showed normocytic anemia. She was diagnosed with CSA due to family history; however, gene mutation analysis was not performed because DNA samples were not available. *SF3B1* gene mutation was examined in nine cases including five XLSA, however, no mutation was identified (Tables 3 and 4). On the other hand, *SF3B1* gene mutation was frequently detected in MDS-RS (Table 5).

Discussion

Because of its rarity, there have been few clinical and pathological investigations focusing on sideroblastic anemia. This study was performed to investigate the epidemiological and pathological characteristics of sideroblastic anemia. Based on the data of 137 patients, it was revealed that hemoglobin level in CSA was significantly lower than those seen in MDS, and serum iron level was higher in CSA compared to MDS. These results revealed that anemia in CSA is more severe than that in MDS at onset, although significant cases improved by Vit.B6 treatment. Reflecting the high incidence of XLSA in CSA, MCV level was significantly lower in CSA than MDS. These findings suggest that CSA should be strongly suspected rather than MDS, at least in Japan, in male patients exhibiting microcytic anemia and an elevated serum iron level.

MDS-RCMD is the most common form of acquired sideroblastic anemia. Chromosomal abnormalities were observed in 39.7 % of RCMD cases and 36.2 % of RARS cases. The types of chromosomal abnormality frequently observed in RCMD and RARS did not differ from those reported previously, such as +8, −7, 20q− and −5. Among them, +8 was observed in nine cases of RCMD (33.3 %). As the frequency of +8 in MDS was reported to be 6.5–16.7 %,

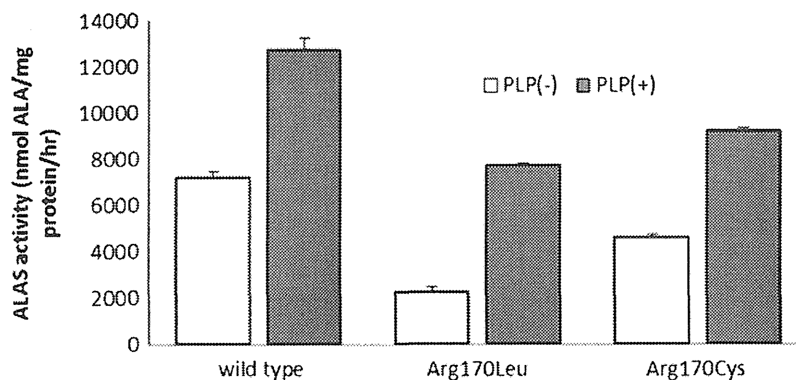


Fig. 2 Enzymatic activity of mutant ALAS2 proteins. Enzymatic activity of wild-type and mutant ALAS2 proteins was measured as described in Materials and Methods. Both of R170L and R170C ALAS2 mutant proteins showed decreased enzymatic activity; however, the activity was partially restored by the addition of PLP

Table 4 Congenital sideroblastic anemia (other than XLSA)

Case number	Age at diag (y.o.)	Gender	Family history	Gene mutation							Hb (g/dl)	MCV (fl)	Response to Vit.B6
				ALAS2	*SLC25A38*	*GLRX5*	*ABCB7*	*SLC19A2*	*PUS1*	*SF3B1*			
11	19	M	–	–	–	–	–	–	–	–	7.8	73.9	–
12	4	M	–	–	–	–	–	–	–	–	6.6	73.6	–
13	0	M	+	–	–	–	–	–	–	–	3.9	65.0	–
14	20	M	+	–	–	–	–	–	–	–	7.6	82.0	+
15	0	M	–	N/D	N/D	N/D	N/D	N/D	N/D	N/D	6.8	88.1	N/D[a]
16	32	M	–	N/D	N/D	N/D	N/D	N/D	N/D	N/D	11.2	69	+
17	36	M	–	N/D	N/D	N/D	N/D	N/D	N/D	N/D	10.8	67.3	+
18	18	F	+	N/D	N/D	N/D	N/D	N/D	N/D	N/D	9.3	96.2	+

N/D not done

[a] Vit.B6 was not administered due to PMPS

+8 appeared to be more common in RCMD. In addition, −7 was identified in four patients with RCMD (14.8 %), whereas it was not identified in RARS. This difference may be related to the poor prognosis of RCMD.

Regarding the responsiveness to pyridoxine treatment among XLSA, 6 of 10 cases responded to Vit.B6 treatment in this study, although the magnitude of response varied among individuals. Thus, as the benefit of treatment of Vit.B6 for XLSA is obvious, a precise diagnosis of XLSA is important. As late-onset XLSA cases have been reported and two patients over 60 years old were found in this study, genetic analysis in sideroblastic anemia patients with microcytic anemia is essential regardless of age.

Focusing on *ALAS2* mutation in XLSA, two patients with the same mutation (c.509G>T), which results in R170L, showed distinct responses to Vit.B6. Edgar et al. [22] reported a Vit.B6 responsive pedigree with XLSA carrying the p.R170L mutation of ALAS2 gene. Furthermore, the crystal structure analysis of ALAS from *Rhodobacter capsulatus* [23] suggests that a missense mutation at Arg170 destabilizes PLP binding, which might be partially restored with excess amounts of PLP. Together with the findings of biochemical analysis in this study, it is strongly suggested that R170L mutation causes pyridoxine-responsive XLSA. However, in consistent with the data of in vitro analysis and clinical course of other R170L patients, case 10 was unresponsive to Vit.B6 treatment. Thus, onset and severity of the disease may be defined by not only the type of mutation but also the environmental and physiological status of the patients. This speculation may be supported by the results that there is a discrepancy between in vitro and in vivo response to Vit.B6 in some cases (Table 3).

The high incidence of XLSA among CSA in the present study was consistent with a previous report in the USA. Bergmann et al. [24] reported genetic analysis of CSA in the USA. In this study, mutations of *ALAS2*, *SLC25A38*, mitochondria DNA, and *PUS1*, were identified in 37, 15, 2.5, and 2.5 % of CSA cases, respectively. The most significant difference from our study was that mutations of the *SLC25A38* gene were frequently found in the USA. Since *SLC25A38* is thought to be a transporter of glycine, which is a substrate for ALAS2 in the first step of heme synthesis, the

Table 5 Mutation of *SF3B1* gene in MDS-RS

Case number	Diagnosis	Age at diagnosis (y.o.)	Gender	Chromosome anomaly	position of *SF3B1* mutation
1	RARS	82	M	–	E622D
2	RARS	57	M	–	N626S
3	RARS	60	M	Complex karyotype, including +8	K700E
4	RARS	60	M	–	K700E
5	RARS	73	F	–	No mutation
6	RARS	74	F	–	H662Q
7	RARS	76	M	–	K700E
8	RARS	67	F	–	K700E
9	RARS	66	M	–	K666E
10	RCMD	50	F	–	No mutation

(–) normal karyotype

pathology of CSA due to mutation of this gene is similar to that of XLSA. Therefore, CSA patients with microcytic anemia, in whom mutations of *ALAS2* gene were not identified, were expected to harbor *SLC25A38* mutation; however, it was not detectable in this study. To date, it has not been reported in Asia, although mutation of the *SLC25A38* gene has been widely reported in the USA, Canada, and Europe. Together with the results of the present study, it is suggested that the causative genes of CSA differ among races and regions.

Recently, mutations of genes involved in splicing machinery were reported in MDS [6]. Among them, *SF3B1*, which is a component of the U2-small nuclear ribonucleoprotein (U2-snRNP) complex [25], was found to be highly mutated in MDS with ring sideroblasts [6]. In this study, *SF3B1* mutation was examined in nine cases of CSA; however, its mutation was not detectable in CSA. These findings suggest that the mechanism for sideroblasts formation may be different between CSA and MDS.

In conclusion, our data showed that XLSA is the most frequent type of CSA; however, onset and severity of the disease may be affected by the environmental and physiological status of the patients. The data, including clinical and genetic analysis, further suggest that genetic background is different between CSA and MDS.

Acknowledgements We thank the doctors participating in this study for providing clinical data and samples. This research was supported by a Grant-in-Aid for Scientific Research from the Ministry of Health, Labor, and Welfare of Japan.

Conflict of interest disclosure The authors declare no competing financial interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Cotter PD, Baumann M, Bishop DF (1992) Enzymatic defect in "X-linked" sideroblastic anemia: molecular evidence for erythroid delta-aminolevulinate synthase deficiency. Proc Natl Acad Sci USA 89(9):4028–4032
2. Camaschella C, Campanella A, De Falco L, Boschetto L, Merlini R, Silvestri L, Levi S, Iolascon A (2007) The human counterpart of zebrafish shiraz shows sideroblastic-like microcytic anemia and iron overload. Blood 110(4):1353–1358
3. Allikmets R, Raskind WH, Hutchinson A, Schueck ND, Dean M, Koeller DM (1999) Mutation of a putative mitochondrial iron transporter gene (ABC7) in X-linked sideroblastic anemia and ataxia (XLSA/A). Hum Mol Genet 8(5):743–749
4. Guernsey DL, Jiang H, Campagna DR, Evans SC, Ferguson M, Kellogg MD, Lachance M, Matsuoka M, Nightingale M, Rideout A, Saint-Amant L, Schmidt PJ, Orr A, Bottomley SS, Fleming MD, Ludman M, Dyack S, Fernandez CV, Samuels ME (2009) Mutations in mitochondrial carrier family gene SLC25A38 cause nonsyndromic autosomal recessive congenital sideroblastic anemia. Nat Genet 41(6):651–653
5. Pearson HA, Lobel JS, Kocoshis SA, Kocoshis SA, Naiman JL, Windmiller J, Lammi AT, Hoffman R, Marsh JC (1979) A new syndrome of refractory sideroblastic anemia with vacuolization of marrow precursors and exocrine pancreatic dysfunction. J Pediatr 95(6):976–984
6. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, Shiosaka M, Kawahata R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Ishiyama K, Mori H, Nolte F, Hofmann WK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih LY, Haferlach T, Chiba S, Nakauchi H, Miyano S, Ogawa S (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478(7367):64–69
7. Riddle RD, Yamamoto M, Engel JD (1989) Expression of delta-aminolevulinate synthase in avian cells: separate genes encode erythroid-specific and nonspecific isozymes. Proc Natl Acad Sci USA 86(3):792–796
8. Furuyama K, Harigae H, Heller T, Hamel BC, Minder EI, Shimizu T, Kuribara T, Blijlevens N, Shibahara S, Sassa S (2006) Arg452 substitution of the erythroid-specific 5-aminolaevulinate synthase, a hot spot mutation in X-linked sideroblastic anaemia, does not itself affect enzyme activity. Eur J Haematol 76(1):33–41
9. Harigae H, Furuyama K, Kimura A, Neriishi K, Tahara N, Kondo M, Hayashi N, Yamamoto M, Sassa S, Sasaki T (1999) A novel mutation of the erythroid-specific delta-aminolaevulinate synthase gene in a patient with X-linked sideroblastic anaemia. Br J Haematol 106(1):175–177
10. Harigae H, Furuyama K, Kudo K, Hayashi N, Yamamoto M, Sassa S, Sasaki T (1999) A novel mutation of the erythroid-specific gamma-Aminolevulinate synthase gene in a patient with non-inherited pyridoxine-responsive sideroblastic anemia. Am J Hematol 62(2):112–114
11. Harigae H, Furuyama K (2010) Hereditary sideroblastic anemia: pathophysiology and gene mutations. Int J Hematol 92(3):425–431
12. Furuyama K, Harigae H, Kinoshita C, Shimada T, Miyaoka K, Kanda C, Maruyama Y, Shibahara S, Sassa S (2003) Late-onset X-linked sideroblastic anemia following hemodialysis. Blood 101 (11):4623–4624
13. Cotter PD, May A, Fitzsimons EJ, Houston T, Woodcock BE, Al-Sabah AI, Wong L, Bishop DF (1995) Late-onset X-linked sideroblastic anemia. Missense mutations in the erythroid delta-aminolevulinate synthase (ALAS2) gene in two pyridoxine-responsive patients initially diagnosed with acquired refractory anemia and ringed sideroblasts. J Clin Invest 96(4):2090–2096
14. Surinya KH, Cox TC, May BK (1997) Transcriptional regulation of the human erythroid 5-aminolevulinate synthase gene. Identification of promoter elements and role of regulatory proteins. J Biol Chem 272(42):26585–26594
15. Surinya KH, Cox TC, May BK (1998) Identification and characterization of a conserved erythroid-specific enhancer located in intron 8 of the human 5-aminolevulinate synthase 2 gene. J Biol Chem 273(27):16798–16809
16. Furuyama K, Fujita H, Nagai T, Yomogida K, Munakata H, Kondo M, Kimura A, Kuramoto A, Hayashi N, Yamamoto M (1997) Pyridoxine refractory X-linked sideroblastic anemia caused by a point mutation in the erythroid 5-aminolevulinate synthase gene. Blood 90(2):822–830
17. Sato K, Torimoto Y, Hosoki T, Ikuta K, Takahashi H, Yamamoto M, Ito S, Okamura N, Ichiki K, Tanaka H, Shindo M, Hirai K, Mizukami Y, Otake T, Fujiya M, Sasaki K, Kohgo Y (2011) Loss of ABCB7 gene: pathogenesis of mitochondrial iron accumulation in erythroblasts in refractory anemia with ringed sideroblast with isodicentric (X)(q13). Int J Hematol 93(3):311–318

18. Cheson BD, Bennett JM, Kantarjian H, Pinto A, Schiffer CA, Nimer SD, Löwenberg B, Beran M, de Witte TM, Stone RM, Mittelman M, Sanz GF, Wijermans PW, Gore S, Greenberg PL, World Health Organization(WHO) international working group (2000) Report of an international working group to standardize response criteria for myelodysplastic syndromes. Blood 96(12):3671–3674
19. Furuyama K, Uno R, Urabe A, Hayashi N, Fujita H, Kondo M, Sassa S, Yamamoto M (1998) R411C mutation of the ALAS2 gene encodes a pyridoxine-responsive enzyme with low activity. Br J Haematol 103(3):839–841
20. Furuyama K, Fujita H, Nagai T, Yomogida K, Munakata H, Kondo M, Yomogida K, Munakata H, Kondo M, Kimura A, Kuramoto A, Hayashi N, Yamamoto M (1997) Pyridoxine refractory X-linked sideroblastic anemia caused by a point mutation in the erythroid 5-aminolevulinate synthase gene. Blood 90:822–830
21. Kadirvel S, Furuyama K, Harigae H, Kaneko K, Tamai Y, Ishida Y, Shibahara S (2012) The carboxyl-terminal region of erythroid-specific 5-aminolevulinate synthase acts as an intrinsic modifier for its catalytic activity and protein stability. Exp Hematol 40:477–86
22. Edgar AJ, Vidyatilake HM, Wickramasinghe SN (1998) X-linked sideroblastic anaemia due to a mutation in the erythroid 5-aminolaevulinate synthase gene leading to an arginine170 to leucine substitution. Eur J Haematol 61:55–58
23. Astner I, Schulze JO, van den Heuvel J, Jahn D, Schubert WD, Heinz DW (2005) Crystal structure of 5-aminolevulinate synthase, the first enzyme of heme biosynthesis, and its link to XLSA in humans. EMBO J 24:3166–3177
24. Bergmann AK, Campagne DR, McLoughlin EM, Agarwal S, Fleming MD, Bottomley SS, Neufeld EJ (2010) Systemic molecular genetic analysis of congenital sideroblastic anemia: evidence for genetic heterogeneity and identification of novel mutations. Pediatr Blood Cancer 54(2):271–278
25. Das BK, Xia L, Palandjian L, Gozani O, Chyung Y, Reed R (1999) Characterization of a protein complex containing spliceosomal proteins SAPs 49, 130, 145, and 155. Mol Cell Biol 19(10):6796–6802

Identification of a novel erythroid-specific enhancer for the *ALAS2* gene and its loss-of-function mutation which is associated with congenital sideroblastic anemia

Kiriko Kaneko,[1,2] Kazumichi Furuyama,[1,6] Tohru Fujiwara,[3] Ryoji Kobayashi,[4] Hiroyuki Ishida,[5] Hideo Harigae,[3] and Shigeki Shibahara[1]

[1]Department of Molecular Biology and Applied Physiology, [2]Department of Endocrinology and Applied Medical Science, [3]Department of Hematology and Rheumatology, Tohoku University Graduate School of Medicine, Sendai, Miyagi; [4]Department of Pediatrics, Sapporo Hokuyu Hospital, Sapporo; [5]Department of Pediatrics, Kyoto Prefectural University of Medicine, Graduate School of Medical Science, Kyoto; and [6]Laboratory of Molecular Biochemistry, Iwate Medical University, Yahaba, Iwate, Japan

ABSTRACT

Erythroid-specific 5-aminolevulinate synthase (ALAS2) is the rate-limiting enzyme for heme biosynthesis in erythroid cells, and a missense mutation of the *ALAS2* gene is associated with congenital sideroblastic anemia. However, the gene responsible for this form of anemia remains unclear in about 40% of patients. Here, we identify a novel erythroid-specific enhancer of 130 base pairs in the first intron of the *ALAS2* gene. The newly identified enhancer contains a *cis*-acting element that is bound by the erythroid-specific transcription factor GATA1, as confirmed by chromatin immunoprecipitation analysis *in vivo* and by electrophoretic mobility shift assay *in vitro*. A promoter activity assay in K562 human erythroleukemia cells revealed that the presence of this 130-base pair region increased the promoter activity of the *ALAS2* gene by 10-15-fold. Importantly, two mutations, each of which disrupts the GATA-binding site in the enhancer, were identified in unrelated male patients with congenital sideroblastic anemia, and the lower expression level of ALAS2 mRNA in bone marrow erythroblasts was confirmed in one of these patients. Moreover, GATA1 failed to bind to each mutant sequence at the GATA-binding site, and each mutation abolished the enhancer function on ALAS2 promoter activity in K562 cells. Thus, a mutation at the GATA-binding site in this enhancer may cause congenital sideroblastic anemia. These results suggest that the newly identified intronic enhancer is essential for the expression of the *ALAS2* gene in erythroid cells. We propose that the 130-base pair enhancer region located in the first intron of the *ALAS2* gene should be examined in patients with congenital sideroblastic anemia in whom the gene responsible is unknown.

Introduction

The *ALAS2* gene encodes for erythroid-specific 5-aminolevulinate synthase (ALAS-E, EC 2.3.1.37), which is the rate-limiting enzyme of the heme biosynthetic pathway in erythroid cells.[1] It has been reported that the human *ALAS2* gene is mapped on the X chromosome,[2] and that a loss-of-function mutation of this gene causes X-linked sideroblastic anemia (XLSA),[3,4] which is the most common genetic form of congenital sideroblastic anemia (CSA). Moreover, a missense mutation of *ALAS2* was identified in a patient with non-familial CSA (nfCSA),[5] in which no family history of sideroblastic anemia was identified. In addition to *ALAS2*, several other genes were recently identified as causative genes for CSA, including *SLC25A38*,[6] *GLRX5*,[7] *ABCB7*,[8] *PUS1*,[9] and *SLC19A2*,[10] but the cause of sideroblastic anemia still remains undefined in more than 40% of patients with CSA.[11]

GATA1 transcription factor regulates the expression of several erythroid-specific genes, such as erythropoietin receptor gene,[12,13] α- and β-globin genes,[14,15] *ALAS2*[16] and the *GATA1* gene itself,[17] during erythroid differentiation.[18,19] Ablation of the *Gata1* gene in mice resulted in embryonic death because of anemia,[20] suggesting that GATA1 is essential for erythroid differentiation *in vivo*. It has been reported that GATA1 regulates transcription of human *ALAS2* through the proximal promoter region[16] and the erythroid-specific enhancer located in the eighth intron of *ALAS2*.[21] However, Fujiwara *et al.* demonstrated that the GATA1 protein binds to the *ALAS2* gene only in the middle of its first intron, where no regulatory region had so far been identified, by genome-wide analysis of K562 human erythroleukemia cells using chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq).[22]

In the present study, we have identified a novel erythroid-specific enhancer region in the first intron of the *ALAS2* gene. Moreover, we describe two mutations in the newly identified enhancer of *ALAS2*: a T-to-C transition, which changes G<u>A</u>TA to G<u>G</u>TA at the GATA element in the antisense strand, in a pedigree with XLSA and one proband with nfCSA, and a 35-base pair (bp) deletion including the above-mentioned GATA element in a proband with nfCSA.

Methods

Polymerase chain reaction

DNA polymerases used for polymerase chain reaction (PCR) analysis were purchased from TAKARA BIO Inc. (Shiga, Japan). The

©2013 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2013.085449
The online version of this article has a Supplementary Appendix.
Manuscript received on February 4, 2013. Manuscript accepted on August 1, 2013.
Correspondence: furuyama@iwate-med.ac.jp

sequence of primers and probes used in this study are listed in the *Online Supplementary Tables*.

Polymerase chain reaction-based quantitative chromatin immunoprecipitation

Real-time PCR-based quantitative chromatin immunoprecipitation (ChIP-qPCR) analysis was conducted essentially as previously described.[22]

Electrophoretic mobility shift assay

Electrophoretic mobility shift assay (EMSA) was performed using "DIG Gel Shift Kit, 2nd Generation" (Roche Diagnostics GmbH, Mannheim, Germany), according to the manufacturer's protocol. Sequences of oligonucleotides for probes are indicated by the horizontal bar in the relevant figures. Nuclear extracts were prepared, as described previously,[23] from K562 cells or HEK293 human embryonic kidney cells that were transfected with a GATA1-FLAG fusion protein expression vector or its backbone vector.

Promoter/enhancer activity assays

Each target DNA fragment was prepared from genomic DNA from normal volunteers (WT) or patients with CSA (referred to as "GGTA" or "delGATA" in each reporter construct) and was cloned into pGL3basic plasmid (Promega Corporation, Madison, WI, USA). The human ALAS2 proximal promoter region (g.4820_5115, between -267 and +29 from the transcription start site)[16,24] was cloned into the multiple-cloning site of pGL3basic [referred to as pGL3-AEpro(-267)]. A single DNA fragment (5.2 kbp), carrying the ALAS2 proximal promoter, first exon, first intron and the untranslated region of the second exon, was subcloned into the multiple cloning site of pGL3basic [referred to as pGL3-AEpro(-267)+intron1]. A DNA fragment containing the GATA1-binding region in the first intron of the *ALAS2* gene (corresponding to g.7488_7960), which was defined by ChIP-seq analysis,[22] is referred to as the ChIP-peak. The length of the WT ChIP-peak is 473 bp. In addition, a 130-bp fragment containing ALAS2int1GATA, the consensus sequence for the GATA1-binding site in the ChIP-peak, is referred to as ChIPmini. Several deletion mutants of ChIPmini were prepared using pGL3-AEpro(-267)+ChIPmini(WT) as a template. The pGL3-TKpro plasmid was constructed by cloning herpes simplex virus thymidine kinase promoter into the multiple cloning site of pGL3basic plasmid. Each reporter vector and pEF-RL25 were introduced into K562 cells or HEK293 cells. Luciferase activity was determined using a dual-luciferase reporter system (Promega).

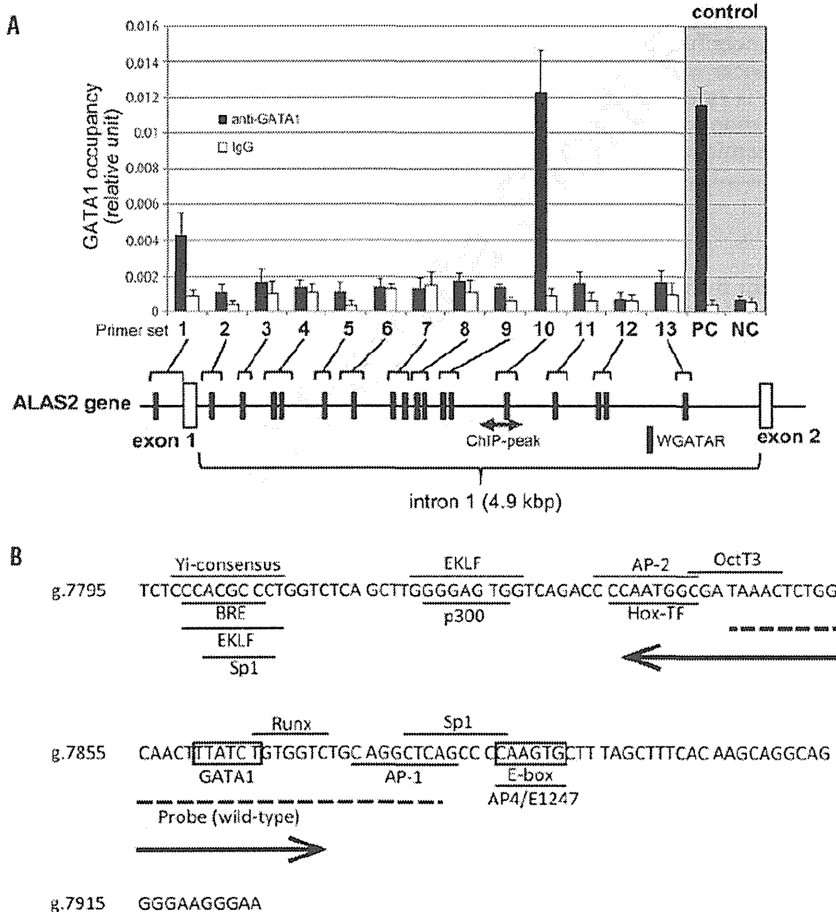


Figure 1. Identification of a functional GATA1 element in the first intron of the *ALAS2* gene. (A) Chromatin immunoprecipitation assay. Fragmented genomic DNA segments were immunoprecipitated with anti-GATA1 antibody or control IgG, and then precipitated fragments were quantified using real-time PCR as described in the *Online Supplementary Methods*. PC or NC indicates positive control or negative control, respectively, for the ChIP assay using anti-GATA1 in K562 cells.[22] One GATA element is present in the proximal promoter region and 17 GATA elements in the first intron (black symbols). The shaded double arrow indicates the region corresponding to ChIP-peak. (B) Nucleotide sequence of ChIPmini. The GATA binding site, ALAS2int1GATA, is located in the center of ChIPmini (boxed). A box also indicates the consensus for E-box that is bound by Scl/TAL1.[22] The sequence of ChIPmini was further analyzed for putative transcription factor binding sites using GeneQuest software (DNASTAR Inc., Madison, WI, USA), and the results are indicated by the horizontal bar. Yi-consensus, Yi transcription factor consensus site;[33] BRE, transcription factor IIB binding site;[34] EKLF, erythroid/Kruppel-like factor consensus site;[35] Sp1, stimulatory protein 1 binding site;[36] P300, P300 transcriptional coactivator consensus site;[37] AP-2, AP-2 beta consensus site;[38] Hox-TF, C1 element binding factor binding site;[39] OctT3, OctT3 binding site;[40] Runx, Runx proteins binding site;[41] AP-1, activator protein 1 binding site;[42] and AP4/E1247, AP4/E1247 binding site.[43] The sequence for the wild-type probe used in the EMSA is indicated by a dashed line. A double arrow indicates the deleted region of the delGATA mutation.

Identification of mutations of the ALAS2 gene

All exons including exon-intron boundaries, the proximal promoter region, and intron 1 and intron 8 of the *ALAS2* gene (GeneBank: NG_8983.1) were directly sequenced according to previously reported methods.[26]

Measurement of ALAS2 mRNA in purified erythroblasts

Total RNA was extracted from glycophorin A-positive bone marrow mononuclear cells, and was used for cDNA synthesis. ALAS2 expression was measured by real-time PCR, and was normalized to that of GAPDH mRNA.

Statistical analysis

Multiple comparisons between groups were made using the Tukey-Kramer test.

Patients

Eleven probands (eight pedigrees) with CSA of unknown cause were selected to determine the nucleotide sequence of the first intron of *ALAS2* gene. In these patients no disease-causative mutation was identified in the coding regions or reported regulatory regions in *ALAS2, SLC25A38, GLRX5, ABCB7, PUS1* and *SLC19A2*, which have been reported to be genes causing CSA[11] (see the *Online Supplementary Methods* for full details of the methods).

The genetic analyses performed in this project were approved by the ethical committee of Tohoku University School of Medicine. Blood samples were withdrawn from the probands and the family members after informed consent.

Results

Polymerase chain reaction-based quantitative chromatin immunoprecipitation analysis of the first intron of the ALAS2 gene

To identify the novel regulatory region for ALAS2 tran-

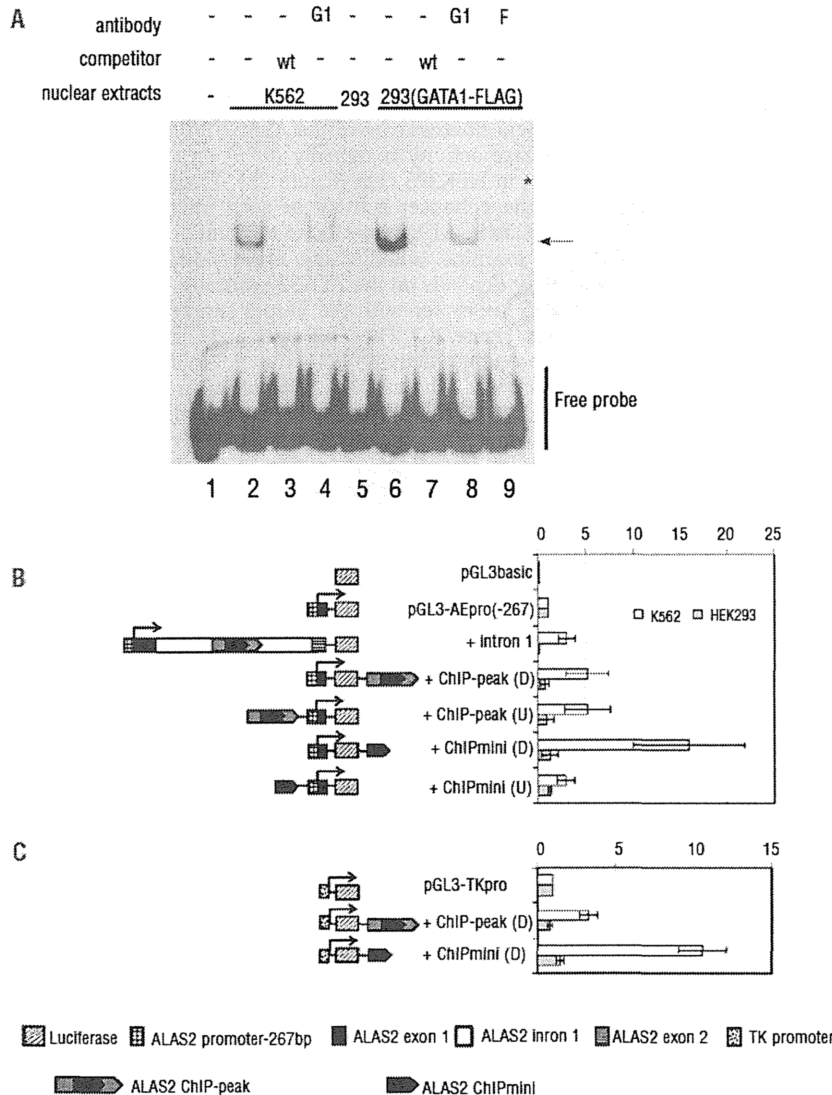


Figure 2. **Functional analyses of ChIPmini present in the first intron of the *ALAS2* gene.** (A) Electrophoretic mobility shift assay (EMSA). Wild-type (wt) probe was incubated with nuclear extracts prepared from K562 cells (lanes 2-4) or HEK293 cells expressing GATA1-FLAG (lanes 6-9). HEK293 cells were transfected with mock vector (lane 5) or FLAG-fused GATA1 expression vector before preparation of nuclear extracts. The protein-probe complex was detected as a retarded band (arrow). An excess amount of unlabeled probe (lanes 3, 7), anti-GATA1 antibody (G1) (lanes 4, 8) or anti-FLAG antibody (F) (lane 9) was included in the reaction mixture. Lane 1 shows the control without nuclear extracts. The asterisk indicates the super-shifted band (lane 9). (B) Functional analysis of ChIPmini as an enhancer for the *ALAS2* gene. Details of the fragments for each plasmid, such as Intron1, ChIP-peak and ChIPmini, are described in the *Methods* section. Each DNA fragment was inserted upstream of the *ALAS2* proximal promoter or downstream of luciferase cDNA, indicated as (U) or (D), respectively. Results are expressed as a relative activity compared to that of pGL3-AEpro(-267), and are presented as the mean ± standard deviation (SD) of three independent experiments. (C) Functional analysis of ChIPmini as an enhancer for non-erythroid gene promoter. The enhancer activity of the first intron was examined using the herpes simplex virus TK promoter as a non-erythroid promoter. ChIP-peak or ChIPmini was inserted downstream of the luciferase gene of pGL3-TKpro, yielding pGL3-TKpro+ChIP-peak(D) or pGL3-TKpro+ChIPmini(D). Each of these reporter vectors was introduced into K562 cells or HEK293 cells to measure enhancer activity. Results are expressed as a relative activity compared to that of pGL3-TKpro, and are presented as the mean ± SD of three independent experiments.

scription, we first performed ChIP-qPCR analysis in K562 cells to localize the GATA1-binding region of the *ALAS2* gene *in vivo*, which was determined by genome-wide ChIP-seq analysis.[22] In fact, ChIP-qPCR enabled us to examine the GATA1-binding activity of an individual GATA element or two adjacent GATA elements in the first intron of the *ALAS2* gene. Based on a search of NCBI Reference Sequence (NG_8983.1) using SeqBuilder software (DNASTAR Inc., Madison, WI, USA), we identified 17 GATA elements (16 out of 17 GATA elements are present in the antisense orientation) in the first intron of human *ALAS2* (Figure 1A), which is compatible with the previous report.[21] We also included the proximal promoter region that contains a functional GATA-binding site (g.4961_4966).[16] Overall 13 primer sets were designed to amplify the GATA elements located in the proximal promoter region and the first intron of *ALAS2* (Figure 1A and *Online Supplementary Table S1*). Among the 12 primer sets targeting the first intron, using primer set 10, we could amplify genomic DNA that was precipitated with anti-GATA1 antibody at a similar level to that of the positive control, but not with other primer sets. We refer to this region amplified with primer set 10 as ChIPmini (g.7795_7924), the sequence of which is shown in Figure 1B. *In silico* analysis identified only one GATA element (g.7860_7865, boxed in Figure 1B) in ChIPmini, termed ALAS2int1GATA. In addition, primer set 1 which targets the proximal promoter region yielded notable amounts of amplified genome DNA. These results indicate that GATA1 protein bound to the regions amplified with primer sets 1 and 10 in K562 cells; that is, GATA1 protein could bind to the proximal promoter region as well as to ALAS2int1GATA in the first intron of the *ALAS2* gene *in vivo*. Since the GATA element located in the proximal promoter has been well examined *in vitro*,[16] we further determined the functional features of ALAS2int1GATA.

GATA1 protein binds to ALAS2int1GATA located in ChIPmini

We then examined whether GATA1 protein binds to ALAS2int1GATA present in the center of ChIPmini using EMSA (Figure 2A). The WT probe contains ALAS2int1GATA (Figure 1B). The incubation of labeled WT probe with nuclear extracts of K562 cells yielded the retarded band that represents the protein-probe complex (lane 2), whereas this retarded band was undetectable with an excess amount of non-labeled WT probe (lane 3). Moreover, the addition of anti-GATA1 antibody reduced the intensity of the retarded band (lane 4), suggesting that GATA1 protein may bind to the WT probe. In fact, the retarded band was not detected when the labeled probe was incubated with nuclear extracts of mock-transfected HEK293 cells (lane 5). In contrast, the retarded band was observed when the labeled probe was incubated with the nuclear extracts of HEK293 cells expressing FLAG-fused GATA1 (lane 6). Importantly, the retarded band observed in lane 6 was not detectable in the presence of an excess amount of non-labeled probe (lane 7). The formation of the retarded band was partially inhibited by anti-GATA1 antibody (lane 8). Likewise, the inclusion of anti-FLAG antibody (lane 9) resulted in the disappearance of the retarded band and instead generated the super-shifted band (indicated by an asterisk). These results suggest that GATA1 protein binds to the WT probe containing ALAS2int1GATA.

Enhancement of ALAS2 promoter activity by the DNA segment containing ALAS2int1GATA

To examine the functional importance of ALAS2int1GATA in the promoter activity of the *ALAS2* gene (Figure 2B), we constructed the pGL3-AEpro(-267) vector, in which the expression of firefly luciferase gene is controlled under the proximal promoter of the *ALAS2* gene (g.4820_5115). The presence of the first intron of *ALAS2* (pGL3-AEpro(-267)+intron1) increased luciferase activity about 3-fold in K562 cells, whereas luciferase activity was decreased to 10% of pGL3-AEpro(-267) in HEK293 cells. When the ChIP-peak, the region determined by ChIP-seq analysis (g.7488_7960),[22] was present downstream [+ChIP-peak(D)] or upstream [+ChIP-peak(U)] of the ALAS2 proximal promoter, luciferase activity was increased about 5-fold, irrespective of the location, compared to that of pGL3-AEpro(-267) in K562 cells. Moreover, the presence of the ChIPmini fragment downstream of the luciferase gene [+ChIPmini (D)] resulted in a 16-fold increase of luciferase activity. However, when the same fragment was inserted upstream of the ALAS2 promoter [+ChIPmini(U)], luciferase activity increased only 3-fold. Thus, the enhancer activity of the ChIPmini fragment varies, depending on its location. Moreover, among the constructs examined, the ChIPmini fragment showed maximum enhancer activity downstream of the luciferase gene. The ChIP-peak or ChIPmini fragment downstream of the ALAS2 promoter influenced luciferase activity marginally (0.73- or 1.25-fold, respectively) in HEK293 cells (Figure 2B). These results suggest that the enhancer activity of each fragment containing ALAS2int1GATA is specific to erythroid cells.

To examine whether the erythroid-specific enhancer activity depends on the ALAS2 promoter, we replaced the ALAS2 promoter with the herpes simplex virus TK promoter (Figure 2C). The ChIP-peak and ChIPmini enhanced TK promoter activity 3.4- and 9.8-fold in K562 cells, respectively, whereas they did not enhance TK promoter activity in HEK293 cells. These results indicate that the erythroid-specific enhancer is present in the ChIP-peak and ChIPmini fragments. In addition, the erythroid-specific enhancer is functional in the non-erythroid gene promoter.

Identification of mutations in the first intron of the ALAS2 gene in patients with congenital sideroblastic anemia

Considering the newly identified enhancer in the first intron of the *ALAS2* gene, we examined whether some CSA patients carry the mutation in ChIP-peak or ChIPmini of *ALAS2*. We determined the nucleotide sequence of the first intron of *ALAS2* in 11 probands (eight pedigrees), and found two distinct mutations in the newly identified enhancer region in five Japanese patients (three pedigrees). The clinical features and hematologic status of the probands at diagnosis of the disease are summarized in Table 1.

Proband 1 in a pedigree with XLSA

The first male Japanese proband was referred to hospital at the age of 3 months to investigate the cause of his pale face. No problems were reported during the birth. Investigations showed microcytic/hypochromic anemia, an increased concentration of serum iron and raised serum ferritin level. Bone marrow aspiration revealed the presence of ring sideroblasts. Two maternal relatives – male cousins of the proband's mother – have sideroblastic ane-

mia (Figure 3A). The pedigree of this family suggested X chromosome-linked inheritance of the disease. The proband's anemia was not improved by pyridoxine administration (5 mg/kg/day for 3 months), and the boy required once monthly transfusions of one unit of concentrated red blood cells to maintain an adequate hemoglobin level. At the age of 7 months, this proband died of sepsis caused by alpha-streptococcus.

Proband 2 with nfCSA

The second male Japanese proband visited hospital at the age of 4 years because of the paleness of his complexion. Investigations showed microcytic/hypochromic anemia, mild thrombocytosis, and a high serum iron concentration with a normal serum ferritin concentration. Bone marrow aspiration revealed the presence of ring sideroblasts (38% of the erythroblasts). Giant platelets were observed in the bone marrow, although dysplasia of the megakaryocytes was not clear. There was no family history of sideroblastic anemia (Figure 3B).

Proband 3 with nfCSA

The third male Japanese proband was noted to have anemia at the age of 2 years, but details are not available. Without any treatment, serum hemoglobin level was maintained at 70 g/L, and increased to 100 g/L at the age of 10. Accordingly, the proband stopped visiting the hospital. At the age of 19, however, the proband was admitted to hospital because of general fatigue. Investigations revealed microcytic, hypochromic anemia with systemic iron overload. The presence of ring sideroblasts was confirmed in his bone marrow by Prussian blue staining (36% of erythroblasts). Although this proband was treated with pyridoxine (150 mg/day) for 8 months, his anemia did not improve. There was no family history of sideroblastic anemia (Figure 3C).

In proband 1 from the pedigree with XLSA (Figure 3A), we identified a single nucleotide mutation (Figure 4, upper panel, g.7863T>C), which alters the core sequence of ALAS2int1GATA in the antisense strand from GATA to GGTA (referred to as "GGTA mutation"). The same mutation of the *ALAS2* gene was also identified in two cousins of the proband's mother, both of whom were diagnosed as having sideroblastic anemia (Figure 3A). Clinical specimens for genetic analysis were not available from either the parents or the elder brother of proband 1.

The same GGTA mutation was identified at ALAS2int1GATA in proband 2 with CSA (Figure 4, middle panel). There was no known consanguinity between proband 1 and proband 2. Genomic DNA from the parents of proband 2 was not available, because they did not agree to provide their clinical specimens for genetic analysis. Since proband 2 was also noted to have thrombocytosis (Table 1), we searched for a *JAK2* mutation in the genomic DNA extracted from the peripheral blood of this patient. However, no V617F mutation or any missense

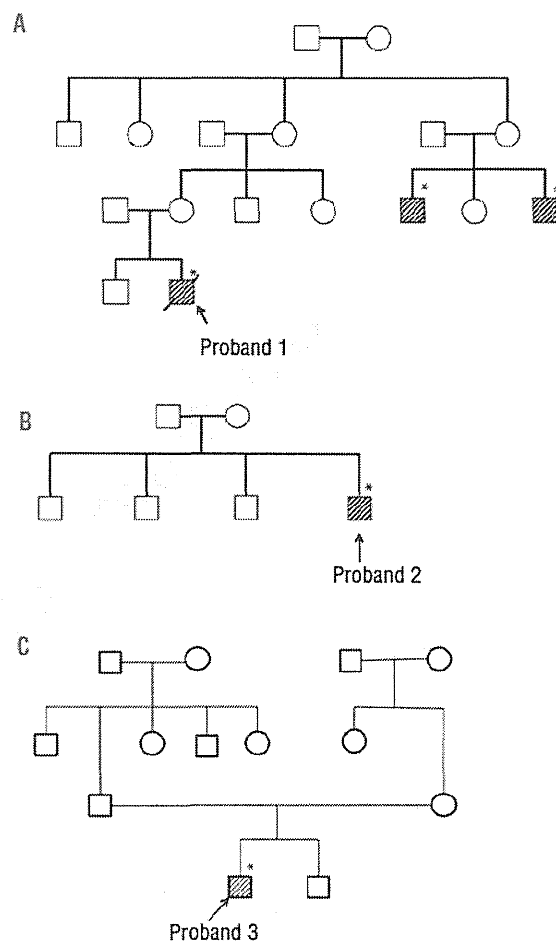


Figure 3. **Family trees of three unrelated probands. Family tree of: (A) proband 1 with XLSA, (B) proband 2 with nfCSA, and (C) proband 3 with nfCSA. Shaded boxes indicate affected individuals in each pedigree. Asterisks indicate the individuals in whom a mutation in the first intron of the *ALAS2* gene was detected.**

Table 1. **Hematologic status of each proband at diagnosis of the disease.**

	Onset of the anemia	Age at diagnosis of SA	Family history of XLSA	Hb (g/L)	MCV (fL)	MCH (pg)	Platelets ($x10^9$/L)	Serum iron (μmol/L)	Ferritin (pmol/L)
Proband 1	4 months	4 months	yes	39 [136-183]	65 [83-101]	18.7 [28-35]	246 [140-379]	63.9 [10.7-37.6]	399.7 [49.4-270]
Proband 2	4 years	4 years	no	84 [126-165]	73.4 [87-104]	22 [29-35]	610 [138-309]	49.1 [12.5-25.0]	670.1 [67.4-725]
Proband 3	2 years	19 years	no	78 [120-165]	73.9 [80-100]	22.2 [28-34]	373 [160-420]	39.6 [14.3-21.5]	2489.7 [40.4-288]

The normal value of each clinical examination is shown in brackets. SA: sideroblastic anemia; XLSA: X-linked sideroblastic anemia; Hb: hemoglobin; MCV: mean corpuscular volume; MCH: mean corpuscular hemoglobin.

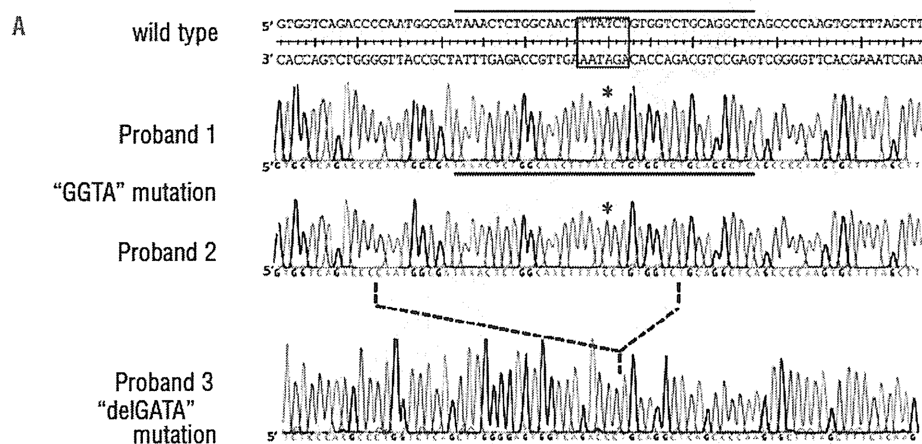
mutation in exon 12, each of which is frequently observed in patients with refractory anemia with ring sideroblasts and thrombocytosis (RARS-T),[27] was detected (*data not shown*). Thus, the GGTA mutation at ALAS2int1GATA may be responsible for the sideroblastic anemia in proband 2.

In proband 3 with CSA, a deletion of 35 bp was identified in the first intron of the *ALAS2* gene (Figure 4A, lower panel, g.7836_7870del, referred to as "delGATA mutation"). The delGATA mutation results in the loss of ALAS2int1GATA. However, the delGATA mutation was not identified in the *ALAS2* gene of the parents of proband 3 (*data not shown*). Thus, the delGATA mutation may be a *de novo* mutation or a somatic mutation. Accordingly, we compared the relative ALAS2 mRNA level in the erythroid progenitor cells isolated from proband's bone marrow with those of normal subjects. The ALAS2 mRNA level was more than 7-fold lower in the proband's erythroblasts than in those of three independent, normal subjects (Figure 4B), suggesting that the delGATA mutation may lead to decreased transcription of the *ALAS2* gene.

Lastly, we examined the sequence of the region corresponding to g.7513_8165 of the *ALAS2* gene, which contains ChIPmini, in 103 healthy, Japanese volunteers (44 males and 59 females, total 162 alleles) using PCR followed by direct sequencing. No mutation was found in this region (*data not shown*). In addition, no single nucleotide polymorphism was reported in this GATA element, based on the single nucleotide polymorphism database available at the NCBI home page (*http://www.ncbi.nlm.nih.gov/snp*, current assembly is GRCh37.p5). Thus, the GGTA mutation and delGATA mutation at ALAS2int1GATA may be unique to patients with sideroblastic anemia. Taken together, we suggest that the newly identified mutations at ALAS2int1GATA are responsible for sideroblastic anemia.

The mutation at ALAS2int1GATA impairs GATA1-binding activity and enhancer function

We examined the effect of the GGTA mutation or the delGATA mutation on the binding of GATA1 protein to ALAS2int1GATA using each mutant probe (Figure 5A). The delGATA probe represents the 5'- and 3'-flanking sequences of the deleted 35-bp segment (see Figure 4A). As shown in Figure 5B, the incubation of labeled WT probe with nuclear extracts from HEK293 cells expressing FLAG-fused GATA1



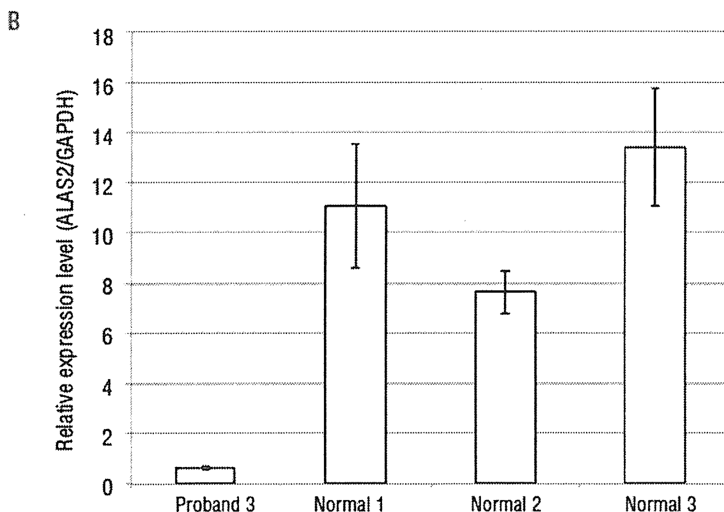


Figure 4. **Identification of mutations in the first intron of the *ALAS2* gene in a patient with XLSA and two patients with nfCSA. (A) *ALAS2* mutations in three probands. Upper, middle and lower panels show the sequences of the flanking regions of ALAS2int1GATA (boxed in the wild-type sequence) in the *ALAS2* gene of probands 1, 2 and 3, respectively. Asterisks indicate the T to C transition in the sense strand identified in the *ALAS2* gene of proband 1 and proband 2 with CSA. The broken line between the middle and lower panels indicates the deleted region identified in proband 3 with CSA. The solid horizontal bar in each panel indicates the sequence of the sense strand of each probe used for the EMSA (see Figures 3A and 5B). (B) ALAS2 mRNA expression in erythroblasts of proband 3. ALAS2 mRNA levels were determined in purified erythroblasts isolated from proband 3 and three independent normal individuals using real-time PCR. Results are expressed as the mean ± SD of three independent experiments.**