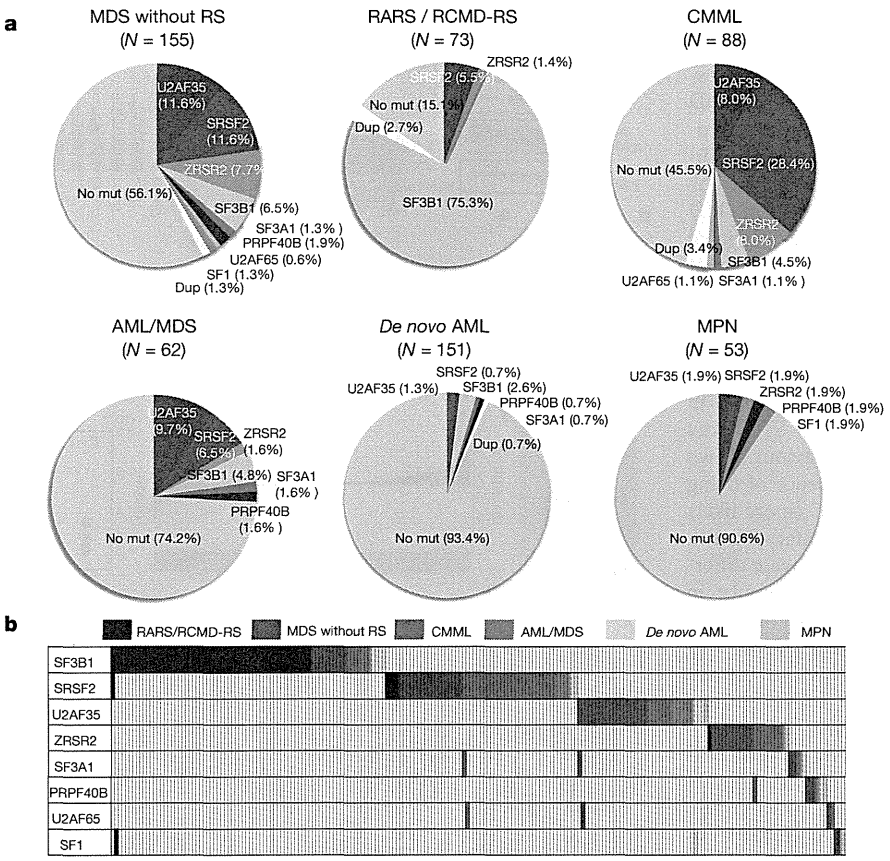


**Figure 3 | Frequencies and distribution of spliceosome pathway gene mutations in myeloid neoplasms. a,** Frequencies of spliceosome pathway mutations among 582 cases with various myeloid neoplasms. **b,** Distribution of mutations in eight spliceosome genes, where diagnosis of each sample is shown by indicated colours.

for 4.3% and 12.9% of MDS cases, respectively, where deregulated iron metabolism has been implicated in the development of refractory anaemia[23]. With such high mutation frequencies and specificity, the *SF3B1* mutations were thought to be almost pathognomonic to these MDS subtypes characterized by increased ring sideroblasts, and strongly implicated in the pathogenesis of MDS in these categories. Less conspicuously but significantly, *SRSF2* mutations were more frequent in CMML cases (Fig. 3 and Supplementary Table 4). Thus, although commonly involving the E/A splicing complexes, different mutations may still have different impacts on cell functions, contributing to the determination of discrete disease phenotypes. For example, studies have demonstrated that *SRSF2* was also involved in the regulation of DNA stability and that depletion of SRSF2 can lead to genomic instability[24]. Of interest in this context, regardless of disease subtypes, samples with *SRSF2* mutations were shown to have significantly more mutations of other genes compared with *U2AF35* mutations ($P = 0.001$, multiple regression analysis) (Supplementary Table 6 and Supplementary Fig. 7).

Notably, with a rare exception of A26V in a single case, the mutations of U2AF35 exclusively involved two highly conserved amino acid positions (S34 or Q157) within the amino- and the carboxyl-terminal zinc finger motifs flanking the U2AF homology motif (UHM) domain. SRSF2 mutations exclusively occurred at P95 within an intervening sequence between the RNA recognition motif (RRM) and arginine/serine-rich (RS) domains (Fig. 2 and Supplementary Figs 8 and 9). Similarly, SF3B1 mutations predominantly involved K700 and, to a lesser extent, K666, H662 and E622, which are also conserved across species (Fig. 2 and Supplementary Fig. 10). The involvement of recurrent amino acid positions in these spliceosome genes strongly indicated a gain-of-function nature of these mutations, which has been a well-documented scenario in other oncogenic mutations[25]. On the other hand, the 23 mutations in *ZRSR2* (Xp22.1) were widely distributed along the entire coding region (Fig. 2). Among these, 14 mutations were nonsense or frameshift changes, or involved splicing donor/acceptor sites that caused either a premature truncation or a large structural change of the protein, leading to loss-of-function. Combined with their strong male preference for the mutation (14/14 cases), *ZRSR2* most likely acts as a tumour suppressor gene with an X-linked recessive mode of genetic action. The remaining nine *ZRSR2* mutations were missense changes and found in both males (six cases) and females (three cases), whose somatic origin was only confirmed in two cases. However, neither the dbSNP database (build131 and 132) nor the 1000 Genomes database (May 2011 snp calls) contained these missense nucleotides, suggesting that many, if not all, of these missense changes are likely to represent functional somatic changes, especially those found in males. Interrogation of these hot spots for mutations in *U2AF35* and *SRSF2* found no mutations among lymphoid neoplasms, including acute lymphoblastic leukaemia ($N = 24$) or non-Hodgkin's lymphoma ($N = 87$) (data not shown).

## RNA splicing and spliceosome mutations

Because the splicing pathway mutations in myelodysplasia widely and specifically affect the major components of the splicing complexes E/A in a mutually exclusive manner, the common consequence of these mutations is logically the impaired recognition of 3′SSs that would lead to the production of aberrantly spliced mRNA species. To appreciate this and also to gain an insight into the biological/biochemical impact of these splicing mutations, we expressed the wild-type and the mutant (S34F) U2AF35 in HeLa cells using retrovirus-mediated gene transfer with enhanced green fluorescent protein (EGFP) marking (Fig. 4a and Supplementary Methods III) and examined their effects on gene expression in these cells using GeneChip Human genome U133 plus 2.0 arrays (Affymetrix), followed by gene set enrichment analysis (GSEA) (Supplementary Methods IV)[26]. Intriguingly, the GSEA disclosed a significant enrichment of the genes on the nonsense-mediated mRNA decay (NMD) pathway among the significantly upregulated genes in mutant *U2AF35*-transduced HeLa cells (Fig. 4b, Supplementary Fig. 11a and Supplementary Table 7), which was

confirmed by quantitative polymerase chain reactions (qPCR) (Fig. 4c and Supplementary Methods 5V). A similar result was also observed for the gene expression profile of an MDS-derived cell line (TF-1) transduced with the S34F mutant (Supplementary Figs 11b, c). The NMD activation by the mutant U2AF35 was suppressed significantly by the co-overexpression of the wild-type protein (Supplementary Fig. 11d), indicating that the effect of the mutant protein was likely to be mediated by inhibition of the functions of the wild-type protein. Given that the NMD pathway, known as mRNA surveillance, provides a post-transcriptional mechanism for recognizing and eliminating abnormal transcripts that prematurely terminate translation[27], the result of the GSEA analyses indicated that the mutant U2AF35 induced abnormal RNA splicing in HeLa and TF-1 cells, leading to the generation of unspliced RNA species having a premature stop codon and induction of the NMD activity.

To confirm this, we next performed whole transcriptome analysis in these cells using the GeneChip Human exon 1.0 ST Array (Affymetrix), in which we differentially tracked the behaviour of two discrete sets of probes showing different level of evidence of being exons, that is, 'Core' (authentic exons) and 'non-Core' (more likely introns) sets (Supplementary Methods IV and Supplementary Fig. 12). As shown in Fig. 4d, the Core and non-Core set probes were differentially enriched among probes showing significant difference in expression between wild-type and mutant-transduced cells (false discovery rate (FDR) = 0.01). The Core set probes were significantly enriched in those probes significantly downregulated in mutant *U2AF35*-transduced cells compared with wild-type *U2AF35*-transduced cells, whereas the non-Core set probes were enriched in those probes significantly upregulated in mutant *U2AF35*-transduced cells (Fig. 4e). The significant differential enrichment was also demonstrated, even when all probe sets were included (Fig. 4f). Moreover, the significantly differentially expressed Core set probes tended to be up- and downregulated in wild-type and mutant *U2AF35*-transduced cells compared with mock-transduced cells, respectively, and vice versa for the differentially expressed non-Core set probes (Fig. 4e). Combined, these exon array results indicated that the wild-type U2AF35 correctly promoted authentic RNA splicing, whereas the mutant U2AF35 inhibited this processes, rendering non-Core and therefore, more likely intronic sequences to remain unspliced.

The abnormal splicing in mutant *U2AF35*-transduced cells was more directly demonstrated by sequencing mRNAs extracted from HeLa cells, in which expression of the wild-type and mutant (S34F) U2AF35 were induced by doxycycline. First, after adjusting by the total number of mapped reads, the wild-type *U2AF35*-transduced cells showed an increased read counts in the exon fraction, but reduced counts in other fractions, compared with mutant *U2AF35*-transduced cells (Fig. 4g). The reads from the mutant-transduced cells were mapped to broader genomic regions compared with those from the wild-type *U2AF35*-transduced cells, which were largely explained by non-exon reads (Fig. 4h). Finally, the number of those reads that encompassed the authentic exon/intron junctions was significantly increased in mutant *U2AF35*-transduced cells compared with wild-type *U2AF35*-transduced cells (Fig. 4i and Supplementary Methods VI). These results clearly demonstrated that failure of splicing ubiquitously occurred in mutant *U2AF35*-transduced cells. A typical example of abnormal splicing in mutant-transduced cells and the list of significantly unspliced exons are shown in Supplementary Fig. 13 and Supplementary Table 8, respectively.

## Biological consequence of U2AF35 mutations

Finally, we examined the biological effects of compromised functions of the E/A splicing complexes. First, TF-1 and HeLa cells were transduced with lentivirus constructs expressing either the S34F U2AF35 mutant or wild-type U2AF35 under a tetracycline-inducible promoter (Fig. 5a and Supplementary Figs 14a and 15a), and cell proliferation was examined after the induction of their expression. Unexpectedly, after the induction of gene expression with

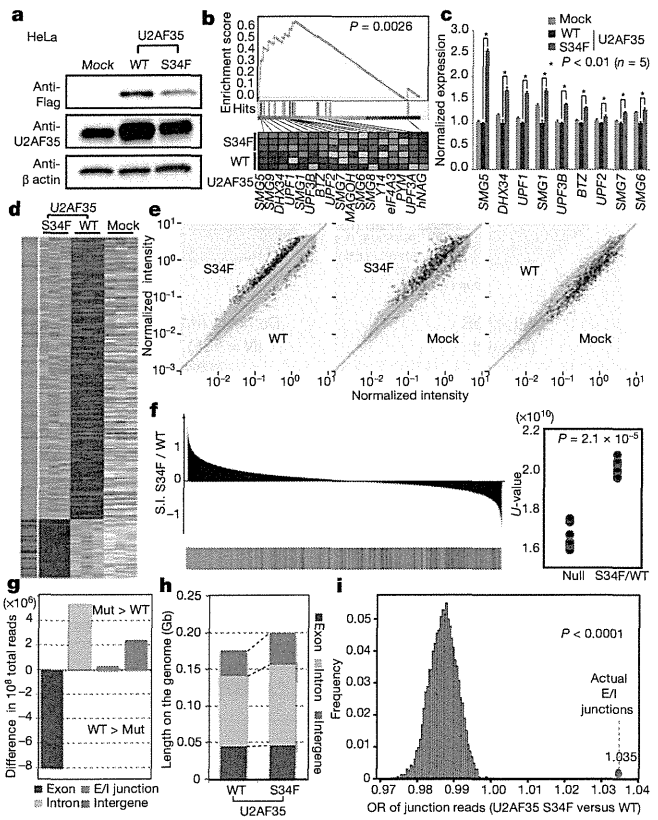


**Figure 4 | Altered RNA splicing caused by a U2AF35 mutant.** **a**, Western blot analyses showing expression of transduced wild-type or mutant (S34F) U2AF35 in HeLa cells used for the analyses of expression and exon microarrays. **b**, The GSEA demonstrating a significant enrichment of the set of 17 NMD pathway genes among significantly differentially expressed genes between wild-type and mutant *U2AF35*-transduced HeLa cells. The significance of the gene set was empirically determined by 1,000 gene-set permutations. **c**, The confirmation of the microarray analysis for the expression of nine genes that contributed to the core enrichment in the NMD gene set. Means ± s.e. are provided for the indicated NMD genes. *P* values were determined by the Mann–Whitney *U* test. **d**, Significantly upregulated and downregulated probe sets (FDR = 0.01) in mutant *U2AF35*-transduced cells compared with wild-type *U2AF35*-transduced cells in triplicate exon array experiments are shown in a heat map. The origin of each probe set is depicted in the left lane, where red and green bars indicate the Core and non-Core sets, respectively. **e**, Pair-wise scatter plots of the normalized intensities of entire probe sets (grey) across different experiments. The Core and non-Core set probes that were significantly differentially expressed between the wild-type and mutant *U2AF35*-transduced cells are plotted in red and green, respectively. **f**, Distribution of the Core (red) and non-Core (green) probe sets within the entire probe sets ordered by splicing index (S.I.; Supplementary Methods IV), calculated between wild-type and mutant *U2AF35*-transduced cells. In the right panel, the differential enrichment of both probe sets was confirmed by Mann–Whitney *U* test. **g**, Difference in read counts for the indicated fractions per $10^8$ total reads in RNA sequencing between wild-type and mutant U2AF35-expressing HeLa cells analysis. Increased/decreased read counts in mutant U2AF35-expressing cells are plotted upward/downward, respectively. **h**, Comparison of the genome coverage by the indicated fractions in wild-type- and mutant-U2AF35-expressing cells. The genome coverage was calculated for each fraction within the $10^8$ reads randomly selected from the total reads and averaged for ten independent selections. **i**, The odds ratio of the junction reads within the total mapped reads was calculated between the two experiments (red circle), which was evaluated against the 10,000 simulated values under the null hypothesis (histogram in blue).

doxycycline, the mutant *U2AF35*-transduced cells, but not the wild-type *U2AF35*-transduced cells, showed reduced cell proliferation (Fig. 5b and Supplementary Fig. 15b) with a marked increase in the G2/M fraction (G2/M arrest) together with enhanced apoptosis as

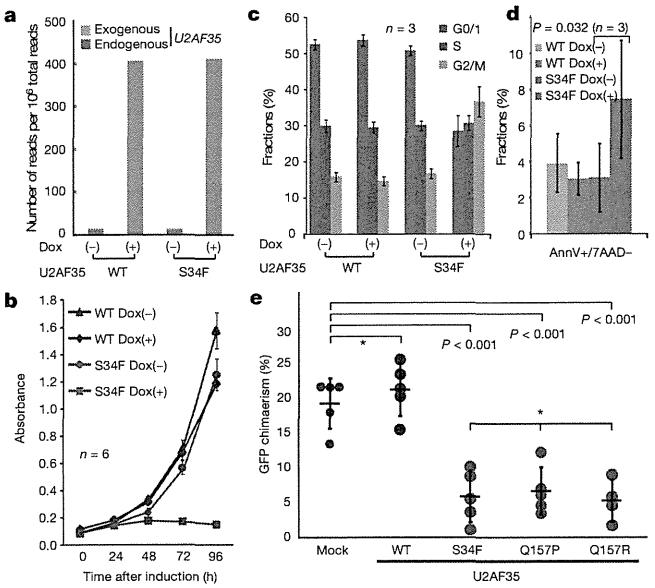


**Figure 5 | Functional analysis of mutant U2AF35.** **a**, Expression of endogenous and exogenous *U2AF35* transcripts in HeLa cells before and after induction determined by RNA sequencing. *U2AF35* transcripts were differentially enumerated for endogenous and exogenous species, which were discriminated by the Flag sequence. **b**, Cell proliferation assays of *U2AF35*-transduced HeLa cells, where cell numbers were measured using cell-counting apparatus and are plotted as mean absorbance ± s.d. **c**, The flow cytometry analysis of propidium iodide (PI)-stained HeLa cells transduced with the different U2AF35 constructs. Mean fractions ± s.d. in G0/G1, S and G2/M populations after the induction of U2AF35 expression are plotted. **d**, Fractions of the annexin V-positive (AnnV+) populations among the 7-amino-actinomycin D (7AAD)-negative population before and after the induction of U2AF35 expression are plotted as mean ± s.d. for indicated samples. The significance of difference was determined by paired *t*-test. **e**, Competitive reconstitution assays for CD34-negative KSL cells transduced with indicated U2AF35 mutants. Chimaerism in the peripheral blood 6 weeks after transplantation are plotted as mean %EGFP-positive Ly5.1 cells ± s.d., where outliers were excluded from the analysis. The significance of differences was evaluated by the Grubbs test with Bonferroni's correction for multiple testing. *not significant.

indicated by the increased sub-G1 fraction and annexin V-positive cells (Fig. 5c, d, Supplementary Fig. 14b and Supplementary Methods VI). To confirm the growth-suppressive effect of U2AF35 mutants *in vitro*, a highly purified haematopoietic stem cell population ($CD34^-$c-$Kit^+Sca1^+Lin^-$, $CD34^-$KSL) prepared from C57BL/6 (B6)-Ly5.1 mouse bone marrow[28] was retrovirally transduced with either the mutant (S34F, Q157P and Q157R) or wild-type *U2AF35*, or the mock constructs, each harbouring the EGFP marker gene (Supplementary Fig. 16). The ability of these transduced cells to reconstitute the haematopoietic system was tested in a competitive reconstitution assay. The transduced cells were mixed with whole bone marrow cells from B6-Ly5.1/5.2 F1 mice, transplanted into lethally irradiated B6-Ly5.2 recipients, and peripheral blood chimaerism derived from EGFP-positive cells was assessed 6 weeks after transplantation by flow cytometry. We confirmed that each recipient mouse received comparable numbers of EGFP-positive cells among the different retrovirus groups by estimating the percentage of EGFP-positive cells and overall proliferation in transduced cells by *ex vivo* tracking. Also no significant difference was observed in their homing capacity to bone marrow as assessed by transwell migration assays (Supplementary Fig. 17). As shown in Fig. 5e, the wild-type *U2AF35*-transduced cells showed a slightly higher reconstitution capacity than the mock-transduced cells. On the other hand, the recipients of the cells transduced with the various *U2AF35* mutants showed significantly lower EGFP-positive cell chimaerism than those of either the mock- or the wild-type *U2AF35*-transduced cells, indicating a compromised reconstitution capacity of the haematopoietic stem/progenitor cells expressing the U2AF35 mutants. In summary, these mutants lead to loss-of-function of U2AF35 most probably by acting in a dominant-negative fashion to the wild-type protein.

## Discussion

Our whole-exome sequencing study unexpectedly unmasked a complexity of novel pathway mutations found in approximately 45% to 85% of myelodysplasia patients depending on the disease subtypes, which affected multiple but distinctive components of the splicing machinery and, as such, demonstrated the unquestionable power of massively parallel sequencing technologies in cancer research.

The RNA splicing system comprises essential cellular machinery, through which eukaryotes can achieve successful transcription and guarantee the functional diversity of their protein species using alternative splicing in the face of a limited number of genes[29]. Accordingly, the meticulous regulation of this machinery should be indispensable for the maintenance of cellular homeostasis[30], deregulation of which causes severe developmental abnormalities[31,32]. The current discovery of frequent mutations of the splicing pathway in myelodysplasia, therefore, represents another remarkable example that illustrates how cancer develops by targeting critical cellular functions. It also provides an intriguing insight into the mechanism of 'cancer specific' alternative splicing, which have long been implicated in the development of cancer, including MDS and other haematopoietic neoplasms[33,34].

In myelodysplasia, the major targets of spliceosome mutations seemed to be largely confined to the components of the E/A splicing complex, among others to *SF3B1*, *SRSF2*, *U2AF35* and *ZRSR2*, and to a lesser extent, to *SF3A1*, *SF1*, *U2AF65* and *PRPF40B*. The broad coverage of the wide spectrum of spliceosome components in our exome sequencing was likely to preclude frequent involvement of other components on this pathway (Supplementary Fig. 18). The surprising frequency and specificity of these mutations in this complex, together with the mutually exclusive manner they occurred, unequivocally indicate that the compromised function of the E/A complex is a hallmark of this unique category of myeloid neoplasms, playing a central role in the pathogenesis of myelodysplasia. The close relationship between the mutation types and unique disease subtypes also support their pivotal roles in MDS.

Given the critical functions of the E/A splicing complex on the precise 3′SS recognition, the logical consequence of these relevant mutations would be the impaired splicing involving diverse RNA species. In fact, when expressed in HeLa cells, the mutant U2AF35 induced global abnormalities of RNA splicing, leading to increased production of transcripts having unspliced intronic sequences. On the other hand, the functional link between the abnormal splicing of RNA species and the phenotype of myelodysplasia is still unclear. Mutant U2AF35 seemed to suppress cell growth/proliferation and induce apoptosis rather than confer a growth advantage or promote clonal selection. *ZRSR2* knockdown in HeLa cells has been reported to also result in reduced viability, arguing for the common consequence of these pathway mutations[35]. These observations suggested that the oncogenic actions of these splicing pathway mutations are distinct from what is expected for classical oncogenes, such as mutated kinases and signal transducers, but could be more related to cell differentiation. Of note in this regard, the commonest clinical presentation of MDS is severe cytopenia in multiple cell lineages due to ineffective haematopoiesis with increased apoptosis rather than unlimited cell proliferation[1]. In this regard, lessons may be learned from the recent findings on the pathogenesis of the 5q− syndrome, where haploinsufficiency of *RPS14* leads to increased apoptosis of erythroid progenitors, but not myeloproliferation[36,37].

A lot of issues remain to be answered, however, to establish the functional link between these splicing pathway mutations and the

pathogenesis of MDS, where the broad spectrum of RNA species affected by impaired splicing hampers identification of responsible gene targets. Moreover, the mutated components of the splicing machinery have distinct function of their own other than direct regulation of RNA splicing, involved in elongation and DNA stability, which may be important to determine specific disease phenotypes. Clearly, more studies are required to answer these questions through understanding of the molecular basis of their oncogenic actions.

## METHODS SUMMARY

Whole-exome sequencing of paired tumour/normal DNA samples from the 29 patients was performed after informed consent was obtained. SNP array-based copy number analysis was performed as previously described[17,18]. Mutation analysis of the splicing pathway genes in a set of 582 myeloid neoplasms were performed by first screening mutations in PCR-amplified pooled targets from 12 individuals, followed by validation/identification of the candidate mutations within the corresponding 12 individuals by Sanger sequencing. Flag-tagged cDNAs of the wild-type and mutant *U2AF35* were generated by *in vitro* mutagenesis, constructed into a murine stem cell virus-based retroviral vector as well as a tetracycline-inducible lentivirus-based expression vector, and used for gene transfer to $CD34^{-}$KSL cells and cultured cell lines, with EGFP marking, respectively. Total RNA was extracted from wild-type or mutant *U2AF35*-transduced HeLa and TF-1 cells, and analysed on microarrays. RNA sequencing was performed according to the manufacturer's instructions (Illumina). Cell proliferation assays (MTT assays) on HeLa and TF-1 cells stably transduced with lentivirus U2AF35 constructs were performed in the presence or absence of doxycycline. For competitive reconstitution assays, $CD34^{-}$KSL cells collected from C57BL/6 (B6)-Ly5.1 mice were retrovirally transduced with various *U2AF35* constructs with EGFP marking, and transplanted with competitor cells (B6-Ly5.1/5.2 F1 mouse origin) into lethally irradiated B6-Ly5.2 mice 48 h after gene transduction. Frequency of EGFP-positive cells was assessed in peripheral blood by flow cytometry 6 weeks after the transplantation (Supplementary Methods VII). The primer sets used for validation of gene mutations and qPCR of NMD gene expression are listed in Supplementary Tables 9–11. A complete description of the materials and methods is provided in the Supplementary Information. This study was approved by the ethics boards of the University of Tokyo, Munich Leukaemia Laboratory, University Hospital Mannheim, University of Tsukuba, Tokyo Metropolitan Ohtsuka Hospital and Chang Gung Memorial Hospital. Animal experiments were performed with approval of the Animal Experiment Committee of the University of Tokyo.

**Received 7 June; accepted 24 August 2011.**
**Published online 11 September 2011.**

1. Corey, S. J. *et al.* Myelodysplastic syndromes: the complexity of stem-cell diseases. *Nature Rev. Cancer* **7,** 118–129 (2007).
2. Ma, X., Does, M., Raza, A. & Mayne, S. T. Myelodysplastic syndromes: incidence and survival in the United States. *Cancer* **109,** 1536–1542 (2007).
3. Bejar, R., Levine, R. & Ebert, B. L. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J. Clin. Oncol.* **29,** 504–515 (2011).
4. Sanada, M. *et al.* Gain-of-function of mutated *C-CBL* tumour suppressor in myeloid neoplasms. *Nature* **460,** 904–908 (2009).
5. Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40,** 722–729 (2008).
6. Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471,** 467–472 (2011).
7. Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465,** 473–477 (2010).
8. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456,** 66–72 (2008).
9. Metzker, M. L. Sequencing technologies — the next generation. *Nature Rev. Genet.* **11,** 31–46 (2010).
10. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nature Biotechnol.* **26,** 1135–1145 (2008).
11. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461,** 809–813 (2009).
12. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469,** 539–542 (2011).
13. Ley, T. J. *et al. DNMT3A* mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363,** 2424–2433 (2010).
14. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361,** 1058–1066 (2009).
15. Yan, X. J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene *DNMT3A* in acute monocytic leukemia. *Nature Genet.* **43,** 309–315 (2011).
16. Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475,** 101–105 (2011).
17. 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).
18. 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).
19. Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136,** 701–718 (2009).
20. Tronchère, H., Wang, J. & Fu, X. D. A protein related to splicing factor $U2AF^{35}$ that interacts with $U2AF^{65}$ and SR proteins in splicing of pre-mRNA. *Nature* **388,** 397–400 (1997).
21. Bevilacqua, L. *et al.* A population-specific *HTR2B* stop codon predisposes to severe impulsivity. *Nature* **468,** 1061–1066 (2010).
22. Calvo, S. E. *et al.* High-throughput, pooled sequencing identifies mutations in *NUBPL* and *FOXRED1* in human complex I deficiency. *Nature Genet.* **42,** 851–858 (2010).
23. Haase, D. *et al.* New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood* **110,** 4385–4395 (2007).
24. Xiao, R. *et al.* Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol. Cell. Biol.* **27,** 5393–5402 (2007).
25. Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature Genet.* **42,** 181–185 (2010).
26. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
27. Maquat, L. E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nature Rev. Mol. Cell Biol.* **5,** 89–99 (2004).
28. Ema, H. *et al.* Adult mouse hematopoietic stem cells: purification and single-cell assays. *Nature Protocols* **1,** 2979–2987 (2007).
29. Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Rev. Mol. Cell Biol.* **10,** 741–754 (2009).
30. Ni, J. Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21,** 708–718 (2007).
31. He, H. *et al.* Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science* **332,** 238–240 (2011).
32. Edery, P. *et al.* Association of TALS developmental disorder with defect in minor splicing component *U4atac* snRNA. *Science* **332,** 240–243 (2011).
33. David, C. J. & Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* **24,** 2343–2364 (2010).
34. Pajares, M. J. *et al.* Alternative splicing: an emerging topic in molecular and clinical oncology. *Lancet Oncol.* **8,** 349–357 (2007).
35. Shen, H., Zheng, X., Luecke, S. & Green, M. R. The U2AF35-related protein Urp contacts the 3′ splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes Dev.* **24,** 2389–2394 (2010).
36. Barlow, J. L. *et al.* A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q− syndrome. *Nature Med.* **16,** 59–66 (2010).
37. Ebert, B. L. *et al.* Identification of *RPS14* as a $5q^{-}$ syndrome gene by RNA interference screen. *Nature* **451,** 335–339 (2008).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** This work was supported by Grant-in-Aids from the Ministry of Health, Labor and Welfare of Japan and from the Ministry of Education, Culture, Sports, Science and Technology, and also by the Japan Society for the Promotion of Science (JSPS) through the 'Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)', initiated by the Council for Science and Technology Policy (CSTP). pGCDNsamIRESEGFP vector was a gift from M. Onodera. We thank Y. Mori, O. Hagiwara, M. Nakamura and N. Mizota for their technical assistance. We are also grateful to K. Ikeuchi and M. Ueda for their continuous encouragement throughout the study.

**Author Contributions** Y.Sh., Y.Sa., A.S.-O., Y.N., M.N., G.C., R.K. and S.Miyano were committed to bioinformatics analyses of resequencing data. M.Sa., A.S.-O. and Y.Sa. performed microarray experiments and their analyses. R.Y., T.Y., M.O., M.Sa., A.K., M.Sh. and H.N. were involved in the functional analyses of U2AF35 mutants. N.O., M.S.-Y., K.I., H.M., W.-K.H., F.N., D.N., T.H., C.H., S.Miyawaki, S.C., H.P.K. and L.-Y.S. collected specimens and were also involved in planning the project. K.Y., Y.N., Y.Su., A.S.-O. and S.S. processed and analysed genetic materials, library preparation and sequencing. K.Y., M.Sa., Y.Sh., A.S.-O., Y. Sa. and S.O. generated figures and tables. S.O. led the entire project and wrote the manuscript. All authors participated in the discussion and interpretation of the data and the results.

**Author Information** Sequence data have been deposited in the DDBJ repository under accession number DRA000433. Microarray data have been deposited in the GEO database under accession numbers GSE31174 (for SNP arrays), GSE31171 (for exon arrays) and GSE31172 (for expression arrays). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.O. (sogawa-tky@umin.ac.jp).

# Frequent loss of HLA alleles associated with copy number-neutral 6pLOH in acquired aplastic anemia

*Takamasa Katagiri,[1,2] *Aiko Sato-Otsubo,[3] Koichi Kashiwase,[4,5] Satoko Morishima,[6] Yusuke Sato,[3] Yuka Mori,[3] Motohiro Kato,[3] Masashi Sanada,[3] Yasuo Morishima,[7] Kohei Hosokawa,[2] Yumi Sasaki,[2] Shigeki Ohtake,[1] †Seishi Ogawa,[3,5] and †Shinji Nakao,[2] on behalf of the Japan Marrow Donor Program

[1]Clinical Laboratory Science, Division of Health Sciences, and [2]Cellular Transplantation Biology, Kanazawa University Graduate School of Medical Science, Ishikawa, Japan; [3]Cancer Genomics Project, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; [4]Tokyo Metropolitan Red Cross Blood Center, Tokyo, Japan; [5]Core Research for Evolutional Science and Technology, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, Saitama, Japan; [6]Department of Hematology, Fujita Health University, Aichi, Japan; and [7]Department of Hematology and Cell Therapy, Aichi Cancer Center Hospital, Nagoya, Japan

**Idiopathic aplastic anemia (AA) is a common cause of acquired BM failure. Although autoimmunity to hematopoietic progenitors is thought to be responsible for its pathogenesis, little is known about the molecular basis of this autoimmunity. Here we show that a substantial proportion of AA patients harbor clonal hematopoiesis characterized by the presence of acquired copy number-neutral loss of heterozygosity (CNN-LOH) of the 6p arms (6pLOH). The 6pLOH commonly involved the HLA locus, leading to loss of one HLA haplotype. Loss of HLA-A expression from multiple lineages of leukocytes was confirmed by flow cytometry in all 6pLOH(+) cases. Surprisingly, the missing HLA-alleles in 6pLOH(+) clones were conspicuously biased to particular alleles, including HLA-A*02:01, A*02:06, A*31:01, and B*40:02. A large-scale epidemiologic study on the HLA alleles of patients with various hematologic diseases revealed that the 4 HLA alleles were over-represented in the germline of AA patients. These findings indicate that the 6pLOH(+) hematopoiesis found in AA represents "escapes" hematopoiesis from the autoimmunity, which is mediated by cytotoxic T cells that target the relevant autoantigens presented on hematopoietic progenitors through these class I HLAs. Our results provide a novel insight into the genetic basis of the pathogenesis of AA. (*Blood.* 2011;118(25):6601-6609)**

## Introduction

Acquired aplastic anemia (AA) is a rare condition associated with BM failure and pancytopenia.[1] A series of classic observations and experiments have unequivocally supported that the autoimmunity to hematopoietic stem/progenitor cells (HSPCs) critically underlies the pathogenesis of the BM failure in the majority of AA cases. According to the widely accepted model of immune-mediated BM failure, activated cytotoxic T cells (CTLs) that recognize an auto-antigen(s) presented on HSPCs through their class I HLA molecules have a major role in initiating the autoimmune reactions.[2-4] However, no definitive evidence exists that supports this model or the presence of such CTL repertoires. Moreover, little information is available about their target antigens or about the way by which they are recognized by effector T cells.

Another long-standing issue on AA is its close relationship with clonal hematopoiesis.[5,6] It was first suspected from an apparent overlap between AA and paroxysmal nocturnal hemoglobinuria (PNH)[7,8] and was also implicated by the frequent development of late clonal disorders in AA, such as myelodysplastic syndromes, PNH, or even acute myeloid leukemia (AML).[9-11] Clonal hematopoiesis can be explicitly demonstrated by conventional clonality assays at presentation in a substantial proportion of newly diagnosed typical AA cases.[12] Although it has been expected that the inciting autoimmune insult somehow confers selective pressures on the evolution of clonal hematopoiesis,[5] the exact mechanism for such immunologic selection or escape is still unclear.

The objectives of this study, therefore, were to characterize the clonal nature of the hematopoiesis that is maintained even under the severe autoimmune insult in AA, and to explore the genetic/immunologic mechanism that could underlie the pathogenesis of AA. To achieve these aims, we performed single nucleotide polymorphism (SNP) array-based analysis of genomic copy numbers and/or allelic imbalances in peripheral blood (PB) specimens obtained from 306 patients with AA. Initially, we found that AA patients frequently showed clonal/oligoclonal hematopoiesis that lost specific HLA alleles as a result of copy number-neutral loss of heterozygosity (CNN-LOH) of the 6p arms, which led us to further analyses of the contribution of 6pLOH(+) clones to residual hematopoiesis and a large-scale epidemiologic study on the HLA alleles that are over-represented in AA, involving a total of 6,613 transplants registered in the Japan Marrow Donor Program (JMDP).

## Methods

### Subjects

PB specimens from a total of 306 patients with AA were analyzed for the presence of genetic alterations using SNP arrays (see Figure 1). The clinical

Submitted July 1, 2011; accepted September 18, 2011. Prepublished online as *Blood* First Edition paper, September 30, 2011; DOI 10.1182/blood-2011-07-365189.

*T.K. and A.S.-O. contributed equally to this study.

†S. Ogawa and S.N. contributed equally to this study.

The online version of this article contains a data supplement.

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

**Table 1. Patient characteristics**

| | Newly diagnosed (n = 107) | Previously treated (n = 199) |
|---|---|---|
| Median age at diagnosis, mo (range) | 64 (9-88) | 24 (2-80) |
| Sex, male/female, no. | 58/49 | 110/89 |
| **Severity of AA at onset, no. (%) of patients** | | |
| Severe | 79 (74) | 185 (93) |
| Nonsevere | 28 (26) | 14 (7) |
| History, mo, median (range) | 19 (0.1-251) | 51 (0.1-372) |
| **Past treatment, no. (%) of patients** | | |
| ATG + CsA | — | 39 (20) |
| CsA alone | — | 51 (26) |
| Anabolic steroid alone | — | 13 (7) |
| Unknown* | — | 96 (48) |

ATG indicates antithymocyte globulin; CsA, cyclosporine A; and —, not applicable.

*Information regarding previous therapies of 96 cases (from Japan Marrow Donor Program) was unavailable.

characteristics of these patients are summarized in Table 1 and supplemental Table 1 (available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article). Among the 306 patients, 107 were newly diagnosed and 199 were previously treated. Ninety-six patients received allogeneic BM transplantation from unrelated donors through the JMDP, and their HLA information was available from the JMDP. The other 210 were newly genotyped for HLA-A, -B, -C, -DRB1, -DQB1, and -DPB1 alleles as described elsewhere.[13] A total of 103 patients had been treated with anti–thymocyte globulin plus cyclosporine, cyclosporine alone, or anabolic steroids at the time of sampling. All patients and healthy persons provided their informed consent before sampling in accordance with the Declaration of Helsinki. The study protocol was approved by the ethics committee of the Graduate School of Medical Science, Kanazawa University and also by that of the Graduate School of Medicine, University of Tokyo.

### Analysis of genomic copy numbers and detection of 6pLOH

Genomic copy numbers, as well as allele-specific copy numbers, were analyzed by using GeneChip 500K arrays (Affymetrix) as previously described.[14,15] Briefly, genomic DNA from AA patients and normal controls were analyzed on GeneChip 500K arrays separately. After adjusting several biases introduced during experiments, signal ratios of the corresponding probes between test (patient) and controls were calculated across the genome to obtain genome-wide copy numbers. Genetic lesions, including copy number gains and losses, as well as CNN-LOHs, were first detected using a hidden Markov model-based algorithm implemented in the CNAG software.[14,15] Known copy number variations were carefully excluded by referring to the Database of Genomic Variants (www.projects.tcag.ca/variation). CNN-LOH in 6p involving the HLA locus was more specifically and sensitively detected by statistically evaluating the mean differences in allele-specific copy numbers between heterozygous SNPs on 6p (N = ~ 1400) that were telomeric from the 5′-end of the HLA-A locus (rs1655927) and all non-6p heterozygous SNPs (N = ~ 105 000) using the Mann-Whitney *U* test with the R package (www.r-project.org). Possible false-positive findings arising from multiple testing involving the 306 samples were evaluated by maintaining the false discovery rate under 0.01 as previously described,[16] where the microarray data of 1000 JMDP donor specimens obtained from an ongoing whole genome association study (unpublished data) were used to calculate an empiric null distribution.[17,18]

### Determination of the missing HLA alleles in 6pLOH(+) clones in patients with AA

The 500K SNP data of the 1800 JMDP donor-recipient pairs (JMDP dataset), together with their HLA genotyping information, was used to generate an HLA SNP haplotype table on the GeneChip 500K platform, which contains the consensus SNPs of the 3 major haplotypes (P1, P2, and P3) in Japanese subjects[18] and the SNP sequences of all observed HLA haplotypes complementary to P1 to P3 within the JMDP set (N = 1576; data not shown). To determine the missing HLA haplotype in each 6pLOH(+) patient, those "HLA" haplotypes were first selected from the aforementioned HLA haplotype table that were compatible with the observed HLA genotypes of that patient. Among these, a candidate haplotype was selected such that it contained the minimum number of SNPs that were incompatible with the patient's genotype. For each candidate haplotype, genomic copy numbers were inferred at the heterozygous SNPs along that haplotype using the circular binary segmentation algorithm,[19,20] which divided the haplotype into one or more discrete segments with different mean copy numbers. Finally, each copy number segment was thought to be "missing," when the alternative hypothesis ($H_a$: $S_i \neq \overline{S_i}$, for $^{\forall}_i$) was supported against the null hypothesis ($H_0$: $S_i = \overline{S_i}$, for $^{\forall}_i$) using the Wilcoxon signed rank test with a significance level of .05, where $S_i$ represents the allele-specific copy number at the *i*th heterozygous SNP site within the segment of the candidate haplotype with $\overline{S_i}$ being the corresponding value for the complementary haplotype (supplemental Figure 1). Finally, for those HLA types that appeared more than 8 times among 6pLOH(+) cases, their contribution to the observed allelic loss of HLA haplotypes was evaluated by multivariate logistic regression analysis with stepwise backward selection

### Flow cytometry

Heparinized PB and BM were collected from the patients at diagnosis and/or after treatment. HLA-A expression on granulocytes, monocytes, B and T cells, and BM $CD34^+$ cells was analyzed by flow cytometry using a FACSCanto II instrument (BD Biosciences) with the FlowJo 7.6.1 program (TreeStar). The monoclonal antibodies used for this study are provided in supplemental Table 2.

### Human androgen receptor assay

The human androgen receptor gene was amplified from genomic DNA of 23 female patients, including 3 6pLOH(+) patients, as described by Ishiyama et al[21] with some modifications. Clonality was assessed using an "S value" as a marker of skewing in granulocytes and T lymphocytes.

### Association of HLA types with AA

A total of 6613 patients who had received allogeneic BM transplantation through the JMDP between 1992 and 2008 were investigated to see whether the HLA alleles frequently missing in CNN-LOH in 6p with the development of AA could represent risk alleles for the development of AA. Thus, the frequencies of patients with each of the candidate risk alleles (HLA-A*31:01, B*40:02, A*02:01, and A*02:06) and those having none of these alleles were compared between 407 patients with AA and those with other hematopoietic disorders (1827 with AML, 1606 with acute lymphocytic leukemia, 1014 with chronic myeloid leukemia, 825 with myelodysplastic syndrome, 566 with non-Hodgkin lymphoma, and 368 with other hematopoietic neoplasms; supplemental Table 3) by calculating the Fisher *P* values in the corresponding 2 × 2 contingency tables.

## Results

### Genetic lesions in AA detected by SNP array analysis

After excluding known or suspected copy number variations, a total of 50 genetic lesions were identified in 46 of the 306 (15%) PB specimens of our AA case series (Table 1; Figure 1). Among these by far, the most conspicuous was the recurrent CNN-LOH involving the 6p arm, which was detected in 28 cases as a significant dissociation of allele-specific copy number graphs in 6p regions using a hidden Markov model–based algorithm implemented in the CNAG software[2,14,15] (Figure 2A-2B). Of particular interest was that all CNN-LOH in 6p commonly affected the HLA locus, causing a haploid loss of HLA alleles and uniparental HLA expression. In some cases, the breakpoint of the 6pLOH was

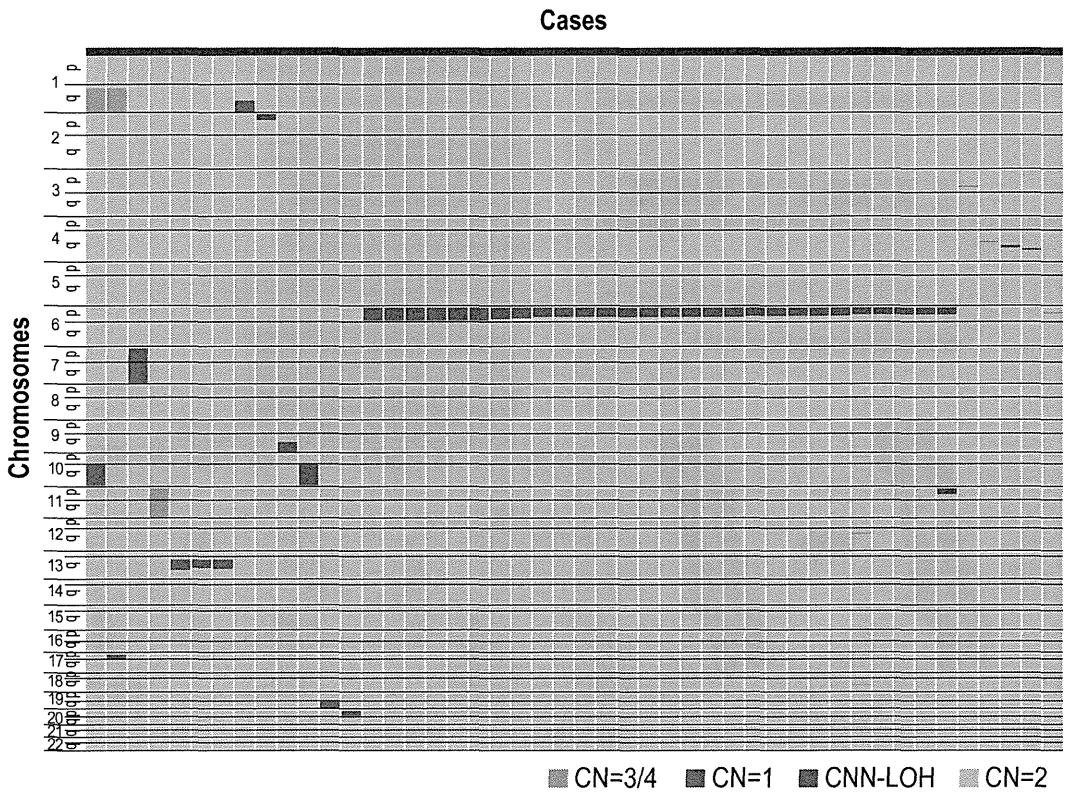


**Figure 1. Copy number changes and allelic imbalances in 46 of the 306 AA cases.** The copy number changes and allelic imbalances (or CNN-LOHs) in each case are summarized in the chromosomal order vertically for 46 AA cases with copy number abnormalities. Gains and losses, as well as CNN-LOHs, are shown in the indicated colors.

predicted to fall within the HLA locus (Figure 2B). These findings strongly indicated that the HLA locus was the genetic target of these 6pLOHs. Also supporting this was the finding that, in half of the cases, the dissociations in the allele-specific copy number graphs were gradually attenuated to the baseline over several mega base pair regions rather than showing a discrete breakpoint, indicating the presence of multiple 6pLOH(+) clones within a single case that had different breakpoints but still shared the same missing HLA alleles (Figure 2C). Moreover, the 6pUPDs existing only in a minor population were more sensitively detected by statistically evaluating the size of dissociation of allele-specific copy numbers in the 6p arm. With this improved statistical test, CNN-LOH in 6p was found in a total of 40 cases (13%; Figure 2D; supplemental Figure 2), where the false discovery rate was maintained at 0.01 to avoid too many false positive findings. In all 6pLOH(+) cases, substantial numbers of heterozygous SNP calls were retained within the affected regions, thus indicating that the CNN-LOHs in 6p were not constitutional but represented acquired genetic events only found in the affected subclones (Figure 1). Indeed, all 6pLOH(+) cases were shown to have "heterozygous" HLA alleles in high-resolution HLA typing of their PB (Table 2). Moreover, 6pLOH was not detected in the CD3-positive T cells in selected cases (cases 25 and 26, supplemental Figure 3). By quantitatively comparing the observed differences in allele-specific copy numbers in the 6pLOH segments with what were expected assuming 100% LOH(+) components, the 6pLOH(+) clones were estimated to account for 0.2% to 53.9% of the PB leukocytes (Table 2). The trend of the lower percentages of the 6pLOH(+) fraction in newly diagnosed patients compared with those in patients at remission was thought to reflect the fact that the former patients tended to have lower counts of granulocytes and monocytes, which were the predominant targets of 6pLOH (see supplemental Table 1).

The disease status of the 40 patients at the sampling was before treatment in 16 cases, during remission for 1 to 16 years after therapies in 15, and before BM transplantation for refractory disease in 9. All evaluable 6pLOH(+) AA cases responded to immunosuppressive therapy (IST) (23 of 23), whereas 101 of 126 evaluable cases with 6pLOH(−) responded ($P = .014$; Table 3).

### Uniparental expression of HLA-A in multilineage hematopoietic cells

The genetic loss of one HLA haplotype in SNP array analysis was further confirmed by expression analysis of HLA-A in PB leukocytes using flow cytometry in 19 eligible cases with 6pLOH(+), in which the HLA-A alleles were heterozygous and fresh PB samples were available. Loss of expression of one HLA-A antigen was confirmed in all 19 6pLOH(+) cases (Figure 3A; supplemental Figure 4). The HLA-A missing cells in the PB were shown to have appeared shortly after the onset or before the initiation of treatments in 2 cases, and were confirmed to persist for 1 to 16 months (median, 6 months) in 14 patients (supplemental Table 1; supplemental Figure 5). The percentage of granulocytes lacking HLA-A antigens in the 2 patients who were responsive to IST remained almost the same during the convalescent period of 2 to 3 months (supplemental Figure 6). Importantly, uniparental expression of HLA-A alleles was detected in multiple cell lineages, including granulocytes, monocytes, B cells, and, to a lesser extent, in T cells. Moreover, uniparental HLA-A expression was demonstrated in BM $CD34^+$ cells in 5 patients whose BM samples were available for flow cytometry. All 5 patients possessed various proportions of BM $CD34^+$ cells (49.7%-71.3%), which had lost the expression of one

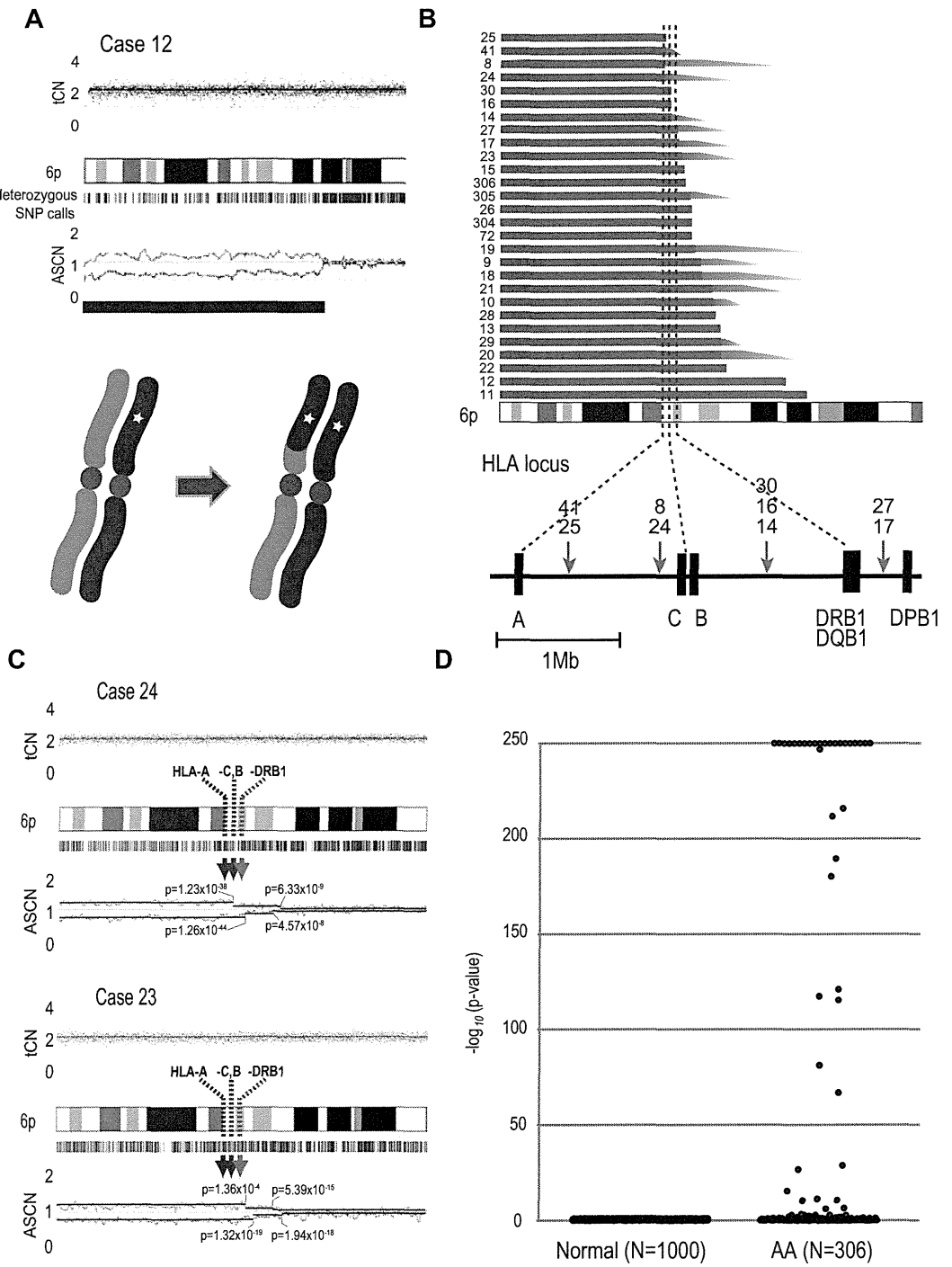


**Figure 2. Acquired 6pLOHs in AA patients that target the HLA locus.** (A) Typical CNAG outputs in SNP array analysis showing CNN-LOH (purple line) that appears as significant dissociation in allele-specific copy number graphs (red and green lines) from the baseline with normal total copy numbers (tCN; top panel). As a result of an allelic conversion, the affected segment causes LOH (* indicates 1; bottom panel). The "acquired" origin of these lesions is indicated by the retention of substantial numbers of heterozygous SNP calls (green bars below the chromatogram) that would otherwise mostly disappear. (B) The breakpoints of 6pLOHs found in a total of 28 AA cases, all involving the HLA locus in common. In more than half of cases (indicated by arrowheads in panel B), the exact location of the breakpoint was difficult to uniquely determine, where dissociation of the allele-specific copy number graphs continuously tapered along the 6p arm, indicating the presence of multiple 6pLOH(+) clones with common missing alleles (C). Indeed, the breakpoint containing regions are separated into multiple segments having significantly different copy numbers in the circular binary segmentation model, as indicated by solid lines with *P* values. Note that the most telomeric breakpoint is located within (case 24) or centromeric to (case 23) the HLA locus in each case. (D) A skewed distribution of the logarithm of *P* values in AA cases compared with normal persons. The *P* values were calculated in the Mann-Whitney *U* test, with which the difference in the mean allele-specific copy numbers between 6p and other chromosomal regions was evaluated (see "Analysis of genomic copy numbers and detection of 6pLOH"). A total of > 250 values are plotted as 250.

HLA-A antigen; and in each case, the missing HLA-A allele was identical to that in the PB leukocytes (Figure 3B). The uniparental expression of HLA-A in case 13 was also observed in the $CD34^+$ compartment of the archived BM specimen obtained 2 years before analysis (supplemental Figure 7). Together, these findings suggested that the 6pLOH involved early HSPCs and that the 6pLOH occurred at the level of long-term repopulating stem cells.

**Table 2. 6pLOH(+) AA cases and imputed allelic status of HLA alleles**

| UID | 6pUPD(+) fraction,* % | Missing alleles | | | | | | Retained alleles | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | DRB1 | DQB1 | DPB1 | A | B | C | DRB1 | DQB1 | DPB1 |
| 19 | 53.9 | 31:01†‡ | 40:02† | 03:04† | 12:01 | 03:01 | 05:01 | 24:02 | 52:01 | 12:02 | 15:02 | 06:01 | 05:01 |
| 12 | 51.8 | 02:01†‡ | 40:02† | 03:03 | 15:01 | 06:02 | 05:01 | 26:02 | 40:06 | 08:01 | 09:01 | 03:03 | 05:01 |
| 17 | 51.6 | 24:02 | 13:01 | 03:04† | 12:02 | 03:01 | 04:02 | 24:02 | 52:01 | 12:02 | 15:02 | 06:01 | 09:01 |
| 304 | 49.3 | 31:01†‡ | 55:02 | 01:02 | 12:02 | 03:01 | 41:01 | 24:02 | 07:02 | 07:02 | 01:01 | 05:01 | 04:02 |
| 11 | 48.0 | 02:06†‡ | 40:02† | 03:04† | 15:01 | 06:02 | ND‖ | 11:01 | 67:01 | 07:02 | 16:02 | 05:02 | ND‖ |
| 21 | 46.2 | 31:01‡§ | 51:01 | 14:02 | 14:05 | 05:03 | 03:01 | 24:02 | 07:02 | 07:02 | 01:01 | 05:01 | 04:02 |
| 24 | 44.9 | 31:01† | 40:02† | 03:04† | 11:01 | 03:01 | 02:01 | 24:02 | 40:06 | 08:01 | 09:01 | 03:03 | 05:01 |
| 26 | 44.3 | 31:01†‡§ | 40:01 | 03:04† | 04:05 | 04:01 | 03:01 | 26:03 | 52:01 | 12:02 | 15:02 | 06:01 | 09:01 |
| 27 | 43.5 | 02:06† | 40:02† | 03:04† | 04:10 | 04:02 | 02:01 | 11:01 | 52:01 | 12:02 | 15:02 | 06:01 | 09:01 |
| 10 | 42.1 | 31:01† | 40:02† | 03:04† | 08:03 | 06:01 | 02:01 | 24:02 | 51:01 | 14:02 | 09:01 | 03:03 | 02:01 |
| 8 | 40.8 | 02:06†‡ | 40:02† | 03:03 | 12:01 | 03:01 | 05:01 | 24:02 | 52:01 | 12:02 | 15:02 | 06:01 | 04:02 |
| 23 | 35.2 | 02:01† | 40:02† | 03:04† | 09:01 | 03:03 | 02:01 | 24:02 | 54:01 | 01:02 | 04:05 | 04:01 | 04:02 |
| 25 | 32.1 | 02:06†‡ | | | No LOH | | | 01:01 | | | No LOH | | |
| 9 | 23.5 | 02:06†‡ | 39:01 | 07:02 | 08:02 | 04:02 | 02:01 | 24:02 | 15:18 | 07:04 | 04:01 | 03:01 | 14:01 |
| 20 | 21.7 | 26:01‡ | 40:02† | 03:03 | 15:01 | 06:02 | 05:01 | 02:18 | 46:01 | 01:02 | 08:03 | 06:01 | 05:01 |
| 14 | 21.7 | 31:01†‡ | 51:01 | 14:02 | 09:01 | 03:03 | 05:01 | 24:02 | 52:01 | 12:02 | 15:02 | 06:01 | 09:01 |
| 22 | 20.6 | 02:01† | 39:01 | 07:02 | 08:03 | 06:01 | 05:01 | 24:02 | 52:01 | 12:02 | 15:02 | 06:01 | 09:01 |
| 18 | 17.6 | 02:01†‡ | 40:06 | 08:01 | 09:01 | 03:03 | 02:01 | 24:02 | 35:01 | 03:03 | 15:01 | 06:02 | 04:02 |
| 15 | 17.4 | 02:06† | 40:06 | 08:01 | 09:01 | 03:03 | 02:01 | 24:02 | 07:02 | 07:02 | 01:01 | 05:01 | 02:01 |
| 41 | 15.2† | 31:01†‡ | 35:01 | 03:03 | 09:01 | 03:03 | 03:01 | 26:01 | 39:01 | 07:02 | 08:03 | 06:01 | 05:01 |
| 28 | 12.8 | 24:02 | 54:01 | 01:02 | 01:01 | 05:01 | 04:02 | 24:02 | 52:01 | 12:02 | 15:02 | 06:01 | 09:01 |
| 29 | 11.7 | 31:01† | 40:02† | 03:04† | 15:01 | 06:02 | 02:01 | 24:02 | 54:01 | 01:02 | 04:05 | 04:01 | 05:01 |
| 305 | 10.3 | 02:06†‡ | 40:02† | 15:02 | 15:02 | 06:01 | 04:01 | 24:02 | 51:01 | 14:02 | 09:01 | 03:03 | 02:01 |
| 13 | 9.6 | 24:02‡ | 40:02† | 03:04† | 15:01 | 06:02 | 02:01 | 02:01‡ | 35:01 | 08:01 | 09:01 | 03:03 | 02:01 |
| 306 | 8.5 | 24:02‡ | 40:02† | 03:04† | 09:01 | 03:03 | 02:01 | 26:02 | 40:06 | 08:01 | 09:01 | 03:03 | 02:01 |
| 16 | 8.1 | 11:01 | 40:06 | 08:01 | | No LOH | | 24:02 | 46:01 | 01:02 | | No LOH | |
| 30 | 8.0 | 02:06† | 39:01 | 07:02 | | No LOH | | 24:02 | 40:06 | 08:01 | | No LOH | |
| 72 | 5.6 | 02:01† | 40:02† | 03:04† | 09:01 | 03:03 | 05:01 | 02:07 | 46:01 | 01:02 | 08:03 | 06:01 | 02:02 |
| 36 | 4.0 | 02:01†‡ | ND¶ | ND# | 15:02 | 06:01 | 09:01 | 24:02 | ND¶ | ND# | 15:02 | 06:01 | 09:01 |
| 124 | 3.5 | 24:02 | 40:02† | 03:04† | 12:01 | 03:01 | 02:01 | 24:02 | 52:01 | 12:02 | 15:02 | 06:01 | 09:01 |
| 223 | 2.8 | 31:01†‡ | 48:01 | 03:04† | 09:01 | 03:03 | 05:01 | 02:06† | 39:01 | 07:02 | 15:01 | 06:02 | 02:01 |
| 215 | 2.8 | 31:01† | 51:01 | 14:02 | 08:02 | 04:02 | 04:02 | 03:01 | 44:02 | 05:01 | 13:01 | 06:03 | 05:01 |
| 181 | 1.3 | 02:06† | 13:01 | 03:04† | 12:02 | 03:01 | 05:01 | 24:02 | 52:01 | 12:02 | 15:02 | 06:01 | 09:01 |
| 97 | 1.0 | 24:02 | 07:02 | 07:02 | 01:01 | 05:01 | 05:01 | 02:01† | 39:01 | 07:02 | 15:01 | 06:02 | 02:01 |
| 252 | 0.9 | ND** | 40:02† | 03:04† | 09:01 | 03:03 | 05:01 | ND** | 46:01 | 01:02 | 04:05 | 04:01 | 05:01 |
| 118 | 0.9 | 02:06‡§ | 40:02† | 03:04† | 08:02 | 03:02 | 05:01 | 24:02 | 52:01 | 12:02 | 15:02 | 06:01 | 09:01 |
| 298 | 0.8 | 24:02 | 40:02† | 03:04† | 15:01 | 06:02 | 05:01 | 24:02 | 52:01 | 12:02 | 15:02 | 06:01 | 09:01 |
| 188 | 0.7 | 24:02 | 52:01 | 12:02 | 15:02 | 06:01 | 09:01 | 02:01† | 52:01 | 12:02 | 11:01 | 03:01 | 05:01 |
| 291 | 0.7 | 31:01† | 51:01 | 14:02 | 15:01 | 06:02 | 02:01 | 24:02 | 40:01 | 03:04† | 11:01 | 03:01 | 05:01 |
| 196 | 0.2 | ND†† (A*02:06/24:02, B*35:01/51:01, C*03:03/15:02, DRB1*04:03/15:01, DQB1*03:02/06:02, DPB1*0:201/02:01) | | | | | | | | | | | |

UID indicates unique ID.

*The percentage of 6pUPD(+) fraction is derived from total peripheral blood leukocytes that include lymphoid as well as myeloid element.

†HLA types significantly deviated to missing alleles.

‡The allelic loss was confirmed by flow cytometry.

§The missing haplotype was determined by flow cytometry.

‖DPB1*04:02/05:01.

¶B*15:18/52:01.

#C*08:01/12:02.

**A*02:01/02:07.

††Missing allele was not determined because copy number changes in these segments were not statistically significant.

### Clonality of the HLA-missing granulocytes

The human androgen receptor-based clonality assays in granulocytes were performed in 3 6pLOH(+) and 20 6pLOH(−) patients, in which all 3 6pLOH(+) and 4 (20%) of the 6pLOH(−) patients showed evidence of clonality in granulocyte populations (supplemental Figure 8).

### Missing HLA alleles in 6pLOH

Given that the HLA is the genetic target of 6pLOH in AA, the missing HLA alleles in 6pLOH are of particular interest because in this context they are thought to be directly involved in the presentation of the target auto-antigens to CTLs and, therefore, to be critically important in the pathogenesis of AA. We determined the missing HLA alleles in each 6pLOH(+) AA patient by the haplotype imputation of HLA alleles based on the large data of HLA haplotypes observed in the JMDP set, followed by statistical evaluation of allele-specific copy numbers along the imputed haplotypes (Figure 4). The imputed haplotypes were confirmed in 4 cases by the family studies on the HLA. The allelic status was imputed at least partially in 39 of the 40 6pLOH(+) cases. The imputed results were consistent with the patterns of uniparental expression of HLA-A in flow cytometry in 18 cases with 6pLOH (Table 2; Figure 4), except for those in case 26, in which no valid SNP haplotype

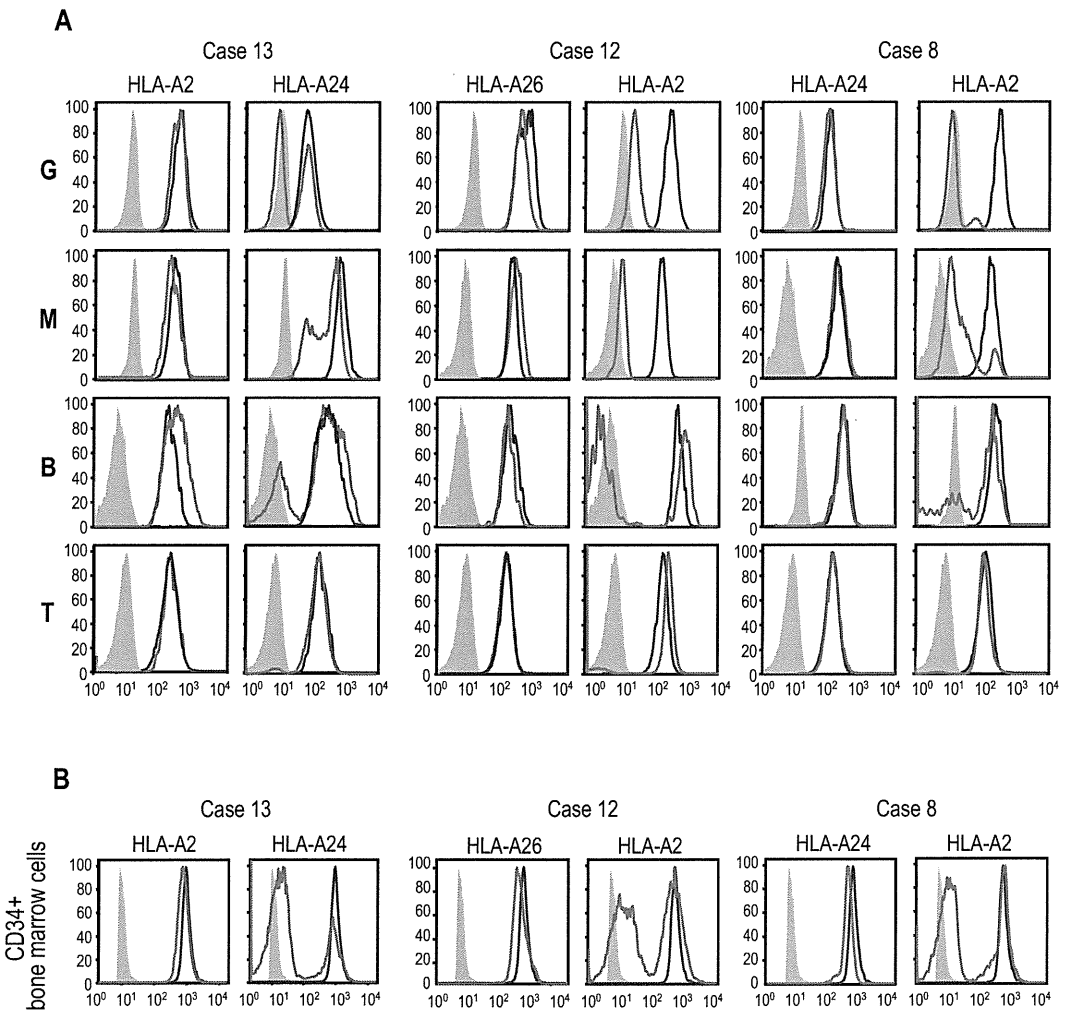


**Figure 3. Uniparental expression of HLA in AA cases with CNN-LOH in 6p.** Allele-specific expression of HLA-A antigens in AA specimens was examined by flow cytometry using monoclonal antibodies that specifically recognize the indicated HLA types (red lines), where leukocytes from healthy persons were used as a control (blue lines). (A-B) The uniparental expression of HLA-A antigens in PB leukocytes and BM CD34$^+$ cells obtained from 3 AA cases with CNN-LOH in 6p. Different leukocyte compartments were separately examined, including granulocytes (G), monocytes (M), B-lymphocytes (B), and T-lymphocytes (T).

around the HLA-A locus was identified and the status of HLA-A was determined by flow cytometry. The missing HLA alleles in 6pLOH(+) AA showed a conspicuous deviation to some selected HLA alleles, including HLA-A*31:01, B*40:02, C*03:04, and, to a lesser extent, HLA-A*02:01 and A*02:06. After the effects of linkage disequilibrium between individual HLA alleles were taken into consideration by multivariate analysis, 4 HLA alleles were shown to remain as the principal determinants of the missing haplotypes, HLA-A*31:01, B*40:02, A*02:01, and A*02:06 (supplemental Table 4).

### Over-representation of frequently missing HLAs in AA populations

Because these missing HLA alleles in 6pLOH could be involved in the pathogenesis of AA, we next tested whether these relevant HLA alleles are associated with the risk of the development of AA among the 6,613 JMDP registrants. As shown in Table 4, the 4 major missing HLA alleles, HLA-A*31:01, B*40:02, A*02:01, and A*02:06, were more frequently observed in AA cases compared with nonsignificant HLA alleles (ie, all HLA alleles other

**Table 3. Response rate (CR + PR) according to the Camitta criteria**

| | Newly diagnosed (n = 107) | | Previously treated (n = 103) | |
|---|---|---|---|---|
| | 6pLOH(−) (n = 91), no. (%) | 6pLOH(+) (n = 16), no. (%) | 6pLOH(−) (n = 88), no. (%) | 6pLOH(+) (n = 15), no. (%) |
| **Immunosuppressive therapies (all)** | 36/49 (73) | 11/11 (100) | 65/77 (84) | 12/12 (100) |
| ATG + CsA | 14/19 (74) | 7/7 (100) | 27/33 (82) | 5/5 (100) |
| CsA alone | 22/30 (73) | 4/4 (100) | 38/44 (86) | 7/7 (100) |
| Anabolic steroid alone | 0/0 (0) | 0/0 (0) | 7/11 (64) | 2/2 (100) |
| Unknown/not evaluable | 42 | 5 | 0 | 1 |

CR indicates complete remission; PR, partial remission; ATG, antithymocyte globulin; and CsA, cyclosporine A.

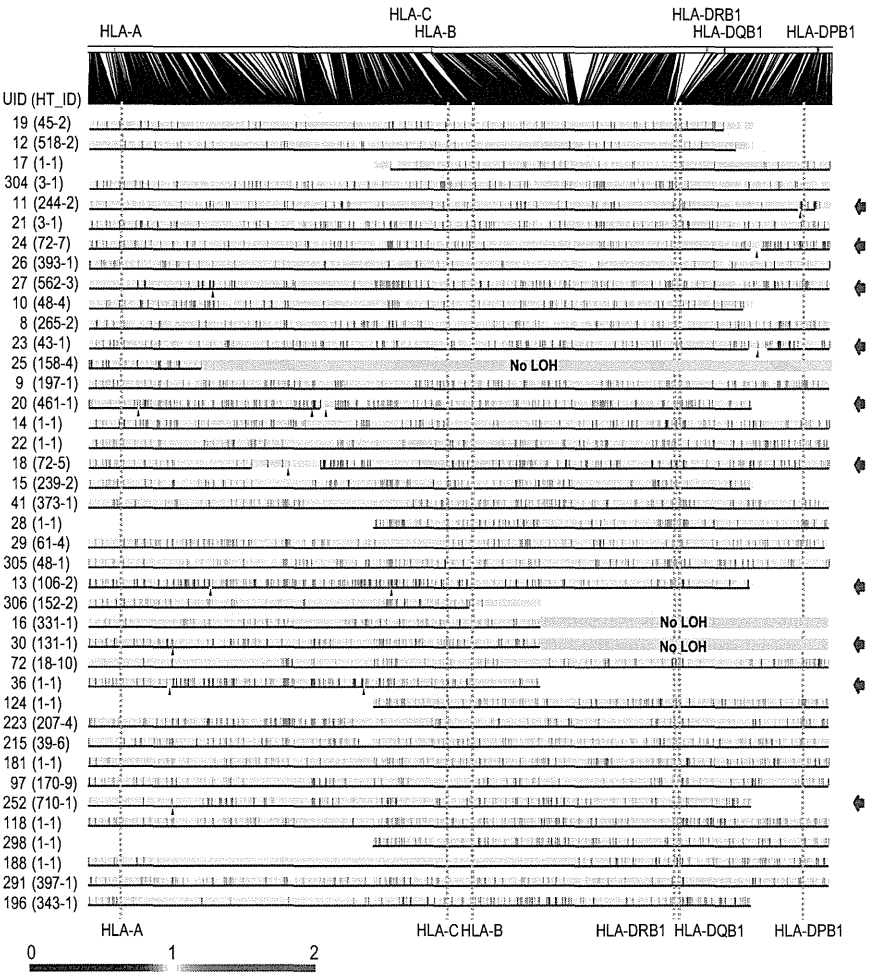


**Figure 4. Imputation of missing HLA haplotypes.** The observed allelic copy numbers at heterozygous SNP sites along each candidate SNP haplotype are color-coded as indicated at the bottom. Green bars showed the SNPs that are incompatible with the patient's genotype. Case IDs and haplotype ID (HT_ID) are indicated on the left. The locations of the 500K SNPs and HLA-A, C, B, DRB1, DQB1, and DPB1 are indicated in the figure. For each allele, genomic copy numbers were imputed using *the circular binary segmentation algorithm*. This divided each haplotype into one or more segments having discrete mean allelic copy numbers (blue arrows on the right). The positions of breakpoints are indicated by arrowheads. Finally, the mean allelic copy number of each segment was statistically compared with that of the corresponding segment on the other haplotype using the Wilcoxon signed rank test. Missing HLA haplotypes were determined based on the result of the statistic tests. Purple and blue lines indicated the retained and missing segments, respectively, whereas the allelic status was not determined statistically for those segments shown by green lines.

than these 4 alleles), where the odds ratios for the risk of the development of AA between each of these alleles and nonsignificant alleles were 1.87 (95% confidence interval [CI], 1.43-2.43) for A*02:01, 2.22 (95% CI, 1.70-2.90) for A*02:06, 1.37 (95% CI, 1.00-1.88) for A*31:01, and 1.95 (1.48-2.58) for B*40:02 (Table 4). The combined relative risk for all these alleles was 1.75 (1.42-2.17; $P = 1.3 \times 10^{-7}$).

## Discussion

The origin of clonal hematopoiesis in AA is a focus of long-standing disputes, in which a profoundly reduced hematopoietic stem cell pool and/or escape from the autoimmune insults have been implicated in the evolution of the clonal hematopoiesis in AA.[5,22,23] Our findings on 6pLOH in AA provide an intriguing insight not only into the underlying mechanism of the clonal hematopoiesis in AA but also into the origin of the autoimmunity that is responsible for the pathogenesis of AA. A recent study from the United States also reported 3 cases with 6pLOH.[24] With a sensitive detection algorithm, the presence of the 6pLOH(+) components was demonstrated in as many as 13% of typical cases with AA, and the evidence from the subsequent studies strongly indicated that the HLA genes are the genetic targets of 6pLOH in AA patients. First, the HLA locus was commonly and critically involved in all 6pLOHs found in AA. Second, some AA patients carried multiple 6pLOH(+) subclones with different breakpoints, but in all cases, the 6pLOH involved the HLA locus and occurred in a manner that targeted the same parental HLA allele. Moreover, particular class I HLA alleles were over-represented among 6pLOH(+) cases and consistently found in the missing haplotypes. Finally, many of these HLA alleles were shown to be tightly

**Table 4. Association of missing HLA alleles with AA in Japanese patients**

| Risk allele | AA (N = 407) | Other diseases (N = 6206) | Total (N = 6613) | $P$ ($\chi^2$ test) | Odds ratio (95% CI) (vs no risk alleles) |
|---|---|---|---|---|---|
| A*02:01 | 103 | 1173 | 1276 | $2.5 \times 10^{-6}$ | 1.87 (1.43–2.43) |
| A*02:06 | 100 | 957 | 1057 | $< 1.0 \times 10^{-7}$ | 2.22 (1.70–2.90) |
| A*31:01 | 58 | 899 | 957 | 0.048 | 1.37 (1.00–1.88) |
| B*40:02 | 86 | 938 | 1024 | $1.8 \times 10^{-6}$ | 1.95 (1.48–2.58) |
| All risk alleles | 268 | 3250 | 3518 | $1.3 \times 10^{-7}$ | 1.75 (1.42–2.17) |
| No risk alleles | 139 | 2956 | 3095 | — | — |

— indicates not applicable.

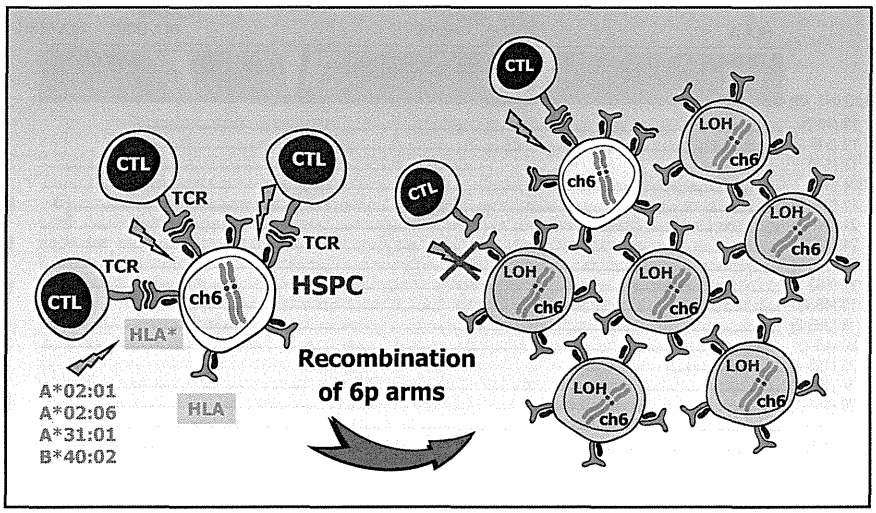


**Figure 5. A proposed mechanism for escape hematopoiesis in 6pLOH(+) AA.** In AA, the targets of CTLs are the HSPCs that present some auto-antigen through particular class I HLA molecules, including HLA-A*02:01, A*02:06, A*31:01, and B*40:02. In the presence of these autoimmune insults, the HSPCs that lose their expression of the antigen-presenting HLA molecule as a result of CNN-LOH in 6p would acquire a growth advantage over other HSPCs expressing the relevant HLA, leading to clonal outgrowth of the 6pLOH(+) progenies.

associated with the development of AA in Japanese patients in case-control studies using the large JMDP registry.

The conspicuous bias of the missing HLA alleles in 6pLOH to particular HLA types and the significant association of AA with those HLA types strongly suggest that the recurrent 6pLOH in AA is a phenomenon tightly related to the pathogenesis of AA rather than mere secondary event during the course of AA. Based on these observations, it is well reasoned that, in 6pLOH(+) AA cases, the autoimmunity to HSPCs is mediated by the CTLs that target the antigens presented via specific class I HLA molecules and that the 6pLOH(+) cells found in AA could be explained as escape hematopoiesis that survives the autoimmune insult by genetically deleting the relevant HLA species that are required for antigen presentation (Figure 5). These scenarios are further supported by the recent reports showing that the CNN-LOH in 6p provides a common mechanism of leukemic relapse after HLA haploidentical stem cell transplantations, in which leukemic cells that lost the mismatched HLA haplotype through CNN-LOH in 6p are thought to escape the immunologic surveillance of the engrafted donor T cells.[25,26] Importantly, it was experimentally demonstrated by immunologic assays that the 6pLOH(+) leukemic cells actually escaped GVL by CTLs, whereas 6pLOH(−) leukemic cells were effectively killed by the same CTLs. Although the immunologic targets of CTLs are different between relapse after haploidentical transplants (mismatched HLAs themselves) and AA (still unknown autoantigens presented on missing HLAs), the prominent similarities found in both cases further support that CNN-LOH in 6p confers an escape mechanism from autoreactive CTLs in AA.

In light of the above considerations, the chronologic behavior of the 6pLOH(+) components in PB is also interesting and worth discussing. Despite the assumption that 6pLOH is an effective escape mechanism from CTLs, the 6pLOH(+) stem cells were unable to repopulate the BM to cure AA, unless effective IST was applied (supplemental Figure 6). This is most probably explained by the presence of inflammatory cytokines, such as IFN-γ and TNF-α, which have also been shown to play an important role in the BM failure in AA and are thought to be responsible for the continued prevention of the 6pLOH(+) stem cells from fully expanding and reconstituting the BM (supplemental Figure 9A-B).[27,28]

When the autoimmune insults are removed after IST, no further injury of normal stem cells would occur. However, this does not necessarily mean the surviving normal stem cells can eventually outnumber the 6pLOH(+) stem cells over time. Note that, once the autoimmune insults disappear, nothing could biologically or immunologically discriminate a 6pLOH(+) stem cell from a 6pLOH(−) stem cell (supplemental Figure 9A). In particular, a 6pLOH(+) stem cell and a 6pLOH(−) stem cell will produce the same number of progeny on average and feed the same number of mature blood cells. As a consequence, once established, the predominance of 6pLOH(+) stem cells over 6pLOH(−) stem cells should be maintained, after the severely reduced hematopoietic stem cell pool has been re-expanded with removal of the inciting autoimmunity. It is also of note that the recovery of myeloid components after IST, which are affected more strongly by 6pLOH than lymphoid cells, contributes to an apparent increase in 6pLOH components in the SNP array analysis in PB (supplemental Figure 6A).

One of the most significant findings in the current study is the identification of the HLA alleles that are over-represented in the Japanese AA populations, including HLA-A*31:01, B*40:02, A*02:01, and A*02:06. All of these HLA alleles belong to class I MHCs and thus are thought to be involved in the antigen presentation to CTLs. This provides another prominent example, in which specific HLA types play a critical role in the development of a human disease, and the information about these particular HLA types provides a solid basis on which we can ultimately isolate the relevant antigens responsible for the development of AA. Of particular note, there was a previous report indicating that HLA-B*40:02 and A*02:06 were over-represented in PNH as well as AA, although the study size was much smaller than the current study.[29] Combined with our study, these findings support the hypothesis that AA and PNH are the different outcomes of the same immunologic insult[5,30] and may also provide the genetic basis of the high prevalence of AA and PNH in East Asia.[31,32]

In some AA cases, hematopoiesis could be maintained over years by the progenitors that escaped and survived the inciting autoimmune insult by deleting the target HLA through CNN-LOH in 6p. Given that the 6pLOH was detected in only 13% of our series, it is probable that other escape mechanisms may also operate to maintain hematopoiesis in AA. Indeed, clonality was clearly demonstrated in 20% of the 6pLOH(−) cases in the human androgen receptor assay study (supplemental Figure 8). In addition, our SNP array analysis also revealed a variety of clonal abnormalities in AA cases (Figure 1), although it is still open to question

whether these abnormalities actually represent the mechanism of escape hematopoiesis or were related to some neoplastic process. Further studies on the genetic basis of the escape mechanisms would contribute to our understanding of the molecular pathogenesis of AA.

## Acknowledgments

The authors thank the patients and donors and their physicians, including K. Kawakami of Suzuka General Hospital and A. Okamoto of Nagoya Daini Red Cross Hospital, for contributing to this study.

This work was supported in part by the Core Research for Evolutional Science and Technology, the Japan Science and Technology Agency, the Ministry of Education, Culture, Sports, Science and Technology of Japan (Grant-in-Aids for Scientific Research), and the Ministry of Health, Labor and Welfare of Japan (Grant-in-Aids).

## Authorship

Contribution: S. Ohtake, S. Ogawa, and S.N. developed the concept of the study and supervised the project; T.K., S. Ohtake, and S.N. designed the experiments; T.K., A.S.-O., Y. Sato, Y. Mori, M.K., M.S., K.H., and Y. Sasaki performed the experiments and analyzed the data; K.K. performed high-resolution HLA typing; S.M. and Y. Morishima provided the information of JMDP donor-recipient pairs (JMDP dataset); T.K., A.S.-O., S. Ogawa, and S.N. wrote the paper; and all authors approved the final version of the manuscript.

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

Correspondence: Shinji Nakao, Cellular Transplantation Biology, Kanazawa University Graduate School of Medical Science, 13-1 Takaramachi, Kanazawa, Ishikawa 920-8640 Japan; e-mail: snakao@med3.m.kanazawa-u.ac.jp.

## References

1. Young NS, Calado RT, Scheinberg P. Current concepts in the pathophysiology and treatment of aplastic anemia. *Blood.* 2006;108(8):2509-2519.
2. Nakao S, Takami A, Takamatsu H, et al. Isolation of a T-cell clone showing HLA-DRB1*0405-restricted cytotoxicity for hematopoietic cells in a patient with aplastic anemia. *Blood.* 1997;89(10):3691-3699.
3. Chen J, Ellison FM, Eckhaus MA, et al. Minor antigen h60-mediated aplastic anemia is ameliorated by immunosuppression and the infusion of regulatory T cells. *J Immunol.* 2007;178(7):4159-4168.
4. Risitano AM, Maciejewski JP, Green S, Plasilova M, Zeng W, Young NS. In-vivo dominant immune responses in aplastic anaemia: molecular tracking of putatively pathogenetic T-cell clones by TCR beta-CDR3 sequencing. *Lancet.* 2004;364(9431):355-364.
5. Young NS. The problem of clonality in aplastic anemia: Dr Dameshek's riddle, restated. *Blood.* 1992;79(6):1385-1392.
6. Tiu R, Gondek L, O'Keefe C, Maciejewski JP. Clonality of the stem cell compartment during evolution of myelodysplastic syndromes and other bone marrow failure syndromes. *Leukemia.* 2007;21(8):1648-1657.
7. Lewis SM, Dacie JV. The aplastic anaemia-paroxysmal nocturnal haemoglobinuria syndrome. *Br J Haematol.* 1967;13(2):236-251.
8. Dameshek W. Riddle: what do aplastic anemia, paroxysmal nocturnal hemoglobinuria (PNH) and "hypoplastic" leukemia have in common? *Blood.* 1967;30(2):251-254.
9. Socie G, Rosenfeld S, Frickhofen N, Gluckman E, Tichelli A. Late clonal diseases of treated aplastic anemia. *Semin Hematol.* 2000;37(1):91-101.
10. Tichelli A, Gratwohl A, Wursch A, Nissen C, Speck B. Secondary leukemia after severe aplastic anemia. *Blut.* 1988;56(2):79-81.
11. de Planque MM, Kluin-Nelemans HC, van Krieken HJ, et al. Evolution of acquired severe aplastic anaemia to myelodysplasia and subsequent leukaemia in adults. *Br J Haematol.* 1988;70(1):55-62.
12. van Kamp H, Landegent JE, Jansen RP, Willemze R, Fibbe WE. Clonal hematopoiesis in patients with acquired aplastic anemia. *Blood.* 1991;78(12):3209-3214.
13. Kawase T, Morishima Y, Matsuo K, et al. High-risk HLA allele mismatch combinations responsible for severe acute graft-versus-host disease and implication for its molecular mechanism. *Blood.* 2007;110(7):2235-2241.
14. Nannya Y, Sanada M, Nakazaki K, et al. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res.* 2005;65(14):6071-6079.
15. Yamamoto G, Nannya Y, Kato M, 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.* 2007;81(1):114-126.
16. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 2003;100(16):9440-9445.
17. Ogawa S, Matsubara A, Onizuka M, et al. Exploration of the genetic basis of GVHD by genetic association studies. *Biol Blood Marrow Transplant.* 2009;15(1 suppl):39-41.
18. Morishima S, Ogawa S, Matsubara A, et al. Impact of highly conserved HLA haplotype on acute graft-versus-host disease. *Blood.* 2010;115(23):4664-4670.
19. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics.* 2004;5(4):557-572.
20. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics.* 2007;23(6):657-663.
21. Ishiyama K, Chuhjo T, Wang H, Yachie A, Omine M, Nakao S. Polyclonal hematopoiesis maintained in patients with bone marrow failure harboring a minor population of paroxysmal nocturnal hemoglobinuria-type cells. *Blood.* 2003;102(4):1211-1216.
22. Murakami Y, Kosaka H, Maeda Y, et al. Inefficient response of T lymphocytes to glycosylphosphatidylinositol anchor-negative cells: implications for paroxysmal nocturnal hemoglobinuria. *Blood.* 2002;100(12):4116-4122.
23. Bessler M, Mason PJ, Hillmen P, et al. Paroxysmal nocturnal haemoglobinuria (PNH) is caused by somatic mutations in the PIG-A gene. *EMBO J.* 1994;13(1):110-117.
24. Afable MG 2nd, Wlodarski M, Makishima H, et al. SNP array-based karyotyping: differences and similarities between aplastic anemia and hypocellular myelodysplastic syndromes. *Blood.* 2011;117(25):6876-6884.
25. Vago L, Perna SK, Zanussi M, et al. Loss of mismatched HLA in leukemia after stem-cell transplantation. *N Engl J Med.* 2009;361(5):478-488.
26. Villalobos IB, Takahashi Y, Akatsuka Y, et al. Relapse of leukemia with loss of mismatched HLA resulting from uniparental disomy after haploidentical hematopoietic stem cell transplantation. *Blood.* 2010;115(15):3158-3161.
27. Zoumbos NC, Gascon P, Djeu JY, Trost SR, Young NS. Circulating activated suppressor T lymphocytes in aplastic anemia. *N Engl J Med.* 1985;312(5):257-265.
28. Hinterberger W, Adolf G, Aichinger G, et al. Further evidence for lymphokine overproduction in severe aplastic anemia. *Blood.* 1988;72(1):266-272.
29. Shichishima T, Noji H, Ikeda K, Akutsu K, Maruyama Y. The frequency of HLA class I alleles in Japanese patients with bone marrow failure. *Haematologica.* 2006;91(6):856-857.
30. Karadimitris A, Manavalan JS, Thaler HT, et al. Abnormal T-cell repertoire is consistent with immune process underlying the pathogenesis of paroxysmal nocturnal hemoglobinuria. *Blood.* 2000;96(7):2613-2620.
31. Issaragrisil S, Kaufman DW, Anderson T, et al. The epidemiology of aplastic anemia in Thailand. *Blood.* 2006;107(4):1299-1307.
32. Montane E, Ibanez L, Vidal X, et al. Epidemiology of aplastic anemia: a prospective multicenter study. *Haematologica.* 2008;93(4):518-523.

Analytical Biochemistry 415 (2011) 145–150

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

# Detection method for quantifying global DNA methylation by fluorescence correlation spectroscopy

Tomohiro Umezu [a], Kazuma Ohyashiki [b], Junko H. Ohyashiki [c,*]

[a] Department of Molecular Science, Tokyo Medical University, Tokyo 160-0023, Japan
[b] First Department of Internal Medicine, Tokyo Medical University, Tokyo 160-0023, Japan
[c] Institute of Medical Science, Tokyo Medical University, Tokyo 160-0023, Japan

ARTICLE INFO

*Article history:*
Received 28 January 2011
Received in revised form 16 April 2011
Accepted 21 April 2011
Available online 27 April 2011

*Keywords:*
Methyl-CpG-binding domain (MBD)
Fluorescence correlation spectroscopy (FCS)
Global DNA methylation

ABSTRACT

A method for quantifying global DNA methylation using fluorescence correlation spectroscopy (FCS) has been established. The single-molecule methylation assay (SMMA) is based on two methodologies. One methodology, FCS, estimates the translational diffusion coefficient of molecules in solution, whereas the other methodology uses the high affinity of methyl-CpG-binding domain protein 2 (MBD2) to bind specifically to methylated DNA. We studied the specific binding rates of fluorescence-labeled MBD2 and methylated DNA from biological samples using the automated FCS system. Using a standard curve with methylated control DNA, we developed the SMMA index to assess the global DNA methylation level of the biological samples. A marked decrease in the SMMA index was observed when human leukemia cell lines (U937 and K562) were cultured with DNA demethylating agents. Our findings clearly indicate the applicability of SMMA as a simple and rapid tool for quantifying global DNA methylation. SMMA may prove useful for genome-wide comparative methylation analyses of malignancies and as an indicator of the demethylation effects of epigenetic drugs.



Genomic DNA methylation plays an important role both during normal cell development and in tumorigenesis. The methylation occurs almost exclusively in CpG dinucleotides. Although the CpG dinucleotides constitute only 1% of the human genome, CpG-rich stretches, so-called CpG islands, are located in the promoter regions of more than 70% of all known human genes [1–3]. In normal cells, CpG islands are unmethylated, reflecting a transcriptionally active state of the respective genes. Epigenetic silencing of tumor suppressor genes by hypermethylation of CpG islands is a very early and stable characteristic of tumorigenesis [4–6]. In contrast, global DNA hypomethylation seems to occur in early neoplasia and is a feature of genomic DNA derived from solid tumor tissues; such hypomethylation leads to chromosomal instability and increased tumor frequency, as well as oncogene activation [7,8]. Accordingly, aberrant methylation patterns, either global DNA hypomethylation or hypermethylation of tumor suppressor genes, are considered a hallmark of human cancers.

Methyl-CpG-binding protein appears to be the central player in the process of DNA methylation-dependent gene silencing. MeCP2, MBD1, MBD2, MBD3, and MBD4 compose a family of nuclear proteins, each of which contains a methyl-CpG-binding domain (MBD).[1] Each of these proteins is capable of binding specifically to a single symmetrically methylated CpG pair in any sequence context and of recruiting chromatin remodeling and transcriptional repressor complexes, thereby establishing a repressive chromatin state [9]. The MBD has been used to separate and purify the methylated DNA in methylation analyses such as CpG island microarray and methyl-CpG immunoprecipitation [10].

Here we describe a new method for quantifying global DNA methylation using fluorescence correlation spectroscopy (FCS) and the high affinity of MBD for methylated DNA. The single-molecule methylation assay (SMMA) is based on fluorescence colocalization imaging of two single molecules, fluorescence-labeled MBD and methylated DNA (Fig. 1). The concept of FCS was developed more than 30 years ago [11,12], and with recent advances in laser and microscopic technologies, it now has single-molecule sensitivity. FCS measures fluorescence intensity fluctuations of fluorescently labeled molecules that, as a result of Brownian motion in a solution, cross through the immobile confocal region formed by a laser. The translational diffusion time in which a

* Corresponding author. Address: Institute of Medical Science (Oncology Division), Tokyo Medical University, 6-7-1 Nishi-shinjuku, Shinjuku, Tokyo 160-0023, Japan. Fax: +81 3 3345 0185.
*E-mail addresses:* t_umezu@tokyo-med.ac.jp (T. Umezu), ohyashik@rr.iij4u.or.jp (K. Ohyashiki), junko@hh.iij4u.or.jp (J.H. Ohyashiki).

[1] *Abbreviations used:* 5-AzaC, 5-azacytidine; DAC, 5-aza-2′-deoxycytidine; FCS, fluorescence correlation spectroscopy; MBD, methyl-CpG-binding domain; SMMA, single-molecule methylation assay; TAMRA-MBD2, TAMRA-labeled MBD2.

0003-2697/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2011.04.035

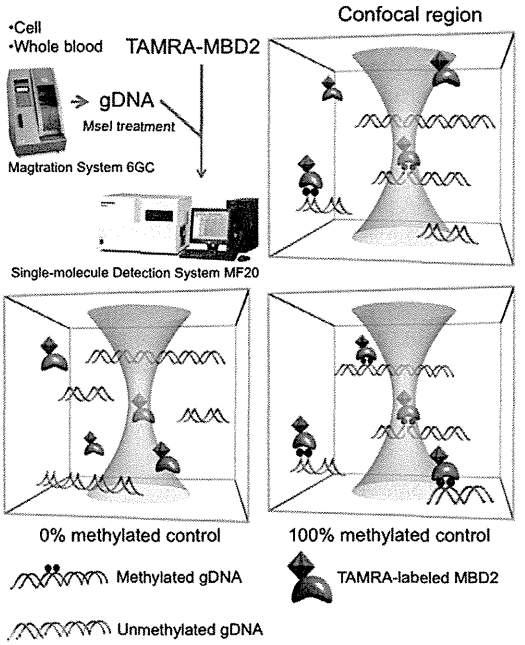


**Fig.1.** Schematic diagram of the SMMA method. The confocal region formed by the laser is depicted by the green area; genomic DNA fragment is represented by a spiral line; methylated CpG is represented by black spheres; recombinant MBD2 is depicted by a blue shape; nonexcited TAMRA is represented by a black octahedron; and excited TAMRA is a pink octahedron. Genomic DNA was purified with the robotic workstation (Magtration System 6GC, Precision System Science), and the crude genomic DNA was fragmented with MseI. Using the automated FCS system (MF20, Olympus), we measured the changes in diffusion time of TAMRA-MBD2. To quantify the binding of TAMRA-MBD2 with methylated DNA, we used FCS to analyze the binding of TAMRA-MBD2 with whole genome amplification (WGA) control DNA (0% methylated) or M.SssI-treated DNA (100% methylated). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

molecule stays in the focal volume depends on its hydrodynamic radius. Recently, DNA–DNA interactions [13], DNA–RNA interactions, and protein–protein interactions [14,15] have been analyzed by FCS. Therefore, FCS is ideally suited for analyzing specific interactions [16] and single-nucleotide polymorphism (SNP) [17]. Our findings clearly indicate the applicability of FCS as a reliable, simple, and rapid tool for quantifying the binding of TAMRA-labeled MBD2 (TAMRA-MBD2) with methylated DNA.

With SMMA, the difference in diffusion time of TAMRA-MBD2, resulting from methylated DNA binding, could be quantified, and the specificity of the binding was shown even in crude biological samples. The SMMA technique may prove useful for genome-wide comparative methylation analyses of malignancies and as an indicator of the demethylation effects of epigenetic drugs.

## Materials and methods

### *Cell culture*

The human leukemia cell lines U937 and K562 were cultured in RPMI 1640 supplemented with 10% heat-inactivated fetal bovine serum, penicillin (100 U/ml), and streptomycin (100 mg/ml) at 37 °C in a humidified atmosphere of 95% air–5% $CO_2$. For demethylation experiments, U937 and K562 cells, $5 \times 10^5$ each, were grown for 72 h with 3 μM 5-azacytidine (5-AzaC) (Sigma, St. Louis, MO, USA) or 5 μM 5-aza-2′-deoxycytidine (Decitabine, DAC; Wako, Osaka, Japan). The DNA demethylating agents were replenished each day.

### *Purification and preparation of genomic DNA*

Purification of genomic DNA was performed with the robotic workstation (Magtration System 6GC, Precision System Science, Chiba, Japan), for automated purification of nucleic acids, and with the EZ1 DNA Blood 350 μl kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. For each subject, at least 500 ng of genomic DNA was digested with MseI (New England Biolabs, Beverly, MA, USA). Completion of the digest was confirmed with agarose gel electrophoresis, and the digested DNA was quantified with NanoDrop ND-1000 (Thermo Fisher Scientific, Waltham, MA, USA).

### *CpG unmethylated or methylated control DNA*

To generate CpG-unmethylated DNA (0% methylated control), we amplified human control DNA (Qiagen) using the REPLI-g mini kit (Qiagen) according to the manufacturer's instructions. Briefly, amplification was carried out in two steps. The step 1 reaction mixture contained 5–10 ng of DNA in 1 μl of sterile water and 9 μl of sample buffer. This mixture was heated at 95 °C for 3 min and then chilled on ice. Step 1 resulted in denaturation of the genomic DNA template. The step 2 reaction (amplification) mixture contained 9 μl of reaction buffer, 1 μl of enzyme mix, and 10 μl of the product from step 1. The amplification reaction was incubated at 30 °C for 16–18 h. Step 2 allows binding of the exonuclease-resistant random hexamers and subsequent isothermal amplification. The enzyme was inactivated by heating at 65 °C for 10 min, followed by cooling to 4 °C. The reaction was purified with the QIAquick PCR purification kit (Qiagen).

We generated CpG-methylated DNA (100% methylated control) by methylating CpG motifs within the whole genome amplification (WGA) DNA using the CpG methylase M.SssI kit (New England Biolabs) according to the manufacturer's instructions. Briefly, 1.5 μg of WGA DNA was combined with 2 μl of 10× NEBuffer 2, 0.1 μl of *S*-adenosylmethionine, 5 units of M.SssI, and sterile water up to a final volume of 20 μl. The reaction was incubated at 37 °C for 4 h and then purified with the QIAquick PCR purification kit (Qiagen). The DNA concentration was determined with a NanoDrop ND-1000 spectrophotometer.

### *Bisulfite genomic sequencing*

Genomic DNA was chemically modified with EpiTect Bisulfite Kits (Qiagen), a treatment that changes the unmethylated but not the methylated cytosine into uracil. Bisulfite-treated DNA was subjected to polymerase chain reaction (PCR) amplification with the use of primers designed to recognize both methylated and unmethylated forms of the *p16/INK4a* promoter region: p16-F: 5′-GAGGAGGGGTTGGTTGGTTATTAGAG-3′ and p16-R: 5′-TACCTA ATTCCAATTCCCCTACAAAC-3′, and the *p15/INK4b* promoter region: p15-F: 5′-TGGGGATTAGGAGCTGAGGG-3′ and p15-R: 5′-TCTGGCA GAGTGGGGAGCCAGCC-3′. The resulting PCR products were fractionated in low-melting agarose (Sigma), purified with the QIAEX II kit (Qiagen), and cloned into the pGEM-T Easy vector (Promega, Madison, WI, USA). DNA of individual clones was then sequenced with an ABI Prism 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

### *Preparation of TAMRA-labeled MBD2*

Fluorescently labeled and unlabeled MBD2 was purchased from Protein Express, Inc. (Chiba, Japan). The procedures for protein expression and purification of MBD2 domain were described in previous reports [18,19]. Using the pROX-FL vector included in the coding sequence of MBD2, we synthesized TAMRA-labeled

MBD2 with an RTS 100 *Escherichia coli* HY kit (Roche Diagnostics, Basel, Switzerland) and the TAMRA-tRNA (TAMRA-X aminophenyl-alanyl-$tRNA_{cccg}$) included in the In vitro Pin-point fluorescence labeling kit 543 (Olympus, Tokyo, Japan). The reaction mixture for in vitro translation (150 μl) contained 15 μl of the plasmid DNA (100 ng/μl), 15 μl of TAMRA-tRNA, 36 μl of amino acid, 3 μl of methionine, 30 μl of reaction mix, and 36 μl of *E. coli* lysate. Amino acids, methionine, reaction mix, and 36 μl of *E. coli* lysate were included in the RTS 100 *E. coli* HY kit. The reaction mixture was incubated at 30 °C for 2 h. Expressed TAMRA-labeled substrates were purified with His Spin Trap columns (GE Healthcare, Piscataway, NJ, USA) according to the supplier's instructions. Substrates were concentrated with an UltraFree-0.5 centrifuge (Millipore, Bedford, MA, USA), and the substrate sizes and concentrations were confirmed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

### *Binding reaction of DNA and TAMRA-labeled MBD2*

TAMRA-MBD2 (5 nM) was incubated at room temperature for 20 min with 10 ng DNA (MseI-treated gDNA) in 40 μl of 50 mM Tris buffer (pH 7.5) containing 70 mM KCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, 0.2 mg/ml bovine serum albumin, and 4% glycerol. For competition analysis, 0–500 nM unlabeled MBD2 (WT-MBD2) was incubated with 10 ng of methylated control DNA (100% methylated control) for 20 min prior to the addition of TAMRA-MBD2.

### *FCS measurements*

The MF20 single-molecule fluorescence detection system (Olympus) was used for FCS measurements. Each 35-μl sample was added to a 384-well glass-bottomed microplate. These samples were sequentially and automatically loaded into the device, the optical system of which was also automatically adjusted for each measurement. For the detection of TAMRA-MBD2, a He–Ne laser (543 nm) and 580DF30 filter were used. All experiments were performed under identical conditions with a data acquisition time of 5 s per measurement, and measurements were repeated 15 times per sample. A cutoff at chi square distribution of >600 was set to improve the reproducibility of the mean value of 15 measurements. The standard deviations of diffusion times were calculated from the average values of the three tests for each sample.

### *Statistical analysis*

We used GraphPad Prism 5.0 software (GraphPad Software Inc., San Diego, CA, USA) for statistical analysis. A Mann–Whitney test was used to determine statistical significances between the control and test groups. *P* values less than 0.05 were considered to indicate statistically significant differences.

## Results

### *Detection of MBD2-genomic DNA complex by FCS analysis*

The schematic chart of the SMMA procedure is shown in Fig. 1. We assumed that recombinant MBD2 could be used to discriminate methylated and unmethylated DNA in crude biological samples [20–23]. First, genomic DNA was purified with the robotic workstation (Magtration System 6GC, Precision System Science), and the crude genomic DNA was fragmented with MseI. After mixing fragmented DNA and TAMRA-MBD2, we measured the changes in diffusion time of TAMRA-MBD2 using the automated FCS system (MF20, Olympus). Thus, the binding of TAMRA-MBD2 with methylated DNA can be quantified simply and rapidly.

To explore this type of application at the single-molecular level, we fragmented crude genomic DNA with MseI, which was chosen because it is known to preferentially cleave in regions of low CpG content while leaving many CpG islands intact. The electrophoretic images of MseI-treated genomic DNA are shown in Fig. 2. The intact genomic DNA purified from U937 cells was larger than 8 kbp (Fig. 2A), whereas MseI cleaved genomic DNA into fragments that were spread densely from 200 to 1000 bp (Fig. 2B). Using the automated FCS system, we measured the diffusion time that TAMRA-MBD2 was reacted with MseI-treated DNA or nontreated DNA (see Tables S1 and S2 in Supplementary material). The 15 measurements were inconsistent when TAMRA-MBD2 was reacted with nontreated DNA, whereas the reproducibility of repeated measurements was confirmed with MseI-treated DNA.

### *Definition of the SMMA index*

With the use of the standard curve for methylated control DNA, the diffusion time can be converted into percentage values to assess the global DNA methylation level of sample DNA. First, to generate the methylated control DNA, we performed WGA and M.SssI treatments. When WGA control DNA of the *p15/INK4b* and the *p16/INK4a* promoter regions was analyzed by bisulfite sequencing, all CpGs were demethylated. In contrast, all CpGs were methylated in M.SssI-treated DNA of the *p15/INK4b* and the *p16/INK4a* promoter regions (Fig. 3A).

Next, the binding of TAMRA-MBD2 with methylated DNA was examined by using 0% and 100% methylated control DNA (Fig. 3B). We used different concentrations of control DNA (0–40 ng) and measured the changes in diffusion time of TAMRA-MBD2. The diffusion time of unbound free TAMRA-MBD2 was determined as 697.2 ± 8.4 μs. When 100% methylated control DNA was added, a dose-dependent prolongation of the diffusion time, resulting from the formation of an MBD2-methylated DNA complex, was observed (Fig. 3B, solid bars). These binding reactions appeared to be saturated in the presence of <40 ng methyl-

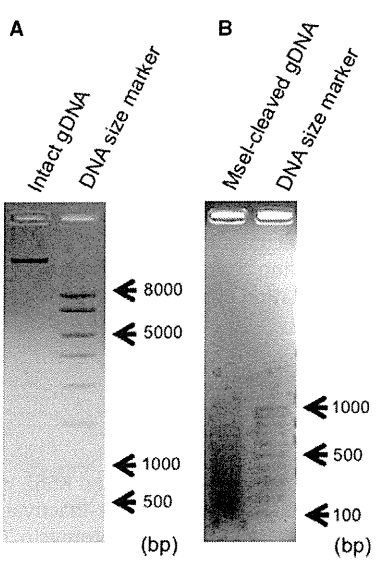


**Fig.2.** The electrophoretic images of MseI-treated genomic DNA. For each subject, at least 500 ng of genomic DNA was digested with MseI. Completion of the digest was confirmed with agarose gel electrophoresis. (A) The intact genomic DNA purified from U937 cells was larger than 8 kbp. (B) MseI cleaved genomic DNA into fragments that were spread densely from 200 to 1000 bp. The right lanes in (A) and (B) indicate the DNA size marker.

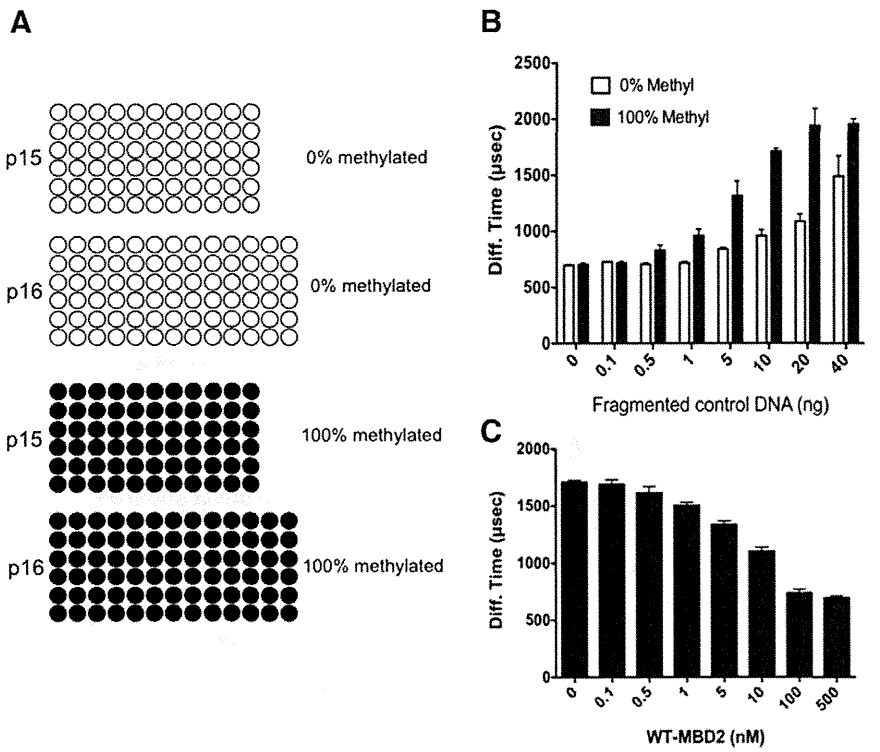


**Fig.3.** Detection of MBD2-methylated genomic DNA complex by FCS analysis. To generate control DNA to quantify the binding of TAMRA-MBD2 with methylated genomic DNA, we performed WGA and M.SssI treatment. (A) Bisulfite sequencing results of the WGA-treated (0% methylated control) and M.SssI-treated (100% methylated control) genomic DNA of the *p15/INK4b* and the *p16/INK4a* promoter regions. Each circle indicates a CpG site in the primary sequence, and each line of circles represents analysis of a single cloned allele. Open and closed circles represent unmethylated and methylated C residues, respectively. (B) The relationship between diffusion time and genomic DNA contents. TAMRA-MBD2 was incubated at room temperature for 20 min with MseI-treated control DNA (0, 0.1, 0.5, 1, 5, 10, 20, or 40 ng). The concentration of TAMRA-MBD2 was fixed at 5 nM. Closed bars indicate the diffusion time of TAMRA-MBD2 with 100% methylated control DNA. Open bars indicate the diffusion time of TAMRA-MBD2 with 0% methylated control DNA. Values represent the mean ± SD of three tests. (C) MBD2 competition assay for specificity of methylated DNA binding. Different concentrations (0–100 nM) of unlabeled MBD2 (WT-MBD2) were incubated with 100% methylated control DNA (10 ng) for 20 min prior to the addition of TAMRA-MBD2 (5 nM). Values represent the mean ± SD of 3 tests. Diff., diffusion.

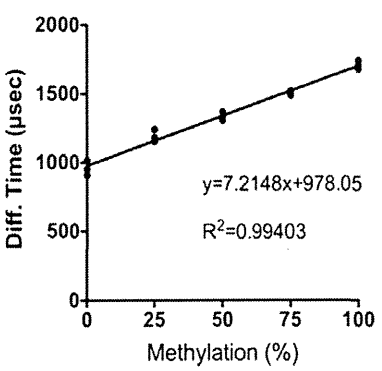


**Fig.4.** Establishment of standard curve for the SMMA index. The binding of TAMRA-MBD2 with 0% methyl control (WGA-treated DNA/M.SssI-treated DNA proportions of 4:0), 25% methyl control (WGA-treated DNA/M.SssI-treated DNA proportions of 3:1), 50% methyl control (WGA-treated DNA/M.SssI-treated DNA proportions of 2:2), 75% methyl control (WGA-treated DNA/M.SssI-treated DNA proportions of 1:3), and 100% methyl control (WGA-treated DNA/M.SssI-treated DNA proportions of 0:4) were analyzed by FCS. Diff., diffusion.

ated DNA. On the other hand, unspecific binding between MBD2 and unmethylated DNA increased in a manner dependent on the concentration of 0% methyl control DNA (Fig. 3B, open bars). Thus, to further detect the specificity of the binding reaction between MBD2 and methylated DNA, the density of DNA used to react with MBD2 was fixed to 10 ng.

We next estimated the specificity of the binding reaction between MBD2 and methylated DNA by competition assay, in which an excess amount of unlabeled MBD2 (WT-MBD2) was added to the reaction mixture of TAMRA-MBD2 and 100% methylated control DNA. The binding of 100% methylated DNA (10 ng) with TAMRA-MBD2 (5 nM) was completely blocked in the presence of more than a 20-fold amount of WT-MBD2 (100 nM) (Fig. 3C). These results indicate that the binding of TAMRA-MBD2 with methylated DNA was a sequence-specific interaction.

We next tried to establish the standard curve for quantifying global DNA methylation of crude biological samples. A series of mixtures of WGA-treated control DNA (0% methylated control) and M.SssI-treated control DNA (100% methylated control) in diluted proportions of 4:0, 3:1, 2:2, 1:3, and 0:4 was used. Fig. 4 shows the standard curve that can measure global DNA methylation status. As shown in this figure, a linear relationship was found ($y = 7.2148x + 978.05$, $R^2 = 0.99403$). The linear correlation shows that measurements of the percentages of global DNA methylation (0%, 25%, 50%, 75%, and 100%) are well distinguished and can be used to determine an SMMA index for test samples.

### *SMMA index of genome methylation in leukemia cell lines treated with DNA demethylating agents*

We used the SMMA index to attempt to quantify the demethylation effect of DNA demethylating agents. U937 and K562 cells were grown for 72 h with 3 μM 5-AzaC or 5 μM DAC, and the crude genomic DNA was collected and fragmented with MseI. The binding of TAMRA-MBD2 with MseI-treated genomic DNA was analyzed by FCS, and the diffusion time was converted to the SMMA

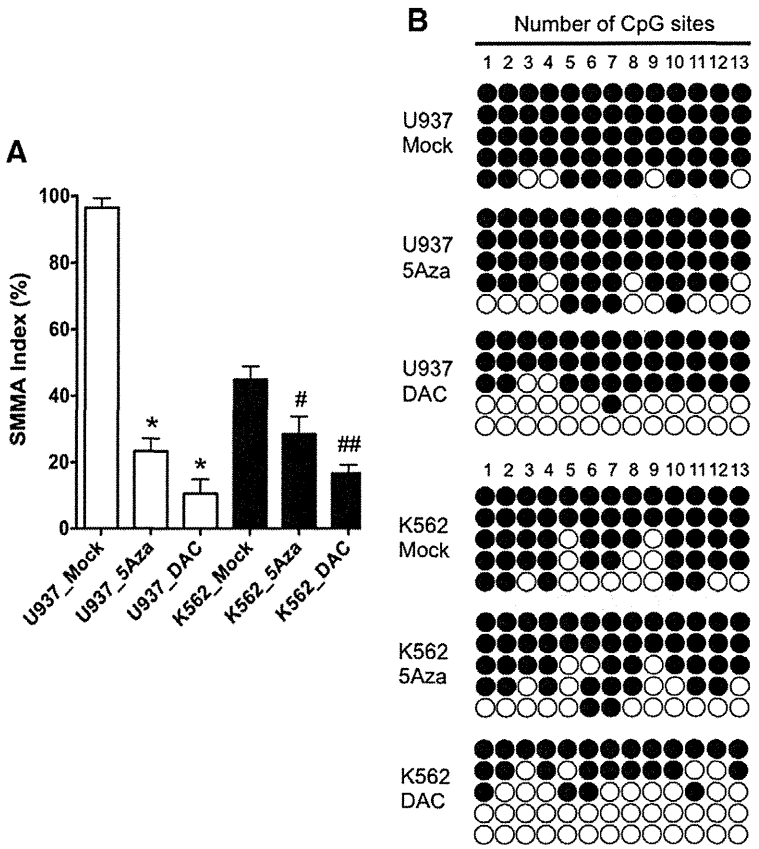


**Fig.5.** Quantification of the demethylation effect of DNA demethylating agents. For demethylation experiments, U937 and K562 cells were grown for 72 h with 3 μM 5-AzaC or 5 μM DAC. (A) Diffusion times of DNA methylation in cell lines treated with DNA demethylating agents. The binding of TAMRA-MBD2 with MseI-treated genomic DNA was analyzed by FCS. Demethylation effects by 5-AzaC and DAC: open bars indicate diffusion time of the 5-AzaC- or DAC-treated U937 cells; solid bars indicate diffusion time of the 5-AzaC- or DAC-treated K567 cells. Values represent the mean ± SD of three tests. $^{*}P < 0.0001$ versus U937_Mock. $^{\#}P = 0.0432$ versus K562_Mock. $^{\#\#}P < 0.01$ versus K562_Mock. (B) Bisulfite sequencing results of the p16/INK4a promoter in U937 and K562 cells cultured for 72 h with 3 μM 5-AzaC or 5 μM DAC. Each circle indicates a CpG site in the primary sequence, and each line of circles represents analysis of a single cloned allele. Open and closed circles represent unmethylated and methylated C residues, respectively.

index. The SMMA index decreased significantly in the 5-AzaC- or DAC-treated U937 cells compared with mock cells ($P < 0.0001$, Fig. 5A, open bar, and Table S3). We found a modest decrease in the SMMA index in 5-AzaC-treated K562 cells; however, no significant change was seen because of the originally low SMMA index (Fig. 5A, solid bar and Table S3).

Furthermore, to compare the SMMA method, we measured the demethylation effect of DNA demethylating agents using a currently available method. The genomic DNA derived from 5-AzaC- and DAC-treated U937 and K562 cells of the *p16/INK4a* promoter region was analyzed by bisulfite sequencing (promoter-associated analysis). As shown in Fig. 5B, the *p16/INK4a* promoter is highly methylated in U937 cells (94%) and K562 cells (80%). When cells were treated with DAC, the methylation level decreased to 58% in U937 cells and to 40% in K562 cells. However, no significant change was seen in the 5-AzaC-treated cells (82% in U937 cells and 71% in K562 cells), suggesting that the demethylation effect of 5-AzaC differed between the SMMA method (genome-wide analysis) and the bisulfite sequencing method (promoter-associated analysis).

## Discussion

We have established a novel approach to quantitatively determine the global DNA methylation levels using FCS, namely the SMMA method. This method provides an alternative method for evaluating global DNA methylation with a small amount of input DNA. This study clearly indicates the applicability of SMMA as a robust, simple, and fast tool for quantifying the global DNA methylation level in biological samples, such as genomic DNA extracted from human cells.

Global DNA methylation can be detected with bisulfite conversion, methylation-sensitive restriction enzymes, methyl-binding proteins, and anti-methylcytosine antibodies [24,25]. However, the results obtained using techniques based on these principles likely will be biased, because the procedures are complicated. In contrast, SMMA using the FCS technique is less time consuming, and the measurements can be carried out within 1–2 min per sample. Thus, an assay based on FCS could facilitate the analysis of DNA–protein interactions tens of times faster than previous methods; the total assay time is only 4 h. Moreover, no primary modification of genomic DNA is needed, and the assay requires only 500 ng of genomic DNA.

Members of the MBD protein family bind to methylated DNA by recognizing methylated cytosines, and they regulate protein biosynthesis by recruitment of transcriptional repression complexes to silence gene expression. SMMA is based on the selective enrichment of methylated DNA fragments by fluorescently labeled MBD2. The binding of MBD2 to methylated DNA is relatively sequence independent, in contrast to other methylated CpG-binding proteins (MeCP2, MBD3, and MBD4) [20,21,24]. In human colon

cancer, the *p14/ARF* and *p16/INK4a* tumor suppressor genes are commonly inactivated by MBD2-mediated aberrant methylation of their promoter regions [26]. Moreover, a deficiency of MBD2 strongly inhibits intestinal tumorigenesis in the Min mouse, suggesting that the protein might be involved in transcriptional repression of methylated tumor suppressor genes [22]. Therefore, while it is logical that we choose MBD2 as the fluorescently labeled molecule, unfortunately, we did not have a chance to compare MBD2 with other fluorescently labeled methylated CpG-binding proteins.

DNA methylation is a reversible epigenetic process [27], which makes it an attractive potential therapeutic target of cancer. The only approved way to target DNA methylation is to inhibit the DNMT enzymes. DNMT inhibitors that are clinically available and indicated for myelodysplastic syndrome (MDS) include 5-AzaC and DAC [27]. Although both 5-AzaC and DAC show clinical response to MDS, it is unknown to what extent the clinical activity of DNMT inhibitors depends on reversal of methylation, or if a clinical response can be detected by reversal of methylation early in the course of treatment. In the current study, we compared the DNA demethylation effects induced by 5-AazC and DAC in U937 cells, using both genome-wide analysis by SMMA and conventional methylation analysis of *p16/INK4a* by bisulfite sequencing. With conventional methylation analysis, we failed to detect the effect of demethylation of a single gene; however, we found a marked demethylation effect in 5-AzaC-treated U937 cells by SMMA. These findings strongly suggest that the effect of demethylating agents on genomic methylation cannot be judged from an assessment of limited regions.

Although we could not simply compare the results obtained with the SMMA method and a genome-wide methylation array that contains promoter-associated CpG islands, we found a similar drug-dependent decrease of methylation in both assays (Fig. S1). This indicates that the global DNA methylation level could be more useful than the methylation of single promoter-associated CpG sites to estimate the demethylation status in human cells, especially in patient samples with heterogeneous methylation profiles.

In conclusion, our results show that SMMA might be a robust and simple method for quantifying the level of global genome methylation, even if the source of the DNA is crude cellular extracts. SMMA may be useful for genome-wide comparative methylation analyses of clinical samples (e.g., cancer tissue versus noncancer tissue, patients versus healthy subjects, drug-treated patients versus treatment-naïve patients). Taken together, our results indicate that SMMA is a powerful method for studying global genome methylation in molecular biological and clinical research.

## Acknowledgments

This work was supported by the "Promotion of Science and Technology" project for private universities, with a matching fund subsidy from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), 2009–2014, and by the "University-Industry Joint Research Project" for private universities with a matching fund subsidy from MEXT, 2007–2009. Thanks are also due to Dr. N. Kato (Olympus Corp.) for her advice, and Ms. C. Kobayashi, R. Soya-Hamamura, A. Hirota, and Mr. Y. Kamimura for their technical assistance.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ab.2011.04.035.

## References

[1] R.L. Momparler, Cancer epigenetics, Oncogene 22 (2003) 6479–6483.
[2] C. Plass, Cancer epigenomics, Hum. Mol. Genet. 11 (2002) 2479–2488.
[3] A.G. Tsai, H. Lu, S.C. Raghavan, M. Muschen, C.L. Hsieh, M.R. Lieber, Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity, Cell 135 (2008) 1130–1142.
[4] E.E. Cameron, S.B. Baylin, J.G. Herman, p15(INK4B) CpG island methylation in primary acute leukemia is heterogeneous and suggests density as a critical factor for transcriptional silencing, Blood 94 (1999) 2445–2451.
[5] P.A. Jones, P.W. Laird, Cancer epigenetics comes of age, Nat. Genet. 21 (1999) 163–167.
[6] C.L. Hsieh, Dependence of transcriptional repression on CpG methylation density, Mol. Cell. Biol. 14 (1994) 5487–5494.
[7] P.A. Jones, S.B. Baylin, The fundamental role of epigenetic events in cancer, Nat. Rev. Genet. 3 (2002) 415–428.
[8] M. Ehrlich, Cancer-linked DNA hypomethylation and its relationship to hypermethylation, Curr. Top. Microbiol. Immunol. 310 (2006) 251–274.
[9] E. Ballestar, A.P. Wolffe, Methyl-CpG-binding proteins. Targeting specific gene repression, Eur. J. Biochem. 268 (2001) 1–6.
[10] E. Ballestar, M.F. Paz, L. Valle, S. Wei, M.F. Fraga, J. Espada, J.C. Cigudosa, T.H. Huang, M. Esteller, Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer, EMBO J. 22 (2003) 6335–6345.
[11] M. Ehrenberg, E. Cronvall, R. Rigler, Fluorescence of proteins interacting with nucleic acids. Correction for light absorption, FEBS Lett. 18 (1971) 199–203.
[12] D. Magde, E.L. Elson, W.W. Webb, Fluorescence correlation spectroscopy. II. An experimental realization, Biopolymers 13 (1974) 29–61.
[13] M. Kinjo, R. Rigler, Ultrasensitive hybridization analysis using fluorescence correlation spectroscopy, Nucleic Acids Res. 23 (1995) 1795–1799.
[14] K. Kuroki, S. Kobayashi, M. Shiroishi, M. Kajikawa, N. Okamoto, D. Kohda, K. Maenaka, Detection of weak ligand interactions of leukocyte Ig-like receptor B1 by fluorescence correlation spectroscopy, J. Immunol. Methods 320 (2007) 172–176.
[15] G. Maertens, J. Vercammen, Z. Debyser, Y. Engelborghs, Measuring protein–protein interactions inside living cells using single color fluorescence correlation spectroscopy. Application to human immunodeficiency virus type 1 integrase and LEDGF/p75, FASEB J. 19 (2005) 1039–1041.
[16] M. Eigen, R. Rigler, Sorting single molecules: application to diagnostics and evolutionary biotechnology, Proc. Natl. Acad. Sci. USA 91 (1994) 5740–5747.
[17] M. Bannai, K. Higuchi, T. Akesaka, M. Furukawa, M. Yamaoka, K. Sato, K. Tokunaga, Single-nucleotide-polymorphism genotyping for whole-genome-amplified samples using automated fluorescence correlation spectroscopy, Anal. Biochem. 327 (2004) 215–221.
[18] H. Nakata, T. Ohtsuki, M. Sisido, A protease inhibitor discovery method using fluorescence correlation spectroscopy with position-specific labeled protein substrates, Anal. Biochem. 390 (2009) 121–125.
[19] R. Abe, K. Shiraga, S. Ebisu, H. Takagi, T. Hohsaka, Incorporation of fluorescent non-natural amino acids into N-terminal tag of proteins in cell-free translation and its dependence on position and neighboring codons, J. Biosci. Bioeng. 110 (2010) 32–38.
[20] J. Berger, A. Bird, Role of MBD2 in gene regulation and tumorigenesis, Biochem. Soc. Trans. 33 (2005) 1537–1540.
[21] M. Fodermayr, J. Proll, O. Zach, C. Wechselberger, D. Lutz, In vitro detection of methylated DNA via recombinant protein MBD2b, Mol. Biol. Rep. 36 (2009) 1859–1862.
[22] O.J. Sansom, J. Berger, S.M. Bishop, B. Hendrich, A. Bird, A.R. Clarke, Deficiency of Mbd2 suppresses intestinal tumorigenesis, Nat. Genet. 34 (2003) 145–147.
[23] P.A. Wade, Methyl CpG-binding proteins and transcriptional repression, Bioessays 23 (2001) 1131–1137.
[24] T.A. Rauch, G.P. Pfeifer, The MIRA method for DNA methylation analysis, Methods Mol. Biol. 507 (2009) 65–75.
[25] M. Karimi, S. Johansson, T.J. Ekstrom, Using LUMA: a luminometric-based assay for global DNA-methylation, Epigenetics 1 (2006) 45–48.
[26] V. Martin, H.F. Jorgensen, A.S. Chaubert, J. Berger, H. Barr, P. Shaw, A. Bird, P. Chaubert, MBD2-mediated transcriptional repression of the p14ARF tumor suppressor gene in human colon cancer cells, Pathobiology 75 (2008) 281–287.
[27] J.G. Herman, S.B. Baylin, Gene silencing in cancer in association with promoter hypermethylation, N. Engl. J. Med. 349 (2003) 2042–2054.

Leukemia (2011) 25, 110–120

www.nature.com/leu

ORIGINAL ARTICLE

# Impact of adjunct cytogenetic abnormalities for prognostic stratification in patients with myelodysplastic syndrome and deletion 5q

M Mallo[1,2,3,12], J Cervera[1,3,12], J Schanz[4], E Such[1,3], G García-Manero[5], E Luño[1,3], C Steidl[4], B Espinet[1], T Vallespí[1,3,6], U Germing[4], S Blum[7], K Ohyashiki[8,9], J Grau[1,3], M Pfeilstöcker[4], JM Hernández[1,3], T Noesslinger[4], A Giagounidis[4], C Aul[4], MJ Calasanz[1,3], ML Martín[1,3], P Valent[10], R Collado[1,3], C Haferlach[9,11], C Fonatsch[4], M Lübbert[4], R Stauder[4], B Hildebrandt[4], O Krieger[4], C Pedro[3], L Arenillas[3], MÁ Sanz[3], A Valencia[1,3], L Florensa[3], GF Sanz[3,13], D Haase[4,9,13] and F Solé[1,2,3,9,13]

*[1]Spanish Haematological Cytogenetics Working Group, Spain; [2]Faculty of Life Sciences, Department of Cell Biology, Physiology, and Immunology, Autonomous University of Barcelona, Bellaterra, Spain; [3]Spanish MDS Registry Group, Spain; [4]German–Austrian MDS Study Group, Germany and Austria; [5]Department of Leukemia, The University of Texas MD Anderson Cancer Center, Houston, USA; [6]International Working Group on Morphology of MDS (MDS Foundation); [7]Hematology Service, University Hospital Vaudois, Lausanne, Switzerland; [8]Division of Hematology, The First Department of Internal Medicine, Tokyo Medical University, Tokyo, Japan; [9]International Working Group on MDS Cytogenetics (MDS Foundation); [10]Department of Medicine I, Medical University of Vienna, Vienna, Austria and [11]Munich Leukemia Laboratory, Munich, Germany*

**This cooperative study assessed prognostic factors for overall survival (OS) and risk of transformation to acute myeloid leukemia (AML) in 541 patients with *de novo* myelodysplastic syndrome (MDS) and deletion 5q. Additional chromosomal abnormalities were strongly related to different patients' characteristics. In multivariate analysis, the most important predictors of both OS and AML transformation risk were number of chromosomal abnormalities ($P<0.001$ for both outcomes), platelet count ($P<0.001$ and $P=0.001$, respectively) and proportion of bone marrow blasts ($P<0.001$ and $P=0.016$, respectively). The number of chromosomal abnormalities defined three risk categories for AML transformation (del(5q), del(5q)+1 and del(5q)+ ⩾2 abnormalities) and two for OS (one group: del(5q) and del(5q)+1; and del(5q)+ ⩾2 abnormalities, as the other one); with a median survival time of 58.0 and 6.8 months, respectively. Platelet count ($P=0.001$) and age ($P=0.034$) predicted OS in patients with '5q−syndrome'. This study demonstrates the importance of additional chromosomal abnormalities in MDS patients with deletion 5q, challenges the current '5q−syndrome' definition and constitutes a useful reference series to properly analyze the results of clinical trials in these patients.**
*Leukemia* (2011) **25,** 110–120; doi:10.1038/leu.2010.231; published online 30 September 2010
**Keywords:** '5q−syndrome'; cytogenetics; deletion 5q; myelodysplastic syndromes

## Introduction

Myelodysplastic syndromes (MDS) are a group of clonal hematopoietic stem cell diseases characterized by dysplasia and ineffective hematopoiesis in one or more myeloid cell lines. MDS is associated with a variable overall survival (OS) and a relatively high risk of progression to acute myeloid leukemia (AML). Evolution to AML and the clinical consequences of cytopenias are main causes of morbidity and mortality in MDS.[1–3]

Correspondence: Dr F Solé, Laboratori de Citogenètica Molecular, Servei de Patologia, Hospital del Mar, Passeig Marítim, 25-29, 08003 Barcelona, Spain.
E-mail: fsole@parcdesalutmar.cat
[12]These authors equally contributed to this work.
[13]These authors equally contributed to the coordination of the project.
Received 29 April 2010; revised 11 August 2010; accepted 26 August 2010; published online 30 September 2010

Although many specific chromosomal abnormalities have been associated with MDS, partial or complete deletion of the long arm of chromosome 5 (deletion 5q), with or without additional karyotypic abnormalities, is present in 10–15% of patients with *de novo* MDS, and thus is the most frequently documented recurrent cytogenetic abnormality in MDS.[4–8] Outcomes among MDS patients with deletion 5q vary greatly, both in terms of OS and risk of transformation to AML.[5,8–11] The presence of additional chromosomal abnormalities or an excess of blasts shortens OS and increases the risk of AML transformation.[5,8,10,11] The '5q−syndrome' is the only MDS group considered to represent a separate cytogenetically defined disease category in the World Health Organization (WHO) classification. Patients with this syndrome, mostly women, are characterized by the presence of isolated deletion 5q, a blast count below 5%, favorable prognosis and a low rate of AML transformation.[2,3] So far, no other characteristic besides the proportion of bone marrow (BM) blasts and the existence of additional chromosomal abnormalities has been recognized and universally accepted as a predictor of outcome for patients with MDS and deletion 5q.[10,11] Further, no variable has been shown to impact the clinical course of patients with WHO-defined '5q−syndrome'. Lenalidomide therapy has activity in single-arm clinical trials in patients with International Prognostic Scoring System (IPSS) low or intermediate-1 risk, red blood cell transfusion dependency and deletion 5q,[12–14] leading to approval by the US Food and Drug Administration for this indication. In contrast, the European Medicines Agency refused approval of lenalidomide for these patients, because there were no historical data against which the safety of lenalidomide could be compared, especially regarding the expected risk of AML transformation.[15] Thus, the analysis of further prognostic parameters for OS and AML transformation in large series of MDS patients with deletion 5q is of importance.

The major aim of this global cooperation study was to assess the characteristics and natural history of a large series of 541 patients with *de novo* MDS and deletion 5q to identify prognostic factors of outcome.

## Materials and methods

### *Patients and diagnostic criteria*

A total of 541 patients with primary MDS and deletion 5q, included in the Spanish Haematological Cytogenetic Working