Hybridization with SNP-typing arrays. Each DNA sample (250 ng) was digested with *Xba*I, ligated to Adaptor-Xba (Affymetrix), amplified by polymerase chain reaction (PCR), and subjected to hybridization with Mapping 50K Xba 240 arrays (Affymetrix). SNP genotyping calls were generated using GDAS software version 3.0 (Affymetrix) with a confidence score threshold of 0.05. Chromosome copy number and allele-specific copy number at each SNP site were calculated from the hybridization signal intensity for both CRC and paired normal mucosa specimens with the use of CNAG software (http://www.genome.umin.jp).[8] Only the CNAG data for autosomes were analyzed in the present study. Genotype-call data and original CEL files are available at the Gene Expression Omnibus website (http://www.ncbi.nlm.nih.gov/geo) under the accession number GSE11417, and CNAG output data are available upon request. We considered chromosome copy number changes or LOH data reliable only when contiguous SNP probes presented the same data.

Quantitative real-time PCR. RNA was isolated from the samples with the use of an RNeasy Mini column (Qiagen) and was used to synthesize cDNA with PowerScript reverse transcriptase (Clontech, Palo Alto, CA, USA). Portions of genomic DNA or cDNA were subjected to PCR with the QuantiTect SYBR Green PCR Kit (Qiagen). The amplification protocol comprised incubations at 94°C for 15 s, 60°C for 30 s, and 72°C for 60 s. Incorporation of the SYBR Green dye into the PCR products was monitored in real time with the ABI PRISM 7700 sequence detection system (Applied Biosystems, Foster City, CA, USA), thereby allowing determination of the threshold cycle (C_T) at which exponential amplification of products begins. To quantitate the genomic DNA, the C_T values for genomic DNA corresponding to the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene and the target regions were used to calculate the abundance of target regions relative to that of *GAPDH* DNA. The primer sequences used for PCR were: 5′-AGGACATTTGTAATCAGTATCTGTG-3′ and 5′-AGGGCAGTCAATAAGCTAAGGAA-3′ for period homolog 3 (*PER3*); 5′-CTCAACTTCCTTGAGCACCTCCTG-3′ and 5′-TACCTTGGACAGCTTGCTCTGTTG-3′ for invasion inhibitory protein 45 (*IIP45*); 5′-ACTGGTGCTCTCACTGTCCAAAAC-3′ and 5′-CGCAGAGTAGACATCCTGGGTAAA-3′ for FAT tumor suppressor homolog (*FAT*); 5′-AGCGAATGGAAGTTAAATTTTGG-3′ and 5′-TGCATCTGTCCTAACTCACTCCT-3′ for breast cancer cell 2 (*BRCC2*); 5′-AGAGACTGTATTGCAGGGTGAAGA-3′ and 5′-CTTTCCATTATATGTCCCGACTCC-3′ for v-maf musculoaponeurotic fibrosarcoma oncogene homolog K (*MAFK*); 5′-CTACTCTCTTGCCAGCATTTTCAC-3′ and 5′-ACCTAAGCCTTATCCACACCTCAC-3′ for protein tyrosine phosphatase, non-receptor type 1 (*PTPN1*); 5′-GCTCATAGCCCTGCCTTCCT-3′ and 5′-GGTCCCCAAAACGCACACTC-3′ for *CSMD1*; and 5′-CTGACCTGCCGTCTAGAAAAACCT-3′ and 5′-CAGGAAATGAGCTTGACAAAGTGG-3′ for *GAPDH*.

Similarly, the relative quantity of cDNA was calculated using the C_T value of PCR for each cDNA and that for the *GAPDH* cDNA. The primer sequences for reverse transcription (RT)-PCR were: 5′-CGGTTTTCTCACAACACATTAGCA-3′ and 5′-ACTGGAAGGTGGGAAATCAATAGG-3′ for *PER3* cDNA; 5′-CTGGAACTCAGGCAGCAGACAAG-3′ and 5′-GACTCCTGGGGGAGAACAGCATT-3′ for *IIP45* cDNA; 5′-GTGAGTAATCCGCGCTGTTCTTT-3′ and 5′-CAGTAGTTGGGACACTGGAAATGG-3′ for *FAT* cDNA; 5′-GACAGATTTCGCCCATTATTCAGG-3′ and 5′-TGTTTCTCTGCACAATTTGAACCA-3′ for *BRCC2* cDNA; 5′-GCCATATACCACTCTCCCTTCCAC-3′ and 5′-TGGAGTGTGCCTTGATTTCATACA-3′ for *CSMD1*; and 5′-GTCAGTGGTGGACCTGACCT-3′ and 5′-TGAGCTTGACAAAGTGGTCG-3′ for *GAPDH* cDNA. The primer sets for *MAFK* and *PTPN1* cDNA were the same ones used for genomic amplification of the corresponding genes.

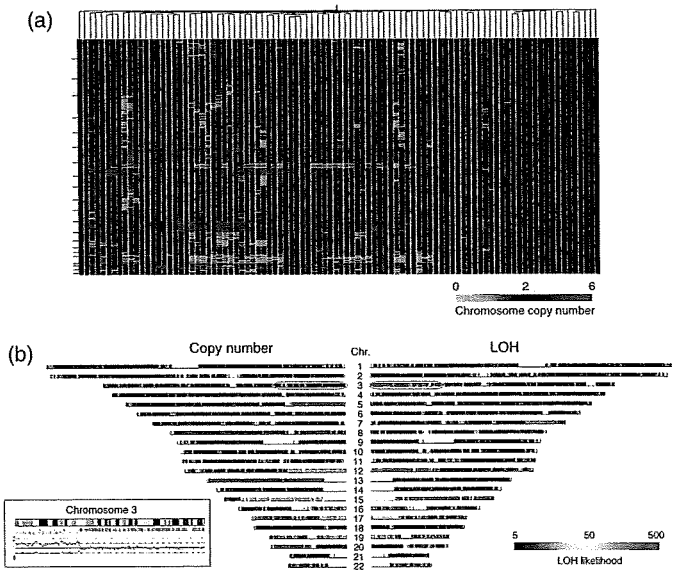

Fig. 1. Chromosomal copy number alterations and loss of heterozygosity (LOH) in the colorectal carcinoma genome. (a) The study subjects ($n = 94$) were subjected to a hierarchical clustering analysis based on the inferred copy number for all autosomal single nucleotide polymorphism (SNP) sites. Copy number is color coded according to the indicated scheme at the bottom. SNP sites are ordered by their physical position from top to bottom, and the borders between chromosomes are indicated by small bars at the left. (b) Chromosome copy number (left panel) and LOH likelihood score (right panel) are demonstrated for patient ID#002 in a chromosome view in a symmetrical manner. Copy number value is color coded as in (a), and LOH likelihood score is colored according to the scheme indicated at the bottom. Chromosome numbers are shown at the center. The allele-specific copy number data for the 3p region (indicated by a blue circle) is demonstrated in the inset as pink and green lines. Below the cytoband figure, the positions of SNP sites with a hetero- or discordant-call are indicated by green or pink bars, respectively.

Statistical analysis. Hierarchical clustering of the dataset and Student's *t*-test were carried out using GeneSpring 7.0 software (Agilent Technologies, Santa Clara, CA, USA), and survival analyses were carried out with SAS software (version 8.0.2: SAS Inc., Cary, NC, US) and the 'Survival' package in R version 2.6.0 (http://www.R-project.org.). The *q*-values for the false discovery rate were calculated directly from the ordered *P*-values above using the '*Q*-value' software (http://genomics.princeton.edu/storeylab/qvalue/) developed by Storey *et al.*[16] with parameters defined by Jones *et al.*[17]

Results

Frequent CAN. Genomic DNA was extracted from both CRC specimens and normal mucosa obtained from the same study subjects ($n = 94$). Both data were integrated into the CNAG software to infer chromosome copy number at every SNP site for each CRC sample. Incorporation of the data for paired normal mucosa markedly increased the accuracy of the calculation; the mean probe-signal intensity at diploid chromosomes in CRC was inferred from the data of control samples (where the majority of the chromosomes were expected to be diploid). Chromosome copy number data at each SNP probe site ($n = 57\,290$ for all autosomal SNP) was thus calculated for all CRC specimens, and a hierarchical clustering analysis for the study subjects was conducted based on the overall CNA profile. As shown in Figure 1a and Suppl. Fig. S1, approximately one-quarter of the subjects (the right side branch in the figure) had stable chromosomes, but the remaining samples had

doi: 10.1111/j.1349-7006.2008.00881.x

Table 1. Frequent regions of chromosomal copy number alterations or loss of heterozygosity (LOH) in colorectal carcinoma patients

Change	Chromosome	Nucleotide position	Mapped RefSeq gene	GenBank accession no.
Gain (chromosome copy no. ≥5 in ≥15 subjects)				
	6	16,176,003–16,176,549	None	
	8	70,887,465–71,089,425	*SLCO5A1*	NM_030958.1
	20	31,768,314–31,919,527	*PXMP4*	NM_007238.4
			ZNF341	NM_032819.3
			CHMP4B	NM_176812.3
Decrease (chromosome copy no. ≤1 in ≥35 subjects)				
	18	60,114,744–61,522,755	None	
	18	64,600,350–65,380,261	*CCDC102B*	NM_024781.1
			DOK6	NM_152721.2
	18	67,791,010–68,366,009	*CBLN2*	NM_182511.2
Homozygous deletion (common in two subjects)				
	3	60,393,402–60,490,818	*FHIT*	NM_002012.1
	20	14,796,659–15,040,864	*C20orf133*	NM_080676.5
LOH (common in ≥55 subjects)				
	5	108,765,615–112,484,272	*APC*	NM_000038.3
	17	5,265,130–8,883,455	*TP53*	NM_000546.3
			XAF1	NM_017523.2
			DVL2	NM_004422.2
	17	11,076,427–12,490,201	Others	

frequent CNA of various sizes. For instance, gross amplification was found commonly in chromosomes 7, 8q, 13, and 20, whereas large deletions of chromosomes were identified in 8p and 18.

Further, in-depth analysis of the dataset identified amplifications of various magnitudes at various frequencies. For instance, a high-grade amplification of the genome (copy number of five or greater) was found at three different loci in the genome of ≥15 subjects (Table 1), the size of which ranged from 547 to 201 961 bp. Surprisingly, amplification of one of these loci at chromosome 8q was found among as many as 25 patients (the most common, highly amplified region in our dataset). As expected, low-grade amplifications of the genome were found more commonly; a region of ~2.7 Mbp at chromosome 20q was, for example, amplified to four or more copies in more than 30 subjects, and this grade of amplification was also identified at many loci throughout the genome. For instance, genome regions with a copy number of four or greater in ≥10% of the patients were mapped to choromosmes 7p, 8q, 13, 20q, and others, comprising a total of 1921 SNP sites (3.4% of all sites).

Similarly, a decrease in chromosome copy number ($n \leq 1$) was also frequently identified throughout the genome; three distinct loci had such decreases in ≥35 subjects (Table 1). Further, a less-frequent decrease (found in ≥10% of patients) was mapped to chromosomes 1p, 5q, 8p, 14q, 17p, and others, comprising 3899 SNP sites in total (6.8% of all sites).

In our dataset, common homozygous deletions were unexpectedly rare. Only two loci demonstrated a chromosome copy number of zero in two individuals (Table 1). Interestingly, one such loci on chromosome 3 is known to be a common fragile region containing the fragile histidine triad gene (*FHIT*, GenBank accession no. NM_002012.1), a putative tumor suppressor.[18] The other homozygous deletion site at chromosome 20 spans 244 206 bp containing only one unknown gene, *C20orf133* (GenBank accession no. NM_080676.5).

Frequent LOH. With the SNP-typing array platform, we can carry out SNP genotyping by comparing the signal intensity between two alleles, which reflects the DNA amount of each allele. In the present study, with a moving window for 21 contiguous SNP, allele-specific copy number decreases were examined to identify LOH regions. Three most common LOH loci (found in 55 cases) were thus mapped to chromosomes 5 and 17 (Table 1). Other frequent LOH (found in ≥20% of patients) were identified on chromosomes 1p, 4q, 5q, 8p, 11q, 14q, 15q, 17p, 18, and 22. Less-frequent LOH were seen in two large loci (chromosomes 10 and 16), which contain a tumor necrosis factor receptor superfamily member (*FAS*, GenBank accession no. NM_000043.3) and ataxin-2 binding protein 1 (*A2BP1*, GenBank accession no. NM_018723.2).

Genome regions with UPD may contain tumor-suppressor genes (where both alleles carry a mutated, inactivated tumor-suppressor gene) or oncogenes (where cancer cells have two copies of a mutated and activated oncogene). CRC may have UPD at specific loci as demonstrated by Andersen *et al.*[19] In our dataset, we readily identified UPD regions that were characterized by a chromosome copy number of two and an LOH likelihood score of ≥50 defined by the CNAG software. In the data for patient ID# 002, for instance, a very high LOH likelihood score was inferred on chromosomes 3p and 15q (right panel of Fig. 1b). Although the latter region of the genome had a decreased copy number (left panel), the former was supposed to be diploid, indicating the presence of UPD. SNP array-based analysis can measure in detail changes in copy number in an allele-specific manner. With such analysis, as shown in the inset in Figure 1b, one allele at 3p was indeed amplified to a copy number of two (pink line), but the other allele was deleted (green line) in the same region, thus confirming the presence of UPD. Similar UPD was also identified on chromosomes 5q, 8p, 11, 14, 15, 17p, and 18q in our dataset.

Verification of the CNA data. The inferred copy number of chromosomes was then verified by quantitative real-time PCR. First, four genes (*BRCC2*, GenBank accession no. NM_001001786.1; *FAT*, GenBank accession no. NM_005245.3; *IIP45*, GenBank accession no. NM_021933.2; and *PER3*, GenBank accession no. NM_016831.1) mapped to independent loci with a frequent copy number loss were chosen to measure DNA quantity. The amount of DNA of each gene relative to that of *GAPDH* was examined in the patients with an inferred copy number of two and those with a copy number of one. As shown in the left panel of Figure 2a, the relative DNA amount of each gene was decreased to 0.7-0.5 in the patients with the copy number loss; 0.65 ± 0.52 (mean ± SD), 0.56 ± 0.26, 0.65 ± 0.22, and 0.73 ± 0.33 for *BRCC2*, *FAT*, *IIP45*, and *PER3*, respectively. The correlation coefficients between inferred copy number by

Table 2. **Prognosis-related regions of copy number alterations (CNA) or loss of heterozygosity (LOH) in colorectal carcinoma patients**

Change	Chromosome	Position	Size (Mbp)	*P*-value	*q*-value	Mapped RefSeq gene	GenBank accession no.
CNA	5	113,733,368–117,078,267	3.34	<0.001	0.095	*SEMA6A* and others	NM_20796.3 and others
	5	121,427,436–122,773,632	1.35	<0.001	0.095	*LOX* and others	NM_002317.3 and others
	5	123,233,993–126,057,451	2.82	<0.0005	0.095	Others	
	5	142,509,574–142,681,049	0.17	<0.001	0.095	Others	
	5	160,137,590–160,786,796	0.65	<0.001	0.098	Others	
	5	162,659,919–162,863,325	0.20	<0.0005	0.095	*CCNG1* and others	NM_004060.3 and others
	6	109,126,299–109,435,125	0.31	<0.0005	0.095	*SESN1* and others	NM_014454.1 and others
	10	110,674,541–111,338,259	0.66	<0.0005	0.095	None	
	18	51,345,876–52,332,836	0.99	<0.0005	0.095	*TCF4* and others	NM_003199.2 and others
	18	53,401,262–55,536,937	2.14	<0.0005	0.095	*RAX* and others	NM_013435.2 and others
LOH	16	4,858,366–6,679,934	1.82	<0.0001	0.046–0.224	*UBN1* and others	NM_002705.4 and others
	16	7,010,644–7,608,397	0.60	<0.0005	0.181	*A2BP1*	NM_018723.2

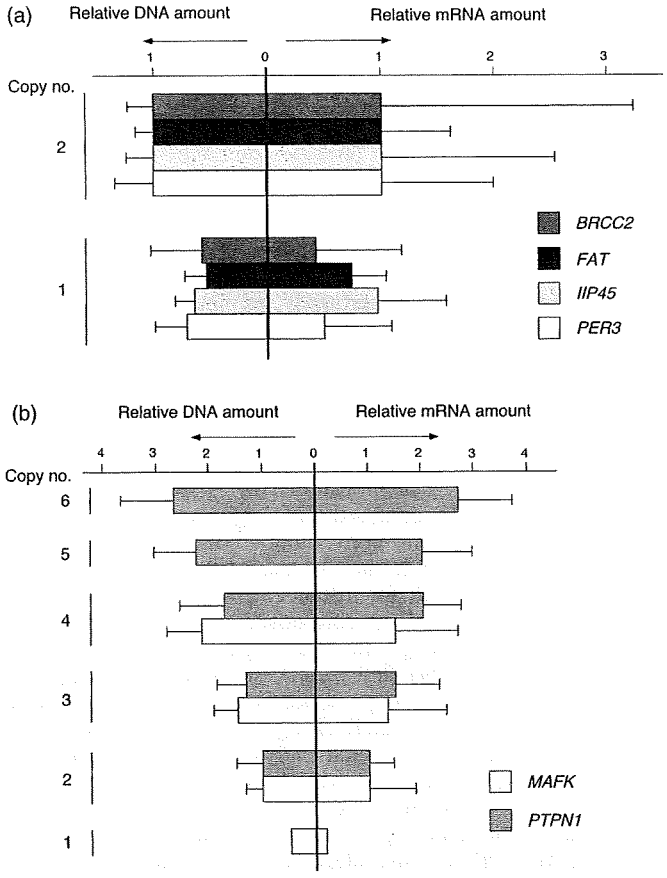

Fig. 2. Verification of copy number changes. (a) The DNA quantities of *PER3*, *IIP34*, *FAT*, and *BRCC2* (relative to that of *GAPDH*) were measured by real-time polymerase chain reaction in the subjects with inferred copy number two ($n = 2$) or one ($n = 1$). The mean + SD value for each gene was normalized to the corresponding mean value for the group with diploid chromosomes, and is shown in the left panel. The mRNA amount for each gene (relative to that of GAPDH) was also quantitated by real-time reverse transcription–polymerase chain reaction and is shown in a similar way. (b) The relative DNA (left panel) or mRNA (right panel) of *MAFK* and *PTPN1* was calculated as in (a).

array hybridization and DNA quantification by PCR were 0.219, 0.383, 0.216, and 0.314, respectively.

For the same gene set, we also examined how copy number changes affect mRNA level. Quantitative real-time reverse transcription-PCR was used to quantify the relative amount of each mRNA to that of GAPDH (right panel of Fig. 2a). Similar to the chromosome copy number, the mean mRNA level in the samples with copy number loss was decreased compared to the level in those without the loss; 0.47 ± 075, 0.80 ± 0.29, 0.97 ± 0.64, and 0.54 ± 0.69 for *BRCC2*, *FAT*, *IIP45*, and *PER3*, respectively. However, a relatively large SD in each mRNA amount indicates that transcriptional level was also influenced significantly by other factors such as epigenetic regulation and transcriptional factors.

We also measured the DNA amount of two genes (*MAFK*, GenBank accession no. NM_002360.3, and *PTPN*, GenBank accession no. NM_002827.2) that showed various levels of copy number amplification in our dataset. As shown in the left panel of Figure 2b, the calculated DNA amount of *MAFK* relative to that of *GAPDH* by quantitative PCR paralleled the copy number inferred from SNP arrays. Similarly, DNA quantity measured by real-time PCR for *PTPN1* generally followed the inferred copy number ($n = 2$–6). Again, the mRNA from each gene was quantified by real-time reverse transcription-PCR, revealing that amount of DNA significantly affects mRNA level (right panel).

Prognosis-related CNA and LOH. To directly search for CNA and LOH linked to the survival of patients, we utilized Cox's proportional-hazard regression analysis[20] coupled with the false-discovery rate correction on the copy number profile of all autosomal SNP sites.

From the CNA dataset, several loci at chromosomes 5, 6, 10, and 18 were proved to be significantly related to prognosis ($P < 0.001$ and $q < 0.1$) (Table 2). For all loci except one at chromosome 10, chromosomal loss had a negative impact on the outcome of the patients (Fig. 3a). One locus at chromosome 5 contains the cyclin G1 gene (*CCNG1*, GenBank accession no. NM_004060), which belongs to the cyclin gene superfamily. In contrast to the other cyclins, expression of *CCNG1* is stable throughout the cell cycle, and becomes activated in mouse cells by exposure to ionizing radiation, which also induces cell cycle arrest.[21] Further, disruption of WT1 function is linked to the downregulation of *CCNG1* expression.[22] These data together indicate a pro-apoptotic role for *CCNG1*, and our discovery of a relationship between loss of *CCNG1* and poor prognosis may imply a function of *CCNG1* as a tumor suppressor in CRC.

In addition to CNA analysis, we further searched for prognosis-related LOH with the following approach. There were many recurrent LOH regions in our dataset at various frequencies. We thus examined whether some of those recurrent alterations (observed in five or more samples) were preferentially present in the patients who died of CRC compared to those who survived in our observation period. For these potentially outcome-related genomic regions, prognosis was compared statistically between the two subject groups by the log-rank test with the false-discovery

doi: 10.1111/j.1349-7006.2008.00881.x

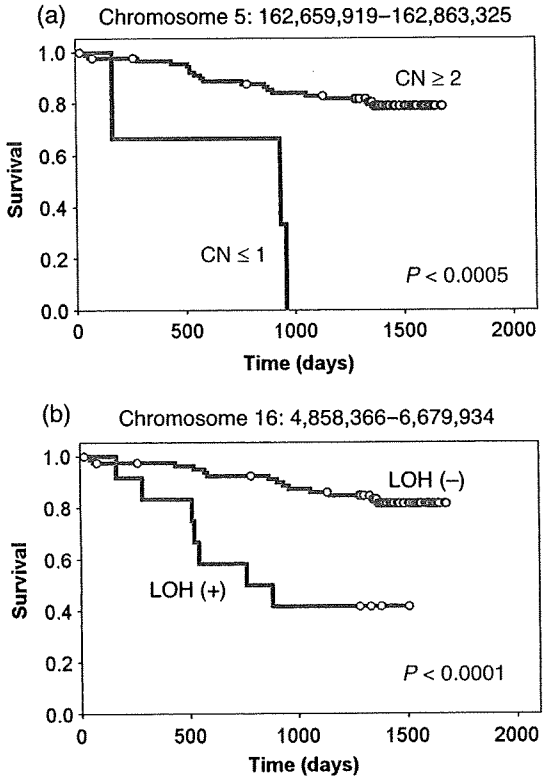

Fig. 3. Prognosis-related copy number (CN) loss and loss of heterozygosity (LOH). The survival of the subjects with or without copy number loss of a locus at (a) chromosome 5 or (b) chromosome 16 was compared using Kaplan–Meier analysis. The *P*-value for each comparison was calculated using the log rank test.

rate correction. We finally isolated two loci of LOH where the presence of LOH was related to a short survival time ($P < 0.0005$) (Table 2; Fig. 3b). One such prognosis-related LOH locus contains the ubinuclein 1 gene (*UBN1*, GenBank accession no. NM_016936). Because *UBN1* associates physically with *AP1* and interferes with its DNA-binding activity,[23] *UBN1* may also function to suppress tumor development.

Discussion

We have here calculated chromosome copy number as well as LOH likelihood throughout the genome of 94 CRC specimens. Together with the clinical information for the study subjects, we identified many loci whose DNA quantity or LOH is associated with the survival and various characteristics of CRC subjects. Some of the RefSeq genes mapped on such loci are well-known cancer-related genes. One frequent LOH was mapped to a genomic region of approximately 235 kb only containing the *MCC* gene, which had already been shown to be prone to somatic mutations and deletions in CRC and other cancers.[24,25] Overexpression of *MCC* suppresses the G_1 to S transition of the cell cycle, whereas such activity is lost for an *MCC* mutant identified in CRC,[26] supporting the tumor-suppressor activity of *MCC*.

In addition to the analysis presented in the present manuscript, our large dataset can also be utilized to characterize other aspects of CRC. CRC may be subdivided into microsatellite-stable cancer and MSI-high cancer. Comparison of our copy number data between the two subgroups has identified a locus of only 56 kb long, the copy number of which was statistically different between the subgroups ($P < 0.001$). This region contains only one RefSeq gene, ribosomal protein S6 kinase 90-kDa 5 (*RPS6KA5*, GenBank accession no. NM_004755.2), discovering another unexpected linkage between MSI and mitogen-activated protein kinase (MAPK) functions. A similar comparison of our data between the CRC with or without lymph node metastasis has identified two distinct loci in the genome (Suppl. Table S2). Also, a narrow genomic region was identified, the LOH of which is linked to the presence of liver metastasis ($P < 0.001$). However, that locus does not contain any RefSeq genes. Given the high resolution of SNP-typing arrays for CNA and LOH analysis, many genomic regions identified in this manuscript are <100 kb and contain only a few RefSeq genes per locus (Tables 1,2). Thus, our analysis is highly useful in narrowing down the list of genes associated with various characteristics of CRC.

Copy number alterations of CRC specimens have been studied with bacterial artificial chromosome array-based CGH,[13,27,28] and large segmental changes of chromosomes in such reports and publicly available databases match well with those identified in our study (see, for example, http://www.cghtmd.jp/CGHDatabase/tumor?lang=en). Although SNP-typing array-based CNA and LOH analyses have been reported recently for CRC, information for genes involved directly in such CNA and LOH is scarce.[11–14] Lips *et al.* examined the LOH status of paraffin-embedded CRC specimens ($n = 4$) and found recurrent LOH at chromosomes 5q, 17p, 18, and 20,[14] the former three of which were indeed identified in our study. However, Gaasenbeek identified LOH at the *TP53* locus in MSI-positive CRC.[13] In our cohort, however, there was only one MSI-positive case among 55 cases with LOH at *TP53*, whereas four were positive for MSI among 39 individuals without LOH at the locus, indicating no significant linkage between MSI and LOH at *TP53* (Fisher's exact test, $P = 0.186$).

It should, however, be noted that the RefSeq genes may not be the sole players in carcinogenesis. Long non-coding RNA is known to be involved in methylation of the genome,[29] and short non-coding RNA such as microRNA may be involved directly in cell growth and differentiation.[30] These transcripts, despite their inability to synthesize proteins, may thus contribute to the characteristics of CRC. As the discovery and annotation of these non-coding RNAs is still in its infancy,[31,32] many loci identified through our analysis may contain yet-undiscovered non-coding RNA, and these transcripts, not protein-coding mRNA, may play an important role in carcinogenesis as well. Indeed, one of the loci linked to lymph node metastasis has no RefSeq genes but only one non-coding RNA (Suppl. Table S2).

Our analysis provides a large-scale, accurate CNA and LOH dataset together with detailed information of clinical characteristics (including survival information in Suppl. Table S1) for the subjects. These data may become a framework for further analysis on structural alterations of the cancer genome in CRC.

Acknowledgments

The present study was supported in part by a Grant-in-Aid for Third-Term Comprehensive Control Research for Cancer from the Ministry of Health, Labor, and Welfare of Japan, and by a grant for 'High-Tech Research Center' Project for Private Universities: Matching Fund Subsidy, from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (2002-06) to HM.

References

1 Jemal A, Siegel R, Ward E *et al.* Cancer statistics, 2006. *CA Cancer J Clin* 2006; **56**: 106–30.

2 Portier G, Elias D, Bouche O *et al.* Multicenter randomized trial of adjuvant fluorouracil and folinic acid compared with surgery alone after resection of colorectal liver metastases: FFCD ACHBTH AURC 9002 trial. *J Clin Oncol* 2006; **24**: 4976–82.

3 Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990; **61**: 759–67.
4 Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* 1998; **396**: 643–9.
5 Kralovics R, Passamonti F, Buser AS *et al.* A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* 2005; **352**: 1779–90.
6 Jiang JK, Chen YJ, Lin CH, Yu IT, Lin JK. Genetic changes and clonality relationship between primary colorectal cancers and their pulmonary metastases – an analysis by comparative genomic hybridization. *Genes Chromosomes Cancer* 2005; **43**: 25–36.
7 Kleivi K, Teixeira MR, Eknaes M *et al.* Genome signatures of colon carcinoma cell lines. *Cancer Genet Cytogenet* 2004; **155**: 119–31.
8 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**: 6071–9.
9 Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* 2004; **20**: 1233–40.
10 Redon R, Ishikawa S, Fitch KR *et al.* Global variation in copy number in the human genome. *Nature* 2006; **444**: 444–54.
11 Andersen CL, Wiuf C, Kruhoffer M, Korsgaard M, Laurberg S, Orntoft TF. Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis* 2007; **28**: 38–48.
12 Tsafrir D, Bacolod M, Selvanayagam Z *et al.* Relationship of gene expression and chromosomal abnormalities in colorectal cancer. *Cancer Res* 2006; **66**: 2129–37.
13 Gaasenbeek M, Howarth K, Rowan AJ *et al.* Combined array-comparative genomic hybridization and single-nucleotide polymorphism–loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers. *Cancer Res* 2006; **66**: 3471–9.
14 Lips EH, Dierssen JW, van Eijk R *et al.* Reliable high-throughput genotyping and loss-of-heterozygosity detection in formalin-fixed, paraffin-embedded tumors using single nucleotide polymorphism arrays. *Cancer Res* 2005; **65**: 10 188–91.
15 Miyakura Y, Sugano K, Konishi F *et al.* Extensive methylation of *hMLH1* promoter region predominates in proximal colon cancer with microsatellite instability. *Gastroenterology* 2001; **121**: 1300–9.
16 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003; **100**: 9440–5.
17 Jones HE, Ohlssen DI, Spiegelhalter DJ. Use of the false discovery rate when comparing multiple health care providers. *J Clin Epidemiol* 2008; **61**: 232–40.
18 Zanesi N, Fidanza V, Fong LY *et al.* The tumor spectrum in FHIT-deficient mice. *Proc Natl Acad Sci USA* 2001; **98**: 10 250–5.
19 Andersen CL, Wiuf C, Kruhoffer M, Korsgaard M, Laurberg S, Orntoft TF. Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis* 2006; **28**: 38–48.
20 Cox DR. Regression models and life tables. *J R Stat Soc* 1972; **34**: 187–220.
21 Sugihara T, Magae J, Wadhwa R *et al.* Dose and dose-rate effects of low-dose ionizing radiation on activation of Trp53 in immortalized murine cells. *Radiat Res* 2004; **162**: 296–307.
22 Wagner KJ, Patek CE, Miles C, Christie S, Brookes AJ, Hooper ML. Truncation of WT1 results in downregulation of cyclin G1 and IGFBP-4 expression. *Biochem Biophys Res Commun* 2001; **287**: 977–82.
23 Aho S, Buisson M, Pajunen T *et al.* Ubinuclein, a novel nuclear protein interacting with cellular and viral transcription factors. *J Cell Biol* 2000; **148**: 1165–76.
24 Kinzler KW, Nilbert MC, Vogelstein B *et al.* Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. *Science* 1991; **251**: 1366–70.
25 Cawkwell L, Lewis FA, Quirke P. Frequency of allele loss of DCC, p53, RBI, WT1, NF1, NM23 and APC/MCC in colorectal cancer assayed by fluorescent multiplex polymerase chain reaction. *Br J Cancer* 1994; **70**: 813–18.
26 Matsumine A, Senda T, Baeg GH *et al.* MCC, a cytoplasmic protein that blocks cell cycle progression from the G_0/G_1 to S phase. *J Biol Chem* 1996; **271**: 10 341–6.
27 Fijneman RJ, Carvalho B, Postma C, Mongera S, van Hinsbergh VW, Meijer GA. Loss of 1p36, gain of 8q24, and loss of 9q34 are associated with stroma percentage of colorectal cancer. *Cancer Lett* 2007; **258**: 223–9.
28 Jones AM, Douglas EJ, Halford SE *et al.* Array-CGH analysis of microsatellite-stable, near-diploid bowel cancers and comparison with other types of colorectal carcinoma. *Oncogene* 2005; **24**: 118–29.
29 Chang SC, Tucker T, Thorogood NP, Brown CJ. Mechanisms of X-chromosome inactivation. *Front Biosci* 2006; **11**: 852–66.
30 Carrington JC, Ambros V. Role of microRNAs in plant and animal development. *Science* 2003; **301**: 336–8.
31 Carninci P, Kasukawa T, Katayama S *et al.* The transcriptional landscape of the mammalian genome. *Science* 2005; **309**: 1559–63.
32 Takada S, Berezikov E, Yamashita Y *et al.* Mouse microRNA profiles determined with a new and sensitive cloning method. *Nucleic Acids Res* 2006; **34**: e115.

Supporting Information

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

Fig. S1. Hierarchical clustering tree in Figure 1a is demonstrated with subject ID indicated at the bottom.

Table S1. Clinical characteristics of the study subjects

Table S2. Chromosomal copy number alterations (CNA) and loss of heterozygosity (LOH) related to clinical characteristics of colorectal carcinoma

Please note: Blackwell Publishing are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

doi: 10.1111/j.1349-7006.2008.00881.x

Identification of Novel Isoforms of the *EML4-ALK* Transforming Gene in Non–Small Cell Lung Cancer

Young Lim Choi,[1] Kengo Takeuchi,[3] Manabu Soda,[1,2] Kentaro Inamura,[3] Yuki Togashi,[3] Satoko Hatano,[3] Munehiro Enomoto,[1,2] Toru Hamada,[1] Hidenori Haruta,[1] Hideki Watanabe,[1] Kentaro Kurashina,[1] Hisashi Hatanaka,[1] Toshihide Ueno,[1] Shuji Takada,[1] Yoshihiro Yamashita,[1] Yukihiko Sugiyama,[2] Yuichi Ishikawa,[3] and Hiroyuki Mano[1,4]

Divisions of [1]Functional Genomics and [2]Pulmonary Medicine, Jichi Medical University, Tochigi, Japan; [3]Department of Pathology, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan; and [4]CREST, Japan Science and Technology Agency, Saitama, Japan

Abstract

The genome of a subset of non–small-cell lung cancers (NSCLC) harbors a small inversion within chromosome 2 that gives rise to a transforming fusion gene, *EML4-ALK*, which encodes an activated protein tyrosine kinase. Although breakpoints within *EML4* have been identified in introns 13 and 20, giving rise to variants 1 and 2, respectively, of *EML4-ALK*, it has remained unclear whether other isoforms of the fusion gene are present in NSCLC cells. We have now screened NSCLC specimens for other in-frame fusion cDNAs that contain both *EML4* and *ALK* sequences. Two slightly different fusion cDNAs in which exon 6 of *EML4* was joined to exon 20 of *ALK* were each identified in two individuals of the cohort. Whereas one cDNA contained only exons 1 to 6 of *EML4* (variant 3a), the other also contained an additional 33-bp sequence derived from intron 6 of *EML4* (variant 3b). The protein encoded by the latter cDNA thus contained an insertion of 11 amino acids between the EML4 and ALK sequences of that encoded by the former. Both variants 3a and 3b of EML4-ALK exhibited marked transforming activity *in vitro* as well as oncogenic activity *in vivo*. A lung cancer cell line expressing endogenous variant 3 of EML4-ALK underwent cell death on exposure to a specific inhibitor of ALK catalytic activity. These data increase the frequency of *EML4-ALK*–positive NSCLC tumors and bolster the clinical relevance of this oncogenic kinase. [Cancer Res 2008;68(13):4971–6]

Introduction

Lung cancer is the leading cause of cancer deaths in the United States, with >160,000 individuals dying of this condition in 2006 (1). The efficacy of conventional chemotherapeutic regimens with regard to improving clinical outcome in lung cancer patients is limited. Activating mutations within the epidermal growth factor receptor gene (*EGFR*) have been identified in non–small-cell lung cancer (NSCLC), the major subtype of lung cancer (2, 3), and chemical inhibitors of the kinase activity of EGFR have been found to be effective in the treatment of a subset of NSCLC patients harboring such mutations. However, these somatic mutations of *EGFR* are prevalent only among young women, nonsmokers, and Asian populations (3, 4).

We recently identified a novel transforming fusion gene, *EML4* (echinoderm microtubule–associated protein–like 4)-*ALK* (anaplastic lymphoma kinase), in a clinical specimen of lung adenocarcinoma from a 62-year-old male smoker (5). This fusion gene was formed as the result of a small inversion within the short arm of chromosome 2 that joined intron 13 of *EML4* to intron 19 of *ALK* (transcript ID ENST00000389048 in the Ensembl database[5]). The EML4-ALK protein thus contained the amino-terminal half of EML4 and the intracellular catalytic domain of ALK. Replacement of the extracellular and transmembrane domains of ALK with this region of EML4 results in constitutive dimerization of the kinase domain of ALK and a consequent increase in its catalytic activity (5).

Whereas this *EML4-ALK* fusion gene was detected in 3 of 75 individuals with NSCLC, we further identified another isoform of *EML4-ALK* in two patients of the same cohort (5). In these two individuals, intron 20 of *EML4* was disrupted and joined to intron 19 of *ALK*, with the fusion protein thus consisting of the amino-terminal two thirds of EML4 and the intracellular domain of ALK. This larger version of EML4-ALK was referred to as variant 2, with the original smaller version being termed variant 1. A total of 5 of the 75 (6.7%) patients in the cohort were thus positive for *EML4-ALK*.

Given that detection of *EML4-ALK* cDNA by the PCR would be expected to provide a highly sensitive means for diagnosis of lung cancer, and given that inhibition of the catalytic activity of EML4-ALK may be an effective approach to treatment of this disorder, we have examined whether other isoforms of EML4-ALK are associated with NSCLC. We now describe a third isoform of EML4-ALK (variant 3) that is smaller than variants 1 and 2.

Materials and Methods

PCR. This study was approved by the ethics committees of Jichi Medical University and The Cancer Institute of the Japanese Foundation for Cancer Research. Total cDNA of NSCLC specimens was synthesized with PowerScript reverse transcriptase (Clontech) and an oligo(dT) primer from total RNA purified with the use of an RNeasy Mini RNA purification kit (Qiagen). Reverse transcription-PCR (RT-PCR) to amplify the fusion point of *EML4-ALK* variant 3 mRNA was done with a QuantiTect SYBR Green kit (Qiagen) and the primers 5′-TACCAGTGCTGTCTCAATTGCAGG-3′ and 5′-TCTTGCCAGCAAAGCAGTAGTTGG-3′. A full-length cDNA for *EML4-ALK*

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Hiroyuki Mano, Division of Functional Genomics, Jichi Medical University, 3311-1 Yakushiji, Shimotsukeshi, Tochigi 329-0498, Japan. Phone: 81-285-58-7449; Fax: 81-285-44-7322; E-mail: hmano@jichi.ac.jp.

doi:10.1158/0008-5472.CAN-07-6158

[5] http://www.ensembl.org/index.html

variant 3 was amplified from total cDNA of a NSCLC specimen (ID no. 2075) with PrimeSTAR HS DNA polymerase (Takara Bio) and the primers 5′-ACTCTGTCGGTCCGCTGAATGAAG-3′ and 5′-CCACGGTCTTAGGGATCCCAAGG-3′; PCR was done for 35 cycles of 98°C for 10 s and 68°C for 6 min. The fusion point of *EML4-ALK* in the genome was amplified by PCR with genomic DNA of NSCLC specimens, PrimeSTAR HS DNA polymerase, and the primers 5′-GGCATAAAGATGTCATCATCAACCAAGG-3′ and 5′-AGCTTGCTCAGCTTGTACTCAGGG-3′. The nucleotide sequences of the *EML4-ALK* variant 3a and 3b cDNAs have been deposited in DDBJ/EMBL/GenBank under accession nos. AB374361 and AB374362, respectively.

Fluorescence *in situ* hybridization. Fluorescence *in situ* hybridization (FISH) analysis of the fusion gene was done with archival pathology specimens and with bacterial artificial chromosomes containing genomic DNA corresponding to *EML4* or *ALK* and their flanking regions as probes. In brief, surgically removed lung cancer tissue was fixed in 20% neutral buffered formalin, embedded in paraffin, and sectioned at a thickness of 3 μm. The sections were placed on glass slides and processed with a Histology FISH Accessory Kit (DakoCytomation) before hybridization with the *EML4* and *ALK* probes and examination with a fluorescence microscope (BX61, Olympus).

Transforming activity of EML4-ALK variant 3. Analyses of the function of EML4-ALK variant 3 were done as described previously (5). In brief, the cDNA for EML4-ALK variant 3a or 3b was fused with an oligonucleotide encoding the FLAG epitope tag and then inserted into the retroviral expression plasmid pMXS (6). The resulting plasmids as well as similar pMXS-based expression plasmids for EML4-ALK variant 1, variant 1 (K589M), or variant 2 were individually introduced into mouse 3T3 fibroblasts by the calcium phosphate method for a focus formation assay and assay of tumorigenicity in nu/nu mice. The same set of EML4-ALK proteins was expressed in HEK293 cells and assayed for kinase activity *in vitro* with the YFF peptide (7).

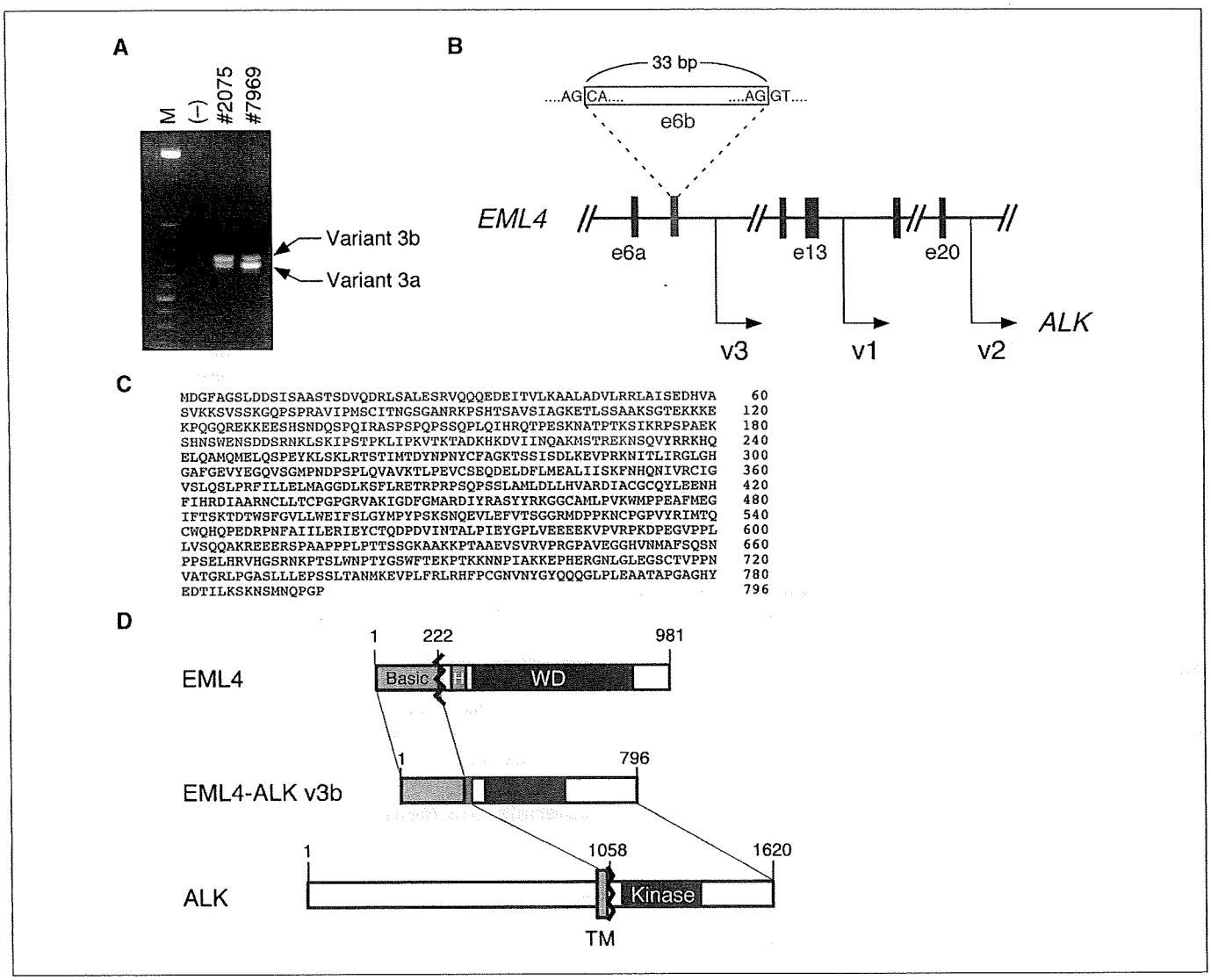

Figure 1. Identification of *EML4-ALK* variant 3. *A*, detection of fusion cDNAs linking exon 6 of *EML4* to exon 20 of *ALK* by RT-PCR analysis. Two RT-PCR products of 548 bp (corresponding to variant 3b) and 515 bp (corresponding to variant 3a) were detected by agarose gel electrophoresis with total RNA from two NSCLC specimens (tumor ID nos. 2075 and 7969). *Lane (–)*, no-template control; *lane M*, size markers (50-bp ladder). *B*, genomic organization of *EML4*. Intronic sequences downstream of exons (*e*) 6, 13, and 20 of *EML4* are fused to intron 19 of *ALK* to generate variants (*v*) 3, 1, and 2 of *EML4-ALK*, respectively. Exon-intron boundary sequences as well as the size of exon 6b are indicated. *C*, predicted amino acid sequence of EML4-ALK variant 3b. *Blue, green,* and *red,* amino acids corresponding to exons 1 to 6a of *EML4*, exon 6b of *EML4*, and *ALK*, respectively. Amino acid number is indicated on the right. *D*, fusion of an amino-terminal portion of EML4 [which consists of a basic region (*Basic*), HELP domain (*H*), and WD repeats] to the intracellular region of ALK (containing the tyrosine kinase domain) generates EML4-ALK variant 3b. *Green*, the region of the fusion protein encoded by exon 6b of *EML4*. *TM*, transmembrane domain.

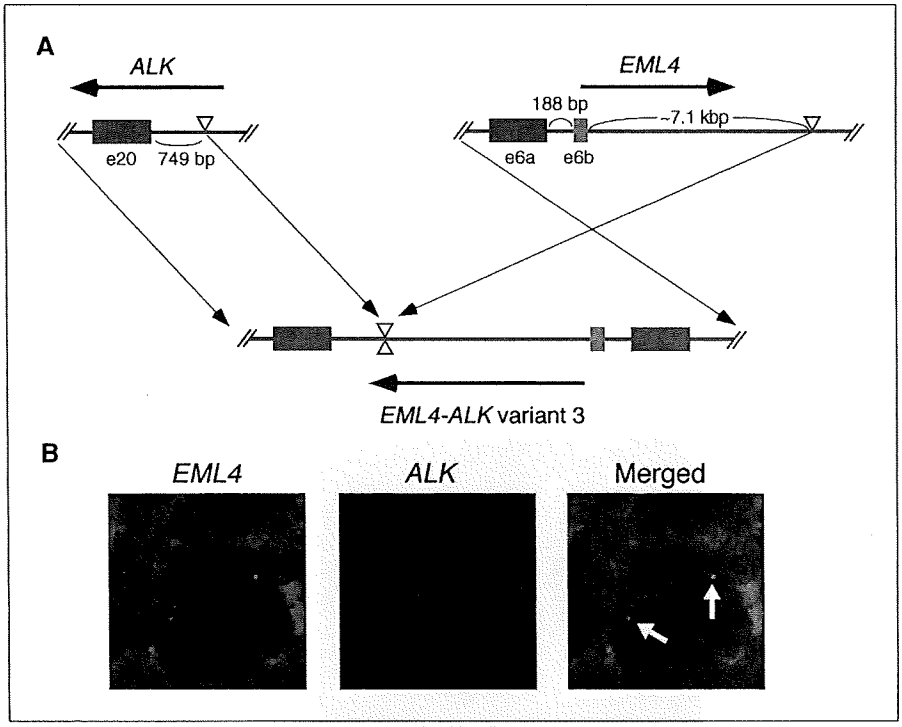

Figure 2. Chromosomal rearrangement responsible for generation of *EML4-ALK* variant 3. *A*, schematic representation of the chromosomal rearrangement underlying the generation of *EML4-ALK* variant 3. Exon 6b of *EML4* is located 188 bp downstream of exon 6a. In NSCLC specimen ID no. 7969, *EML4* is disrupted at a position ~7.1 kbp downstream of exon 6b and is ligated to a position 749 bp upstream of exon 20 of *ALK*, giving rise to the *EML4-ALK* (variant 3) fusion gene. *Horizontal arrows*, direction of transcription. *B*, FISH analysis of a representative cancer cell in a histologic section of lung adenocarcinoma (ID no. 7969) with differentially labeled probes for *EML4* (*left*) and *ALK* (*center*). Two fusion signals (*arrows*) and a pair of green (corresponding to *EML4*) and red (corresponding to *ALK*) signals are present in the merged image (*right*).

The cDNA for FLAG-tagged EML4-ALK variant 3b was also inserted into pMX-iresCD8 for the expression of both EML4-ALK and mouse CD8 (8), and the resulting recombinant retroviruses were used to infect mouse BA/F3 cells (9). CD8-positive cells were then purified with the use of a miniMACS magnetic bead–based separation system (Miltenyi Biotec) and cultured in the absence or presence of mouse interleukin-3 (IL-3; Sigma) or 2,4-pyrimidinediamine (Example 3-39, a specific inhibitor of ALK enzymatic activity that was developed by Novartis[6] and synthesized by Astellas Pharma).

Mouse 3T3 fibroblasts and NCI-H2228 lung cancer cells (both from American Type Culture Collection) as well as 3T3 cells expressing v-Ras were plated in 96-well spheroid culture plates (Celltight Spheroid, Sumilon) at a density of 1×10^3 per well. Cell growth was examined with the WST-1 Cell Proliferation Reagent (Clontech) after culture for 5 d with 2,4-pyrimidinediamine.

Luciferase reporter assays. The promoter fragments of *Fos*, *Myc*, and *Bcl-x_L* genes were ligated to a luciferase cDNA to generate pFL700 (10), pHXLuc (11), and pBclxl-Luc (12) reporter plasmids, respectively. Luciferase cDNA ligated to the DNA binding sequence for nuclear factor κB (NF-κB) or to the GAS sequence was obtained from Stratagene. HEK293 cells were transfected with these various reporter plasmids together with the expression plasmid for EML4-ALK variant 3b or the empty vector, as described previously (13). The pGL4 plasmid (Promega) for expression of *Renilla* luciferase was also included in each transfection mixture. After culture of the cells for 2 d, luciferase activity in cell lysates was measured with a Luciferase Assay system (Promega).

Results and Discussion

Detection of *EML-ALK* variant 3. The EML4-ALK variant 1 and 2 proteins are produced as a result of genomic rearrangements that lead to the juxtaposition of exons 13 and 20 of *EML4*, respectively, to exon 20 of *ALK*. It is theoretically possible that exon 2, 6, 18, or 21 of *EML4* also could undergo in-frame fusion to exon 20 of *ALK*. We therefore examined whether transcripts of any such novel *EML4-ALK* fusion genes are present in NSCLC cells by RT-PCR analysis with primers that flank each putative fusion point (data not shown). With the primer set for amplification of the *EML4* (exon 6)-*ALK* (exon 20) fusion cDNA, we detected a pair of PCR products in two individuals with lung adenocarcinoma (Fig. 1*A*). Although one of the patients (tumor ID no. 7969) had a smoking index of 540, the other patient (tumor ID no. 2075) had never smoked. Nucleotide sequencing of each PCR product from both patients revealed that the smaller product of 515 bp corresponded to a fusion cDNA linking exon 6 of *EML4* to exon 20 of *ALK*, whereas the larger product of 548 bp contained an additional sequence of 33 bp that was located between these exons of *EML4* and *ALK* and which mapped to intron 6 of *EML4* (Fig. 1*B*). The larger cDNA would thus be expected to encode a fusion protein with an insertion of 11 amino acids between the EML4 and ALK sequences of the protein encoded by the smaller cDNA.

Although we did not detect human mRNAs or expressed sequence tags containing this cryptic exon of *EML4* in the nucleotide sequence databases, it is likely that this exon is physiologic and functional because (*a*) the fusion cDNA containing this exon was identified in two independent patients and in amounts no less than those of the corresponding cDNA without it (Fig. 1*A*); (*b*) the intron-exon boundary sequence for this exon conforms well to the AG-GU rule for mRNA splicing (Fig. 1*B*); and (*c*) *EML4* cDNAs or expressed sequence tags containing this exon were detected in the sequence databases for other species (for instance, GenBank accession no. AK144604 corresponding to a mouse *EML4* cDNA). We thus refer to this cryptic exon as exon 6b and to the original exon 6 as exon 6a (Fig. 1*B*). The novel isoforms of *EML4-ALK* transcripts containing exons 1 to 6a or 1 to 6b of *EML4* were also designated variants 3a and 3b, respectively.

[6] Patent information: Garcia-Echeverria C, Kanazawa T, Kawahara E, Masuya K, Matsuura N, Miyake T, et al., inventors; Novartis AG, Novartis Pharma GmbH, IRM LLC, applicants. 2,4-Pyrimidinediamines useful in the treatment of neoplastic disease, inflammatory and immune system disorders. PCT WO 2005016894. 2005 Feb 24.

To isolate a full-length cDNA for EML4-ALK variant 3, we performed RT-PCR with total cDNA of a positive specimen (ID no. 2075) and with a sense strand primer targeted to the 5′ untranslated region (UTR) of *EML4* mRNA and an antisense strand primer targeted to the 3′ UTR of *ALK* mRNA. One-step PCR analysis yielded cDNA products for both *EML4-ALK* variants 3a and 3b (Fig. 1*C*; Supplementary Fig. S1).

The EML4 protein contains an amino-terminal basic domain followed by a hydrophobic echinoderm microtubule–associated protein–like protein (HELP) domain and WD repeats (14). Given that exons 1 to 6 of *EML4* encode the basic domain, the proteins encoded by the variant 3 cDNAs contain the entire basic domain of EML4 directly linked to the catalytic domain of ALK (Fig. 1*D*). The fact that the basic domain was found to be essential for both the self-dimerization and oncogenic activity of EML4-ALK (5) suggested that the variant 3 isoforms likely also possess transforming activity.

Chromosome rearrangement responsible for generation of *EML4-ALK* variant 3. To show the presence of a chromosome rearrangement responsible for the generation of *EML4-ALK* variant

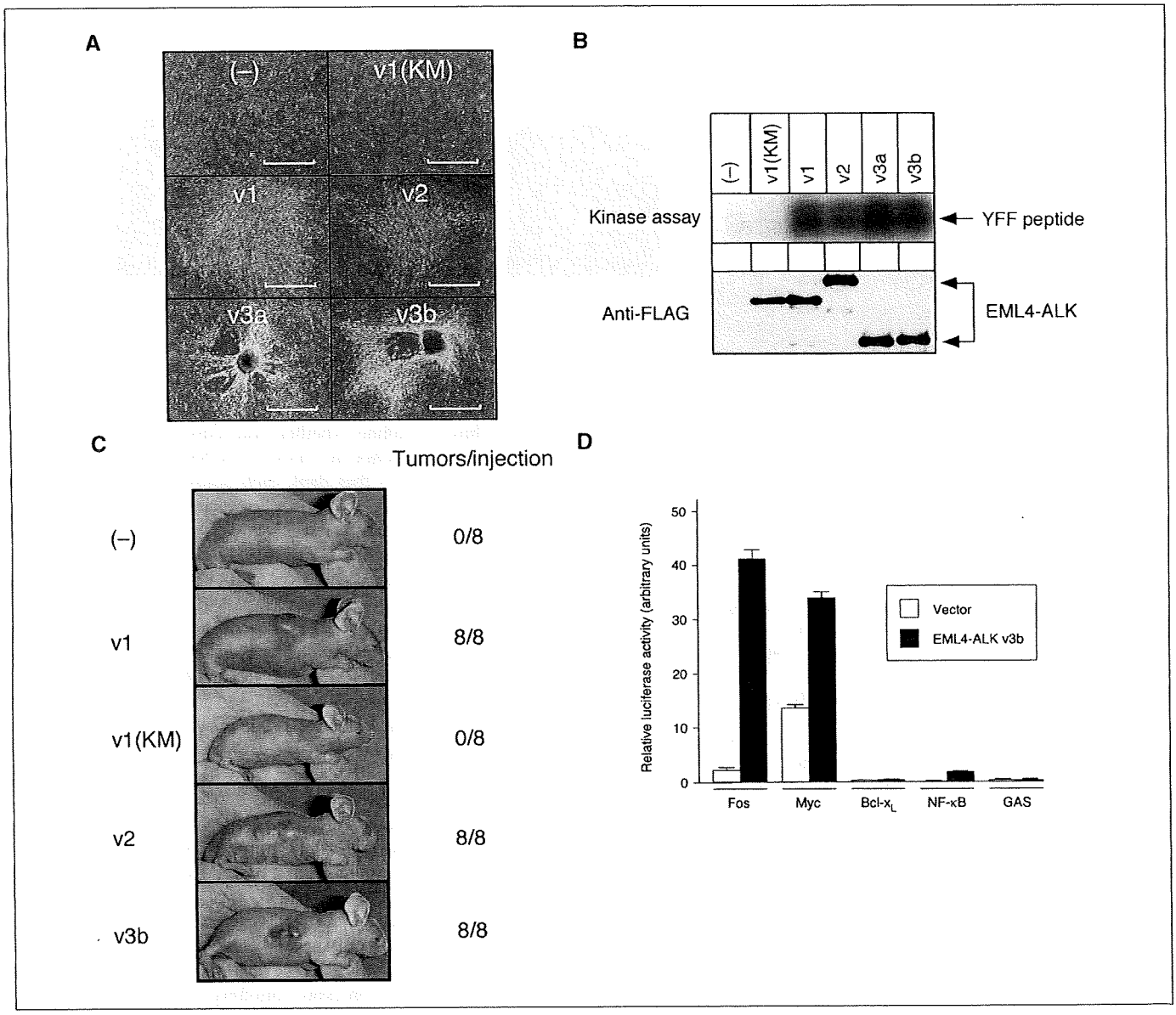

Figure 3. Transforming potential of EML4-ALK variants. *A*, focus formation assay. Mouse 3T3 fibroblasts were transfected with the empty expression plasmid [(–)] or with plasmids for wild-type (*v1*) or K589M mutant [*v1(KM)*] forms of variant 1, variant 2 (*v2*), variant 3a (*v3a*), or variant 3b (*v3b*) of FLAG-tagged EML4-ALK. The cells were photographed after culture for 18 d. *Bar*, 1 mm. *B*, *in vitro* kinase assay. HEK293 cells expressing the various FLAG-tagged variants of EML4-ALK were lysed and subjected to immunoprecipitation with antibodies to FLAG, and the resulting precipitates were assayed for kinase activity with the synthetic YFF peptide (*top*) or subjected to immunoblot analysis with antibodies to FLAG (*bottom*). *C*, *in vivo* assay of tumorigenicity. 3T3 cells expressing the indicated EML4-ALK variants were injected s.c. into nu/nu mice, and tumor formation was examined after 20 d. The number of tumors formed per eight injections is indicated on the right. *D*, analysis of EML4-ALK signaling with luciferase-based reporter plasmids. HEK293 cells were transfected with an expression plasmid for EML4-ALK variant 3b (or with the empty vector) together with reporter plasmids containing the promoter fragment of *Fos*, *Myc*, or *Bcl-x_L* gene; the DNA binding sequence for NF-κB; or the GAS sequence. Cells were cultured for 2 d, lysed, and assayed for luciferase activity. The activity of firefly luciferase was normalized by that of *Renilla* luciferase. *Columns*, mean of three experiments; *bars*, SD.

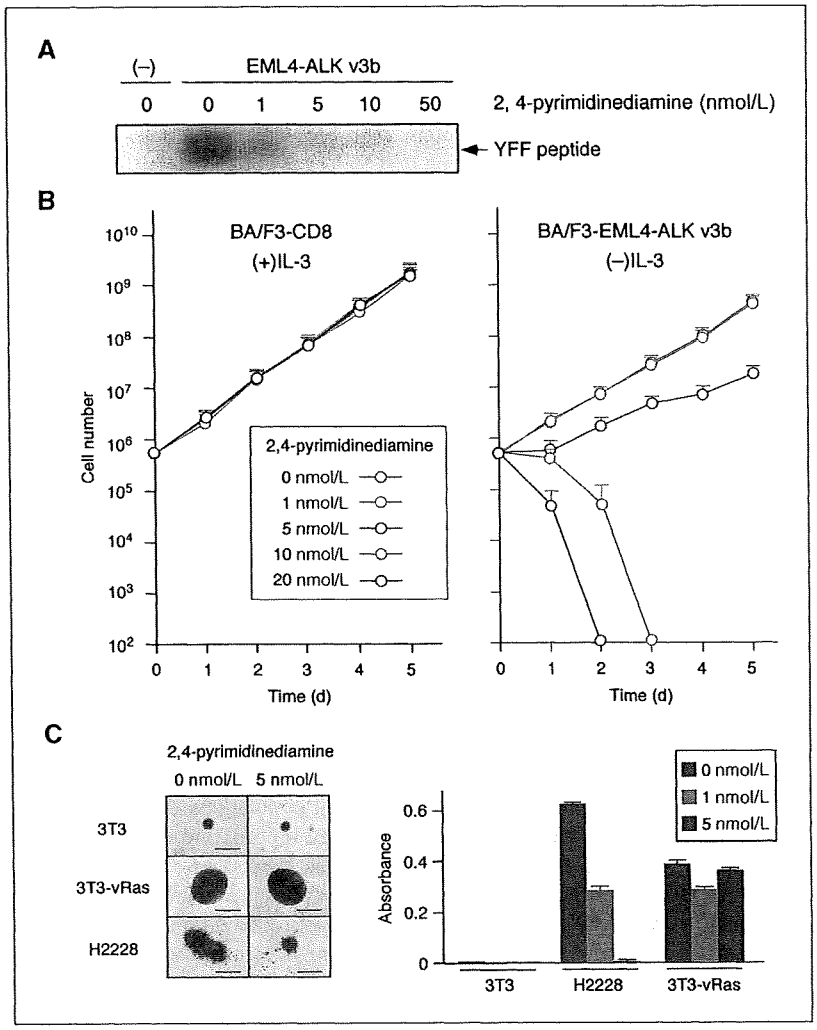

Figure 4. Essential role of EML4-ALK kinase activity in malignant transformation. *A*, lysates of HEK293 cells expressing FLAG-tagged EML4-ALK variant 3b (*v3b*) were divided into five equal portions, and each portion was subjected to immunoprecipitation with antibodies to FLAG. The immunoprecipitates were washed with kinase buffer [10 mmol/L HEPES-NaOH (pH 7.4), 50 mmol/L NaCl, 5 mmol/L $MgCl_2$, 5 mmol/L $MnCl_2$, 0.1 mmol/L Na_3VO_4] containing 0, 1, 5, 10, or 50 nmol/L of 2,4-pyrimidinediamine and then incubated for 30 min at room temperature for assay of kinase activity with the YFF peptide in the continued absence or presence of 2,4-pyrimidinediamine. The same amount of lysate of cells transfected with the empty vector was also subjected to immunoprecipitation and assayed as a negative control (–). *B*, mouse BA/F3 cells expressing CD8 alone were cultured in the presence of IL-3 (1 ng/mL) and the indicated concentrations of 2,4-pyrimidinediamine (*left*). BA/F3 cells expressing both CD8 and EML4-ALK variant 3b were cultured with the indicated concentrations of 2,4-pyrimidinediamine but without IL-3 (*right*). Cell number was counted at the indicated times. *Points*, mean of three separate experiments; *bars*, SD. *C*, mouse 3T3 fibroblasts expressing (or not) v-Ras or NCI-H2228 cells were cultured in a spheroid culture plate for 2 d, after which 2,4-pyrimidinediamine was added to the culture medium at a concentration of 0, 1, or 5 nmol/L. The cells were photographed after culture for an additional 5 d (*left*). *Bar*, 4 mm. Cell number in each well was also assessed at the same time with the use of the WST-1 assay (*right*). *Columns*, mean of three wells from a representative experiment; *bars*, SD.

3, we attempted to amplify the fusion point between the two genes from the genome of positive NSCLC cells. PCR with primers targeted to regions flanking the putative fusion point yielded a product of ~8 kbp with the genomic DNA of tumor ID no. 7969 (data not shown). Our failure to detect an unambiguous PCR product with genomic DNA of tumor ID no. 2075 may indicate that the breakpoint in intron 6 of *EML4* in this specimen is too distant from exon 6 to be readily amplified by PCR (intron 6 of *EML4* is >16 kbp). Nucleotide sequencing of the PCR product for tumor ID no. 7969 revealed that intron 6 of *EML4* was disrupted at a position ~7.1 kbp downstream of exon 6b and was joined to a point 749 bp upstream of exon 20 of *ALK* (Fig. 2*A*).

We also confirmed the chromosome rearrangement involving *EML4* and *ALK* by FISH analysis of cells from tumor ID no. 7969 (Fig. 2*B*) and tumor ID no. 2075 (data not shown) with differentially labeled probes for the two genes. Both genes map to the short arm of chromosome 2 within a distance of ~12 Mbp. The tumor cells exhibited fusion signals (corresponding to *EML4-ALK*) in addition to a pair of isolated green and red signals (corresponding to the two genes on the normal chromosome 2). The chromosome rearrangement involving the *ALK* locus was further verified with a different set of fluorescent probes (Supplementary Fig. S2).

Transforming activity of EML4-ALK variant 3. To compare the transforming potential of variants 1, 2, 3a, and 3b of EML4-ALK, we introduced expression plasmids for each variant into mouse 3T3 fibroblasts for assay of focus formation. No transformed foci were detected for cells transfected with the empty plasmid or with a plasmid for a kinase-inactive mutant (K589M) of EML4-ALK variant 1 (5) in which Lys^{589} in the ATP binding site of the catalytic domain is replaced with Met (Fig. 3*A*). In contrast, variants 3a and 3b of EML4-ALK each exhibited marked transforming activity that was not less than that of variant 1 or 2. To examine directly the tyrosine kinase activity of EML4-ALK variants, we subjected HEK293 cells expressing each of these variants to an *in vitro* kinase assay with a synthetic YFF peptide (7). Again, both variants 3a and 3b exhibited marked kinase activity that was not less than that of variant 1 or 2 (Fig. 3*B*). Similarly, in a tumorigenicity assay with nude mice, 3T3 cells expressing EML4-ALK variant 3b formed large subcutaneous tumors at all injection sites (Fig. 3*C*). Consistent with our previous observations (5), cells expressing variant 1 or 2 of EML4-ALK also formed tumors.

To examine the intracellular signaling pathways activated by EML4-ALK, we linked the luciferase cDNA to the promoter fragment of *Fos*, *Myc*, or *Bcl-x_L* gene (10–12); the DNA binding sequence for NF-κB; or the GAS sequence [a target site of the transcription factors signal transducers and activators of transcription (STAT)-1 and STAT3; ref. 15]. The resulting constructs were then introduced into HEK293 cells together with an

expression plasmid for EML4-ALK variant 3b. EML4-ALK variant 3b markedly activated the promoters of the *Fos* and *Myc* genes (Fig. 3*D*), consistent with the transforming potential of EML4-ALK. In contrast, although STAT3 has been shown to be a downstream target of the NPM-ALK fusion protein (16), EML4-ALK did not activate the GAS sequence, suggesting that STAT3 is unlikely to be a major target of EML4-ALK, as was shown in an EML4-ALK–positive lung cancer cell line by a proteomics approach (17). The distinct subcellular localizations of the two ALK fusion proteins [EML4-ALK in the cytoplasm (5) and NPM-ALK in both the nucleus and cytoplasm (18)] may account for this difference. Whereas EML4-ALK did not activate the *Bcl-x_L* gene promoter, it induced a small but significant increase in the activity of the NF-κB binding sequence ($P = 1.86 \times 10^{-4}$, Student's t test).

Several compounds have recently been identified as specific inhibitors of the kinase activity of ALK and as potential drugs for the treatment of lymphoma positive for *NPM-ALK* (19). We examined the effects of one such inhibitor, 2,4-pyrimidinediamine, on the transforming potential of EML4-ALK. We first determined the effect of this inhibitor on the kinase activity of EML4-ALK variant 3b immunoprecipitated from transfected cells. 2,4-Pyrimidinediamine inhibited the kinase activity of EML4-ALK in a concentration-dependent manner, with a concentration of 1 nmol/L reducing the kinase activity to <50% of the control value (Fig. 4*A*).

We also introduced EML4-ALK variant 3b and CD8 (or CD8 alone) into the IL-3–dependent hematopoietic cell line BA/F3 (9) and then purified the resulting CD8-positive cell populations. 2,4-Pyrimidinediamine, even at a concentration of 20 nmol/L, did not affect the IL-3–dependent growth of BA/F3 cells expressing only CD8 (Fig. 4*B*), indicating that this agent does not inhibit mitogenic signaling mediated by Janus kinase in BA/F3 cells. Expression of EML4-ALK rendered BA/F3 cells independent of IL-3 for growth, but the cells expressing the fusion protein also rapidly underwent cell death on exposure to 2,4-pyrimidinediamine (Fig. 4*B*).

Finally, we examined the effect of 2,4-pyrimidinediamine on lung cancer cells that express endogenous EML4-ALK variant 3. The human lung cancer cell line NCI-H2228 expresses EML4-ALK variants 3a and 3b (data not shown) and forms spheroids in a three-dimensional spheroid culture system (Fig. 4*C*; ref. 20). Whereas 3T3 fibroblasts are unable to form such spheroids, expression of v-Ras in these cells results in the formation of large spheroids in culture. Whereas 2,4-pyrimidinediamine did not affect the proliferation of 3T3 cells expressing v-Ras in this system, it inhibited the growth of NCI-H2228 cells in a concentration-dependent manner (Fig. 4*C*). These data thus indicate that EML4-ALK is essential for the growth of cancer cells expressing this oncokinase.

In conclusion, we have identified novel isoforms of *EML4-ALK* in two patients with NSCLC. A chromosome inversion within 2p was shown to connect intron 6 of *EML4* to intron 19 of *ALK* and to be responsible for the generation of fusion cDNAs connecting exons 1 to 6a or exons 1 to 6b of *EML4* to exon 20 of *ALK*. Given that fusion cDNAs with or without exon 6b of *EML4* were each present in the two patients, EML4-ALK variant 3a and 3b proteins are likely to be coexpressed in NSCLC cells. Although RT-PCR analysis to detect *EML4-ALK* may provide a highly sensitive means to detect lung cancer, it is important that all variant forms of the fusion gene be assayed with appropriately designed primer sets. Given that all the identified variants possess prominent transforming activity, the newly revealed increased incidence of *EML4-ALK* fusion in NSCLC further increases the importance of the fusion gene as a therapeutic target for this intractable disorder.

Disclosure of Potential Conflicts of Interest

K. Takeuchi: Consultant, DAKO. The other authors disclosed no potential conflicts of interest.

Acknowledgments

Received 11/8/2007; revised 3/3/2008; accepted 4/22/2008.

Grant support: Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan; the Japan Society for the Promotion of Science; and grants from the Ministry of Health, Labor, and Welfare of Japan, the Smoking Research Foundation of Japan, the National Institute of Biomedical Innovation of Japan, and the Vehicle Racing Commemorative Foundation of Japan.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Takashi Aoki and Yasunobu Sugiyama for technical assistance.

References

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2006. CA Cancer J Clin 2006;56:106–30.
2. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004;350:2129–39.
3. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304:1497–500.
4. Shigematsu H, Lin L, Takahashi T, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 2005;97:339–46.
5. Soda M, Choi YL, Enomoto M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 2007;448:561–6.
6. Onishi M, Kinoshita S, Morikawa Y, et al. Applications of retrovirus-mediated expression cloning. Exp Hematol 1996;24:324–9.
7. Donella-Deana A, Marin O, Cesaro L, et al. Unique substrate specificity of anaplastic lymphoma kinase (ALK): development of phosphoacceptor peptides for the assay of ALK activity. Biochemistry 2005;44:8533–42.
8. Yamashita Y, Kajigaya S, Yoshida K, et al. Sak serine/threonine kinase acts as an effector of Tec tyrosine kinase. J Biol Chem 2001;276:39012–20.
9. Palacious R, Steinmetz M. IL-3 dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ-line configuration, and generate B lymphocytes *in vivo*. Cell 1985;41:727–34.
10. Hu Q, Milfay D, Williams LT. Binding of NCK to SOS and activation of ras-dependent gene expression. Mol Cell Biol 1995;15:1169–74.
11. Takeshita T, Arita T, Higuchi M, et al. STAM, signal transducing adaptor molecule, is associated with Janus kinase and involved in signaling for cell growth and c-myc induction. Immunity 1997;6:449–57.
12. Grillot DAM, Gonzalez-Garcia M, Ekhterae D, et al. Genomic organization, promoter region analysis, and chromosome localization of the mouse bcl-x gene. J Immunol 1997;158:4750–7.
13. Fujiwara S, Yamashita Y, Choi YL, et al. Transforming activity of purinergic receptor P2Y, G protein coupled, 8 revealed by retroviral expression screening. Leuk Lymphoma 2007;48:978–86.
14. Pollmann M, Parwaresch R, Adam-Klages S, Kruse ML, Buck F, Heidebrecht HJ. Human EML4, a novel member of the EMAP family, is essential for microtubule formation. Exp Cell Res 2006;312:3241–51.
15. Wesoly J, Szweykowska-Kulinska Z, Bluyssen HA. STAT activation and differential complex formation dictate selectivity of interferon responses. Acta Biochim Pol 2007;54:27–38.
16. Marzec M, Kasprzycka M, Ptasznik A, et al. Inhibition of ALK enzymatic activity in T-cell lymphoma cells induces apoptosis and suppresses proliferation and STAT3 phosphorylation independently of Jak3. Lab Invest 2005;85:1544–54.
17. Rikova K, Guo A, Zeng Q, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell 2007;131:1190–203.
18. Duyster J, Bai RY, Morris SW. Translocations involving anaplastic lymphoma kinase (ALK). Oncogene 2001;20:5623–37.
19. Galkin AV, Melnick JS, Kim S, et al. Identification of NVP-TAE684, a potent, selective, and efficacious inhibitor of NPM-ALK. Proc Natl Acad Sci U S A 2007;104:270–5.
20. Kunita A, Kashima TG, Morishita Y, et al. The platelet aggregation-inducing factor aggrus/podoplanin promotes pulmonary metastasis. Am J Pathol 2007;170:1337–47.

H Tamai[1], Y Shioi[1], H Yamaguchi, M Okabe, S Wakita, T Mizuki, K Nakayama, K Inokuchi, K Tajika and K Dan
Department of Hematology, Nippon Medical School, Tokyo, Japan
E-mail: s6056@nms.ac.jp
[1]*These two authors contributed equally to this work.*

References

1 Taksin AL, Legrand O, Raffoux E, de Revel T, Thomas X, Contentin N *et al.* High efficacy and safety profile of fractionated doses of Mylotarg as induction therapy in patients with relapsed acute myeloblastic leukemia: a prospective study of the alfa group. *Leukemia* 2007; **21**: 66–71.
2 Mrozek K, Bloomfield CD. Chromosome aberrations, gene mutations and expression changes, and prognosis in adult acute myeloid leukemia. *Hematology 2006 Education program book.* American Society of Hematology: Washington, DC, 2006, 169–177.
3 Garrido SM, Bryant E, Appelbaum FR. Allogeneic stem cell transplantation for relapsed and refractory acute myeloid leukemia patients with 11q23 abnormalities. *Leuk Res* 2000; **24**: 481–486.
4 Larson RA, Sivers EL, Stadtmauer EA, Lowenberg B, Estey EH, Dombret H *et al.* Final report on the efficacy and safety of gemtuzumab ozogamicin (Mylotarg) in patients with CD33 positive acute myeloid leukemia in first recurrence. *Cancer* 2001; **104**: 1442–1452.
5 Muñoz L, Nomdedéu JF, Villamor N, Guardia R, Colomer D, Ribera JM *et al.* Acute myeloid leukemia with MLL rearrangements: clinicobiological features, prognostic impact and value of flow cytometry in the detection of residual leukemic cells. *Leukemia* 2003; **17**: 76–82.

MicroRNA expression profiles of human leukemias

Leukemia (2008) **22**, 1274–1278; doi:10.1038/sj.leu.2405031; published online 8 November 2007

MicroRNAs (miRNAs) are small noncoding RNAs of 20–24 nucleotides (nt) that negatively regulate the translation of target mRNAs through incomplete base-pairing with their 3′-untranslated regions.[1] Evidence indicates that miRNAs play an important role in the development of human cancers including leukemias, with one of the most well-characterized examples being association of miR-15a and miR-16a with chronic lymphocytic leukemia. Almost half of chronic lymphocytic leukemia patients harbor a chromosome deletion that encompasses 13q14, a region that includes the genes for miR-15a and miR-16a, and the abundance of these miRNAs is reduced in chronic lymphocytic leukemia cells with the chromosome deletion.[2] Several other miRNAs, such as miR-155 and miR-17-92, have also been implicated in the pathogenesis of lymphoma.[3] It is therefore important that the entire miRNA repertoire of clinical specimens be characterized and compared among various hematologic malignancies.

Reliable assessment of the global expression profiles of miRNAs, especially for the small amounts of clinical specimens available, is not straightforward, however. Microarray-based detection of miRNAs is prone to the generation of false-positive data that may result from mishybridization of probes, although improvements have recently been developed for this technology.[4] A large-scale cloning strategy would be an ideal approach to reliable estimation of the expression level of miRNAs, provided that a sufficient number of clones were to be analyzed. However, conventional methods for isolation of miRNAs require $>10\,\mu g$ of total RNA, which is not always obtainable from clinical specimens.

We recently developed a sensitive method, mRAP (micro RNA amplification profiling)[5] that readily allows the isolation of miRNA clones from $\leqslant 1 \times 10^4$ cells. To examine the miRNA expression profiles for leukemias with mRAP, we first purified $CD34^+$ cells from individuals ($n=12$) with *de novo* acute myeloid leukemia, acute myeloid leukemia secondary to myelodysplastic syndrome, acute lymphoid leukemia or biphenotypic acute leukemia (Table 1). Column affinity-chromatography to isolate $CD34^+$ cells yielded 10–50% of the input cells with a purity of $\geqslant 90\%$ as judged by flow cytometry (data not shown). As a normal control, we also purified a $CD34^+$ cell fraction from bone marrow mononuclear cells of a healthy volunteer. Then mRAP procedure was applied to 1.1×10^6–1.0×10^8 of the purified $CD34^+$ cells from each individual in order to obtain short RNA clones.

Sequencing and computer filtering[5] of the mRAP amplicons identified a total of 38 858 qualified reads for the 13 study subjects. BLAST analysis then isolated 32 867 reads that match the human genome sequence (ncbi 36 assembly), among which 2054 reads were mapped to transfer RNA genes, 2720 to ribosomal RNA genes and 9474 to repetitive sequences. From the remaining sequences, we identified 7191 reads corresponding to 143 independent known miRNAs (Supplementary Table 1). We further searched for candidate sequences corresponding to novel miRNAs whose surrounding genome sequences (of ~100 nt) potentially fold into a hairpin structure with a single notch. In this analysis, we did not exclude miRNA candidates that were not detected in the genomes of other

Table 1 Clinical characteristics of the study subjects

ID (no.)	*Age (years)*	*Sex*	*Sample origin*	*Disease*	*Karyotype*
3	64	M	PB	ALL	46,XY,t(9;22)
4	45	M	BM	AML (M4)	46,XY,inv(16)
7	78	F	BM	MDS-derived AML	46,XX
10	21	F	PB	AML (M0)	46,XX,t(9:15)
12	58	M	BM	AML (M2)	46,XY
32	43	M	BM	AML (M2)	46,XY,t(8;21)
33		M	PB	AML (M1)	46,XY
44	71	M	PB	MDS-derived AML	46,XY,t(8:21)
46	61	M	PB	AML (M2)	46,XY
47	61	M	BM	AML (M3)	46,XY,t(15;17)
48	29	M	PB	BAL	46,XY
49	58	M	PB	MDS-derived AML	46,XY

Abbreviations: ALL, acute myeloid leukemia; AML, acute lymphoid leukemia; BAL, biphenotypic acute leukemia; MDS, myelodysplastic syndrome; BM, bone marrow; F, female; M, male; PB, peripheral blood.

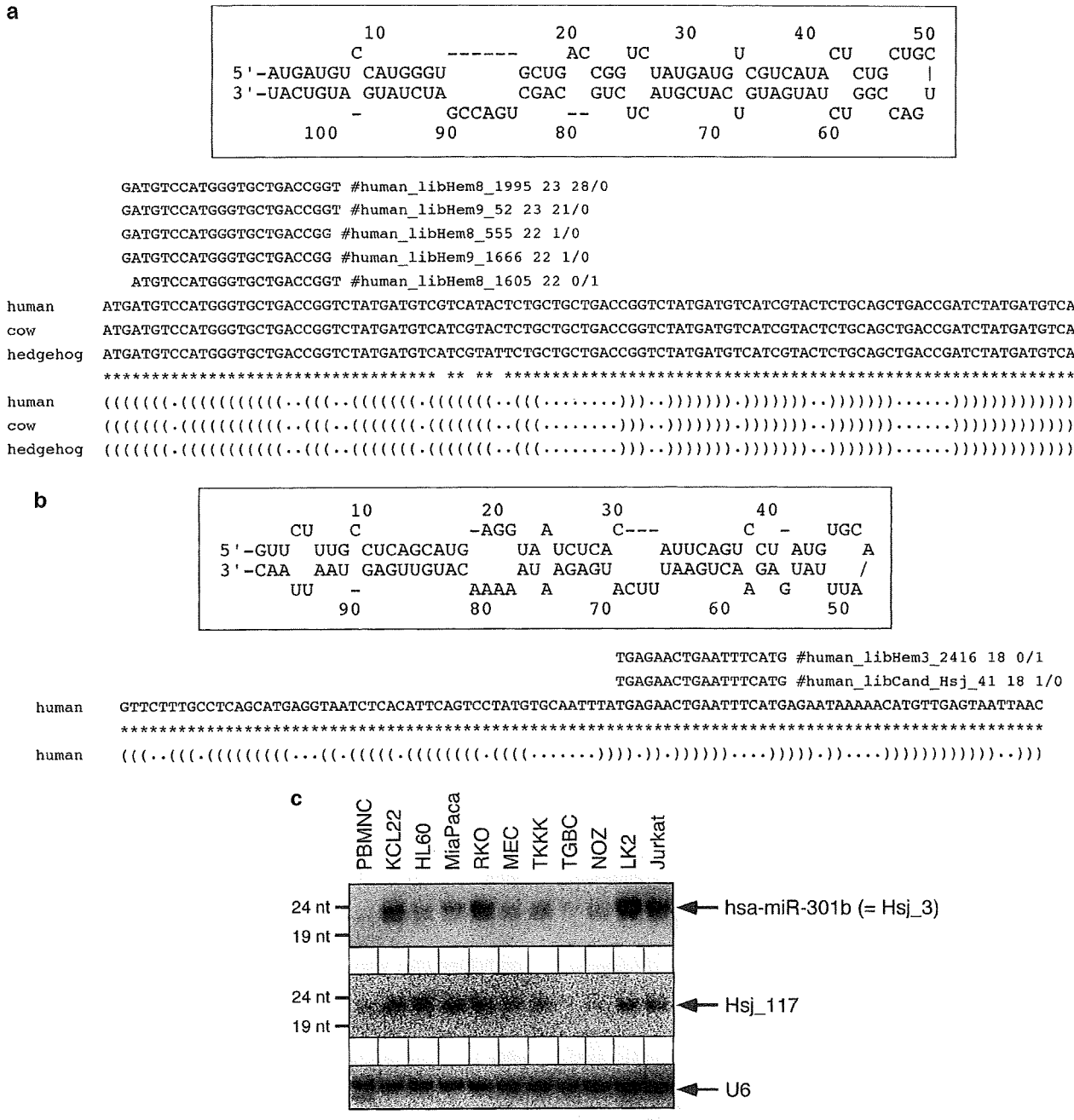

Figure 1 Nucleotide sequence and expression of novel miRNA candidates. Nucleotide sequences (red) of genes for the predicted novel miRNAs Hsj_376 (**a**) and Hsj_41 (**b**) are aligned with genomic sequences of human and other species. Asterisks indicate conserved nucleotides. Possible base-pairing schemes for the respective miRNA precursors are shown in the upper insets. (**c**) Small-RNA fractions (800 ng per lane) purified from the indicated cell lines with the use of a mirVana RNA isolation kit (Ambion, Austin, TX, USA) were subjected to northern blot analysis with 'locked' nucleic acid probes for the candidate miRNAs Hsj_3 or Hsj_117 or for U6 small nuclear RNA (internal control). Hjs_3 has been very recently deposited into the miRBase database as hsa-miR-301b. The positions of 24- and 19-nt size markers are indicated on the left. miRNA, microRNA.

species, given that some miRNAs are species-specific or have arisen recently during evolution.[6]

We isolated an unexpectedly large number ($n=170$) of independent candidates for novel miRNAs among 296 sequence reads (Supplementary Table 1, Supplementary Data). The proportion of reads for such novel candidate miRNAs among all miRNA reads ranged from 1.7 to 9.5% per sample (mean, 4.7%). Of the 170 candidates, 19 were identified in at least two samples, supporting the notion that they are *bona fide* miRNAs. The surrounding genome sequence for one such candidate (designated Hsj_376) is conserved among human, cow and hedgehog (Figure 1a). Hsj_376 was found in two acute myeloid leukemia samples (corresponding to a total of 52 reads) in our data set and folds into a single hairpin (Figure 1a). In contrast, we obtained only one read for a candidate miRNA (Hsj_41) whose surrounding genome sequence also folds into a single

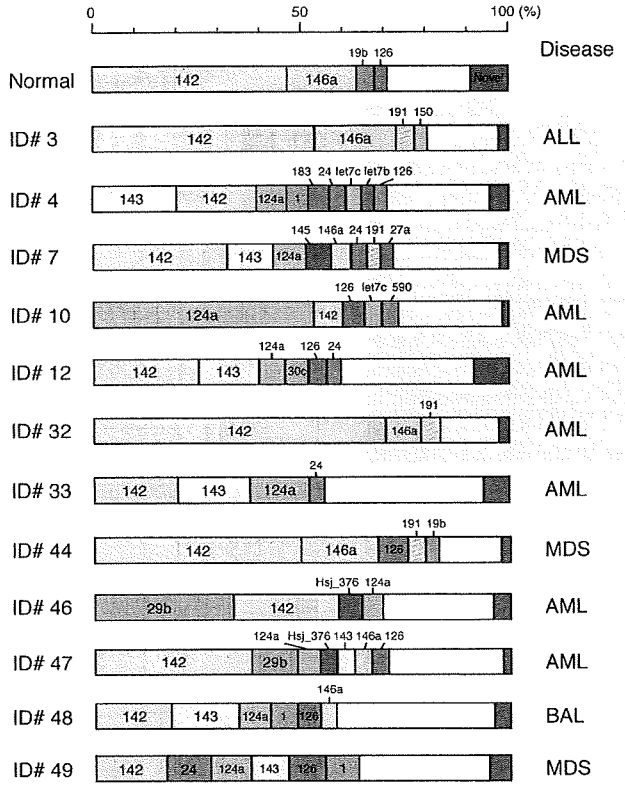

Figure 2 Expression profiles of miRNAs in CD34$^+$ specimens. The percentage contribution of each miRNA to the total miRNA population was calculated for each study subject. Abundant miRNAs are represented as color-coded, with candidates for novel miRNAs shown in red. The disease type of each individual is also indicated on the right. ALL, acute myeloid leukemia; AML, acute lymphoid leukemia; MDS, myelodysplastic syndrome; miRNA, microRNA.

hairpin structure (Figure 1b). However, this read was independently identified in our experiments performed both in Japan and in the Netherlands. The nucleotide sequence of all the miRNA candidates and their flanking sequences are presented in Supplementary Data.

The genomic sequences for some of the candidate miRNAs mapped in the vicinity ($\leqslant$20 kbp) of those for other miRNAs in the human genome. For example, the gene for one candidate (Hsj_360) and hsa-miR-560 are present on the long arm of chromosome 2 separated by a distance of ~1 kbp (Supplementary Figure 1). In this instance, the genome sequences for the two miRNAs are not conserved in other species, indicative of recent evolution.

Expression of some of the candidate miRNAs was confirmed by northern blot analysis with small RNA fractions isolated from a variety of human cancer cell lines, including KCL22 (chronic myeloid leukemia), HL60 (acute myeloid leukemia), MiaPaCa (pancreatic carcinoma), RKO (colorectal carcinoma), MEC (cholangiocarcinoma), TKKK (intrahepatic bile duct carcinoma), TGBC (gallbladder carcinoma), NOZ (gallbladder carcinoma), LK2 (lung squamous cell carcinoma) and Jurkat (T-cell leukemia) (Figure 1c).

The relative expression profile of miRNAs was then calculated for each sample as shown in Figure 2. Whereas some miRNAs, such as miR-124a, miR-142, miR-143 and miR-146a, were expressed in different types of leukemia, most miRNAs were expressed in a sample-specific manner. For instance, miR-29b was abundant in only two samples (ID nos. 46 and 47), with the reads for this miRNA accounting for $<$1% of all miRNA reads in each of the other specimens. Similarly, the novel miRNA candidate Hsj_376 was abundant in the same two samples but not in the others. Both hsa-miR-183 and hsa-miR-590 were detected in only single samples (ID nos. 4 10, respectively).

To examine further the similarities and differences in the miRNA profiles among the study subjects, we performed a hierarchical clustering analysis for the subjects based on the expression patterns of all known and novel miRNAs (Figure 3a). Leukemia specimens with a normal karyotype were clustered in the same branch, indicative of a relative homogeneity of these samples, at least with regard to miRNA expression. Nevertheless, the healthy volunteer was placed in a different branch, suggesting that leukemic blasts with a normal karyotype possess a miRNA profile distinct from that of nonleukemic CD34$^+$ cells with a normal karyotype.

We further attempted to identify miRNAs whose expression level was significantly linked to blast karyotype. Application of Student's *t*-test to the miRNA expression data with a Benjamini and Hochberg false discovery rate [7] of $<$0.05 resulted in the isolation of six miRNAs (hsa-miR-29c, hsa-miR-124a, hsa-miR-150, hsa-miR-183, hsa-miR-382 and hsa-miR-590). Hierarchical clustering of the study subjects based on the expression profiles of these 'karyotype-associated miRNAs' revealed that the healthy volunteer was again placed apart from the leukemic patients with a normal karyotype.

In conclusion, application of the mRAP procedure to CD34$^+$ leukemic blasts yielded 7487 reads for potential miRNA clones. We previously showed that mRAP readily allows the isolation of $>1 \times 10^6$ miRNA concatamers from $\leqslant 1 \times 10^4$ cells and is thus suitable for miRNA profiling of clinical specimens.[5] Indeed, mRAP functioned well with the small number of purified specimens in the present study, with the result that sequencing capacity, rather than specimen quantity, is likely to be the limiting factor for the size of the final data set in most studies.

Although, in the present study, the total number of sequence reads per sample (average = 2989 reads) was not high, we were able to discover a relatively large number (n = 170) of novel miRNA candidates from our sequence reads. Candidates for novel miRNAs continue to be identified, making it likely that the total number of human miRNAs has not yet reached saturation.[8] Our results show that CD34$^+$ leukemic blasts express a wider range of miRNAs than previously appreciated and that overall miRNA expression profiles generally reflect blast karyotype. Such karyotype-specific miRNAs may play a role in the malignant transformation of blasts of the corresponding karyotype, a possibility that needs to be confirmed by analysis of a large number of samples.

It is possible that some of the miRNA candidates identified in our study are not genuine miRNAs but rather degradation products of RNA or DNA. We believe, however, that a substantial proportion of the candidate miRNAs are indeed novel miRNAs because (i) many of them were identified in different samples in different laboratories (in Japan and in the Netherlands), (ii) many of them (together with the surrounding sequences in the genome) are conserved across various species and (iii) the expression of some of them was confirmed by northern blot analysis.

We have identified 170 novel miRNA candidates in, and demonstrated a high level of diversity in miRNA profiles among, leukemic blasts. Our data thus suggest that the miRNA

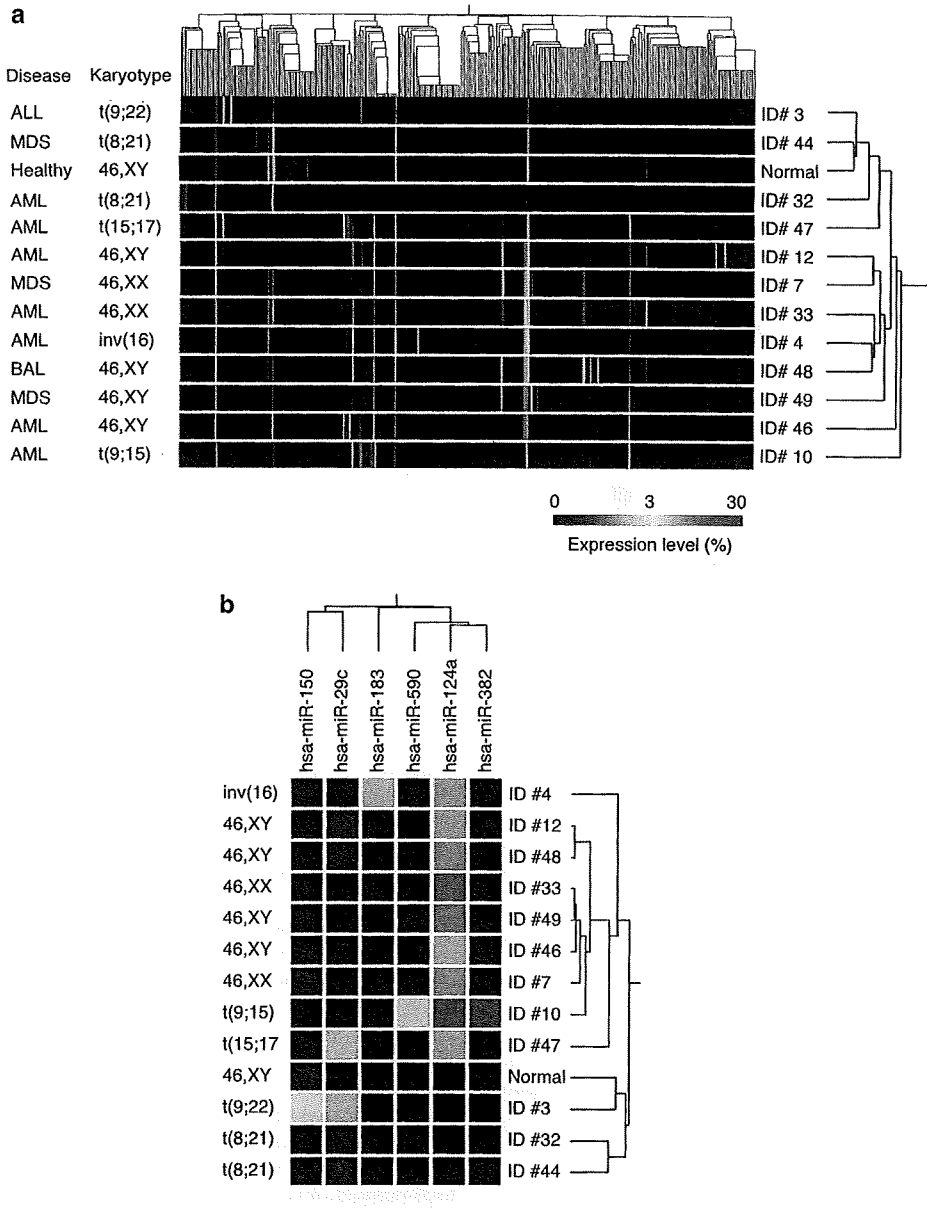

Figure 3 Hierarchical clustering of the study subjects based on miRNA expression profiles. (**a**) Subject tree generated by two-way clustering analysis with the expression profiles of all known and novel miRNAs. Each row corresponds to a separate sample, and each column to a miRNA whose expression is color-coded according to the indicated scale. The disease type and karyotype of each subject are shown at the left. (**b**) Six karyotype-associated miRNAs identified with Student's *t*-test and a false discovery rate of <0.05 were used for two-way clustering analysis as in (**a**). ALL, acute myeloid leukemia; AML, acute lymphoid leukemia; BAL, biphenotypic acute leukemia; MDS, myelodysplastic syndrome; miRNA, microRNA.

repertoire of human leukemias has not yet been exhausted, and they should provide a framework for future studies in this regard.

Note added in proof

Hsj_117 and Hsj_360 have the miRBase accession numbers hsa-miR-590 and hsa-miR-663b, respectively.

Acknowledgements

This study was supported in part by a grant for Third-Term Comprehensive Control Research for Cancer from the Ministry of Health, Labor, and Welfare of Japan as well as by a grant for Scientific Research on Priority Areas 'Applied Genomics' from the Ministry of Education, Culture, Sports, Science and Technology of Japan. The authors declare no competing financial interests.

S Takada[1], Y Yamashita[1], E Berezikov[2], H Hatanaka[1], S-i Fujiwara[1], K Kurashina[1], H Watanabe[1], M Enomoto[1], M Soda[1], YL Choi[1] and H Mano[1,3]
[1]*Division of Functional Genomics, Jichi Medical University, Shimotsukeshi, Tochigi, Japan;*
[2]*Hubrecht Institute, Utrecht, The Netherlands and*
[3]*CREST, Japan Science and Technology Agency, Saitama, Japan*
E-mail: hmano@jichi.ac.jp

References

1 Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004; **116**: 281–297.
2 Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E *et al.* Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci USA* 2002; **99**: 15524–15529.
3 He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S *et al.* A microRNA polycistron as a potential human oncogene. *Nature* 2005; **435**: 828–833.
4 Nelson PT, Baldwin DA, Scearce LM, Oberholtzer JC, Tobias JW, Mourelatos Z. Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat Methods* 2004; **1**: 155–161.
5 Takada S, Berezikov E, Yamashita Y, Lagos-Quintana M, Kloosterman WP, Enomoto M *et al.* Mouse microRNA profiles determined with a new and sensitive cloning method. *Nucleic Acids Res* 2006; **34**: e115.
6 Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E *et al.* Diversity of microRNAs in human and chimpanzee brain. *Nat Genet* 2006; **38**: 1375–1377.
7 Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 2003; **19**: 368–375.
8 Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* 2005; **120**: 21–24.

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu)

Fusion of *ZMIZ1* to *ABL1* in a B-cell acute lymphoblastic leukaemia with a t(9;10)(q34;q22.3) translocation

Leukemia (2008) **22**, 1278–1280; doi:10.1038/sj.leu.2405033; published online 15 November 2007

The *ABL1* gene has been found to be fused to four identified partner genes in haematological malignancies. It is rearranged with *BCR* by the t(9;22)(q34;q11.2) translocation in more than 95% of chronic myeloid leukaemia and in over 25% of adult B-cell acute lymphoblastic leukaemia.[1,2] It is rearranged with *TEL* (also known as *ETV6)* in rare cases of chronic myeloid leukaemia and acute leukaemia.[3] In T-cell acute lymphoblastic leukaemia, *ABL1* can be fused with *NUP214* (a gene located in 9q34) on episomes[4] or with *EML1* by a t(9;14)(q34;q32) translocation.[5] Moreover, a recent publication described a t(1;9)(q24;q34) translocation in a B-cell acute lymphoblastic leukaemia case with a putative RCSD1-ABL1 fusion without molecular confirmation.[6] Here we report the cytogenetic and molecular analysis of a t(9;10)(q34;q23) translocation, from a case of B-lineage ALL, with recombination of *ABL1* to a new partner gene, *ZMIZ1* (zinc-finger MIZ-type containing 1).

The patient is an 18-month-old Japanese girl, with no personal or familial medical history other than bronchiolitis and an episode of atopic dermatitis treated with dermocorticosteroids in November 2006. In December 2006, at the age of 14 months, she presented with pallor, asthenia and fever. Physical examination was found normal. Laboratory investigation showed an abnormal white blood cell count (2.2 $G l^{-1}$ with neutropenia 0 $G l^{-1}$) and non-regenerative anaemia (Hb 3.5 g per 100 ml). Bone marrow analysis showed heterogeneous density with 3–10% of immature cells. Immunophenotyping analysis of the bone marrow sample did not reveal aberrant surface marker expression. No karyotypic or molecular abnormality was detected at the time. Blood culture was positive for alpha-haemolytic streptococcus. The girl was treated for her septicaemia and transfused. The neutropenia (1.4 $G l^{-1}$) persisted for 3 months. In April 2007, she presented with fever. Clinical examination was normal. Blood cell count showed bicytopaenia (neutrophils 0 $G l^{-1}$, Hb 6.6 g per 100 ml). Bone marrow examination showed 90% of CD19+, CD10+, CD34+ CD13−, CD33− lymphoblasts, corresponding to a diagnosis of B-cell acute lymphoblastic leukaemia II, according to the immunological EGIL (European Group for the Immunological Characterization of Acute Leukemias) classification. The patient was then treated abroad, according to the standard risk

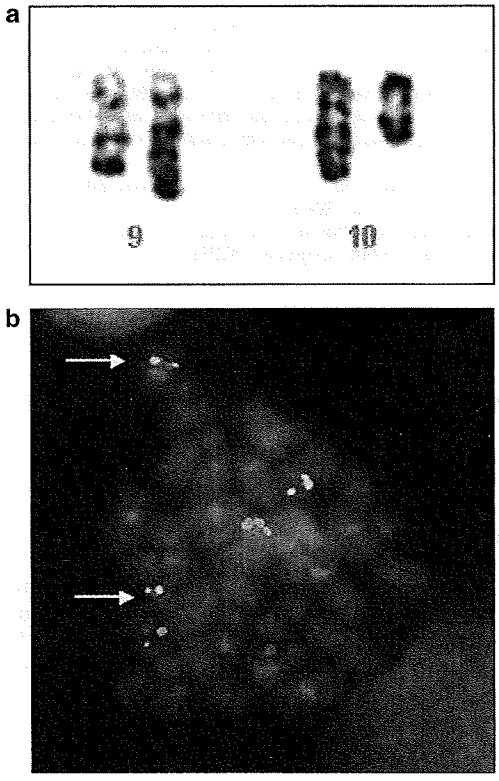

Figure 1 Cytogenetic analysis of patient bone marrow cells. (**a**) Partial karyotype showing the derivative chromosomes 9 and 10. (**b**) FISH using LSI *bcr* (green)/*ABL1* (red) dual-colour probe and the RP11-946M14 BAC clone labelled in coumarin (blue) revealed one normal red signal, one normal blue signal and two red signals fused to two blue signals on derivative chromosomes (arrows).

Leukemia (2008) 22, 1891–1898

www.nature.com/leu

ORIGINAL ARTICLE

High-resolution analysis of chromosome copy number alterations in angioimmunoblastic T-cell lymphoma and peripheral T-cell lymphoma, unspecified, with single nucleotide polymorphism-typing microarrays

S-i Fujiwara[1,2], Y Yamashita[1], N Nakamura[3], YL Choi[1], T Ueno[1], H Watanabe[1], K Kurashina[1], M Soda[1], M Enomoto[1], H Hatanaka[1], S Takada[1], M Abe[4], K Ozawa[2] and H Mano[1,5]

[1]*Division of Functional Genomics, Jichi Medical University, Tochigi, Japan;* [2]*Division of Hematology, Jichi Medical University, Tochigi, Japan;* [3]*Department of Pathology, Tokai University School of Medicine, Kanagawa, Japan;* [4]*Department of Pathology, Fukushima Medical University, Fukushima, Japan and* [5]*CREST, Japan Science and Technology Agency, Saitama, Japan*

Angioimmunoblastic T-cell lymphoma (AILT) and peripheral T-cell lymphoma, unspecified (PTCL-u) are relatively frequent subtypes of T- or natural killer cell lymphoma. To characterize the structural anomalies of chromosomes associated with these disorders, we here determined chromosome copy number alterations (CNAs) and loss of heterozygosity (LOH) at >55 000 single nucleotide polymorphism loci for clinical specimens of AILT ($n=40$) or PTCL-u ($n=33$). Recurrent copy number gain common to both conditions was detected on chromosomes 8, 9 and 19, whereas common LOH was most frequent for a region of chromosome 2. AILT- or PTCL-u-specific CNAs or LOH were also identified at 21 regions, some spanning only a few hundred base pairs. We also identified prognosis-related CNAs or LOH by several approaches, including Cox's proportional hazard analysis. Among the genes that mapped to such loci, a poor prognosis was linked to overexpression of *CARMA1* at 7p22 and of *MYCBP2* at 13q22, with both genes being localized within regions of frequent copy number gain. For a frequent LOH region at 2q34, we also identified IKAROS family zinc-finger 2 cDNAs encoding truncated proteins. Our data indicate that AILT and PTCL-u consist of heterogeneous subgroups with distinct transforming genetic alterations.
Leukemia (2008) **22,** 1891–1898; doi:10.1038/leu.2008.191; published online 17 July 2008
Keywords: T-cell lymphoma; chromosome copy number alterations; loss of heterozygosity; IKZF2

Introduction

Angioimmunoblastic T-cell lymphoma (AILT) and peripheral T-cell lymphoma, unspecified (PTCL-u) are relatively frequent subtypes of T- or natural killer (T/NK) cell lymphoma.[1] Although specific chromosomal translocations[2] and viral infections[3,4] have been associated with subsets of T/NK cell lymphoma, the molecular pathogenesis of these disorders remains obscure in most cases. Furthermore, given that PTCL-u is diagnosed on the basis of patients not having other specific subtypes of PTCL,[5] it likely consists of heterogeneous subgroups of lymphoma. Prognosis of AILT and PTCL-u is generally poor, with a 5-year survival rate of ~30%,[1] and standard treatment strategies for these conditions remain to be established. Characterization of the intrinsic genetic aberrations responsible for these two subtypes of T/NK cell lymphoma and the development of new classification schemes based on such molecular pathogenesis are thus important clinical goals.

In addition to nucleotide mutations and epigenetic abnormalities, structural changes of chromosomes, or chromosomal instability, are important in cancer development.[6] Gene amplification may promote the oncogenic activity of a subset of proto-oncogenes, such as *MYC, ERBB2* and *CCND1*. Conversely, deletion or truncation of tumor suppressor genes may underlie inactivation of their function. Furthermore, loss of heterozygosity (LOH) is frequently observed in the tumor genome; this condition is characterized by the deletion of one allele of a gene either without (copy number (CN) = 1) or with (CN = 2, referred to as uniparental disomy) duplication of the remaining allele. Regions of the genome affected by LOH have been thought to harbor mutated or epigenetically silenced tumor suppressor genes. However, recent evidence indicates that these regions may also harbor activated oncogenes, as demonstrated for mutated *JAK2* in myeloproliferative disorders.[7]

Comparative genomic hybridization (CGH) has been applied to assess chromosome copy number alterations (CNAs) in AILT/PTCL-u. Renedo *et al.*[8] found the most common CN gain on X chromosome in T-cell non-Hodgkin's lymphoma. The same approach for PTCL-u with Zettl *et al.*[9] identified recurrent CN gains on chromosome 7q22-qter, and recurrent CN losses on 5q, 6q, 9p, 10q, 12q and 13q. Array-based CGH with a resolution of >100 kb has also been used to examine AILT/PTCL-u, revealing recurrent CN gains of 11p11–q14, 19 and 22q in AILT, and of 8, 17 and 22q in PTCL-u.[10] Some inconsistency among these data may reflect the genetic heterogeneity in AILT/PTCL-u, and a low-resolution power in conventional or array-based CGH failed to pinpoint the genes essential to these CNAs.

Microarrays originally developed for typing of single nucleotide polymorphisms (SNPs) are now being applied to assess CNAs. Given that SNP-typing arrays are able both to assess heterozygosity or homozygosity along entire chromosomes and to determine the DNA quantity for each chromosome separately,[11] such arrays are able to measure chromosome CN and LOH simultaneously. Furthermore, the recent development of high-density SNP-typing arrays has allowed such measurements to be made at a resolution of <100 kb.

To identify characteristic genomic aberrations for AILT or PTCL-u in a high resolution, we have collected fresh specimens of AILT ($n=40$) and PTCL-u ($n=33$) and subjected them to hybridization with Affymetrix Mapping 50K Hind 240 microarrays (Affymetrix, Santa Clara, CA, USA). Application of

Correspondence: Professor Dr H Mano, Division of Functional Genomics, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke, Tochigi 329-0498, Japan.
E-mail: hmano@jichi.ac.jp
Received 7 November 2007; revised 21 May 2008; accepted 17 June 2008; published online 17 July 2008

bioinformatics to the resulting large data set revealed several novel genomic imbalances and candidate genes that may contribute to the pathogenesis of these two lymphomas.

Patients and methods

Clinical samples

Lymphoma specimens (70 from enlarged lymph nodes; 3 from extranodal tumors) were obtained from 73 patients (40 with AILT, 33 with PTCL-u) who attended Jichi Medical University Hospital or Fukushima Medical University Hospital between 1985 and 2004. The pathology of the specimens was reevaluated on the basis of the revised classification scheme of the World Health Organization (WHO).[5] All 73 specimens, which conformed with the WHO classification of AILT or PTCL-u, were positive for the pan T-cell marker CD3 and negative for the monoclonal integration of human T-cell leukemia virus—I proviral DNA (data not shown). Mean age at diagnosis was 63 years (range, 19–89) and 67% of the patients were men. Most patients had been treated with cyclophosphamide-, doxorubicin-, vincristine- and prednisone-based regimens. Clinical characteristics of the study subjects are summarized in Supplementary Table 1. Informed consent was obtained according to a protocol approved by the ethics committees of Jichi Medical University and Fukushima Medical University Hospital. As normal controls, CD4-positive cells were isolated with the use of CD4 MicroBeads and a Mini-MACS isolation column (Miltenyi Biotec, Auburn, CA, USA) from peripheral blood mononuclear cells of healthy volunteers.

SNP-typing arrays

Genomic DNA was extracted from the lymphoma specimens with the use of a QIAamp DNA Mini kit (Qiagen, Valencia, CA, USA), digested with *Hind*III, ligated to the Adaptor-Hind (Affymetrix) and subjected to hybridization with Mapping 50K Hind 240 arrays (Affymetrix). SNP genotype calls were subsequently determined with GDAS software version 3.0 (Affymetrix) with a confidence score threshold of 0.05. Chromosome CN and the LOH likelihood score at each SNP site were calculated from the hybridization signal intensity and the SNP call with the use of CNAG 2.0 software (http://www.genome.umin.jp).[12] Only CNAG data for autosomes were analyzed, and known copy number variation (CNV) loci[13,14] were excluded from the analysis. We considered chromosome CNAs or LOH reliable only when $\geqslant 2$ contiguous SNP probes yielded the same data. The mean probe signal intensity at diploid chromosomes was inferred from the data of control samples (in which most chromosomes would be expected to be diploid). Chromosome CN and LOH likelihood score data for all autosomal SNP sites are available on request.

Quantitative RT and real-time PCR analysis

Total RNA was isolated from specimens with the use of an RNeasy Mini column (Qiagen) and was subjected to reverse transcription (RT) with PowerScript reverse transcriptase (Clontech, Palo Alto, CA, USA). The amount of specific cDNAs was quantitated by real-time polymerase chain reaction (PCR) analysis with a QuantiTect SYBR Green PCR Kit (Qiagen). The amplification protocol consisted incubations at 94 °C for 15 s, 60 °C for 30 s and 72 °C for 60 s. The incorporation of the SYBR Green dye into the PCR products was monitored in real time with an ABI PRISM 7700 sequence detection system (Applied Biosystems, Foster City, CA, USA), thereby allowing determination of the threshold cycle (C_T) at which exponential amplification of products begins.

The relative abundance of the cDNAs of interest was calculated from the C_T value for each cDNA and that for *ACTB* cDNA. The primer sequences for RT–PCR are shown in Supplementary Table 2.

Nucleotide sequencing

For mutational screening of IKAROS family zinc-finger 2 (*IKZF2*) cDNA, RT–PCR was performed on a subset of lymphoma cDNAs with PrimeSTAR DNA polymerase (Takara Bio, Shiga, Japan) and the primers 5′-AGATCTCCCGACAGAGCTGGA-3′ and 5′-GGTGGGATTGTAAGTGCGGTATT-3′. Amplified PCR products were cloned into the pT7Blue-2 vector (EMD Biosciences, Madison, MI, USA) for nucleotide sequencing. To detect a cDNA for a short isoform of IKZF2, we performed RT–PCR with the primers 5′-ACCTCAAGCACACCCAATGGAC-3′ and 5′-CATCAGCTCAGCCTCCTTCTCA-3′. The resultant *IKZF2* cDNA sequences were compared with the published human *IKZF2* sequence (GenBank accession nos. NM_016260 and NM_001079526).

Statistical analysis

Changes in chromosome CN or gene expression level were evaluated by Student's *t*-test. Hierarchical clustering of the data set was performed with GeneSpring 7.0 software (Agilent Technologies, Santa Clara, CA, USA). Overall survival was estimated by the Kaplan–Meier method and was compared with the logrank test. Multivariate analysis of survival was performed with the Cox proportional hazard model (stepwise regression approach). Unless indicated otherwise, a *P*-value <0.05 was considered statistically significant.

Results

Recurrent chromosome CNAs

Chromosome CN was computationally inferred at 55 700 SNP sites for all autosomes in 73 specimens of AILT or PTCL-u. Hierarchical clustering of all subjects on the basis of these CN profiles revealed that three quarters of the specimens had relatively stable chromosomes, whereas the remaining one quarter had CNAs of various sizes (Figure 1a). Common chromosome gain (CN $\geqslant 3$ in $\geqslant 2$ cases), for example, was identified at 28 243 SNP loci, whereas common chromosome loss (CN $\leqslant 1$ in $\geqslant 2$ cases) was detected at 6479 loci. The prognosis of study subjects with such CNAs (Figure 1a) was significantly worse than that of those without them (Figure 1b), indicative of linkage between these CNAs and the transformation process for AILT or PTCL-u.

In addition, frequent CNAs were readily identified in our data set. Highly recurrent chromosome amplification (CN $\geqslant 4$ in $\geqslant 20$ cases) was apparent at three distinct regions of 8q, 9p and 19q (Table 1). These regions were as small as 175 bp encompassing three contiguous SNP loci at 8q24.11 or 290 bp encompassing another three SNP loci at 19q13.43, demonstrating the high resolution of the SNP array-based CN analysis. Frequent copy number loss (CN of $\leqslant 1$ in $\geqslant 4$ cases), on the other hand, was identified at two distinct regions of 3q and 9p (Table 1). Our CN data further revealed homozygous deletion at these two regions in some individuals (CN = 0 in seven cases at 3q and in three cases at 9p).

Despite the similarity in the profiles for recurrent CNAs between AILT and PTCL-u (Table 1), we examined whether there might be disease-specific CNAs for either of these disorders. Application of Student's *t*-test to the CN profiles for loci with frequent CNAs (those in ⩾10% of subjects) resulted in the isolation of thirteen regions with a disease-dependent CNA (Supplementary Table 3).

Effects of CNAs on gene expression

To examine the relation between CNAs and gene expression, we performed quantitative RT–PCR analysis for genes that mapped within recurrent CNAs. The recurrent loss at 9p21.3 (Table 1) contains the genes for two important inhibitors of cyclin-dependent kinases, *CDKN2A* and *CDKN2B*, which are deleted or epigenetically silenced in a variety of cancer cells.[15] A decrease in DNA content at 9p21.3 was associated with a reduced level of expression of the genes that mapped to this locus: *CDKN2A*, *CDKN2B* and *MTAP* (left panel of Figure 2a).

In addition to the recurrent CNAs shown in Table 1, we also detected a frequent gain in CN at a locus of 7p22.3–22.2 in 30 out of the 73 patients; this locus contains the gene for caspase recruitment domain membrane-associated guanylate kinase protein 1 (*CARMA1*, GenBank accession no. NM_032415). Overexpression of *CARMA1* has been demonstrated in B-cell lymphoma and adult T-cell leukemia or lymphoma.[16,17] As demonstrated in the right panel of Figure 2a, an increase in CN for *CARMA1* was associated with an increase in the amount of the corresponding mRNA, albeit with a marginal statistical significance ($P=0.053$).

We then examined whether the altered expression of these genes influenced the survival of the affected individuals. Consistent with previous results for other hematologic malignancies,[18,19] our data revealed a negative impact of a reduced level of *CDKN2A* expression on the clinical outcome of AILT or PTCL-u (Figure 2b). In addition, individuals with AILT or PTCL-u showing an increase in *CARMA1* expression had a poorer prognosis than did those without such an increase.

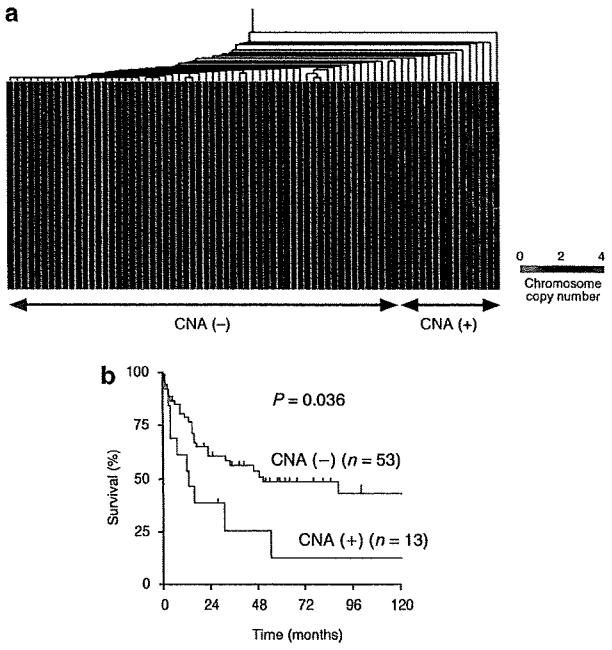

Figure 1 Chromosome copy number alterations (CNAs) in the genome of angioimmunoblastic T-cell lymphoma (AILT) or peripheral T-cell lymphoma, unspecified (PTCL-u). (**a**) Hierarchical clustering analysis of the study subjects ($n=73$) on the basis of the inferred copy number (CN) for all autosomal single nucleotide polymorphism (SNP) sites in the lymphoma specimens. CN is color coded according to the indicated scheme. SNP sites are ordered on the basis of their physical position from top to bottom. The patients could be subdivided into those with or without CNAs as indicated at the bottom. (**b**) The survival of the two groups of patients classified on the basis of the absence or presence of CNAs was compared by Kaplan–Meier analysis, with the *P*-value calculated by the logrank test.

Recurrent LOH

We next calculated the LOH likelihood score[12] at each SNP site. By direct sequencing of some of the genomic regions with a high LOH likelihood score, we determined that a score of ⩾20 was likely to be a reliable indicator of the presence of LOH (data not shown). We therefore used this value as a threshold for LOH in the following analyses.

Many ($n=42\,926$) of the 55 700 SNP sites were found to have an LOH likelihood score of ⩾20 in ⩾2 patients in our cohort. Among these SNP sites, common LOH (LOH likelihood score of ⩾20 in ⩾10% of cases) was apparent at 3093 loci distributed throughout most chromosomes (Figure 3). The most frequent region of LOH (LOH likelihood score of ⩾20 in 22 samples) was an ~440-kb region at 2q32.3 that includes 13 contiguous SNP loci.

We also screened for genomic loci whose LOH status was significantly linked to the diagnosis of AILT or PTCL-u. With a threshold *P*-value 0.001 (Student's *t*-test), we identified eight regions (each consisting of ⩾2 contiguous SNP loci) that mapped to four distinct chromosomes (Table 2). All of these

Table 1 Recurrent CNAs in AILT or PTCL-u specimens

CNA type	*Chromosome*	*Nucleotide position*	*Mapped genes*	*Affected no. of samples*		
				Total (n=73)	*AILT (n=40)*	*PTCL-u (n=33)*
Gain (CN of ⩾4 in ⩾20 cases)						
	8	118 306 152–118 306 326	No genes	20	9	11
	9	10 638 555–10 722 021	No genes	36	24	12
	19	61 752 129–61 752 418	*ZFP28*	20	13	7
Loss (CN of ⩽1 in ⩾4 cases)						
	3	170 709 305–170 709 392	*MDS1*[a]	8	4	4
	9	21 762 317–22 072 375	*MTAP1*, *CDKN2A*, *CDKN2B*	4	1	3

Abbreviations: AILT, angioimmunoblastic T-cell lymphoma; CN, copy number; CNA, copy number alterations; PTCL-u, peripheral T-cell lymphoma, unspecified.
[a]The region with a frequent CN loss at the *MDS1* locus is distinct from the reported CNV region within *MDS1*.[13,14]

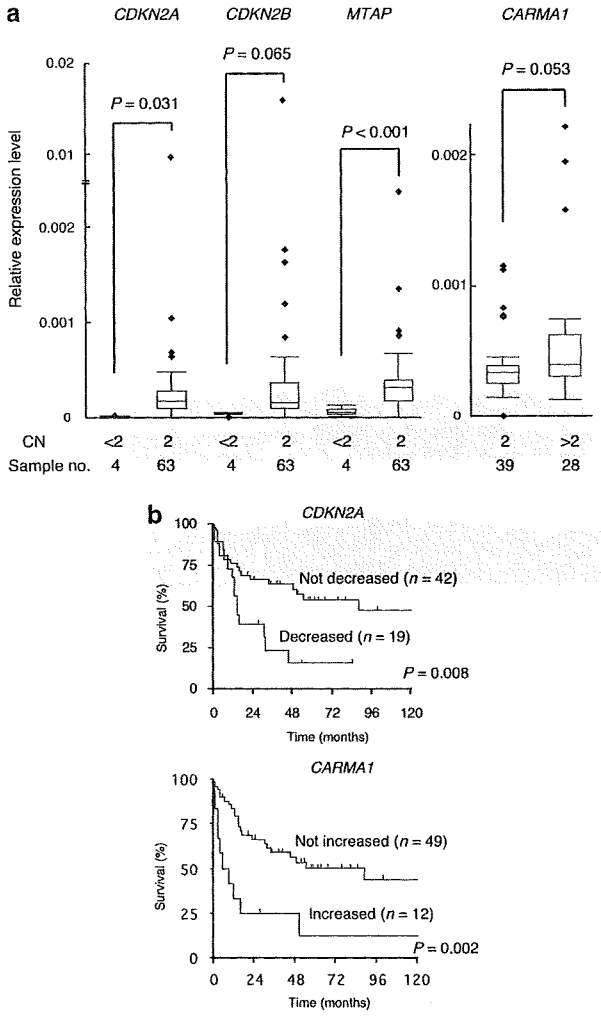

Figure 2 Influence of copy number alterations (CNAs) on candidate gene expression. (**a**) Expression levels of *CDKN2A*, *CDKN2B*, *MTAP* or caspase recruitment domain membrane-associated guanylate kinase protein 1 (*CARMA1*) relative to that of *ACTB* are shown in a box plot for the subjects (in the current cohort for single nucleotide polymorphism (SNP)-typing) with or without CNAs for the corresponding genes. The difference in expression level for each comparison was evaluated by Student's *t*-test. (**b**) The prognosis of patients with or without a reduced level of *CDKN2A* expression (*CDKN2A*/*ACTB* cDNA ratio of <0.0001) was compared by Kaplan–Meier analysis, with the *P*-value calculated by the logrank test (upper panel). Prognosis was similarly compared between individuals with or without an increased level of *CARMA1* expression, with such an increase defined as a relative expression level of more than the mean + 1.0 s.d. of that in the subjects without the copy number (CN) gain at the *CARMA1* locus (lower panel).

disease-associated LOH loci had a normal chromosome CN of 2, indicative of uniparental disomy at these loci.

Recently, aberrant expression of CD10 antigen has been reported for AILT cells.[20] We thus examined whether there are CNA/LOH, in our data set, related to such CD10-postive lymphoma cells. Immunohistostaining for CD10 was conducted among 64 cases in our cohort, revealing 21 cases positive for CD10 (Supplementary Table 1). Statistical analysis to detect CNAs associated with CD10-positive cases have identified one region of ~220 kb at chromosome 7 containing *GPR37* and

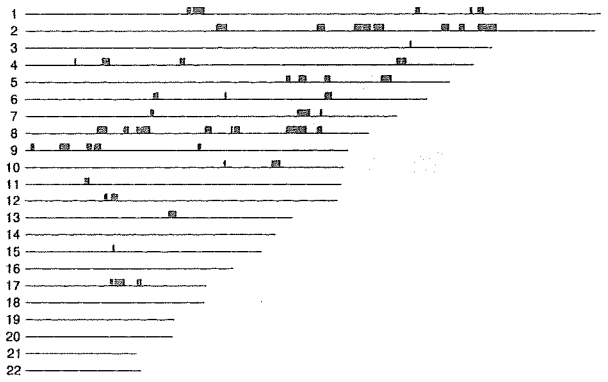

Figure 3 Distribution of recurrent loss of heterozygosity (LOH). Single nucleotide polymorphism (SNP) loci with a recurrent LOH (LOH likelihood score of ⩾20 in ⩾10% of subjects) are indicated by blue bars in chromosome views. Chromosome numbers are shown at the left.

Table 2 Comparison of LOH likelihood profiles between AILT and PTCL-u

Chromosome	*Nucleotide position*	*Mapped genes*	*P-value*
2	31 171 920–31 217 236	*GALNT14*	$<5.4 \times 10^{-4}$
2	140 830 794–140 912 732	*LPR1B*	$<1.6 \times 10^{-4}$
2	141 385 148–141 387 720	*LPR1B*	$<3.2 \times 10^{-4}$
8	10 667 288–10 689 316	*PINX1*	$<9.1 \times 10^{-4}$
8	19 844 621–19 908 967	*LPL*	$<9.8 \times 10^{-4}$
11	80 723 837–80 727 426	No genes	$<1.9 \times 10^{-4}$
11	80 760 044–80 785 374	No genes	$<9.1 \times 10^{-4}$
12	57 759 034–57 781 381	No genes	$<9.9 \times 10^{-4}$

POT1 genes (Student's *t*-test, $P<0.001$), whereas a similar analysis for the LOH status found nine distinct regions on chromosomes 1, 4, 5, 6, 7 and 18 (Supplementary Table 4).

Novel isoforms of IKZF2

To isolate additional candidate genes for AILT or PTCL-u, we performed nucleotide mutation screening of known cancer-related genes located within the identified LOH regions. Extensive cDNA sequencing for these genes revealed a cDNA for a novel isoform of IKZF2, also known as Helios, in a subset of subjects. *IKZF2* maps to chromosome 2q34, for which a high LOH likelihood score (⩾20) was identified in seven specimens (data not shown).

IKZF2 belongs to the IKAROS family of transcriptional factors, which are important regulators of lymphocyte development,[21,22] and short isoforms of IKZF2 have been reported for the malignant cells of adult T-cell leukemia or lymphoma[23] and T-cell acute lymphoblastic leukemia.[24] In our cohort, RT–PCR amplification of the entire coding region of the *IKZF2* mRNA detected a product in five of the seven study subjects with LOH at the *IKZF2* locus (Figure 4a). One of these products (from patient ID no. 1) was ~1.3 kb in size and apparently smaller than the others. Nucleotide sequencing of this cDNA revealed that it did not contain exons 3 and 4 of *IKZF2* (Figures 4b, c) and therefore encodes a protein that lacks 145 amino acids (including the first three zinc-finger domains) compared with the wild-type protein and has a Thr-to-Met substitution at amino-acid position 45 (the exon 2–5 boundary) (Figure 4b).