Table 2. Detection of hyperdiploidy ALL by DNA-index and SNP-chip analysis

	SNP-chip	
	HD	non-HD
DNA Index		
HD	44 cases	4 cases
non-HD	30 cases	200 cases

DNA index was measured by FACS as described in "DNA index, immunophenotyping, molecular analysis of chromosomal abnormalities" and DNA index of 278 ALL samples were available. Normal diploid cells have a DNA index of 1.0. When DNA index is the same as or greater than 1.16, the leukemia is defined as hyperdiploid ALL by DNA index. Hyperdiploid ALL detected by SNP-chip analysis had more than 50 chromosomes, which were counted manually.

HD indicates hyperdiploid ALL; and non-HD, nonhyperdiploid ALL.

estimate survival rates. Differences were compared with the 2-sided log-rank test. Event-free survival (EFS) was calculated from diagnosis to the time of the first event (relapse, secondary malignancy, or death from any cause) or to the date of last follow-up.

Results

Features of samples

Clinical features of 399 pediatric ALL patients are shown in Table 1. Infant ALL (< 1years of age) were excluded from this study, and 77% (307 cases) of the patients were from 1 to 9 years old. Forty-nine cases of T-cell lineage ALL and 339 cases of B-cell lineage ALL were examined. Ninety-six samples (24%) had *ETV6/RUNX1* fusion, and 6 cases had the *BCR/ABL* fusion gene.

Validation of SNP-chip data

Gene dosage, heterozygous SNPs, allele-specific gene dosage, and allelic composition (loss of heterozygosity [LOH]) was visualized as shown in Figure 1 using our novel analysis software, CNAG for SNP-chip.[11,12] Duplication/amplification, deletion, and UPDs of chromosomes were easily detected (Figure 1A). To validate abnormalities found by SNP-chip, genomic quantitative PCR and direct sequencing of SNP sites at duplicated, amplified, deleted, and UPD regions were performed including chromosome 9. Representative results of validation are shown in Figure 1B-D.

Also, hyperdiploid (HD) ALL defined by DNA index and SNP-chip analysis was compared for selected cases (Table 2). Number of total chromosomes was counted manually in SNP-chip analysis, and ALL with more than 50 chromosomes was defined as HD-ALL by SNP-chip. When DNA index of leukemic cells was same as or greater than 1.16, the sample was defined as HD-ALL by DNA index.[16,17] DNA index of 278 ALL samples were available, and 200 cases were defined as non-HD ALL by both methods. SNP-chip detected more cases of HD-ALL (74 cases) than DNA index. As shown in Figure 1Aiv, SNP-chip can precisely detect gene dosage, and this high sensitivity of SNP-chip analysis permitted more accurate detection of HD-ALL than by the DNA index method. Results of these analyses validated that the abnormalities detected by SNP-chip were reliable.

Three common abnormalities in pediatric ALL

Figure 2A summarizes molecular allelokaryotyping profiles of the 399 ALL cases after clustering with regard to the status of copy number alterations as well as copy number neutral LOH, so-called UPD, showing a number of clusters having common genetic lesions.

Among these clusters, 3 genetic abnormalities were frequently detected: hyperdiploidy (HD, > 50 chromosomes), deletion of the 9p region, and deletion of 12p (Figure 2A,B). The common deleted region (CDR) on 9p involved the *p16INK4A* gene (called p16Del, Figure 2B), and the CDR on 12p involved the *ETV6* gene (called ETV6Del, Figure 2B). Concurrent abnormalities of p16Del and HD were rare (*P* < .001); concurrent abnormalities of ETV6Del and HD also were rare (*P* < .001; Figure 2B). No case had all 3 common abnormalities.

The clinical features of cases with each of these 3 genetic abnormalities were analyzed (Table 3). Individuals with p16Del-ALL frequently were older (*P* = .017), had higher WBC (*P* < .001), and T-cell lineage ALL (*P* < .001). Those with ETV6Del-ALL were more often younger (*P* = .009), non–T-cell lineage (*P* = .014), and *ETV6/RUNX1* fusion gene positive (*P* < .001). Patients having HD-ALL were more frequently younger (*P* < .001), showed lower WBC (*P* < .001), non–T-cell lineage (*P* < .001), and *ETV6/RUNX1* negative (*P* < .001).

Numerical chromosomal abnormalities in pediatric ALL

Numerical chromosome changes were frequently detected in pediatric ALL samples, as summarized in Figure 3A. Numerical change of chromosome 21 (trisomy, tetrasomy, and pentasomy) was the most frequent (134 [34%] cases). We had 8 cases with Down syndrome who had trisomy 21 in their leukemic cells and their matched controls. These 8 cases are excluded in Figure 3A. Most of the numerical abnormalities were detected in HD-ALL cases (Figure S1A) except for those with trisomy 21 (Figure S1B). As for trisomy 21, half (21 cases) occurred in patients with subtypes other than HD (Figure S1B). In HD-ALL, gain of chromosomes was restricted to particular chromosomes, involving chromosomes 4, 6, 8, 10, 14, 17, 18, 21, and X (Figures 2A, 3A).

Nonrandom genetic abnormalities in pediatric ALL detected by SNP-chip

All copy number changes (deletions and duplications/amplifications) detected by SNP-chip analysis are shown in Figure 2A and Figure S2. Small deletions that could not be detected by conventional cytogenetics were sensitively identified, including deletions of 3p14.2 (500 kb), 3q26.32 (700 kb), and Xp21.1 (1 Mb) (Table 4 and Figure S2). Nonrandom chromosomal abnormalities (frequency > 1.5% of all cases) are listed in Table 4. Besides the 3 common genetic abnormalities, duplication of 1q (11%) and deletion of 6q (11.5%) were often detected. In 13 cases with 1q duplication, the duplication began at the *PBX1* gene (Figures 2A, S2). Since gain of the entire or part of either chromosome 21 or X was frequently found in non–HD-ALL, these abnormalities were grouped separately (Table 4).

Recently, other groups of investigators performed SNP-chip analysis on pediatric ALL and found deletions of several transcriptional factors associated with B-cell development including *PAX5* (9p13), *EBF* (5q33), *Ikaros* (7p12.2), *Aiolos* (17q12), *LEF1*(4q25), *RAG1*(11p12), and *RAG2* (11p12).[19,20] We also have found deletion of these genes in our study. However, the frequency of deletion of these genes, except for PAX5, was low (fewer than 2%) and/or the deleted regions contained multiple genes (Table 4; Figure S2 and data not shown).

UPD

One of the major advantages of SNP-chip analysis is capability of sensitive detection of UPD, even in samples suffering from small
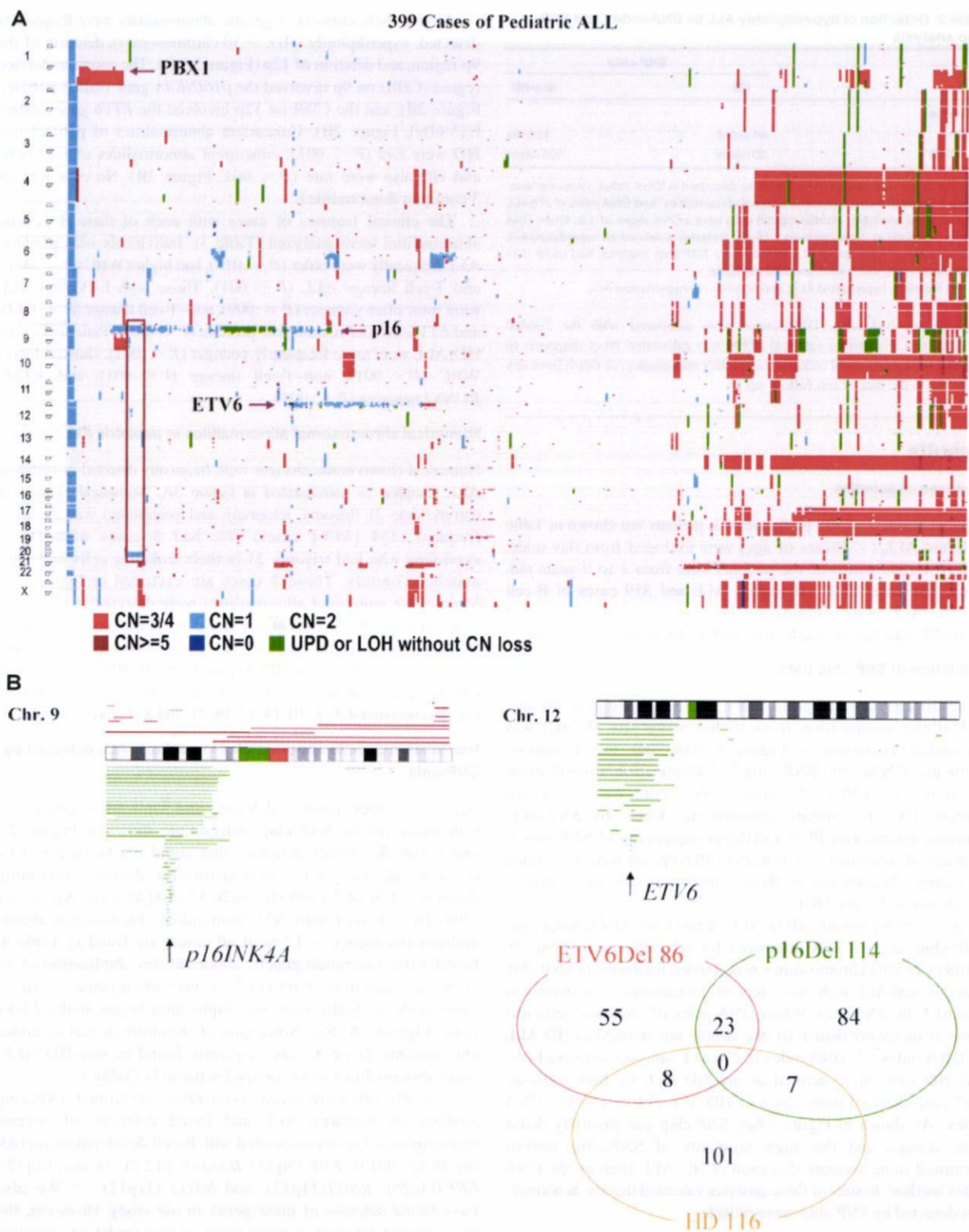
Figure 2. Allelokaryotyping of pediatric ALL. (A) Genetic clustering of 393 ALL samples. Genetic status of each chromosomal region is visualized. Vertical axis: chromosomes, p: short arms, q: long arms. Horizontal axis: individual cases. CN: copy number of alleles. UPD: uniparental disomy. Locations of *PBX1*, *INK4A/ARF*(p16), and *ETV6* genes are indicated. A rectangle indicates cases having 9p and 20q deletions. (B) Three common genomic abnormalities detected in ALL by SNP-chip analysis. Left panel: Deletion of 9p is frequently detected; the arrow indicates the commonly deleted region (CDR) where the p16INK4A gene is located. Right panel: Deletion of 12p often occurs. The arrow indicates that the CDR is where the ETV6 gene is localized. Green lines under the chromosome indicate the deleted regions in individual cases. Brown lines above the chromosome indicate the duplicated regions. Only representative cases are shown. Green and red bands on idiograms indicate centromeres and noncoding regions, respectively. Bottom panel: Venue diagram of 3 common abnormalities detected in this study. Numbers of respective cases in each category are indicated. HD: ALL with hyperdiploid (chromosomes > 50). ETV6Del: ALL with deletion of ETV6 gene. p16Del: ALL with deletion of p16INK4A gene.

Table 3. Clinical features of ALL cases associated with one of three common genetic abnormalities

	Genetic abnormality, no. (%)	Others, no. (%)	P
p16Del-ALL			
Age			
1 to 9 y	76 (68)	231 (80)	—
Older than 9 y	35 (32)	57 (20)	.017
WBC			
Below $10^2 \times 10^9$/L	85 (77)	277 (96)	—
Over $10^2 \times 10^9$/L	26 (23)	11 (4)	.001
Non-T lineage	71 (62)	263 (96)	—
T-lineage	83 (38)	11 (4)	.001
ETV6Del-ALL			
Age			
1 to 9 y	75 (87)	232 (74)	—
Older than 9 y	11 (13)	81 (26)	.009
Non-T lineage	89 (95)	261 (85)	—
T-lineage	4 (5)	45 (15)	.015
ETV6/RUNX1			
Positive	53 (66)	43 (15)	—
Negative	27 (34)	243 (85)	.001
HD-ALL			
Age			
1 to 9 y	101 (89)	206 (72)	—
Older than 9	13 (11)	79 (28)	.001
WBC			
Below $10^2 \times 10^9$/L	112 (98)	250 (88)	—
Over $10^2 \times 10^9$/L	2 (2)	35 (12)	.001
Non-T lineage	110 (100)	229 (82)	—
T-lineage	0 (0)	49 (18)	.001
ETV6/RUNX1	HD-ALL	Others	
Positive	8 (8)	88 (34)	—
Negative	97 (92)	173 (66)	.001

p16Del-ALL indicates ALL with deletion of *p16INK4A* gene; ETV6Del-ALL, ALL with deletion of *ETV6* gene; HD-ALL, ALL with hyperdiploidy (chromosomes >50); —, not applicable; WBC, white blood cell count in peripheral blood ($\times 10^9$/L) at diagnosis; and ETV6/RUNX1, ETV6/RUNX1 fusion was examined by RT-PCR and/or FISH analysis.

tumor content; UPD in samples with as low as 20% of tumor contents are clearly identified (Figure 1Aiv). Whole and partial chromosome UPDs were observed in 95 cases; 6 cases showed both whole and partial chromosome UPD (Figure 3B). Most of the whole chromosome UPD was detected in HD-ALL (Figure 3C). UPD of whole chromosome 9 was the most common whole chromosome UPD (Figure 3B). In contrast, UPD involving part of chromosomes was most often found in non–HD-ALL cases (Figure 3C). Recurrent partial chromosome UPD was detected in many chromosomal regions (Figure 3B). We frequently found whole chromosome 9 UPD (18 cases) and 9p UPD (30 cases) (Figure 3B). *INK4A* deletion was often found in 9p UPD (23 of 30, 77%), while it was rare in whole chromosome 9 UPD (1 of 18, 6%) (Figure 3D).

Relationship between genetic abnormalities

Recurrent abnormalities described above were compared with each other (3 common abnormalities and 26 nonrandom alterations) to detect relationships between these abnormalities (Table S1). Strong correlations between abnormalities of 12p and 21q were detected, duplications of 12p and 21q often occurred simultaneously, and duplications of 21q were accompanied by deletion of the *ETV6* gene that was localized on 12p. ETV6Del ALL frequently had additional changes, including duplications (21q and 1q) and deletions (3p21, 1q, *FHIT*, 15q, and 4q). Abnormalities involving chromosome X, including *DMD* (Xp21.2) deletion, were fre-

quently accompanied by deletions of 8p, 4q, and 6q. Deletion of 20q often occurred with either *p16INK4A* deletion or duplication of 21q (Figure 2A and Table S1).

Impact of nonrandom genetic abnormalities on prognosis

We analyzed prognosis of cases showing nonrandom abnormalities listed in Table 4 and found that the recurrent abnormalities had no impact on the event-free survival (EFS; data not shown) except for amplification/duplication of chromosome 9q. Our initial, early analysis found that EFS of pediatric ALL patients was not impacted by *ETV6* deletion either with or without *ETV6/RUNX1*. Nine cases with 9q amplification/duplication showed a poor prognosis (6 patients relapsed within 3 years; Figure S3A), although the number of cases having this abnormality is too small to reach a significant conclusion. Of these cases, 3 also had duplication of part of chromosome 22 (Figure S3B), suggesting that duplication of 9q is part of an extra copy of the Philadelphia chromosome. These 3 cases showed *BCR/ABL* positively by FISH/RT-PCR analysis (data not shown). Two other cases showed high copy number amplification that encompassed *ABL* and *NUP214* genes (Figure S3C), suggesting these cases had a *NUP214/ABL* fusion.[21] The ALL cells of these 2 cases were steroid-resistant and T-cell lineage phenotype.

Children with HD-ALL without gain of either chromosome 17 or 18 had a worse prognosis (Figure 4). Furthermore, children with HD-ALL and no extra copies of chromosomes 17 and 18 had a significantly worse prognosis ($P < .001$), with 53% EFS at 5 years compared with a 90% 5-year EFS in the other HD-ALL cohort (Figure 4).

Discussion

SNP-chip analysis is a reliable method to detect gene dosage, which was validated by direct sequencing of SNP sites and quantitative PCR in this study. To detect HD-ALL, DNA index is not a sensitive method since contaminated normal cells (DNA index 1.0) decrease the levels of DNA index of hyperdiploid leukemic cells. Although karyotyping is a good method to detect HD-ALL, sufficient number of high-quality chromosomal metaphases is not always obtained from the leukemic cells. SNP-chip analysis may be a more useful and reliable method to detect this subtype of ALL.

Molecular allelokaryotyping of a large series of pediatric ALL samples showed 3 major abnormalities: deletion of *p16INK4A*, deletion of *ETV6*, and hyperdiploidy. Besides these 3 common abnormalities, a number of novel, nonrandom changes were found in ALL. Some of them showed a very narrow commonly deleted region, which was limited to one target gene, including *FHIT* (3p14.2), *TBL1XR1* (3q26.3), and *DMD* (Xp21.2). *DMD* is the causative gene for Duchenne-type muscular dystrophy.[22] While germ-line inactivating mutations of this gene cause the disease, no association has been made between this disease and cancers, including ALL. Since *DMD* is an extremely large gene (2.4 Mb), deletion of it may occur as a result of instability of genomic DNA in ALL cells.

Other investigators and we have found in pediatric ALL a number of deleted genes, including transcriptional factors involved in B-cell differentiation.[19,20] However, since no point mutations of those genes, except PAX5, were detected,[19] it is unclear that these transcriptional factors associated with B-cell development are target genes of these deletions. Mullighan et al[19] and we have found that PAX5 gene is frequently involved in deletions and translocations (N.K., S.O., M.Z., et
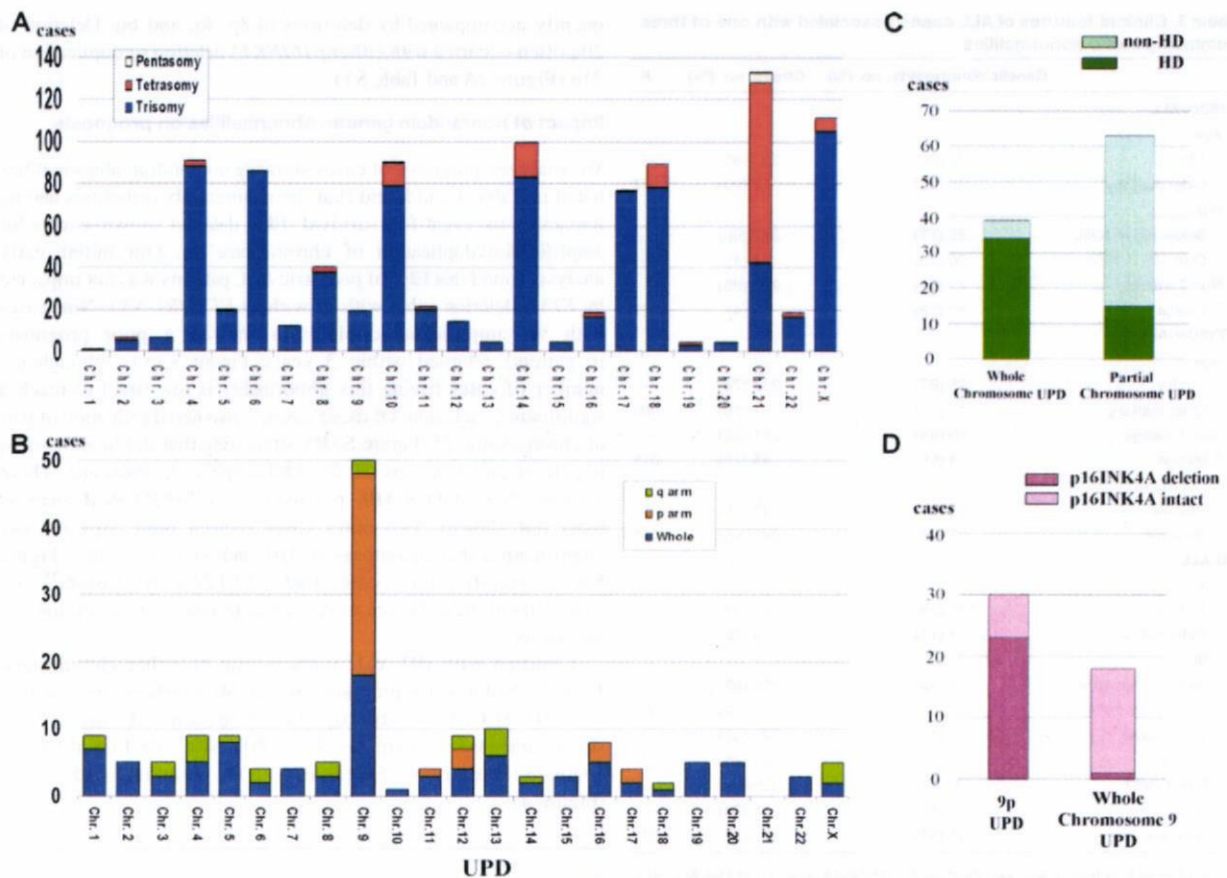
Figure 3. Numerical chromosomal changes and uniparental disomy in pediatric ALL. (A) Frequency of pentasomy/tetrasomy/trisomy affecting each chromosome. For X chromosome, trisomy (105 cases) contains trisomy X in male pateints (67 cases) and disomy X in female patients (38 cases). All tetrasomy X were female patients. (B) Frequency of uniparental disomy (UPD). Whole: cases with whole chromosome UPD; p arm: cases with UPD of short arm; and q arm: cases with UPD of long arm. Chr: chromosome. UPD involving X chromosome was detected only in female cases. (C) Distribution of whole and partial chromosome UPD in HD and non–HD-ALL. Whole chromosome UPD is frequently detected in HD-ALL. Thirty-four cases with whole chromosome UPD were HD-ALL. Partial UPD is frequently detected in non–HD-ALL. Fifteen of 63 cases with partial UPD were HD-ALL. (D) Frequency of deletion of p16INK4A gene in whole chromosome 9 UPD and 9p UPD. Twenty-three cases showed deletion of p16INK4A, out of a total of 30 cases with 9p UPD. One case had p16INK4A deletion from a total 18 ALL samples with whole chromosome 9 UPD.

al, manuscript submitted), suggesting that impairment of PAX5 is associated with leukomogenesis.

This study showed that numerical chromosomal changes and UPD were common genomic abnormalities in pediatric ALL. Interestingly, whereas trisomy 21 was the most common numerical chromosomal abnormality, UPD of chromosome 21 was not detected in our study. Even in 8 cases with Down syndrome who had trisomy 21 in their constitutional DNA, UPD of chromosome 21 was not detected. Although UPD of chromosome 21 in leukemic cells of patients with Down syndrome has been reported,[8] it may be a rare event in pediatric ALL.

Chromosomal mis-segregation occurs when duplicated chromosomes separate improperly during cell division,[23] leading to numerical chromosomal changes in leukemic cells, including HD-ALL. Most of the whole chromosome UPD was detected in HD-ALL, suggesting that whole chromosome UPD is due also to chromosomal mis-segregation. In contrast, UPD involving part of chromosomes was most often found in non–HD-ALL cases. This may suggest that these partial UPDs are not caused by mis-segregation, but by mitotic recombination, which can usually cause chromosomal exchange.[24]

UPD involving chromosome 9 or 9p is a very common abnormality in pediatric ALL. INK4A gene may be a target gene of 9p UPD since most of the cases with 9p UPD had deletion of

INK4A (23 of 30 cases with 9p UPD). For UPD involving whole chromosome 9, INK4A is not a target gene, since only one case with whole chromosome 9 UPD had deletion of this gene (1 of 18, 6%). Which gene is the target of whole chromosome 9 UPD is unclear. Furthermore, 7 cases with 9p UPD had intact INK4A, and the target gene of 9p UPD in these cases is also unknown. This is the first report showing that whole chromosome 9 UPD and 9p UPD are common abnormalities in pediatric ALL. Although Mullighan et al also analyzed a large number of pediatric ALL patients by SNP-chip,[19] they did not report this abnormality.

UPD on 9p is associated with activating point mutations of JAK2 in myeloproliferative disorders (MPDs).[26-28] However, point mutation of JAK2 in ALL is very rare.[29,30] We examined for JAK2 mutations at "hot-spots" (exons 12 and 14)[18,26-28] in these 7 cases of ALL with 9p UPD in which deletion of INK4A was not detected. We found no mutations of JAK2 in these cases. Another unidentified gene(s) in the region may be mutated in these cases.

In this study, we found that absence of gain of chromosomes 17 and 18 had an adverse impact on the prognosis of children with HD-ALL. Another large-scale study found that gain of chromosome 17 was associated with a better prognosis in HD-ALL.[31] Although this previous study reported that gain of chromosomes 4

Table 4. Recurrent genetic abnormalities detected by SNP-chip

Chromosomal sites	Type of abnormality	No. of cases (%)	Candidate genes
1q	Duplication	44 (11)	—
1q	Deletion	11 (3)	—
3p21	Deletion	6 (2)	—
3p14.2	Deletion	6 (2)	FHIT
3q26.3	Deletion	10 (3)	TBL1XR1
4q31	Deletion	7 (2)	—
6q	Deletion	46 (11)	—
7p	Deletion	10 (3)	—
7q34	Deletion	7 (2)	—
8p	Deletion	13 (3)	—
8q	Duplication	9 (2)	—
9q	Dup/amp	9 (2)	ABL
10p	Duplication	7 (2)	—
10q24	Deletion	12 (3)	—
11q	Deletion	24 (6)	—
12p	Duplication	13 (3)	—
13q14.2	Deletion	14 (4)	RB1
15q	Deletion	7 (2)	—
17p	Deletion	8 (2)	TP53
17q	Duplication	10 (3)	—
17q11.2	Deletion	7 (2)	NF1
20p12.2	Deletion	6 (2)	—
20q	Deletion	13 (3)	—
Xp21.2	Deletion	8 (2)	DMD
Gain of Chr. 21 or 21q in non-HD ALL cases	—	37/283 (13)	—
Gain of Chr. X or Xq in non-HD ALL cases	—	23/283 (8)	—

Nonrandom chromosomal abnormalities (frequency >1.5% of all cases) are listed. 9p deletion and 12p deletions are separately shown in Figure 2B.

HD indicates hyperdiploid (>50 chromosomes); dup/amp, duplication and amplification of the region; and —, not applicable.

and 10 also was associated with a better prognosis, our study showed that change in number of these chromosomes did not have an impact on prognosis. Even though the size of our study is relatively large, it might not be able to detect some important factors associated with survival because the number of cases of this ALL subtype enrolled in our study was too small or advances in treatment of pediatric ALL may have eliminated several factors that previously influenced prognosis.

One of the limitations of SNP-chip analysis is that it cannot detect balanced translocations, which are common abnormalities in ALL, since this technique can detect only allelic dosage. In our correlation study of genomic abnormalities, a strong correlation was found between abnormalities involving 12p and

21q as described above. This, in part, reflected translocations of chromosome 12p and 21q (ETV6/RUNX1 fusion). Another correlation that we found between p16Del (on 9p) and deletion of 20q, may reflect dic(p13;q11).[9,20,32,33] These strong correlations of gains or loss of genetic materials may, therefore, sometimes suggest unbalanced translocations in ALL. Combination of SNP-chip and karyotyping will be a very strong technique to examine all genomic abnormalities in malignant cells.

Molecular allelokaryotyping of a series of 399 pediatric ALL samples has defined the range of genetic changes that occur in childhood ALL, including those associated with a poor prognosis. SNP-chip analysis is a novel technique that allows a thorough
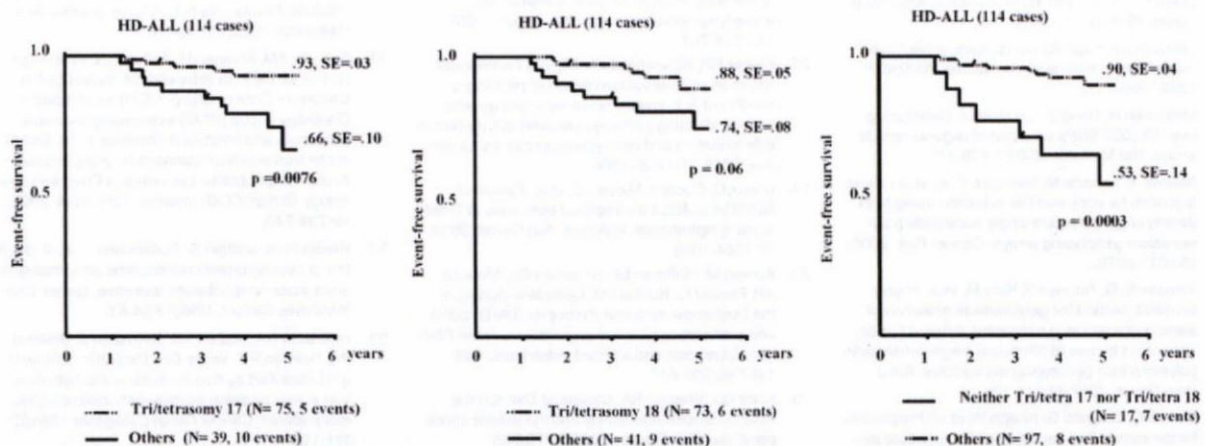

Figure 4. Prognostic impact of chromosomes 17 and 18. Absence of a gain of either chromosome 17 (left panel) or chromosome 18 (middle panel) is associated with poor prognosis in patients with HD-ALL; concurrent absence of a gain of both chromosomes 17 and 18 (right panel) is associated with very poor prognosis. Tri/tetra 18 and 17: HD-ALL with trisomy or tetrasomy 18 or 17. Others: HD-ALL without gain of chromosomes 17 and/or 18.

interrogation of the genome in ALL and identification of clinically significant subgroups of patients.

Acknowledgments

Authorship

Contribution: N.K. and S.O. designed this study. performed experiments, analyzed the data, and wrote the paper. M.Z., K.H., R.K., and W.-D.L. analyzed the data and wrote the paper. M.K., T.F., C.W.M., J.H., and M.S. analyzed the data. M.S., G.Y. and Y.N. performed experiments and analyzed the data. M.S. designed this study and wrote the paper. C.R.B. and H.P.K. designed this study, analyzed the data, and wrote the paper.

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

Correspondence: Seishi Ogawa, Department of Regeneration Medicine for Hematopoiesis, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan; e-mail: sogawa-tky@umin.ac.jp; or Norihiko Kawamata, Hematology/Oncology, Cedars-Sinai Medical Institute, 8700 Beverly Blvd, Los Angeles, CA 90048; e-mail: kawamatan@cshs.org.

References

1. Armstrong SA, Look AT. Molecular genetics of acute lymphoblastic leukemia. J Clin Oncol. 2005;23:6306-6315.

2. Pui CH, Evans WE. Treatment of acute lymphoblastic leukemia. N Engl J Med. 2006;354:166-178.

3. Pui CH, Relling MV, Downing JR. Acute lymphoblastic leukemia. N Engl J Med. 2004;350:1535-1548.

4. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet. 2005;37:549-554.

5. Garraway LA, Widlund HR, Rubin MA, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature. 2005;436:117-122.

6. Raghavan M, Lillington DM, Skoulakis S, et al. Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. Cancer Res. 2005;65:375-378.

7. Teh MT, Blaydon D, Chaplin T, et al. Genomewide single nucleotide polymorphism microarray mapping in basal cell carcinomas unveils uniparental disomy as a key somatic event. Cancer Res. 2005;65:8597-8603.

8. Rogan PK, Close P, Blouin JL, et al. Duplication and loss of chromosome 21 in two children with Down syndrome and acute leukemia. Am J Med Genet. 1995;59:174-181.

9. International HapMap Consortium. A haplotype map of the human genome. Nature. 2005;437:1299-1320.

10. Matsuzaki H, Dong S, Loi H, et al. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. Nat Methods. 2004;1:109-111.

11. 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;65:6071-6079.

12. 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;81:114-126.

13. Bene MC, Castoldi G, Knapp W, et al. Proposals for the immunological classification of acute leukemias: European Group for the Immunological Characterization of Leukemias (EGIL). Leukemia. 1995;9:1783-1786.

14. Ludwig WD, Rieder H, Bartram CR, et al. Immunophenotypic and genotypic features, clinical characteristics, and treatment outcome of adult pro-B acute lymphoblastic leukemia: results of the German multicenter trials GMALL 03/87 and 04/89. Blood. 1998;92:898-909.

15. Schrappe M, Reiter A, Ludwig WD, et al. Improved outcome in childhood acute lymphoblastic leukemia despite reduced use of anthracyclines and cranial radiotherapy: results of trial ALL-BFM 90. German-Austrian-Swiss ALL-BFM Study Group. Blood. 2000;95:3310-3322.

16. Look AT, Melvin SL, Williams DL, et al. Aneuploidy and percentage of S-phase cells determined by flow cytometry correlate with cell phenotype in childhood acute leukemia. Blood. 1982; 60:959-967.

17. Whitehead VM, Vuchich MJ, Lauer SJ, et al. Accumulation of high levels of methotrexate polyglutamates in lymphoblasts from children with hyperdiploid (greater than 50 chromosomes) B-lineage acute lymphoblastic leukemia: a Pediatric Oncology Group study. Blood. 1992;80:1316-1323.

18. Scott LM, Tong W, Levine RL, et al. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. N Engl J Med. 2007;356:459-468.

19. Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature. 2007; 446:758-764.

20. Kuiper RP, Schoenmakers EF, van Reijmersdal SV, et al. High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. Leukemia. 2007;21:1258–1266.

21. Graux C, Cools J, Melotte C, et al. Fusion of NUP214 to ABL1 on amplified episomes in T-cell acute lymphoblastic leukemia. Nat Genet. 2004; 36:1084-1089.

22. Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell. 1987;50:509-517.

23. Kops GJ, Weaver BA, Cleveland DW. On the road to cancer: aneuploidy and the mitotic checkpoint. Nat Rev Cancer. 2005;5:773-785.

24. Kotzot D. Complex and segmental uniparental disomy (UPD): review and lessons from rare chromosomal complements. J Med Genet. 2001; 38:497-507.

25. Baxter EJ, Scott LM, Campbell PJ, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet. 2005;365: 1054-1061.

26. James C, Ugo V, Le Couedic JP, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature. 2005; 434:1144-1148.

27. Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. N Engl J Med. 2005;352:1779-1790.

28. Levine RL, Wadleigh M, Cools J, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell. 2005;7:387-397.

29. Levine RL, Loriaux M, Huntly BJ, et al. The JAK2V617F activating mutation occurs in chronic myelomonocytic leukemia and acute myeloid leukemia, but not in acute lymphoblastic leukemia or chronic lymphocytic leukemia. Blood. 2005;106: 3377-3379.

30. Sulong S, Case M, Minto L, Wilkins B, Hall A, Irving J. The V617F mutation in Jak2 is not found in childhood acute lymphoblastic leukaemia. Br J Haematol. 2005;130:964-965.

31. Sutcliffe MJ, Shuster JJ, Sather HN, et al. High concordance from independent studies by the Children's Cancer Group (CCG) and Pediatric Oncology Group (POG) associating favorable prognosis with combined trisomies 4, 10, and 17 in children with NCI Standard-Risk B-precursor Acute Lymphoblastic Leukemia: a Children's Oncology Group (COG) initiative. Leukemia. 2005; 19:734-740.

32. Rieder H, Schnittger S, Bodenstein H, et al. dic(9; 20): a new recurrent chromosome abnormality in adult acute lymphoblastic leukemia. Genes Chromosomes Cancer. 1995;13:54-61.

33. Heerema NA, Maben KD, Bernstein J, Breitfeld PP, Neiman RS, Vance GH. Dicentric (9;20)(p11; q11) identified by fluorescence in situ hybridization in four pediatric acute lymphoblastic leukemia patients. Cancer Genet Cytogenet 1996;92: 111-115.

Cloning of genes involved in chromosomal translocations by high-resolution single nucleotide polymorphism genomic microarray

Norihiko Kawamata[a,b,c], Seishi Ogawa[b,d], Martin Zimmermann[b,e], Birte Niebuhr[f], Carol Stocking[f], Masashi Sanada[d], Kari Hemminki[g], Go Yamatomo[d], Yasuhito Nannya[d], Rolf Koehler[h], Thomas Flohr[h], Carl W. Miller[a], Jochen Harbott[i], Wolf-Dieter Ludwig[j], Martin Stanulla[e], Martin Schrappe[k], Claus R. Bartram[h,l], and H. Phillip Koeffler[a,l]

[a]Hematology/Oncology, Cedars-Sinai Medical Center/UCLA School of Medicine, Los Angeles, CA 90048; [d]Regeneration Medicine of Hematopoiesis, School of Medicine, University of Tokyo, Tokyo 113-8655, Japan; [e]Department of Pediatric Hematology and Oncology, Children's Hospital, Hannover Medical School, 30625 Hannover, Germany; [f]Molecular Pathology Heinrich-Pette-Institute, 20251 Hamburg, Germany; [g]Division of Molecular Genetic Epidemiology, German Cancer Research Center, 69120 Heidelberg, Germany; [h]Institute of Human Genetics, University of Heidelberg, 69117 Heidelberg, Germany; [i]Department of Hematology and Oncology, Center for Pediatrics, 35390 Giessen, Germany; [j]Department of Hematology, Oncology and Tumor Immunology, Robert-Rössle-Clinic at the HELIOS-Clinic Berlin-Buch, Charité, 13125 Berlin, Germany; and [k]Department of Pediatrics, University of Kiel, 69117 Kiel, Germany

High-resolution single nucleotide polymorphism genomic microarray (SNP-chip) is a useful tool to define gene dosage levels over the whole genome, allowing precise detection of deletions and duplications/amplifications of chromosomes in cancer cells. We found that this new technology can also identify breakpoints of chromosomes involved in unbalanced translocations, leading to identification of fusion genes. Using this technique, we found that the PAX5 gene was rearranged to a variety of partner genes including ETV6, FOXP1, AUTS2, and C20orf112 in pediatric acute lymphoblastic leukemia (ALL). The 3' end of the PAX5 gene was replaced by the partner gene. The PAX5 fusion products bound to PAX5 recognition sequences as strongly as wild-type PAX5 and suppressed its transcriptional activity in a dominant-negative fashion. In human B cell leukemia cells, binding of wild-type PAX5 to a regulatory region of BLK, one of the direct downstream target genes of PAX5, was diminished by expression of the PAX5-fusion protein, leading to repression of BLK. Expression of PAX5-fusion genes in murine bone marrow cells blocked development of mature B cells. PAX5-fusion proteins may contribute to leukemogenesis by blocking differentiation of hematopoietic cells into mature B cells. SNP-chip is a powerful tool to identify fusion genes in human cancers.

chromatin immunoprecipitation | dominant negative | fusion gene | PAX5 | SNP-chip

Pediatric acute lymphoblastic leukemia (ALL) is the most common malignant disease in children (1–3). It is a genetic abnormality resulting from accumulation of mutations in tumor suppressor genes and oncogenes (1–3). Fusion genes including ETV6/RUNX1 and E2A/PBX1 are frequently detected in pediatric ALL (1). Deletion of the INK4A/ARF gene (9p21) is also a common abnormality in ALL (1). However, other genetic changes remain to be elucidated in this disease.

Identification of mutated genes in ALL has evolved with improvements in technology. A very recent approach is single nucleotide polymorphism (SNP) analysis using an array based technology (4–6) that allows identification of amplifications, deletions, and allelic imbalances, such as uniparental disomy (represents doubling of the abnormal allele due to somatic recombination or duplication, and loss of the other normal allele) (7, 8). However, SNP-chip analysis is only able to detect changes of gene dosage and is unable to identify balanced translocations, which commonly occur in ALL.

Previously, we analyzed 399 pediatric ALL cases by SNP-chip analysis and found a number of genomic abnormalities, in addition to well known common alterations (9). This technique is sensitive enough to identify genes involved in start sites of

deletions/duplications. Indeed, this method allowed us to identify that the PBX1 gene was involved in start sites of duplication of 1q23 generated by der(19)t(1;19)(q23;p13) (9). Furthermore, correlation analysis of the individual genomic abnormalities suggested the presence of der(12)t(12;21)(p13;q22) and der(21)t(12;21)(p13;q22), as well as dic(9;20)(p13;q11) (9).

In this study, we found that this new technology permitted us to identify genes involved in well known unbalanced translocations including ETV6/RUNX1. Further, we found previously undetected fusion genes between PAX5 and a number of other partner genes by using this technique.

Results

Genes Involved in Unbalanced Translocations Were Identified by SNP-Chip Analysis. Because SNP-chip analysis can only detect changes of gene dosage including deletions, duplications, and amplifications (Fig. 1A), this technique is unable to identify balanced translocations (Figs. 1Aii). However, when one of a pair of reciprocally translocated chromosomes is lost, SNP-chip analysis can detect this abnormality as partial deletions of involved chromosomes (Fig. 1Aiii). Similarly, when one of a pair of reciprocally translocated chromosomes becomes duplicated, SNP-chip can also detect this abnormality as partial duplication of the involved chromosomes (Fig. 1Aiv). Furthermore, high resolution SNP-chip analysis allows us to identify the genes involved in these unbalanced translocations.

To prove that SNP-chip analysis can detect unbalanced translocations and the genes involved in these translocations, we
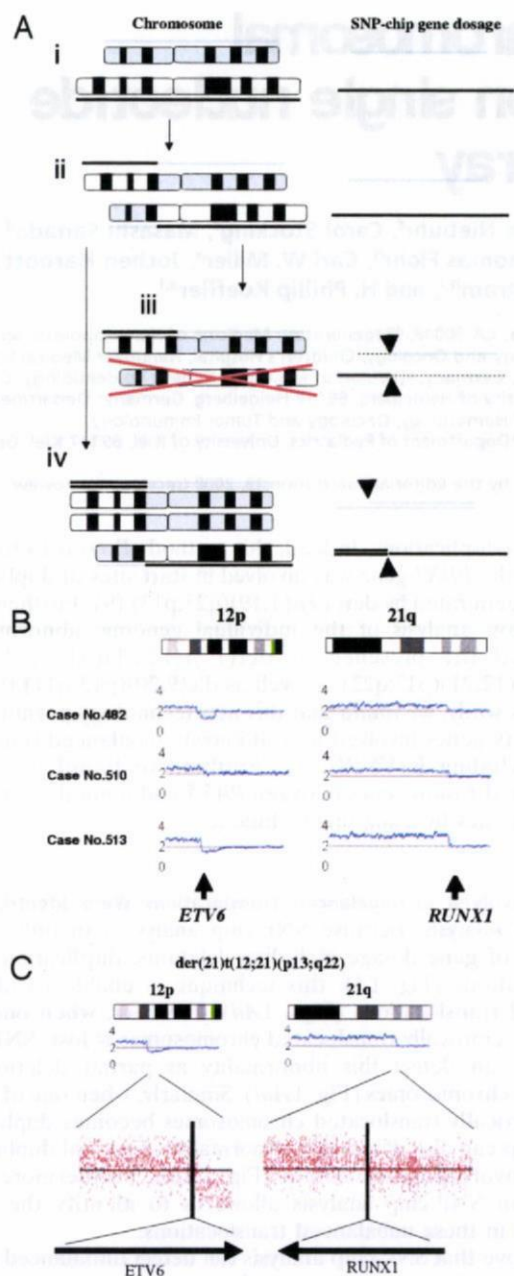
MEDICAL SCIENCES

Fig. 1. SNP-chip analysis detected genes involved in unbalanced transloca-
tions. (*A*) SNP-chip analysis can identify breakpoints of translocations when
one of the paired translocated chromosomes is either lost or duplicated/
amplified. (*Left*) Chromosomal status. Gene dosages are indicated either
above or beneath the chromosomes. (*Right*) Results of SNP-chip analysis. (*Ai*)
Normal chromosomes; gene dosage is normal. (*Aii*) Reciprocal translocation;
gene dosage is normal. (*Aiii*) One of the paired translocated chromosomes is
lost; gene dosage is lower than normal on the left side of the upper chromo-
some and the right side of the lower chromosome. Arrow heads indicate the
breakpoint of the translocation in each chromosome. (*Aiv*) One of the paired
translocated chromosomes is duplicated; gene dosage is higher than normal
on the right side of the upper chromosome and the left side of the lower
chromosome. Arrow heads indicate the breakpoint of this translocation in
each chromosome. (*B*) Representative cases with unbalanced translocation of
der(21)t(12;21)(p13;q22). (*Left*) Start sites of duplication at 12p13 involving
the *ETV6* gene. (*Right*) Start sites of duplication at 21q22 involving the *RUNX1*
gene. SNP-chip data of representative cases with dup(12)(p13) and
dup(21)(q22) are shown. These abnormalities were validated by FISH and/or
RT-PCR (data not shown). Results of SNP-chip data were visualized by CNAG
software. Lines above each chromosome show total gene dosage; level 2
indicates diploid (2N) amount of DNA, which is normal. (*C*) Magnified view of

analyzed cases having extra copies of *ETV6/RUNX1* fusion genes
generated by der(21)t(12;21)(p13;q22) (Fig. 1*B*), which were
initially identified by FISH and/or RT-PCR (data not shown).
SNP-chip was clearly able to identify this abnormality as dupli-
cations involving chromosome 12 and 21 (Fig. 1*B*). Further, the
result of high-resolution (250k) SNP-chip clearly identified *ETV6*
(12p13) and *RUNX1* (21q22) as the target genes involved in this
unbalanced translocation (Fig. 1*C*).

PAX5 Gene Is Frequently Fused to Partner Genes. Our previous data
showed the presence of dic(9;20)(p13;q11) in 11 cases of ALL
(9), 5 of which had deletion 9p13.2-pter. These 5 cases had start
sites of this deletion at 9p13.2 mapping to the *PAX5* gene (Fig.
2*A* and data not shown). This prompted us to reexamine all cases
of B-ALL that had deletion of 9p [supporting information (SI)
Table S1]. We found a total of 9 cases with similar start sites
(9p13.2), mapping to the *PAX5* gene (Fig. 2*A* and data not
shown). In 2 of these cases, simple abnormalities were detected
by SNP-chip: case 514 had only del9p13.2-pter and del7q11.2-
pter; case 458 had only del9p13.2-pter and dup3p13-pter (Table
S1 and Fig. 2*A*). Three cases (536, 543, 572) had complex
abnormalities including del9p13.2-pter and del20q11.21-qter, all
with the *C20ORF112* gene within the start site of del20q (Table
S1 and Fig. 2*A*). The other 2 cases (659, 767) had complex
abnormalities that included *ETV6* on 12p13 (Table S1 and
Fig. 2*A*).

Thus, we found four candidate partner genes fused to *PAX5*
in seven cases by SNP-chip analysis; *ETV6* on 12p13 (two cases)
(12), *C20orf112* on 20q11.1 (three cases), *AUTS2* on 7q11.1 (one
case) and *FOXP1* on 3p13 (one case) (Fig. 2*A*). Because these
translocations could lead to fusion transcripts between *PAX5*
and different partner genes, the presence of the predicted fusion
transcript was examined by RT-PCR using the mapping infor-
mation from the SNP-chip data. RT-PCR and nucleotide se-
quencing data of the PCR products confirmed that the *PAX5*
gene was fused to either the *ETV6* (two cases), *C20orf112* (three
cases), *AUTS2* (one case), or *FOXP1* (one case) gene and
transcribed into aberrant fusion messages (Fig. 2 *B* and *C*). Each
fusion gene was mutually and exclusively detected in the samples
studied. In one case with dic(9;20), exon 5 of *PAX5* was fused to
exon 8 of *C20orf112*, and in two cases with dic(9;20), exon 8 of
PAX5 was fused to exon 3 of *C20orf112*. *PAX5/ETV6* involved
exon 4 of *PAX5* and exon 3 of *ETV6*.

**Cellular Localization and DNA Binding Affinity of PAX5 Fusion Prod-
ucts.** In the *PAX5/FOXP1* fusion transcript, the amino acid
coding frame of the *FOXP1* gene was not identical to that of
PAX5, leading to a frame-shift and an early termination codon
after the fusion point of these two genes (Fig. 2*D*). However, all
other fusion genes were in frame and were predicted to encode
chimeric proteins. Two proteins (a short and long form) with
different breakpoints were predicted from the *PAX5/C20orf112*
fusion genes (Fig. 2*D*).

To confirm cellular localization of PAX5-fusion proteins, we
transfected vectors encoding wild-type *PAX5* and *PAX5* fusion
genes (*PAX5-ETV6*, *PAX5-FOXP1*, *PAX5-C20ORF112S*, and
PAX5-C20ORF112L) into 293T cells, fractionated the cytoplas-
mic and nuclear proteins, and examined the wild-type PAX5 and
PAX5-fusion proteins by Western blot analysis (Fig. 2*E*). PAX5-
ETV6 protein was detected in both the cytoplasm and nucleus;
PAX5-FOXP1 and PAX5-C20ORF112L proteins were predom-

SNP-chip data. (*Upper*) Start sites of duplications at 12p13 and 21q22 are
magnified. Signals of individual probe signals are shown. Vertical lines indi-
cated the positions of start sites of duplications. (*Lower*) Genes involved in the
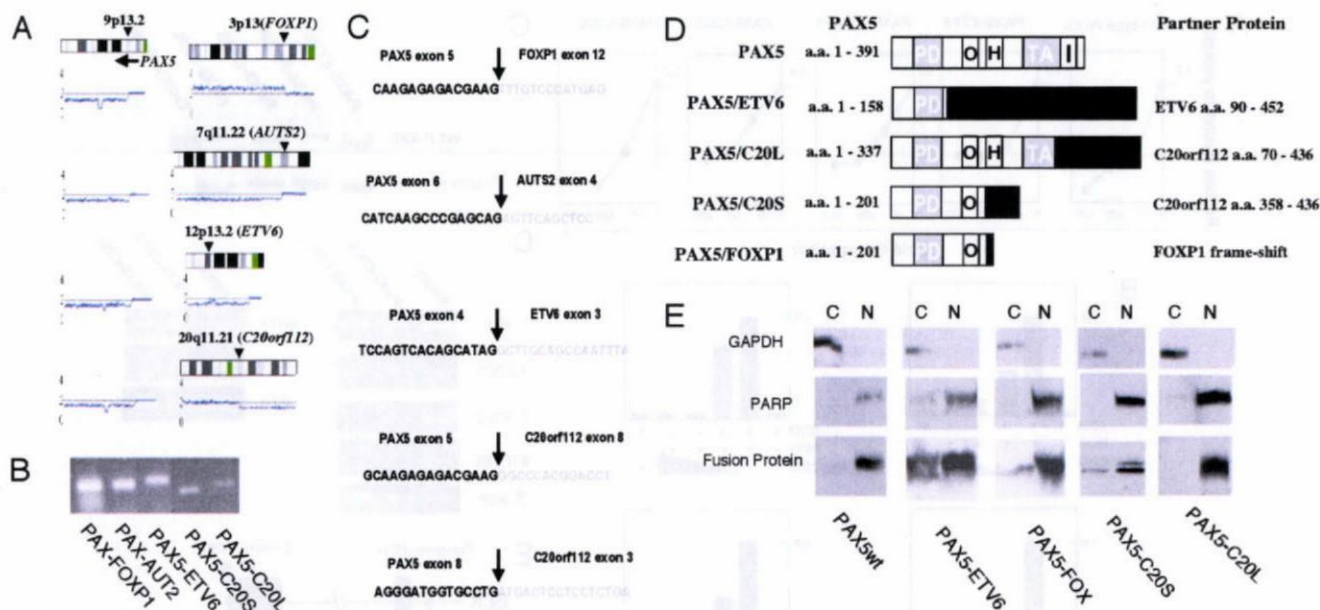start sites of duplications.

Fig. 2. PAX5 gene is fused to partner genes. (A) Start sites of deletion at 9p13.2 involving the PAX5 gene. (Left) SNP-chip data of representative cases with 9p13.2 deletions. A vertical arrow indicates the start sites of 9p deletion that involves the PAX5 gene. A horizontal arrow shows the direction of transcription of the PAX5 gene. (Right) Chromosomal abnormalities of partner chromosomes. Arrow heads indicate the start sites of duplication or deletions. Genes involved in the start sites are shown. (B) Result of RT-PCR. The ALL samples suggesting the presence of PAX5 fusion genes by SNP-chip analysis were examined by RT-PCR using the primers of PAX5 and the respective partner genes. (C) Fusion sequences of the PAX5 and partner genes. Joining sequences of fused transcripts are shown from the indicated exon of the fused gene. (D) Schematic structure of wild-type and mutant PAX5. Amino acid positions (aa) of each protein are indicated. PAX5/FOXP1 fusion construct has an early termination codon caused by a frame-shift. PD, paired domain; TA, transcription activation domain; O, octapeptide; H, homeodomain-like; I, inhibitory domain. (E) Subcellular fractionation of PAX5-fusion proteins. pcDNA vector encoding wild-type PAX5, PAX5-ETV6, PAX5-FOXP1, PAX5-C20ORF112S, or PAX5-C20ORF112L was transfected into 293T cells. Nuclear and cytoplasmic proteins were separated and electrophoresed in the gel. Localization of PAX5-fusion proteins was examined by PAX5 N-terminal specific antibody. Purity of cytoplasmic protein was examined with anti-GAPDH antibody and purity of nuclear proteins with the anti-PARP antibody. C, cytoplasmic fraction; N, nuclear fraction.

inantly localized in the nucleus; and 20% and 80% of PAX5-C20ORF112S proteins were localized in the cytoplasm and the nucleus, respectively (Fig. 2E). Localization of the fusion proteins was also confirmed by immunohistochemical staining (data not shown).

Because PAX5-fusion proteins were localized in the nucleus, we analyzed DNA binding affinity of these PAX5-fusion proteins *in vitro*. DNA binding affinity of the PAX5 wild-type and fusion proteins expressed in 293T cells was analyzed by electrophoretic mobility shift assay (EMSA), and signals of probes bound to the proteins were plotted graphically (Fig. 3A). Binding activity of each protein in the absence of cold competitor oligonucleotide probe was regarded as 1.0, and the binding activity in the presence of cold competitor oligonucleotide probes was measured. All PAX5-fusion proteins showed similar binding activity to the PAX5 recognition sequences as the wild-type PAX5 (Fig. 3A).

PAX5 Fusion Products Suppressed Transcriptional Activity of Wild-Type PAX5 in a Dominant Negative Fashion, Leading to Inhibition of B-Cell Development. To examine the effect of PAX5-fusion proteins on transcriptional activity of wild-type PAX5, we performed a reporter gene assay using 293T cells. Cotransfection reporter gene assays using wild-type and fusion PAX5 expression vectors along with a reporter gene driven by the murine CD19 promoter (which contains three repeats of PAX5 binding sequences) showed that the PAX5 fusion products suppressed transcriptional activity of PAX5 in a dominant-negative fashion (Fig. 3B). Expression of wild-type PAX5 proteins was minimally affected by coexpression of PAX5-fusion proteins (Fig. 3C), suggesting that PAX5-fusion proteins competed with wild-type PAX5 for the PAX5 binding sequences on the reporter gene.

Further, we transfected vectors encoding either PAX5-C20orf112S or PAX5-C20orf112L, each coexpressing the GFP marker, into Nalm 6 cells (a human B cell ALL cell line, which expresses endogenous PAX5) (data not shown). After transfection, GFP-positive cells were sorted by FACS and expression of PAX5-downstream genes was examined by semiquantitative RT-PCR (Fig. 3D and data not shown). We examined 10 downstream target genes (seven positively regulated direct target genes and three negatively regulated genes) of PAX5 (10–12) and found that four, including *ATP1B1*, *BLK*, *NEDD5* and *TCF7L2*, were down-regulated by induction of either PAX5-C20orf112S or PAX5-C20orf112L protein. However, expression of other reported PAX5 downstream target genes, including three positively regulated direct target genes (*IRF8*, *BST1*, *CD19*) and three negatively regulated genes (*CCR2*, *CCR5*, *NOTCH1*) were not affected by the induction of expression of the fusion proteins in these cells.

To examine the effect of PAX5 fusion protein on binding of wild-type PAX5 to the direct target gene *BLK* in the leukemic cells, we performed chromatin-immunoprecipitation (ChIP) assay using Nalm 6 cells transfected with either an empty vector or a construct encoding PAX5-C20orf112S. We used a PAX5 antibody detecting the C-terminal region of the protein, which could detect wild-type PAX5, but not PAX5-C20orf112S, as the C-terminal end of PAX5 was replaced by C20orf112S in this fusion protein. Although wild-type specific PAX5 antibody precipitated the promoter region of *BLK* after transfection of the empty vector, the amount of DNA of the *BLK* promoter region bound to wild-type PAX5 was reduced after transfection of the PAX5-C20orf112S gene (Fig. 3 E and F).

To examine the effect of PAX5-fusion proteins on B cell development in murine hematopoietic cells, we infected murine
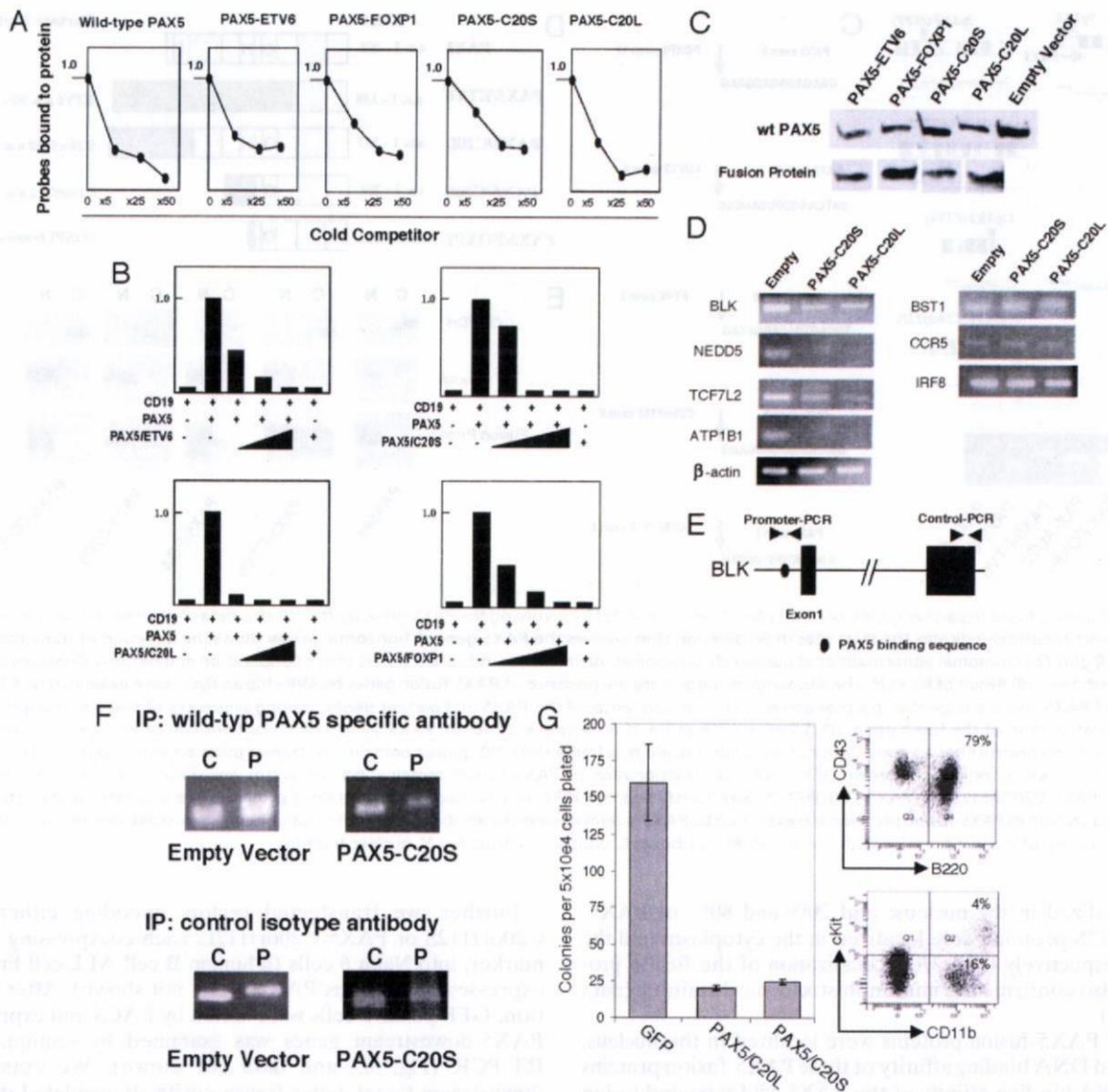
MEDICAL SCIENCES

Fig. 3. *PAX5*-fusion proteins suppress transcriptional activity of PAX5 in a dominant-negative fashion and block the growth of B cells. (*A*) Result of EMSA: Wild-type PAX5 and PAX5 fusion were expressed in 293T cells, and nuclear proteins were purified. The purified nuclear proteins were mixed with radioisotope labeled double-strand oligonucleotide DNA, in either the presence or absence of cold competitor oligonucleotides (5-, 25-, and 50-fold cold competitor probes). Intensity of each shifted band indicating DNA–protein complex was measured and plotted graphically. Intensity of shifted bands in the absence of cold competitor probes was regarded as 1.0. (*B*) Reporter gene assay. Wild-type and mutant PAX5 were mixed at a various ratios (1:0, 1:0.3, 1:1, 1:3, respectively, 1 = 500 ng of construct) and transfected. Forty-eight hours later, relative activity of firefly luciferase was measured and plotted. Results represent the mean values of the three experiments. CD19, PAX5 luciferase reporter construct having PAX5 binding region of CD19 promoter; PAX5, wild-type PAX5; PAX5/ETV6, PAX5/ETV6 fusion; PAX5/C20L, long form of PAX5/C20orf112 fusion in which PAX5 exon 8 is fused to C20orf112 exon 3; PAX5/C20S, short form of PAX5/C20orf112 fusion in which PAX5 exon 5 is fused to C20orf112 exon 8; PAX5/FOXP1, PAX5/FOXP1 fusion with an early termination codon caused by a frame-shift after the site of fusion. (*C*) Results of expression of wild-type PAX5 and PAX5-fusion proteins. After cotransfection of equal amounts of vector encoding either wild-type or fusion PAX5 genes into 293T cells, the expression of respective proteins was examined by Western blot. Levels of expression of wild-type PAX5 protein were minimally affected by coexpression of the PAX5-fusion proteins. (*D*) Semiquantitative RT-PCR of downstream target genes of PAX5. Expression of PAX5 downstream target genes was examined by semiquantitative RT-PCR. Nalm 6, a human B cell ALL cell line expressing endogenous PAX5, was transfected with pMSCV-GFP (Empty), pMSCV-GFP-PAX5-C20orf112S (PAX5-C20S), or pMSCV-GFP-PAX5-C20orf112L (PAX5-C20L). GFP-positive cells were sorted and subject to semiquantify the expression of respective genes. Optimal cycle numbers to semiquantify the expression of respective genes are as follows: BLK: 25 cycles; Nedd5; 25 cycles; TCF7L2: 25 cycles; ATP1B1: 25 cycles; β-actin: 22 cycles; CCR2: 25 cycles; CCR8: 30 cycles; IRF8: 30 cycles. (*E*) Structure of human *BLK* gene. Structure of *BLK* and primers used for ChIP assay within the 5′ regulatory region (Promoter-PCR) and 3′ end (Control-PCR) of the *BLK* gene is schematically shown. PAX5 binding site in the promoter region is indicated. (*F*) ChIP analysis of the PAX5 binding site in the *BLK* gene promoter. pMSCV-GFP (empty vector) or pMSCV-GFP-PAX5-C20S (PAX5-C20S) was transfected into human Nalm 6 B cell leukemia cells expressing endogenous PAX5. GFP-positive cells were subject to ChIP assay. The cells were fixed in formaldehyde solution and sonicated by ultrasound. DNA-protein complex was incubated with wild-type PAX5 specific antibody, which detected the C-terminal region of PAX5 but not the PAX5-C20orf112S protein (*Upper*). As a control, the DNA-protein complex was reacted with isotype nonspecific antibody (*Lower*). Immnoprecipitated DNA was subjected to PCR to amplify either the *BLK* promoter region containing PAX5 binding sequence (P) or, as an internal control, the 3′ end of the *BLK* gene (C). (*G*) Retrovirus infection experiments. Murine bone marrow cells were collected at 5 days after injection of 5FU. The hematopoietic cells were infected by retrovirus containing pMSCV-GFP empty vector (GFP), pMSCV-GFP-C20orf112L (PAX5/C20L), or pMSCV-GFP-C20orf112S (PAX5/C20S). GFP-positive murine hematopoietic cells were sorted and plated at 5 × 10⁴ cells per plate in methylcellulose containing mSCF, mIL7, and hFL. At 8 days after the plating, the colony numbers were counted (*Left;* results represent means and SD of three experiments). Cell surface antigens on the GFP-positive cells infected with pMSCV-GFP (GFP) at Day 11 were examined by FACS using antibodies against CD43 and B220 (*Upper Right*), c-kit and CD11b (*Lower Right*) antibodies, to confirm the development of B cells.

SNP-Chip Analysis. SNP-chip of GeneChip Human mapping 50k array Xbal 240 and/or 250k Nsp were used for this study (Affymetrix Japan). Preparation of samples was reported previously (4, 5). The data were analyzed by CNAG program as previously described (4, 5). All 399 ALL samples and their matched control samples were analyzed by using 50K-SNP chip; selected cases with genomic abnormalities were also analyzed by using 250K SNP-chip.

RT-PCR. RT-PCR was performed by using ThermoScript RT-PCR Systems (Invitrogen) according to the manufacturer's protocol. The primers used for detection of PAX5 fusion transcripts are listed in Table S2. Expression of PAX5 downstream target genes in Nalm 6 cells after transfection was examined by semiquantitative RT-PCR. The gene names and their primer sequences are listed in Table S3.

Reporter Gene Constructs and Expression Vectors. The PAX5 reporter gene construct with the luciferase gene and PAX5 binding region of the CD19 promoter, as well as the human PAX5 cDNA constructs, were kindly provided by Dr. M. Busslinger (Research Institute of Molecular Pathology, Vienna, Austria). PAX5-fusion constructs were generated by using PCR. All coding regions were ligated into the pcDNA3.1 vectors (Stratagene). Wild-type PAX5 cDNA was ligated into pMSCV vector (Clontech), and EGFP cDNA was ligated under the control of pGK promoter as a marker (pMSCV-GFP-wtPAX5). PAX5-C20orf112S and PAX5-C20orf112L cDNA sequences were also ligated into pMSCV-GFP vectors.

Transfection and Reporter Gene Assay. For reporter gene assays, pMSCV-GFP-wtPAX5 and pcDNA vectors encoding PAX5-fusion genes were cotransfected with the PAX5 reporter construct and pRL (*Renilla* luciferase) vector into 293T cells by using the Effecten transfection kit (Qiagen). Firefly and *Renilla* luciferase activities were measured with the Dual-Luciferase Reporter Assay System (Promega). Transfection into Nalm6 human pre-B cell ALL cell line was performed with Amaxa nucleofector. GFP-positive cells were sorted by using the MoFlo cell sorter (Dako). Detailed information about the procedure is described in *SI Text*.

Retrovirus Transduction into Murine Hematopietic Cells. Retrovirus containing pMSCV-GFP (empty), pMSCV-GFP-PAX5-C20orf112S, and pMSCV-GFP-PAX5-C20orf112L was generated. The retrovirus was transfected into murine bone marrow cells as previously reported (33). After the transfection, GFP-positive cells were sorted and plated into methylcellulose cultures (M3231; Stem Cell Technologies) as previously described (33). Surface antigens (CD43, B220, c-kit, and CD11b) of these GFP-positive cells were examined by using FACScan (Becton–Dickinson). Detailed information of the procedure is described in the *SI Text*.

Subcellular Fractionation of Proteins and EMSA. Forty-eight hours after transfection of vectors into 293T cells, the cells were subjected to subcellular fractionation with the CelLytic NuCLEAR Extraction Kit (Sigma-Aldrich). Detailed information of the procedure is described in the *SI Text*.

Purified nuclear proteins from the cells were also subjected to EMSA as previously reported (34). Detailed information of the procedure is described in the *SI Text*.

Chromatin Immunoprecipitation (ChIP) Assay. ChIP assay was performed with the Magna ChIP A kit from Millipore according to the manufacturer's protocol. pMSCV-GFP or pMSCV-GFP-C20orf112S were transfected into Nalm 6 cells as described above, and GFP-positive cells were sorted by MoFlo (Dako). Precipitated DNA was recovered and subjected to PCR to amplify the BLK promoter region and the 3' end of the BLK gene (internal control). The primer sequences used for ChIP assay are listed in Table S4. Detailed information of the procedure is described in the *SI Text*.

1. Armstrong SA, Look AT (2005) Molecular genetics of acute lymphoblastic leukemia. *J Clin Oncol* 23:6306–6315.
2. Pui CH, Evans WE (2006) Treatment of acute lymphoblastic leukemia. *N Engl J Med* 354:166–178.
3. Pui CH, Relling MV, Downing JR (2004) Acute lymphoblastic leukemia. *N Engl J Med* 350:1535–1548.
4. Nannya Y, et al. (2005) A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* 65:6071–6079.
5. Yamamoto G, et al. (2007) 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.
6. Lindblad-Toh K, et al. (2000) Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. *Nat Biotechnol* 18:1001–1005.
7. Raghavan M, et al. (2005) Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res* 65:375–378.
8. Lehmann S, et al. (2008) Molecular allelokaryotyping of early-stage, untreated chronic lymphocytic leukemia. *Cancer* 112:1296–1305.
9. Kawamata N, et al. (2008) Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood* 111:776–784.
10. Lin YH, Shin EJ, Campbell MJ, Niederhuber JE (1995) Transcription of the blk gene in human B lymphocytes is controlled by two promoters. *J Biol Chem* 270:25968–25975.
11. Schebesta A, et al. (2007) Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration, and immune function. *Immunity* 27:49–63.
12. Delogu A, et al. (2006) Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells. *Immunity* 24:269–281.
13. Schröck E, Padilla-Nash H (2000) Spectral karyotyping and multicolor fluorescence in situ hybridization reveal new tumor-specific chromosomal aberrations. *Semin Hematol* 37:334–347.
14. Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts by amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* 85:8998–9002.
15. Cheng S, Fockler C, Barnes WM, Higuchi R (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc Natl Acad Sci USA* 91:5695–5699.
16. Tomlins SA, et al. (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310:644–648.
17. Perner S, et al. (2006) TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res* 66:8337–8341.
18. Urbanek P, Wang ZQ, Fetka I, Wagner EF, Busslinger M (1994) Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* 79:901–912.
19. Xie H, Ye M, Feng R, Graf T (2004) Stepwise reprogramming of B cells into macrophages. *Cell* 117:663–676.
20. Kumaran RI, Thakar R, Spector DL (2008) Chromatin dynamics and gene positioning. *Cell* 132:929–934.
21. Gilliland DG, Griffin JD (2002) The roles of FLT3 in hematopoiesis and leukemia. *Blood* 100:1532–1542.
22. Tenen DG (2003) Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer* 3:89–101.
23. Mullighan CG, et al. (2007) Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446:758–764.
24. Cazzaniga G, et al. (2001) The paired box domain gene PAX5 is fused to ETV6/TEL in an acute lymphoblastic leukemia case. *Cancer Res* 61:4666–4670.
25. Bousquet M, et al. (2007) A novel PAX5-ELN fusion protein identified in B cell acute lymphoblastic leukemia acts as a dominant negative on wild-type PAX5. *Blood* 109:3417–3423.
26. Nebral K, et al. (2007) Identification of PML as novel PAX5 fusion partner in childhood acute lymphoblastic leukaemia. *Br J Haematol* 139:269–274.
27. Busslinger M, Klix N, Pfeffer P, Graninger PG, Kozmik Z (1996) Deregulation of PAX-5 by translocation of the Emu enhancer of the IgH locus adjacent to two alternative PAX-5 promoters in a diffuse large-cell lymphoma. *Proc Natl Acad Sci USA* 93:6129–6134.
28. Iida S, et al. (1996) The t(9;14)(p13;q32) chromosomal translocation associated with lymphoplasmacytoid lymphoma involves the PAX-5 gene. *Blood* 88:4110–4117.
29. Pasqualucci L, et al. (2007) Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 412:341–346.
30. Anderson K, et al. (2007) Ectopic expression of PAX5 promotes self renewal of biphenotypic myeloid progenitors co-expressing myeloid and B-cell lineage associated genes. *Blood* 109:3697–3705.
31. Souabni A, Jochum W, Busslinger M (2007) Oncogenic role of Pax5 in the T-lymphoid lineage upon ectopic expression from the immunoglobulin heavy-chain locus. *Blood* 109:281–289.
32. Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbor Press, Cold Spring Harbor, NY). 3rd Ed.
33. Schwieger M, et al. (2002) AML1-ETO inhibits maturation of multiple lymphohematopoietic lineages and induces myeloblast trnsformation in synergy with ICSBP deficiency. *J Exp Med* 196:1227–1240.
34. Sato H, Wang D, Kudo A (2001) Dissociation of Pax-5 from KI and KII sites during kappa-chain gene rearrangement correlates with its association with the underphosphorylated form of retinoblastoma. *J Immunol* 166:6704–6710.

hematopoietic cells from the bone marrow with retroviral vectors encoding either PAX5-C20orf112S or PAX5-C20orf112L. GFP-positive infected bone marrow cells were sorted by FACS and plated in media containing cytokines that are know to stimulate B cell differentiation (Fig. 3F). Murine hematopoietic cells infected with the empty vector showed abundant colonies (Fig. 3G Left), and 79% of the cells were B220 positive B cells (Fig. 3G Upper Right). In contrast, murine hematopoietic cells infected with either PAX5-C20orf112S or PAX5-C20orf112L formed very few colonies (Fig. 3G Left). Most of these colonies were GFP-negative (data not shown), suggesting that these PAX5-fusion proteins impaired B cell development from murine hematopoietic cells.

Discussion

In this study, we describe a paradigm for discovering fusion genes in malignancy by taking advantage of samples with unbalanced translocations and using high density SNP-chip analysis. This technique allows us to identify genes involved in translocations even if chromosomal analysis is not available, especially in solid tumors.

Steps to identify novel fusion genes using SNP chip analysis include (i) identify either a deletion or duplication that occurs within two genes; (ii) determine whether transcription of both genes is in the same direction; (iii) take advantage of ancillary tests such as standard chromosomal analysis or spectral karyotyping (14), which can grossly show that two chromosomes are fused; and (iv) design primers of candidate genes and perform RT-PCR to clone fusion genes. Rapid amplication of cDNA ends (RACE) (15) or long-distance PCR (12) also help the cloning of genes involved in translocations. In our SNP-chip data, a number of regions of segmental deletions or duplications were detected (9). Although some of them are simple deletions or duplications at the original sites of the chromosomes, the others are deletions that occurred during chromosomal translocations or when duplicated fragments were inserted into chromosomal sites other than the original region (data not shown). Therefore, data of chromosomal analysis help to define translocations, leading to identification of candidate genes in novel fusion genes.

Recently, Tomlins et al. found the fusion genes TMPRSS2/ERG and TMPRSS2/ETV1 in prostate cancers by using expression microarray data (16). They focused on the genes ERG and ETV1, which are highly expressed in this cancer and examined levels of individual exons of these two genes (16). They found differences in expression of 5′ and 3′ regions of the genes, suggesting that these genes are fused to each other (16). In these fusion genes, the 5′ regions were replaced by the TMPRSS2 gene, resulting in the differences in the expression of the 5′ and 3′ region of the ERG/ETV1 genes (16). They also used SNP-chip analysis to identify these fusion genes and found a deletion of a genomic region between TMPRSS2 (21q22.3) and ERG (21q22.2), leading to fusion of these two genes (17). These new technologies, based on oligonucleotide microarrays and bioinformatics, will help to identify fusion genes in cancers.

Our study found that the PAX5 gene was frequently fused to one of a variety of partner genes. PAX5 is a key transcription factor in the development of B cells (18, 19). We found that these PAX5 fusion proteins suppressed the function of wild-type PAX5 in a dominant-negative fashion and suppressed expression of downstream target genes of wild-type PAX5 in leukemic cells.

We found that when PAX5 was joined to one of its fusion partner genes, its C-terminal end was replaced by one of the partner genes. Elimination of the C-terminal end of PAX5 may play an important role in generation of a dominant negative form of mutated PAX5. In in vitro assays, PAX5-fusion proteins showed a similar affinity as wild-type PAX5 for the PAX5 recognition sequences. Although expression of several downstream targets of wild-type PAX5 was repressed by expression of PAX5-fusion proteins, others were not affected. Binding of

transcription factors to DNA can be modulated by cofactors and/or neighboring transcription factors (20). Compared to PAX5, PAX5-fusion proteins may bind more strongly to some target genes and more weakly to others, depending on the contextual environment of the target genes.

Further, our data showed that PAX5-fusion protein inhibited B cell development of hematopoietic cells in a colony formation assay. This result may suggest that PAX5 fusion protein blocked differentiation of hematopoietic cells into mature B cells. PAX5-deficient mice have impairment of B cell differentiation (18). These data suggest that PAX5-fusion proteins may contribute to leukomegenesis by blocking B cell differentiation. It has been suggested that two distinct genetic abnormalities contribute to leukemogenesis in acute myelogenous leukemia (AML); one is mutations promoting cellular proliferation, for example FLT3 or RAS mutations, and the other is mutations blocking differentiation, for example PML-RARA or RUNX1-ETO (21, 22). PAX5-fusion proteins may cooperate with unidentified mutations promoting cellular proliferation in the ALL cells.

Recently, Mullighan et al. have analyzed pediatric ALL samples by high density SNP-chips and found frequent abnormalities of PAX5 gene (23). Their data also showed that PAX5 fusion products suppressed transcriptional activity of PAX5 in a dominant-negative fashion (23). In addition, other researcher have reported PAX5 fusion genes, including PAX5 fused to ETV6 (12p13) (23, 24), FOXP1 (3p14) (23), ZNF521 (18q11) (23), ELN (7q11.23) (25), and PML (15q24) (26). We have found PAX5 fused to either ETV6, FOXP1, C20orf112 (20q11), or AUTS2 (7q11.22).

In our study, the function of PAX5 was attenuated by the dominant-negative forms of the fusion products in B cell lineage ALL, suggesting that PAX5 behaves as a tumor suppressor in early B cells, and that impairment of its function can be associated with the development of ALL. In contrast, translocation of the PAX5 gene to the enhancer region of the Ig heavy chain gene [t(9;14)(p13.2;q32)] or point mutations of the 5′ regulatory region of the PAX5 gene leads to its overexpression, which is associated with B cell lineage lymphomas (27–29). Also, experimental overexpression of wild-type PAX5 can transform lymphocytes (30, 31). Therefore, an aberrant PAX5 may behave in a dominant-negative fashion at the pre-B stage of B cell development, resulting in ALL; its forced expression in a more mature B cell can lead to lymphoma. Our study showed that PAX5-fusion proteins blocked differentiation of B cells but did not transform them. B cells at different stages of differentiation may need alteration of distinct sets of pathways to transform. Why PAX5 can act as a tumor suppressor in ALL and as an oncoprotein in lymphoma is unclear. Further studies are needed to clarify the mechanism of this paradoxical phenomenon in carcinogenesis.

In summary, we identified multiple fusion genes in ALL by SNP-chip analysis, leading to the exploration of a B cell differentiation block as a contributing factor to the development of ALL. This methodology should help researchers to identify oncogenic fusion genes and explore the mechanism of tumorigenesis in other types of cancers as well.

Materials and Methods

Samples and DNA/RNA Preparation. SNP-chip was performed on 399 pediatric ALL patients consecutively enrolled in the ALL-BFM 2000 trial of the Berlin-Frankfurt-Münster (BFM) study at diagnosis and during remission (350 cases were B cell lineage ALL and 49 cases were T cell lineage ALL) (9). Detailed results of the SNP-chip analysis are published separately (9). The ALL-BFM 2000 study was approved by the local ethics committee. DNA and RNA were extracted from the ALL samples and cell lines by using standard techniques (32). Nalm 6, a human pre-B ALL cell line, was generously provided by Dr. G. Crook (Los Angeles Children's Hospital, Los Angeles, CA) and maintained in RPMI medium 1640 with 10% FBS.

MEDICAL SCIENCES

LETTERS

Oncogenic mutations of ALK kinase in neuroblastoma

Yuyan Chen[1,2,3]*, Junko Takita[1,2,3]*, Young Lim Choi[4]*, Motohiro Kato[1,3], Miki Ohira[5], Masashi Sanada[2,3,6], Lili Wang[2,3,6], Manabu Soda[4], Akira Kikuchi[7], Takashi Igarashi[1], Akira Nakagawara[5], Yasuhide Hayashi[8], Hiroyuki Mano[4,6] & Seishi Ogawa[2,3,6]

Neuroblastoma in advanced stages is one of the most intractable paediatric cancers, even with recent therapeutic advances[1]. Neuroblastoma harbours a variety of genetic changes, including a high frequency of *MYCN* amplification, loss of heterozygosity at 1p36 and 11q, and gain of genetic material from 17q, all of which have been implicated in the pathogenesis of neuroblastoma[2-5]. However, the scarcity of reliable molecular targets has hampered the development of effective therapeutic agents targeting neuroblastoma. Here we show that the anaplastic lymphoma kinase (ALK), originally identified as a fusion kinase in a subtype of non-Hodgkin's lymphoma (NPM–ALK)[6-8] and more recently in adenocarcinoma of lung (EML4–ALK)[9,10], is also a frequent target of genetic alteration in advanced neuroblastoma. According to our genome-wide scans of genetic lesions in 215 primary neuroblastoma samples using high-density single-nucleotide polymorphism genotyping microarrays[11-14], the *ALK* locus, centromeric to the *MYCN* locus, was identified as a recurrent target of copy number gain and gene amplification. Furthermore, DNA sequencing of *ALK* revealed eight novel missense mutations in 13 out of 215 (6.1%) fresh tumours and 8 out of 24 (33%) neuroblastoma-derived cell lines. All but one mutation in the primary samples (12 out of 13) were found in stages 3–4 of the disease and were harboured in the kinase domain. The mutated kinases were autophosphorylated and displayed increased kinase activity compared with the wild-type kinase. They were able to transform NIH3T3 fibroblasts as shown by their colony formation ability in soft agar and their capacity to form tumours in nude mice. Furthermore, we demonstrate that downregulation of *ALK* through RNA interference suppresses proliferation of neuroblastoma cells harbouring mutated *ALK*. We anticipate that our findings will provide new insights into the pathogenesis of advanced neuroblastoma and that *ALK*-specific kinase inhibitors might improve its clinical outcome.

To identify oncogenic lesions in neuroblastoma, we performed a genome-wide analysis of primary tumour samples obtained from 215 neuroblastoma patients using high-density single-nucleotide polymorphism (SNP) arrays (Affymetrix GeneChip 250K *Nsp*I) (Supplementary Table 1). Twenty-four neuroblastoma-derived cell lines were also analysed (Supplementary Table 2). Interrogating over 250,000 SNP sites, this platform permits the identification of copy number changes at an average resolution of less than 12 kilobases (kb)[13,14].

Analysis of this large number of samples, consisting of varying disease stages, permitted us to obtain a comprehensive registry of genomic lesions in neuroblastoma (Supplementary Figs 1 and 2). A gain of chromosomes, often triploid or hyperploid (defined by mean copy number of >2.5), was a predominant feature of neuroblastoma genomes in the lower stages. Ploidy generally correlated with the

clinical stage, where non-hyperploid cases were significantly associated with stage 4 disease ($P = 4.13 \times 10^{-5}$, trend test) (Supplementary Fig. 3 and Supplementary Table 3). 17q gains, frequently in multiple copies ($3 \leq$ copy number <5), were a hallmark of the neuroblastoma genome[4] and were found in most neuroblastoma cases. Copy number gains tended to spare chromosomes 3, 4, 10, 14 and 19 (Supplementary Figs 2 and 3). Notably, these chromosomes often had copy number losses including 1p (22.8%), 3p (8.8%), 4p (5.1%), 6q (7.0%), 10q (9.8%), 11q (19.5%), 14q (3.7%), 19p (7.4%) and 19q (5.1%), implicating the pathogenic role of 'relative' gene dosages.

After excluding known copy number variations, we identified a total of 28 loci undergoing high-grade amplifications (copy number ≥ 5) (Supplementary Table 4). These lesions fell into relatively small genomic segments, having a mean size of 361 kb, which accelerated the identification of gene targets in these regions (Supplementary Table 4 and Supplementary Fig. 4). The candidate gene targets included *TERT* (5p15.33), *HDAC3* (5q31.3), *IGF2* (11p15.1), *MYEOV* (11q13.3), *FGF7* (15q21.1) and *CDH13* (16q23.3). However, many of them were not recurrent but found only in a single case. Although the recurrent lesions were mostly explained by the amplification of *MYCN* at 2p24, as found in 50 out of 215 (23%) of the primary cases, we identified another peak of recurrent amplification at 2p23 (Fig. 1a), which consisted of amplicons in five primary cases and in one neuroblastoma-derived cell line, NB-1 (Supplementary Fig. 5). This peak was located at the centromeric margin of the common copy number gains in chromosome 2p, which was created by copy number gains in 109 samples mostly from non-hyperploid stage 4 cases. The minimum overlapping amplification was defined by the amplicons found in the NB-1 cell line (Supplementary Fig. 5) and contained a single gene, the anaplastic lymphoma kinase (*ALK*), which has previously been reported to be overexpressed in neuroblastoma cases[15]. Although five of the six samples showing *ALK* amplification also had *MYCN* amplification, one primary case (NT056) lacked a *MYCN* peak and the amplification was confined to the *ALK*-containing locus. In interphase fluorescent *in situ* hybridization (FISH) analysis of NB-1, *MYCN* and *ALK* loci were amplified in separate amplicons (Fig. 1b), indicating that the 2p23 amplicons containing *ALK* were unlikely to represent merely 'passenger' events of *MYCN* amplification but actively contributed to the pathogenesis of neuroblastoma.

Because an oncogene can be activated by gene amplification and/ or mutation, to search for possible mutations we performed DNA heteroduplex formation analysis[16] and genomic DNA sequencing for the exons 20 to 28 of *ALK*, which encompass the juxtamembrane and kinase domains (Supplementary Table 5). In total, we identified eight nucleotide changes in 21 neuroblastoma samples, 13 out of 215

[1]Department of Pediatrics, [2]Cell Therapy and Transplantation Medicine, [3]Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Tokyo 113-8655, Japan. [4]Division of Functional Genomics, Jichi Medical University, Toguchi 329-0498, Japan. [5]Division of Biochemistry, Chiba Cancer Center Research Institute, Chiba 260-8717, Japan. [6]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama, 332-0012, Japan. [7]Division of Hematology/Oncology, Saitama Children's Medical Center, Saitama 339-8551, Japan. [8]Gunma Children's Medical Center, Shibukawa 377-8577, Japan.
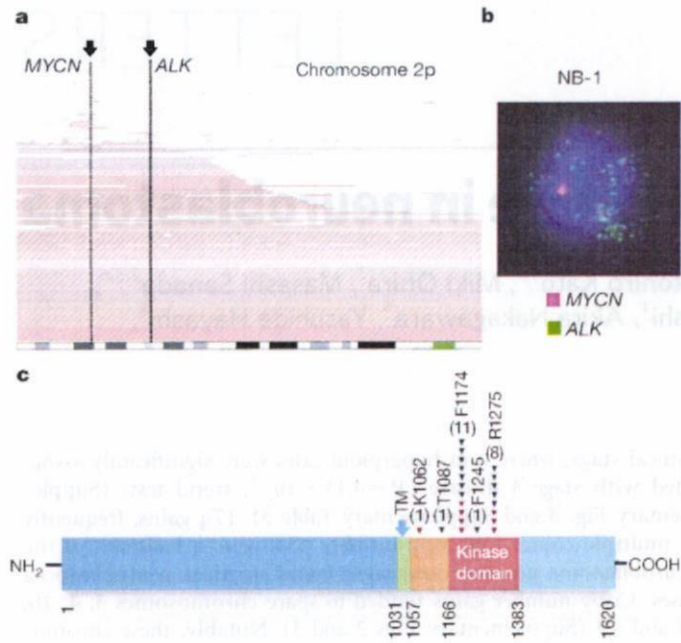*These authors contributed equally to this work.

Figure 1 | Common 2p gains/amplifications and *ALK* mutations in neuroblastoma samples. a, Recurrent copy number gains on the 2p arm. High-grade amplifications are shown by light-red horizontal lines, whereas simple gains are shown by dark-red lines. Two common peaks of copy number gains and amplifications in the *MYCN* and *ALK* loci are indicated by arrows. The cytobands in 2p are shown at the bottom. **b**, Interphase FISH analysis of NB-1 showing high-grade amplification of *MYCN* (red) and *ALK* loci (green). The amplified *MYCN* locus appears as a single large signal. **c**, Distribution of the eight *ALK* mutations found in 21 neuroblastoma samples. The positions of the mutated amino acids are indicated by black (primary samples) and red (cell lines) arrowheads. The number of mutations at each site is shown at the top of the arrowheads. TM, transmembrane.

(6.1%) primary samples and 8 out of 24 (33%) cell lines, which resulted in seven types of amino acid substitutions at five different positions (Table 1 and Supplementary Fig. 6). They were not found in either the genomic DNA collected from 50 healthy volunteers or in the SNP databases at the time of preparing this manuscript. In fact, somatic origins of missense changes were confirmed in 9 out of 13 primary cases, for which DNA was obtained from the peripheral blood or the tumour-free bone marrow specimens (Supplementary Fig. 6). On the other hand, T1087I (ACC>ATC), found in case NT126, had a germline origin and thus it could not be determined whether the T1087I change was a rare non-functional polymorphism or represented a pathogenic germline mutation. For other changes found in three primary cases (NT128, NT217 and NT218) and cell lines, normal DNA was not available but they were likely to represent oncogenic mutations because they were identical to common somatic changes (F1174L or R1275Q) or shown to have oncogenic potential in functional assays (K1062M).

Most mutations occurred within the kinase domain (20 out of 22 or 91%), which clearly showed two mutation hotspots at F1174 and R1275 (Fig. 1c). A neuroblastoma-derived cell line, SJNB-2, had a homozygous *ALK* mutation of R1275Q, which was probably due to uniparental disomy of chromosome 2 (Supplementary Fig. 7a). Another case (NT074) harboured two different mutations, F1174L and R1275Q, but it remains to be determined whether both are on the same allele. ALK mutations within the kinase domain occurred at amino acid positions that are highly conserved across species and during molecular evolution (Supplementary Figs 8 and 9). According to the conserved structure of other insulin receptor kinases we predicted that F1174 is located at the end of the Cα1 helix, whereas the other two are on the two β-sheets: before the catalytic loop (β6, F1245) and within the activation loop (β9, R1275) (Supplementary Fig. 7b, c)[17]. Thus, conformational changes due to amino acid substitutions at these positions might be responsible for the aberrant activity of the mutant kinases.

Table 1 | *ALK* mutations/amplifications in neuroblastoma samples

Sample	Age (months)	Stage	MYCN*	Clinical outcome	Mutations/ amplifications	Nucleotide substitution	Origin of mutations
NT126	99	4	−	Dead	T1087I	ACC>ATC	Germ line
NT218	8	1	−	Alive	F1174L	TTC>TTG	ND
NT074	34	3	+	Dead	F1174L	TTC>TTA	Somatic
					R1275Q	CGA>CAA	Somatic
NT160	12	4	+	Dead	F1174L	TTC>TTA	Somatic
NT217	24	4	+	Dead	F1174L	TTC>TTA	ND
NT190	48	4	+	Alive	F1174L	TTC>TTA	Somatic
NT060	163	3	+	Alive	F1174C	TTC>TGC	Somatic
NT162	28	4	+	Dead	F1174V	TTC>GTC	Somatic
NT195	24	4	+	Alive	F1245L	TTC>TTG	Somatic
NT055	6	3	−	Alive	R1275Q	CGA>CAA	Somatic
NT128	8	4	−	Dead	R1275Q	CGA>CAA	ND
NT164	54	4	−	Dead	R1275Q	CGA>CAA	Somatic
NT200	133	4	−	Dead	R1275Q	CGA>CAA	Somatic
SCMC-N5†			-		K1062M	AAG>ATG	ND
SJNB-4†			-		F1174L	TTC>TTA	ND
LAN-1†			-		F1174L	TTC>TTA	ND
SCMC-N2†			-		F1174L	TTC>TTA	ND
SK-N-SH†			-		F1174L	TTC>TTA	ND
SJNB-2†‡			-		R1275Q	CGA>CAA	ND
LAN-5†			+		R1275Q	CGA>CAA	ND
TGW†			+		R1275Q	CGA>CAA	ND
NT204	12	1	+	Alive	Amplification	-	-
NT056	11	3	−	Dead	Amplification	-	-
NT071	36	3	+	Alive	Amplification	-	-
NT165	19	4	+	Dead	Amplification	-	-
NT169	7	4	+	Dead	Amplification	-	-
NB-1†	-	-	+	-	Amplification	-	-

ND, not determined.
* Presence (+) or absence (−) of *MYCN* amplification in FISH analysis. All cases where there was an absence of *MYCN* amplification (−) were also checked for possible *MYCN* mutations by sequencing of all *MYCN* exons, but no *MYCN* mutations were identified.
† Cell lines.
‡ Homozygous mutation.

ALK mutation highly correlated with *MYCN* amplification ($P = 1.55 \times 10^{-4}$, Fisher's exact test; Supplementary Table 6) where 14 out of 21 mutations coexisted with *MYCN* amplification. Regardless of the status of *MYCN* amplification, 12 of the 13 mutations were found in patients with advanced stage neuroblastoma (Table 1). However, whereas *MYCN* amplification and stage 4 were significant risk factors for poor survival, the mutation/amplification status of *ALK* was not likely to have a major impact on survival (Supplementary Fig. 10 and Supplementary Table 7), although the statistical power of the current analysis was largely limited in order to detect a marginal hazard.

To evaluate the impact of ALK mutations on kinase activity, we generated Flag-tagged constructs of ALK and its mutants, F1174L and K1062M, which were stably expressed in NIH3T3 cells, and examined their phosphorylation status and *in vitro* kinase activity. The ALK mutants stably expressed in NIH3T3 cells were phosphorylated according to western blot analysis using an antibody specific for phosphorylated ALK (anti-pY1604) and a PY20 blot after anti-Flag immunoprecipitation of the mutant kinases (Fig. 2a), whereas the wild-type kinase was not phosphorylated. The immunoprecipitated ALK mutants also showed increased tyrosine kinase activity *in vitro* when compared with wild-type ALK. This was shown using both a universal substrate for tyrosine kinase (poly-GluTyr) and the synthetic YFF peptide[18], which was derived from a sequence of the activation loop of *ALK* (Fig. 2b, c). In accordance with these findings, downstream molecules of ALK signalling including AKT, STAT3 and ERK[15] were activated in cells expressing mutant ALK, as shown by their increased phosphorylation (Fig. 2d).

Next, we investigated the oncogenic potential of these mutants. NIH3T3 cells stably expressing mutant kinases showed increased colony formation in soft agar compared with the wild-type protein (Fig. 3a and Supplementary Fig. 11). The tumorigenicity of these ALK mutants was further assayed by injecting 1.0×10^7 NIH3T3 cells into nude mice. The NIH3T3 cells transfected with the ALK mutants showed focus-forming capacity and developed subcutaneous tumours (6 out of 6 inoculations) 21 days after inoculation, whereas the mock and wild-type ALK-transfected cells did not (0 out of 6 inoculations) (Fig. 3b, c). Finally, we examined the effect of ALK inhibition on the proliferation of neuroblastoma-derived cell lines. RNA interference (RNAi)-mediated *ALK* knockdown resulted in reduced cell proliferation of SK-N-SH cells harbouring the F1174L mutation, but the effects were less clear in wild-type ALK-expressing LAN-2 cells (Fig. 3d, e). Of particular interest is a recent report that 5 out of 17 neuroblastoma-derived cell lines, including SK-N-SH and NB-1, frequently showed high sensitivity to the specific ALK inhibitor TAE684 (ref. 19).
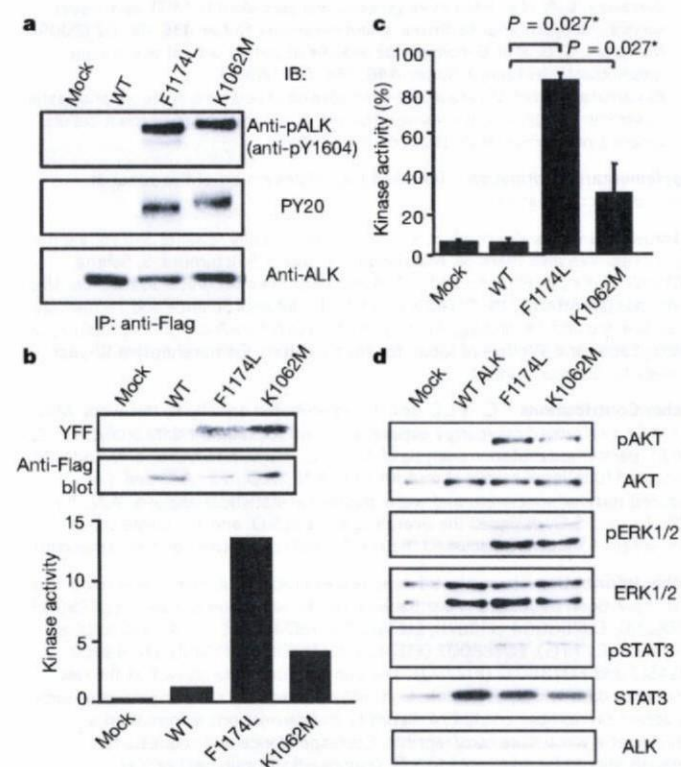

Figure 2 | Kinase activity of ALK mutants and their downstream signalling. a, Stably expressed ALK and its mutants (F1174L and K1062M) were immunoprecipitated with an anti-Flag antibody and subjected to western blot analysis with anti-pY1604 (upper panel) or PY20 (middle panel). An anti-ALK blot of precipitated kinases is also displayed (bottom panel). b, *In vitro* kinase assay for wild-type ALK kinase and its mutants using the synthetic YFF peptide as a substrate, where kinase activity is expressed as relative values to that for wild-type kinase based on the densities in the autoradiogram. c, Kinase activity was also assayed for the poly-GluTyr peptide. Significantly different measurements are indicated by asterisks with *P* values. Bars show mean ($\pm$s.d.) in three independent experiments. d, Western blot analyses of NIH3T3 cells expressing wild-type and mutant ALK for phosphorylated forms of AKT (pAKT), ERK (pERK1/2) and STAT3 (pSTAT3). The total amount of each molecule is also displayed (AKT, ERK1/2, and STAT3) together with an anti-ALK blot (ALK).
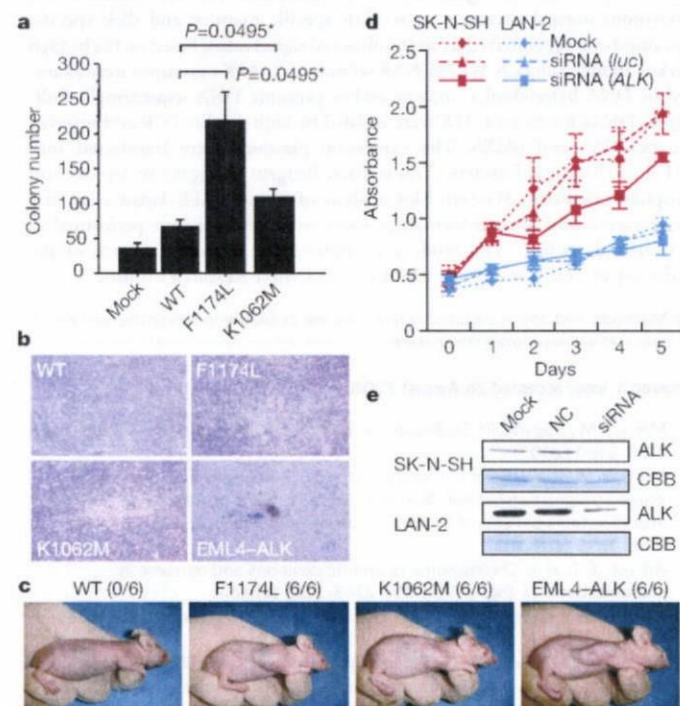

Figure 3 | Oncogenic role of ALK mutations. a, Colony assays for NIH3T3 cells stably expressing wild-type as well as mutant ALK (F1174L and K1062M). The average numbers of colonies in triplicate experiments are plotted and standard deviation is indicated. Results showing statistically significant differences as compared with experiments using wild-type ALK are indicated by asterisks with *P* values. b, c, NIH3T3 cells were transfected with wild-type and mutant ALK (F1174L, K1062M and EML4–ALK) and subjected to a focus forming assay (b) as well as an *in vivo* tumorigenicity assay in nude mice (c). d, Effect of RNAi-mediated ALK knockdown on cell proliferation in neuroblastoma cell lines expressing either the F1174L mutant (SK-N-SH) or wild-type ALK (LAN-2). Cell growth was measured using the Cell Counting Kit-8 after knockdown experiments using *ALK*-specific siRNAs (siRNA *ALK*), control siRNAs (siRNA *luc*), or mock experiments, where absorbance was measured in triplicate and averaged for each assay. To draw growth curves, the mean $\pm$ s.d. of the averaged absorbance in three independent knockdown experiments is plotted. e, Successful knockdown of ALK protein was confirmed by anti-ALK blots (ALK) using Coomassie brilliant blue G-250 (CBB) staining as loading controls. NC, control siRNA; siRNA, *ALK* siRNA.

Through the genome-wide analysis of genetic lesions in neuroblastoma, we identified novel oncogenic *ALK* mutations in advanced neuroblastoma. Combined with the cases having a high-grade amplification of the *ALK* gene, aberrant ALK signalling was likely to be involved in 11% (16 out of 151) of the advanced neuroblastoma cases. Because ALK kinase has been shown to be deregulated only in the form of a fusion kinase in human cancers, including lymphoma and lung cancer, the identification of oncogenic mutations in *ALK* not only increases our understanding of the molecular pathogenesis of advanced neuroblastoma, but also adds a new paradigm to the concept of 'ALK-positive human cancers' in that the mutated *ALK* kinases themselves might participate in human cancers. Our results again highlight the power of genome-wide studies to clarify the genetic lesions in human cancers[20-22]. Given that *ALK* mutations are preferentially involved in advanced neuroblastoma cases having a poor prognosis, our findings implicate that *ALK* inhibitors may improve the clinical outcome of children suffering from intractable neuroblastoma.

METHODS SUMMARY

Genomic DNA from 215 patients with primary neuroblastoma and 24 neuroblastoma-derived cell lines was analysed on GeneChip SNP genotyping microarrays (Affymetrix GeneChip 250K NspI). After appropriate normalization of mean array intensities, signal ratios were calculated between tumours and anonymous normal references in an allele-specific manner, and allele-specific copy numbers were inferred from the observed signal ratios based on the hidden Markov model using CNAG/AsCNAR software[13,14]. *ALK* mutations were examined by DNA heteroduplex analysis and/or genomic DNA sequencing[16]. Full-length cDNAs for mutant *ALK* were isolated by high-fidelity PCR and inserted into pcDNA3 and pMXS. The expression plasmids were transfected into NIH3T3 cells using Effectene Transfection Reagent (Qiagen) or by calcium phosphate methods[9]. Western blot analysis of mutant ALK kinases, *in vitro* kinase assays, and tumour formation assays in nude mice were performed as previously described[9]. This study was approved by the ethics boards of the University of Tokyo and of the Chiba Cancer Center Research Institute.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *Lancet* 369, 2106–2120 (2007).
2. Maris, J. M. *et al.* Loss of heterozygosity at 1p36 independently predicts for disease progression but not decreased overall survival probability in neuroblastoma patients: a Children's Cancer Group study. *J. Clin. Oncol.* 18, 1888–1899 (2000).
3. Attiyeh, E. F. *et al.* Chromosome 1p and 11q deletions and outcome in neuroblastoma. *N. Engl. J. Med.* 353, 2243–2253 (2005).
4. Bown, N. *et al.* Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma. *N. Engl. J. Med.* 340, 1954–1961 (1999).
5. Brodeur, G. M., Seeger, R. C., Schwab, M., Varmus, H. E. & Bishop, J. M. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* 224, 1121–1124 (1984).
6. Shiota, M. *et al.* Anaplastic large cell lymphomas expressing the novel chimeric protein p80NPM/ALK: a distinct clinicopathologic entity. *Blood* 86, 1954–1960 (1995).
7. Morris, S. W. *et al.* Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* 263, 1281–1284 (1994).
8. Fujimoto, J. *et al.* Characterization of the transforming activity of p80, a hyperphosphorylated protein in a Ki-1 lymphoma cell line with chromosomal translocation t(2;5). *Proc. Natl Acad. Sci. USA* 93, 4181–4186 (1996).
9. Soda, M. *et al.* Identification of the transforming *EML4-ALK* fusion gene in non-small-cell lung cancer. *Nature* 448, 561–566 (2007).
10. Rikova, K. *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* 131, 1190–1203 (2007).
11. Kennedy, G. C. *et al.* Large-scale genotyping of complex DNA. *Nature Biotechnol.* 21, 1233–1237 (2003).
12. Matsuzaki, H. *et al.* Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature Methods* 1, 109–111 (2004).
13. 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).
14. 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).
15. Osajima-Hakomori, Y. *et al.* Biological role of anaplastic lymphoma kinase in neuroblastoma. *Am. J. Pathol.* 167, 213–222 (2005).
16. Donohoe, E. Denaturing high-performance liquid chromatography using the WAVE DNA fragment analysis system. *Methods Mol. Med.* 108, 173–187 (2005).
17. Hu, J., Liu, J., Ghirlando, R., Saltiel, A. R. & Hubbard, S. R. Structural basis for recruitment of the adaptor protein APS to the activated insulin receptor. *Mol. Cell* 12, 1379–1389 (2003).
18. Donella-Deana, A. *et al.* Unique substrate specificity of anaplastic lymphoma kinase (ALK): development of phosphoacceptor peptides for the assay of ALK activity. *Biochemistry* 44, 8533–8542 (2005).
19. McDermott, U. *et al.* Genomic alterations of anaplastic lymphoma kinase may sensitize tumors to anaplastic lymphoma kinase inhibitors. *Cancer Res.* 68, 3389–3395 (2008).
20. Garraway, L. A. *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* 436, 117–122 (2005).
21. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446, 758–764 (2007).
22. Kawamata, N. *et al.* Molecular allelokaryotyping of pediatric acute lymphoblastic leukemia by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood* 111, 776–784 (2008).

Author Contributions Y.C., Y.L.C. and J.T. contributed equally to this work. M.K. and M.Sa. performed microarray experiments and subsequent data analyses. Y.C. and J.T. performed mutation analysis of *ALK*. Y.C., Y.L.C., J.T., M.So., L.W. and H.M. conducted functional assays of mutant ALK. A.N., M.O., T.I., A.K. and Y.H. prepared tumour specimens and were involved in statistical analysis. A.N., Y.H., H.M., J.T. and S.O. designed the overall study, and S.O. and J.T. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The nucleotide sequences of *ALK* mutations detected in this study have been deposited in GenBank under the accession numbers EU788003 (K1062M), EU788004 (T1087I), EU788005 (F1174L; TTC/TTA), EU788006 (F1174L; TTC/TTG), EU788007 (F1174C), EU788008 (F1174V), EU788009 (F1245L) and EU788010 (R1275Q). The copy number data as well as the raw microarray data will be accessible from http://www.ncbi.nlm.nih.gov/geo/ with the accession number GSE12494. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.O. (sogawa-tky@umin.net) or Y.H. (hayashiy-tky@umin.ac.jp).

METHODS

Specimens. Primary neuroblastoma specimens were obtained during surgery or biopsy from patients who were diagnosed with neuroblastoma and admitted to a number of hospitals in Japan. In total, 215 primary neuroblastoma specimens were subjected to SNP array analysis after informed consent was obtained from the parents of each patient. The patients were staged according to the International Neuroblastoma Staging System[23]. The clinicopathological findings are summarized in Supplementary Table 1. Twenty-four neuroblastoma-derived cell lines were also analysed by SNP array analysis (Supplementary Table 2). The SCMC-N2, SCMC-N4 and SCMC-N5 cell lines were established in our laboratory[24,25]. The SJNB series of cells and the UTP-N-1[26] cell line were gifts from A. T. Look and A. Inoue, respectively. The other cell lines used were obtained from the Japanese Cancer Resource Cell Bank (http://cellbank.nibio.go.jp/).

Microarray analysis. High-molecular-mass DNA was isolated from tumour specimens as well as from the peripheral blood or the bone marrow as described previously[24]. The DNA was subjected to SNP array analysis using Affymetrix GeneChip Mapping 50K and/or 250K arrays (Affymetrix) according to the manufacturer's suggested protocol. The scanned array images were processed with Gene Chip Operation software (GCOS)[15], followed by SNP calls using GTYE. Genome-wide copy number measurements and loss of heterozygosity detection were performed using CNAG/AsCNAR algorithms[14], which enabled an accurate determination of allele-specific copy numbers.

Confirmation of SNP array data. FISH and/or genomic PCR analysis confirmed the results of SNP array analyses as described previously[15]. PCR primer sets were designed to amplify several adjacent fragments inside and outside of the homozygously deleted regions in tumour samples.

Mutation analysis. Mutations in the *ALK* gene were examined in 239 neuroblastoma samples, including 24 cell lines, by denaturing high-performance liquid chromatography (DHPLC) using the WAVE system (Model 4500; Transgenomic) according to the manufacturer's suggested protocol[16]. The samples showing abnormal conformations were subjected to direct sequencing analysis using an ABI PRISM 3100 Genetic Analyser (Applied Biosystems). Using direct sequencing, mutation analysis of *MYCN* was also performed in seven cases with *ALK* alterations but not *MYCN* amplification. The primer sets used in this study are listed in Supplementary Table 5.

Transforming potential of *ALK* mutants. Total RNA was extracted from SJNB-1 (wild type), SCMC-N2 (F1174L) and SCMC-N5 (K1062M) cells as described previously[26]. First-strand cDNA was synthesized from RNA using Transcriptor Reverse Transcriptase and an oligo-(dT) primer (Roche Applied Science). The resulting cDNA was then amplified by PCR using the KOD-Plus-Ver.2 DNA polymerase (Toyobo) and the primers sense 5'-TCAGAAGCTTTACCAA-GGACTGTTCAGAGC-3' and antisense 5'-AATTGCGGCCGCTACTTGTCA-TCGTCGTCCTTGTAGTCGGGGCCCAGGCTG GTTCATGC-3', thereby introducing a HindIII site at the 5' terminus and a NotI site and a Flag sequence at the 3' terminus. The HindIII–NotI fragments of *ALK* cDNA were subcloned into pcDNA3 to generate expression plasmids. After resequencing to confirm that they had no other mutations, the *ALK* plasmids were used for transfection into NIH3T3 cells using Effectene Transfection Reagent (Qiagen) according to the suggested manufacturer's protocol. The transfected NIH3T3 cells were selected in 800 μg ml^{-1} G418 for 2 weeks to obtain stably expressing clones.

To evaluate the phosphorylation status of ALK mutants, the cell lysates of stable clones were immunoprecipitated with antibodies to Flag (Sigma) and the resulting precipitates were subjected to western blot analysis with the antibody specific to pTyr 1604 (Cell Signaling Technology) of ALK and the generic anti-phosphotyrosine antibody (PY20). The *in vitro* kinase activity of ALK mutants was measured using a non-radioactive isotope solid-phase enzyme-linked immunosorbent assay using the Universal Tyrosine Kinase Assay kit (Takara) according to the manufacturer's suggested protocol. We also performed the *in vitro* kinase assay with the synthetic YFF peptide (Operon Biotechnologies) as described previously[18]. For anchorage-independent growth analysis, 1×10^3 stably transfected NIH3T3 cells were mixed in 0.3% agarose with 10% FBS DMEM and plated on 0.6% agarose-coated 35-mm dishes. After culture for 28 days, the colonies of >0.1 mm in diameter were counted. The quantification of the colonies was from three independent experiments. To investigate the downstream signalling of ALK, western blot analysis was performed using the anti-ERK1/2, anti-phospho-ERK1/2, anti-AKT, anti-phospho-AKT, anti-STAT3 and anti-phospho-STAT3 antibodies (Cell Signaling Technology)[15].

The cDNA mutant of *ALK* was also inserted into the pMXS plasmid and the constructs were introduced into NIH3T3 cells by the calcium phosphate method as described previously[9]. The cells were then either cultured for 21 days or injected subcutaneously at six sites in three nude mice.

Inhibition of ALK through RNAi-mediated knockdown. To suppress the expression of the ALK protein, two different pairs of *ALK* siRNAs (*ALK* siRNA1 and *ALK* siRNA2) were obtained (Qiagen)[15]. The sequences were 5'-GAGUCUGGCAGUUGACUUCdTdT-3' for *ALK* siRNA1 and 5'-GCUCC-GGCGUGCCAAGCAGdTdT-3' for *ALK* siRNA2. A siRNA, targeting a sequence in firefly (*Photinus pyralis*) luciferase mRNA (*luc* siRNA), was used as a negative control (Qiagen)[15]. The sequences of *luc* siRNA were as follow: sense 5'-CGUACGCGGAAUACUUCGAdTdT-3' and antisense 5'-UCGAAGUAUU-CCGCGUACGdTdT-3'. Gene knockdown was achieved in SK-N-SH and LAN-2 cells using HiPerFect transfection reagent following the manufacturer's suggested instructions (Qiagen). To assess the effect of ALK knockdown on cell growth, these cells were seeded in 96-well plates at a concentration of 8.0×10^3 cells per well 24 h before transfection and assayed using the Cell Counting Kit-8 (Wako).

Statistical analysis. The significance of the correlation between *MYCN* amplification and *ALK* mutation was tested according to the conventional 2×2 contingency table using Fisher's exact test. The significance of the differences in kinase activity between wild-type and mutant ALK kinases was examined by the Mann–Whitney U-test based on the measured percentage activity of kinases in the precipitates of the corresponding samples. The significance of the differences in colony formation between wild-type and mutant ALK kinases was also examined by the Mann–Whitney U-test. The size of the hazards from possible risk factors, including International Neuroblastoma Staging System stages, *MYCN* status and *ALK* mutation/amplification were estimated by Cox regression analysis assuming a proportional hazard model using Stata software. Correlation between ploidy and clinical stage was tested by nptrend test.

23. Smith, E. I., Haase, G. M., Seeger, R. C. & Brodeur, G. M. A surgical perspective on the current staging in neuroblastoma—the International Neuroblastoma Staging System proposal. *J. Pediatr. Surg.* 24, 386–390 (1989).
24. Takita, J. *et al.* Allelotype of neuroblastoma. *Oncogene* 11, 1829–1834 (1995).
25. Takita, J. *et al.* Absent or reduced expression of the caspase 8 gene occurs frequently in neuroblastoma, but not commonly in Ewing sarcoma or rhabdomyosarcoma. *Med. Pediatr. Oncol.* 35, 541–543 (2000).
26. Takita, J. *et al.* Allelic imbalance on chromosome 2q and alterations of the caspase 8 gene in neuroblastoma. *Oncogene* 20, 4424–4432 (2001).

The methylation status of *RASSF1A* promoter predicts responsiveness to chemotherapy and eventual cure in hepatoblastoma patients

Shohei Honda[1,2], Masayuki Haruta[1], Waka Sugawara[1], Fumiaki Sasaki[3], Miki Ohira[3], Tadashi Matsunaga[3], Hiroaki Yamaoka[3], Hiroshi Horie[3], Naomi Ohnuma[3], Akira Nakagawara[3], Eiso Hiyama[3], Satoru Todo[2] and Yasuhiko Kaneko[1,3]*

[1]*Department of Cancer Diagnosis, Research Institute for Clinical Oncology, Saitama Cancer Center, Saitama, Japan*
[2]*Department of General Surgery, Hokkaido University Graduate School of Medicine, Sapporo, Japan*
[3]*Japanese Study Group for Pediatric Liver Tumor (JPLT), Hiroshima, Japan*

Despite the progress of therapy, outcomes of advanced hepatoblastoma patients who are refractory to standard preoperative chemotherapy remain unsatisfactory. To improve the mortality rate, novel prognostic markers are needed for better therapy planning. We examined the methylation status of 13 candidate tumor suppressor genes in 20 hepatoblastoma tumors by conventional methylation-specific PCR (MSP) and found hypermethylation in 3 of the 13 genes. We analyzed the methylation status of these 3 genes (*RASSF1A*, *SOCS1* and *CASP8*) in 97 tumors and found hypermethylation in 30.9, 33.0 and 15.5%, respectively. Univariate analysis showed that only the methylation status of *RASSF1A* but not the other 2 genes predicted the outcome, and multivariate analysis showed a weak contribution of *RASSF1A* methylation to overall survival. Using quantitative MSP, we found *RASSF1A* methylation in 44.3% of the 97 tumors. *CTNNB1* mutation was detected in 67.0% of the 97 tumors. While univariate analysis demonstrated *RASSF1A* methylation, *CTNNB1* mutation and other clinicopathological variables as prognostic factors, multivariate analysis identified *RASSF1A* methylation (*p* = 0.043; relative risk 9.39) and the disease stage (*p* = 0.002; relative risk 7.67) but not *CTNNB1* mutation as independent prognostic factors. In survival analysis of 33 patients in stage 3B or 4, patients with unmethylated tumor had better overall survival than those with methylated tumor (*p* = 0.035). *RASSF1A* methylation may be a promising molecular-genetic marker to predict the treatment outcome and may be used to stratify patients when clinical trials are carried out.
© 2008 Wiley-Liss, Inc.

Hepatoblastoma is a rare malignant neoplasm of the liver, with an incidence of 0.5–1.5 per million children.[1] Remarkable progress in clinical outcome has been achieved in the past 20 years due to advances in chemotherapy and surgical procedures; however, the mortality rate remains 20–30% and treatment results in patients in advanced stages who are refractory to standard preoperative chemotherapy regimens are unsatisfactory.[2,3] To improve the mortality of these patients, innovative treatment and potent prognostic markers for better therapy planning are needed. The present clinical factors predicting outcome include the level of alpha-feto protein, histology, disease stage and growth pattern of the tumor.[2–4] Chromosomal gains of 2q, 8q and 20 and high expression of telomerase or *PLK1* were shown to be molecular-genetic markers predicting poor outcome[5–8]; however, none have been proven to be independent prognostic factors by multivariate analysis.

We previously reported that *RASSF1A* (*RAS association domain family protein 1*) methylation, found in 39% of 39 hepatoblastoma tumors, was correlated with poor outcome by univariate analysis.[9] Nevertheless, the article had some limitations that the number of tumors was not enough, the method used to detect the hypermethylation was suboptimal, and the prognostic significance of *RASSF1A* methylation was ambiguous by multivariate analysis.

CTNNB1 (*catenin, beta-1*) mutation was reported in the majority of hepatoblastoma tumors, but reports on alterations of other oncogenes or tumor suppressor genes are rare.[10–12] Thus, we thought that epigenetic silencing of tumor suppressor genes might

be involved in the tumorigenesis of hepatoblastoma and examined the methylation status of 13 candidate tumor suppressor genes, whose aberrant methylation has previously been shown in various cancers.[13–22] Conventional methylation-specific PCR (MSP) analysis showed hypermethylation in only 3 of the 13 genes, *RASSF1A*, *SOCS1* (*suppressor of cytokine signaling 1*) and *CASP8* (*caspase-8*) genes, but not in the remaining 10 genes. We examined the correlation of the methylation status of the 3 genes with various clinical characteristics in a substantial number of hepatoblastoma tumors. Furthermore, we analyzed the methylation status of *RASSF1A* by more sensitive quantitative MSP and verified the prognostic implication of methylation by multivariate analysis. We suggest that *RASSF1A* may be a promising molecular-genetic marker predicting treatment outcome that may be used to stratify hepatoblastoma patients when clinical trials are carried out.

Material and methods

Patients and samples

Tumor tissues were obtained from 97 Japanese children with hepatoblastoma and adjacent normal liver tissues were available from 3 patients. Nonmatched normal liver tissues were also obtained from 5 other hepatoblastoma patients who were not included in the present clinicopathological study. Thirty-five of 39 specimens in the previous report were included; 4 were excluded because of the lack of DNA and 62 were supplied by the Tissue Bank of the Japanese Study Group for Pediatric Liver Tumor (JPLT).[23] The median age of the 97 patients at diagnosis was 16 months (range, 2–177 months).

The clinical stage of the disease was determined at the time of initial biopsy or surgery according to the classification of the Japanese Society of Pediatric Surgeons.[24] While most tumors in stages 1 and 2, and those in 3A, occupying 3 segments of the liver, are completely resectable, tumors in stage 3B, occupying 4 segments of the liver, and those in stage 4 are not. The extent of disease was distributed in stage 1 in 6 tumors, in 2 in 33, in 3A in 25, in 3B in 11 and in 4 in 22. Patients were treated at various hospitals or institutions, mostly under the framework of JPLT-1 (1991–1999) or JPLT-2 (2000–2006) protocols.[23,25] The protocols include pre- and postoperative chemotherapy with cisplatin and THP-adriamy-

cin. Complete response (CR) was defined as the complete disappearance of tumor, and partial response (PR) as at least a 50% reduction of tumor. No change (NC) was defined as a decrease of less than 50% or an increase of tumor.[25] Seventy-two patients underwent preoperative chemotherapy, and one underwent salvage liver transplantation. The median follow-up of survivors was 66 months (range, 9–175 months). The PRETEXT system is based on hepatic surgical anatomy, described elsewhere.[26] The pathological classifications of hepatoblastoma by Haas *et al.* and the Japanese Society of Pathology divide hepatoblastoma into 2 major subtypes, namely the well-differentiated (fetal) type and the poorly differentiated (embryonal) type.[4,24]

Bisulfite treatment and conventional methylation-specific PCR (MSP) analysis

Genomic DNA from tumor samples was treated with sodium bisulfite, and the methylation status of the promoter region in various genes was analyzed by MSP, as previously described.[9,27] The genes examined were *RASSF1A*, *RASSF2A*, *NORE1A*, *SOCS1*, *CASP8*, *RUNX3*, *RIZ1*, *BLU*, *HOXA9*, *HOXB5*, *p16INK4A*, *p14ARF* and *DCR2*.[13–22] The primer sequences and their location in the original genomic sequences are listed in Table I, and the location of the analyzed fragments for *RASSF1A*, *SOCS1* and *CASP8* are shown in Figure 1*a*. While the primer sequences of *RASSF1A* are located in the promoter region, those of *CASP8* and *SOCS1* are derived from the exon 4-intron 4 region and the exon 1, respectively, because the methylation status of these regions is correlated with the expression.[15,20,30] CpGgenome™ Universal Methylated DNA (Chemicon International, Temecula, CA) and normal lymphocyte DNA were used as controls for methylated or unmethylated templates, respectively. PCR products were run on 2% agarose gels and visualized after staining with ethidium bromide.

Quantitative MSP and reverse-transcription (RT)-PCR analyses of RASSF1A

The methylation status of the *RASSF1A* promoter was also examined in all 97 tumor samples by fluorescence-based, real-time quantitative PCR using a LightCycler (Roche Diagnostics). Primers and probes designed to specifically amplify the promoter of *RASSF1A* or a reference gene, *ACTB*, were described elsewhere.[28,29] The primer sequences used for quantitative MSP and those used for conventional MSP share the 17 nucleotides with 1 nucleotide deviation in the forward primer, and the 18 nucleotides with 3 nucleotides deviation in the reverse primer, although they amplified the same *RASSF1A* CpG islands (CGIs) (Fig. 1*a* and Table I).[31] Each amplification reaction included tumor DNA samples, positive and negative controls and water blank. *ACTB* was used as a reference gene to determine the relative level of methylated DNA for *RASSF1A* in each sample. Dividing the methylated *RASSF1A/ACTB* ratio of template amounts in a sample by the methylated *RASSF1A/ACTB* ratio of template amounts in a fully methylated control and multiplying this value by 100 calculated the percentage of methylation.

To determine whether the percentage of *RASSF1A* methylation is correlated with the expression level, we performed RT-PCR analysis of the *RASSF1A* gene in 7 tumor samples with methylated or unmethylated *RASSF1A* and 1 normal liver sample available by the method described previously.[9]

Mutation analysis of the CTNNB1 gene

To detect point mutations and deletions of the *CTNNB1* gene, genomic DNA from each tumor sample was amplified using 2 sets of primers, F1, 5′-TGGCTATCATTCTGCTTTTCTTG-3′ and R1, 5′-CTCTTTTCTTCACCACAACATTTT-3′, and BCAT-3, 5′-AA AATCCAGCGTGGACAATGG-3′, and BCAT-4, 5′-TGTGGCA AGTTCTGCATCATC-3′, respectively (Suppl Fig. 1*a*).[10,32] The PCR products were either directly sequenced or inserted into a

vector [pGEM (R)-T Easy Vector System (Promega, Madison, WI)], and 6 or more clones were sequenced.

Statistical analysis

Patients were grouped according to various biological and clinical aspects of disease. Significance of differences in the characteristics between patient's groups was examined using the chi-square or Fisher's exact test. Overall survival for each group of patients was estimated using the Kaplan-Meier method, and compared using the log-rank test. Time to failure was defined as the interval between surgery or preoperative chemotherapy and death from any cause. The influence of various biological and clinical factors on overall survival was estimated using the Cox proportional-hazards model calculated with Stat Flex software for Windows, version 5.0 (Artec Co., Osaka, Japan).

Results

Conventional MSP analysis of various genes in hepatoblastomas

We first examined the methylation status of 13 genes in 20 tumors, including 2 tumors in stage 1, 6 in stage 2, 6 in stage 3 and 6 in stage 4, by conventional MSP and found no methylation in 10 (*RASSF2A*, *NORE1A*, *RUNX3*, *RIZ1*, *BLU*, *HOXA9*, *HOXB5*, *p16INK4A*, *p14ARF* and *DCR2*); no further analysis was performed on these 10 genes. The remaining 3 genes, including *RASSF1A*, *SOCS1* and *CASP8*, were methylated in a substantial number of tumors. Therefore, we extended the analysis to all 97 tumors and found hypermethylation of *RASSF1A*, *SOCS1* and *CASP8* in 30 (30.9%), 32 (33.0%) and 15 (15.5%) tumors, respectively (Fig. 1*b*). All 3 genes were methylated in 3 tumors. Two of 3 genes, *RASSF1A* and *SOCS1*, *RASSF1A* and *CASP8* and *SOCS1* and *CASP8*, were methylated in 7, 3 and 5 tumors, respectively. Only 1 gene, *RASSF1A*, *SOCS1* or *RASSF1A*, was methylated in 15, 19 or 4 tumors. Conventional MSP detected unmethylated *RASSF1A* in all 8 adjacent normal liver tissues.

Correlation of the methylation status of the 3 genes analyzed by conventional MSP with overall survival

When we analyzed the correlation between the methylation status of any 1 of the 3 genes and overall survival, *RASSF1A* methylation was associated with a poor outcome ($p < 0.001$), but *SOCS1* or *CASP8* methylation was not; however, multivariate analysis using the various factors shown in Table III indicated the significant contribution of disease stage [$p < 0.001$; relative risk (RR) 9.44; 95% confidence interval (CI), 2.51–35.46], but no contribution of *RASSF1A* methylation to overall survival ($p = 0.149$; RR 2.38; 95% CI, 0.73–7.72).

Quantitative MSP analysis of RASSF1A methylation and the correlation between the percentage of the RASSF1A methylation and the expression or clinical outcome

To clarify whether *RASSF1A* methylation is an independent factor predicting outcome, we performed quantitative MSP analysis of *RASSF1A* in 97 tumors. Tumors were classified by the percentage of *RASSF1A* methylation, and about one half of tumors (46) had 0–2.5% of the methylation, and others distributed in various percentages of the methylation (Fig. 2*a*). RT-PCR detected *RASSF1A* expression in 1 normal liver sample and 2 tumor samples with less than 1% of the methylation, but did not detect the expression in tumors with more than 11% of the methylation; 2 tumors with the intermediate incidence of the methylation (4.2 or 4.8%) showed the ambiguous expression (Fig. 2*b*). Thus, there is an inverse relationship between the percentage of the *RASSF1A* methylation and the expression.

Next, we examined the dose-response relationships between the percentage of *RASSF1A* methylation and overall survival analyzed by the Kaplan-Meier method and adopted a cutoff value of 4.8%, which gave the smallest p-value ($p < 0.00001$). We also examined the dose-response relationships between the percentages of

TABLE I – PRIMER SEQUENCE, GENOMIC POSITION, MSP CONDITION AND PRODUCT SIZE

Primer name	Primer sequence	Genomic position[1]	Annealing temp. (°C)	Product size (bp)	Ref.
Quantitative MSP					
ACTB-F	5'-TGGTGATGGAGGAGGTTTAGTAAGT	−1596	60	133	28
ACTB-R	5'-AACCAATAAAACCTACTCCTCCCTTAA				
TaqMan probe	5'-6FAM-TGTGTTTGTTATTGTGTGTTGGGTGGTGGT-TAMRA-3'				
RASSF1A-F	5'-GGTTTTGCGAGAGCGCGT	−72	62	168	29
RASSF1A-R	5'-GCTAACAAACGCGAACCGAAC				
TaqMan probe	5'-6FAM-GGAGGCGTTGAAGTCGGGGTT-TAMRA-3'				
Conventional MSP					
RASSF1A-UF	5'-GGGGTTTTGTGAGAGTGTGTTTAG	−74	63	175	30
RASSF1A-UR	5'-TAAACACTAACAAACACAAACCAAAC				
RASSF1A-MF	5'-GGGTTTTGCGAGAGCGCG	−73	63	169	
RASSF1A-MR	5'-GCTAACAAACGCGAACCG				
BLU-UF	5'-TTGTTTGGATTTAGGTGTGAGTT	−73	58	160	18
BLU-UR	5'-CAAAAACAACAAACCCCAACA				
BLU-MF	5'-CGTTCGGATTTAGGCGCGAGTT	−72	68	158	
BLU-MR	5'-GAAAACGACGAACCCCGACGA				
CASP8-UF	5'-TAGGGGATTTGGAGATTGTGA	+308[2]	55	321	20
CASP8-UR	5'-CCATATATATCTACATTCAAAACAA				
CASP8-MF	5'-TAGGGGATTCGGAGATTGCGA	+308[2]	58	320	
CASP8-MR	5'-CGTATATCTACATTCGAAACGA				
DCR2-UF	5'-TTGGGGATAAAGTGTTTTGATT	+101	58	146	21
DCR2-UR	5'-AAACCAACAACAAAACCACA				
DCR2-MF	5'-GGGATAAAGCGTTTCGATC	+104	59	139	
DCR2-MR	5'-CGACAACAAAACCGCG				
HOXA9-UF	5'-TAATAGTGTGTGGAGTGATTTAT	−124	56	94	22
HOXA9-UR	5'-TAATAAATTACCAACACCCA				
HOXA9-MF	5'-GCGTTTGGTTCGTTCGGTTC	−61[3]	64	123	
HOXA9-MR	5'-CAATAAAAACGCGAACGCCG				
HOXB5-UF	5'-TGAATTGGTTTTAATGATTTTTGGATT	−217	53	117	19
HOXB5-UR	5'-TTAAAAAATCACATACTTTTATTAACCAATCA				
HOXB5-MF	5'-AATCGGTTTTAACGATTTTCGGATC	−215	53	113	
HOXB5-MR	5'-AAAAAATCACGTACTTTTATTAACCAATCG				
NORE1A-UF	5'-ATTTATATTTGTGTAGATGTTGTTTGGTAT	−176		214	14
NORE1A-UR	5'-ACTTTAACAACAACAACTTTAACAACTACA				
NORE1A-MF	5'-CGTCGTTTGGTACGGATTTTATTTTTTTCGGTTC	−159		202	
NORE1A-MR	5'-GACAACTTTAACAACGACGACTTTAACGACTACG				
p14ARF-UF	5'-GGAATAGGGGAGTGGGGAT	−388	60	144	22
p14ARF-UR	5'-AATAACAACCCAAAAACCAAACA				
p14ARF-MF	5'-GGAATAGGGGAGCGGGGAC	−388	60	144	
p14ARF-MR	5'-GATAACGACCCAAAAACCGAACG				
p16INK4A-UF	5'-TTATTAGAGGGTGGGGTGGATTGT	+133	63	151	27
p16INK4A -UR	5'-CAACCCCAAACCACAACCATAA				
p16INK4A -MF	5'-TTATTAGAGGGTGGGGCGGATCGC	+133	63	150	
p16INK4A -MR	5'-GACCCCGAACCGCGACCGTAA				
RASSF2A-UF	5'-GAAGGTGTTTTTATTTTATTTTTGG	+684	59	156	13
RASSF2A-UR	5'-AAAACCTACCTCTAAAAAATCCACC				
RASSF2A-MF	5'-GTTCGTCGTCGTTTTTTAGGCG	+798	60	109	
RASSF2A-MR	5'-AAAAACCAACGACCCCCGCG				
RIZ1-UF	5'-TGGTGGTTATTGGGTGATGGT	−4782	59	177	17
RIZ1-UR	5'-ACTATTTCACCAACCCCAAGA				
RIZ1-MF	5'-GTGGTGGTTATTGGGCGACGGC	−4781		176	
RIZ1-MR	5'-GCTATTTCGCCGACCCCGACG				
RUNX3-UF	5'-ATAATAGTGGTTGTTAGGGTGTTG	+298[3]	60	115	16
RUNX3-UR	5'-ACTTCTACTTTCCCACTTCTCACA				
RUNX3-MF	5'-ATAATAGCGGTCGTTAGGGCGTCG	+298[3]	60	115	
RUNX3-MR	5'-GCTTCTACTTTCCCGCTTCTCGCG				
SOCS1-UF	5'-TTATGAGTATTTGTGTGTATTTTTAGGTTGGTT	+1072	60	175	15
SOCS1-UR	5'-CACTAACAACACAACTCCTACAACAACCA				
SOCS1-MF	5'-TTCGCGTGTATTTTTAGGTCGGTC	+1081	60	160	
SOCS1-MR	5'-CGACACAACTCCTACAACGACCG				

UF, unmethylated forward primer; UR, unmethylated reverse primer; MF, methylated forward primer; MR, unmethylated reverse primer.
[1]The 5' position of the sense unmethylated or sense methylated primer sequences is numbered relative to the transcription start site of the gene concern.–[2]The number indicates the location relative to the transcription start site of *CASP8* transcript variant B (NM_033355.2).–[3]Designed for bottom strand.