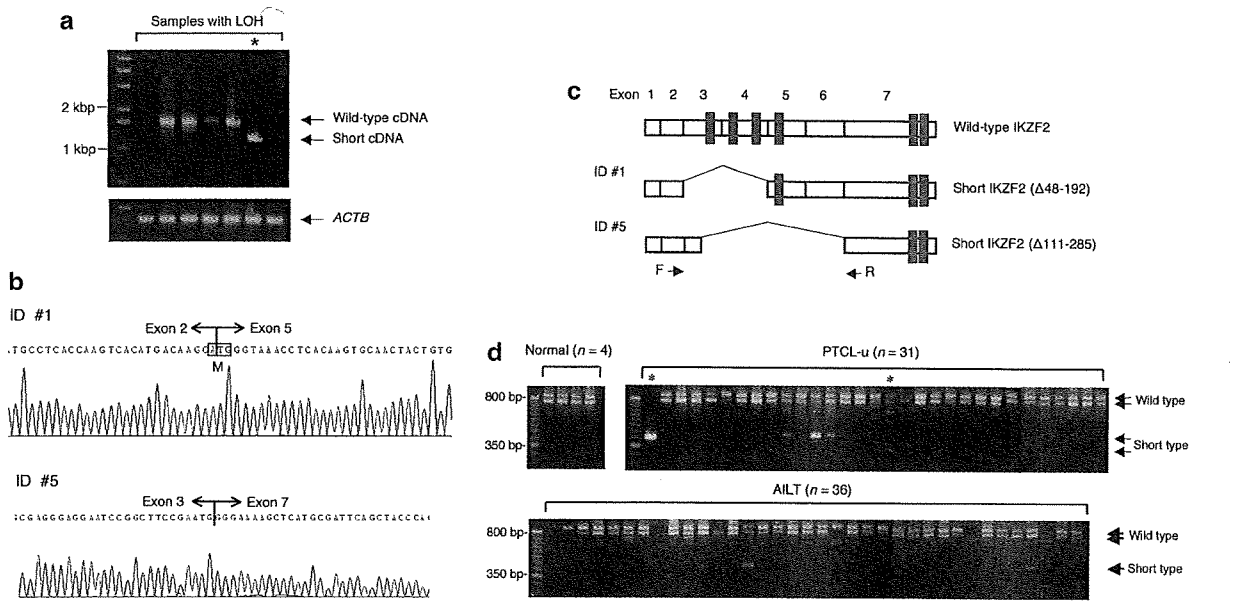

Figure 4 Identification of cDNAs for short isoforms of IKAROS family zinc-finger 2 (IKZF2). (**a**) Reverse transcription (RT)–PCR amplification of the entire coding region of *IKZF2* mRNA (upper panel) or of a portion of *ACTB* mRNA (lower panel) from seven subjects with loss of heterozygosity (LOH) at the *IKZF2* locus. The PCR products were fractionated by electrophoresis on a 0.8% agarose gel. The left lane contains DNA size markers (1-kb ladder). Whereas the predicted ~1.7-kb product of wild-type *IKZF2* cDNA was apparent in four specimens, a single ~1.3-kb product was identified in one patient (red asterisk, ID no. 1). (**b**) Sequence analysis of the exon 2–5 boundary of the *IKZF2* cDNA isolated from patient ID no. 1 or of the exon 3–7 boundary of that isolated from patient ID no. 5. M, methionine. (**c**) Schematic representation of the structure of the wild-type IKZF2 protein and that of the short isoforms identified in patients ID nos. 1 and 5. The positions of zinc-finger domains (red boxes) and of PCR primers (arrows) for amplification of exons 2–7 of *IKZF2* cDNA are also indicated. (**d**) RT–PCR amplification of exons 2–7 of *IKZF2* cDNA from patients with peripheral T-cell lymphoma, unspecified (PTCL-u) or angioimmunoblastic T-cell lymphoma (AILT) or from T cells of normal controls. The products were fractionated by electrophoresis on a 3% agarose gel, with DNA size markers (50-bp ladder) included in the leftmost lanes. Patients ID nos. 1 and 5 are indicated by the red and blue asterisks, respectively. The positions of products corresponding to wild type and short forms of IKZF2 are indicated on the right.

We next examined whether the mRNA for this novel isoform of *IKZF2* identified in the present study was also present in other patients or healthy individuals with the use of RT–PCR to amplify exons 2 to 7 of *IKZF2* cDNA (Figure 4c). Full-length cDNAs for *IKZF2* variant 1 (GenBank accession no. NM_016260) and variant 2 (NM_001079526) were detected in all normal T cells and in most of the lymphoma samples (Figure 4d). However, the cDNA for the short isoform was identified as a 414-bp product in seven patients (four with PTCL-u, three with AILT), including the one in whom this isoform was initially identified. This latter individual (patient ID no. 1) did not yield RT–PCR products corresponding to wild-type *IKZF2* cDNAs. Given that LOH was apparent at the *IKZF2* locus in this patient, the lymphoma cells likely harbor only a single *IKZF2* allele, which produces the truncated mRNA.

A novel cDNA fragment of 321 bp was further detected in the PTCL-u sample from patient ID no. 5 (Figure 4d). Nucleotide sequencing of this product revealed it to encode an IKZF2 protein that lacks 175 amino acids corresponding to a portion of exon 3 and all of exons 4 to 6 (Figures 4b, c).

Prognosis-related CNAs or LOH

To examine whether any of the CNAs or LOH regions identified in our data set are related to clinical outcome, we searched for prognosis-associated changes with several approaches. Given the many recurrent (at various frequencies) CNAs or LOH regions in the data set, we first examined whether some of these changes (observed in $\geqslant 5$ samples) were preferentially present in patients who died within a year after diagnosis compared with those who survived for >1 year. For these potentially outcome-related genomic regions, prognosis was then compared between the individuals with or without each CNA or LOH site with the logrank test. CN gain at chromosomes 2 or 5 was found to be linked to poor prognosis (Supplementary Table 5; Supplementary Figure 1a). In addition, LOH at chromosomes 8 or 9 also had a negative impact on survival (Supplementary Table 5).

We next directly searched for genomic imbalances linked to poor prognosis by applying Cox's proportional hazard regression analysis to the chromosome CN profile for SNP loci with CNAs in $\geqslant 10\%$ of subjects, resulting in the isolation of 12 regions with a *P*-value <0.05 (Supplementary Table 6; Supplementary Figure 1). The β-score in the Cox analysis for all these regions was positive, indicating that CN gain at any of them is linked to a poor prognosis.

Further, Cox's analysis of CNAs only for the AILT data identified a CN gain at 13q22.3 that was significantly related ($P=0.025$) to poor clinical outcome (data not shown). This region spans only two contiguous SNP sites (corresponding to a distance of 237 bp) that map to a position 30 kb upstream of the MYC-binding protein 2 gene (*MYCBP2*, GenBank accession no. NM_015057). An increase in chromosome copy number at these SNP loci was further confirmed by quantitative PCR analysis (data not shown).

MYCBP2 is a large protein that binds specifically to Myc,[25] but whether it is involved in the transformation process of AILT is unknown. We detected a trend of an increase in the level of *MYCBP2* expression in the specimens of AILT patients with a CN gain at this locus compared to that in those without such a gain (Figure 5a), but without a statistical significance (Student's *t*-test,

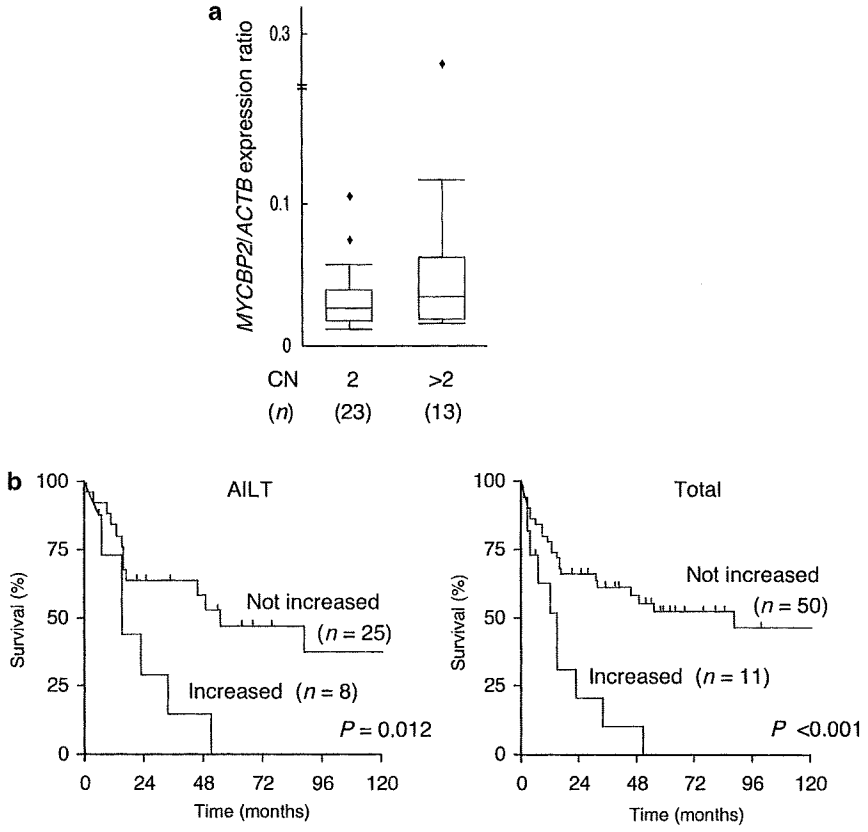

Figure 5 Poor prognosis associated with increased MYC-binding protein 2 (*MYCBP2*) expression. (**a**) The level of *MYCBP2* expression relative to that of *ACTB* expression is shown in a box plot for angioimmunoblastic T-cell lymphoma (AILT) patients ($n=36$) with or without a copy number (CN) gain at the *MYCBP2* locus. (**b**) Comparison of the prognosis of AILT (left panel) or all (right panel) patients with or without an increased level of *MYCBP2* expression (defined by a relative expression level of more than the mean + 1.0 s.d. of that in the corresponding subjects without the CN gain at the *MYCBP2* locus). The *P*-values were calculated by the logrank test.

$P=0.23$). Moreover, comparison of the survival of AILT patients with or without an increased expression of *MYCBP2* revealed a significantly worse prognosis for the former (Figure 5b). A poor prognosis for patients with increased *MYCBP2* expression was also apparent for the entire AILT and PTCL-u cohort (Figure 5b). These results suggest that *MYCBP2* is a candidate for the transformation-associated gene that maps to the 13q22.3 locus.

We also applied Cox's regression analysis to the LOH likelihood score data. For the PTCL-u data set, one region of ~9.2 Mbp at chromosome 8 (nucleotide positions: 41 865 249–51 050 357) was identified as being significantly related to poor clinical outcome (Supplementary Figure 1c).

Discussion

We have determined chromosome CN as well as LOH likelihood throughout the AILT or PTCL-u genome and have identified several novel recurrent and prognosis-related changes, some of which affect candidate genes for lymphomagenesis. In contrast to previous genomic analyses of T-cell lymphoma with CGH,[8–10,26] our study based on SNP-typing arrays was able to identify many small regions (<100 kb) with CNAs or LOH in the genome of lymphoma cells.

We detected CNAs or LOH regions at a similar frequency in AILT and PTCL-u, consistent with previous data showing a common chromosome gain at 11q13 in the two disorders[10] or common CNAs in PTCL-u and ALK-negative anaplastic large-cell lymphoma.[9] Few previous studies have described LOH in AILT or PTCL-u, possibly in part because of a high frequency of uniparental disomy in these conditions, as revealed in our data set (especially in the regions of recurrent LOH), that may be undetectable by conventional array-based CGH.

As tumor cell proportion in affected lymph nodes varies substantially among AILT/PTCL-u specimens, we tried to examine if such tumor cell proportion affects the CNA/LOH data. As shown in Supplementary Table 1, samples were classified into three subgroups; specimens with tumor cells composing <30% of lymph nodes were assigned to the 'L' group, whereas those with 30~60% were to the 'M' group, and those with ⩾60% to the 'H' group. We then tested whether such tumor cell proportion was linked to the detection of chromosome unstable cases where aberrant chromosome copy numbers (other than 2) were found in >10% of the SNP probes. Although detection of such cases was not statistically different between the L and M subgroups, CNA-positive specimens in the H group was significantly more frequent than that in the M or L group (Fisher's exact test, $P=0.002$ for each comparison). Therefore, it is possible that tumor cell proportion in the specimens has significantly affected some parts of data set. However, it should be noted that (1) the high sensitivity of the CNA/LOH-calculation algorithm allows the detection of such changes among tumor samples contaminated with 70–80% of normal cells[27] and (2) tumor cell proportion in the affected lymph nodes

of AILT/PTCL-u may remain stable throughout stage progression.[28,29] Therefore, different frequency of CNA-positive cases may be an intrinsic property to the L/M/H subgroups of AILT/PTCL-u. Large-scale CGH studies with purified AILT/PTCL-u tumor cells (by using laser-capture microdissection system, for instance) would help to address these issues.

One of the goals of our study was to identify novel disease-associated genes in AILT or PTCL-u. Among the recurrent CNA loci, we identified *CARMA1* as a candidate gene for mediating the contribution of a 7p22 gain to lymphomagenesis. CARMA1 interacts with BCL10 and MALT1 and thereby mediates activation of nuclear factor (NF)-κB induced by stimulation of the T- or B-cell receptor.[10] Given that NF-κB frequently is activated and contributes to carcinogenesis in many tumor types,[30] activation of this transcription factor as a result of CARMA1 overexpression may also be important in disease progression and poor outcome of AILT or PTCL-u.

In addition, from a recurrent LOH locus, we identified cDNAs for novel short isoforms of IKZF2. Forced expression of full-length *IKZF2* cDNA results in inhibition of T-cell development at an early stage.[31] A cDNA encoding a truncated, dominant-negative form of IKZF2 with impaired DNA-binding activity has been associated with adult T-cell leukemia or lymphoma[23] and T-cell acute lymphoblastic leukemia.[24] The development of T-cell lymphoma in mice expressing this dominant-negative form of IKZF2 provided further support for its clinical relevance.[31] Although the structure of our isoform of IKZF2 is different from that of the previously described short one, which lacks the second to fourth zinc-finger domains,[23] both forms have only one zinc-finger domain in the N-terminal DNA-binding region of the protein. Given that isoforms of IKAROS family members with fewer than two N-terminal zinc-finger domains act in a dominant-negative manner,[32] both short isoforms of IKZF2 likely function as inhibitors of IKAROS family proteins. Given the transforming potential of a previously identified truncated form of IKZF2,[31] our data support the direct involvement of IKZF2 in transformation for a subset of AILT or PTCL-u.

In addition, an increased level of *MYCBP2* expression was associated with reduced survival time in AILT patients. The region of MYC that mediates binding to MYCBP2 is essential for the transactivation activity of MYC and is frequently mutated in Burkitt's and AIDS-related lymphomas.[25,33] Given that such mutations in MYC impair its ubiquitination and degradation,[33] overexpression of MYCBP2 may similarly hinder the access of ubiquitination enzymes or the proteasome to MYC and thereby promote its accumulation.

As shown in Supplementary Table 1, we have examined the expression of NK cell markers in the tumor cells. Expression of cell-surface CD56 was, for instance, tested with immunohistostaining procedures among 64 samples, and was found in only three cases. Similarly, presence of Epstein–Barr virus genome was analyzed among 65 cases, leading to the isolation of only two cases carrying the genome. It was thus difficult to draw statistically meaningful conclusions for CNA/LOH related to these small subgroups.

RefSeq genes may not be the only potential players in carcinogenesis. Large noncoding RNAs, for example, contribute to methylation of the genome,[34] whereas short noncoding RNAs, such as microRNAs, are implicated in regulation of cell growth and differentiation.[35] Such transcripts, despite their inability to synthesize proteins, may thus be involved in the development of AILT or PTCL-u. Given that the identification and annotation of these noncoding RNAs are still at an early stage,[36,37] many loci identified in our study may contain genes for as yet undiscovered noncoding RNAs, and these transcripts may participate in carcinogenesis.

In conclusion, our study has provided a large-scale, detailed analysis of CNAs and LOH in AILT and PTCL-u, and has identified candidates for lymphomagenesis-related genes. Our data set should prove to be a useful platform for further definition, from the viewpoint of chromosome abnormalities, of these clinical entities.

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.

Disclosure/Conflict of interest

The authors declare no competing financial interests.

References

1 Rudiger T, Weisenburger DD, Anderson JR, Armitage JO, Diebold J, MacLennan KA *et al.* Peripheral T-cell lymphoma (excluding anaplastic large-cell lymphoma): results from the Non-Hodgkin's Lymphoma Classification Project. *Ann Oncol* 2002; **13**: 140–149.
2 Fischer P, Nacheva E, Mason DY, Sherrington PD, Hoyle C, Hayhoe FG *et al.* A Ki-1 (CD30)-positive human cell line (Karpas 299) established from a high-grade non-Hodgkin's lymphoma, showing a 2;5 translocation and rearrangement of the T-cell receptor beta-chain gene. *Blood* 1988; **72**: 234–240.
3 Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci USA* 1980; **77**: 7415–7419.
4 Harabuchi Y, Yamanaka N, Kataura A, Imai S, Kinoshita T, Mizuno F *et al.* Epstein–Barr virus in nasal T-cell lymphomas in patients with lethal midline granuloma. *Lancet* 1990; **335**: 128–130.
5 Jaffe ES, Harris NL, Stein H, Vardiman JW (eds). *Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues.* IARC Press: Lyon, 2001, 225–229.
6 Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* 1998; **396**: 643–649.
7 Kralovics R, Passamonti F, Buser AS, Teo SS, Tiedt R, Passweg JR *et al.* A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* 2005; **352**: 1779–1790.
8 Renedo M, Martinez-Delgado B, Arranz E, Garcia M, Urioste M, Martinez-Ramirez A *et al.* Chromosomal changes pattern and gene amplification in T cell non-Hodgkin's lymphomas. *Leukemia* 2001; **15**: 1627–1632.
9 Zettl A, Rudiger T, Konrad MA, Chott A, Simonitsch-Klupp I, Sonnen R *et al.* Genomic profiling of peripheral T-cell lymphoma, unspecified, and anaplastic large T-cell lymphoma delineates novel recurrent chromosomal alterations. *Am J Pathol* 2004; **164**: 1837–1848.
10 Thorns C, Bastian B, Pinkel D, Roydasgupta R, Fridlyand J, Merz H *et al.* Chromosomal aberrations in angioimmunoblastic T-cell lymphoma and peripheral T-cell lymphoma unspecified: a matrix-based CGH approach. *Genes Chromosomes Cancer* 2007; **46**: 37–44.
11 Matsuzaki H, Dong S, Loi H, Di X, Liu G, Hubbell E *et al.* Genotyping over 100 000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* 2004; **1**: 109–111.
12 Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* 2005; **65**: 6071–6079.

13 Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD *et al.* Global variation in copy number in the human genome. *Nature* 2006; **444**: 444–454.
14 Komura D, Shen F, Ishikawa S, Fitch KR, Chen W, Zhang J *et al.* Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res* 2006; **16**: 1575–1584.
15 Drexler HG. Review of alterations of the cyclin-dependent kinase inhibitor INK4 family genes p15, p16, p18 and p19 in human leukemia-lymphoma cells. *Leukemia* 1998; **12**: 845–859.
16 Oshiro A, Tagawa H, Ohshima K, Karube K, Uike N, Tashiro Y *et al.* Identification of subtype-specific genomic alterations in aggressive adult T-cell leukemia/lymphoma. *Blood* 2006; **107**: 4500–4507.
17 Nakamura S, Nakamura S, Matsumoto T, Yada S, Hirahashi M, Suekane H *et al.* Overexpression of caspase recruitment domain (CARD) membrane-associated guanylate kinase 1 (CARMA1) and CARD9 in primary gastric B-cell lymphoma. *Cancer* 2005; **104**: 1885–1893.
18 Takasaki Y, Yamada Y, Sugahara K, Hayashi T, Dateki N, Harasawa H *et al.* Interruption of p16 gene expression in adult T-cell leukaemia/lymphoma: clinical correlation. *Br J Haematol* 2003; **122**: 253–259.
19 Schmitt CA, McCurrach ME, de Stanchina E, Wallace-Brodeur RR, Lowe SW. INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53. *Genes Dev* 1999; **13**: 2670–2677.
20 Attygalle A, Al-Jehani R, Diss TC, Munson P, Liu H, Du MQ *et al.* Neoplastic T cells in angioimmunoblastic T-cell lymphoma express CD10. *Blood* 2002; **99**: 627–633.
21 Kelley CM, Ikeda T, Koipally J, Avitahl N, Wu L, Georgopoulos K *et al.* Helios, a novel dimerization partner of Ikaros expressed in the earliest hematopoietic progenitors. *Curr Biol* 1998; **8**: 508–515.
22 Hahm K, Cobb BS, McCarty AS, Brown KE, Klug CA, Lee R *et al.* Helios, a T cell-restricted Ikaros family member that quantitatively associates with Ikaros at centromeric heterochromatin. *Genes Dev* 1998; **12**: 782–796.
23 Tabayashi T, Ishimaru F, Takata M, Kataoka I, Nakase K, Kozuka T *et al.* Characterization of the short isoform of Helios overexpressed in patients with T-cell malignancies. *Cancer Sci* 2007; **98**: 182–188.
24 Nakase K, Ishimaru F, Fujii K, Tabayashi T, Kozuka T, Sezaki N *et al.* Overexpression of novel short isoforms of Helios in a patient with T-cell acute lymphoblastic leukemia. *Exp Hematol* 2002; **30**: 313–317.
25 Guo Q, Xie J, Dang CV, Liu ET, Bishop JM. Identification of a large Myc-binding protein that contains RCC1-like repeats. *Proc Natl Acad Sci USA* 1998; **95**: 9172–9177.
26 Melendez B, Diaz-Uriarte R, Cuadros M, Martinez-Ramirez A, Fernandez-Piqueras J, Dopazo A *et al.* Gene expression analysis of chromosomal regions with gain or loss of genetic material detected by comparative genomic hybridization. *Genes Chromosomes Cancer* 2004; **41**: 353–365.
27 Yamamoto G, Nannya Y, Kato M, Sanada M, Levine RL, Kawamata N *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am J Hum Genet* 2007; **81**: 114–126.
28 Niitsu N, Okamoto M, Nakamine H, Aoki S, Motomura S, Hirano M. Clinico-pathologic features and outcome of Japanese patients with peripheral T-cell lymphomas. *Hematol Oncol* 2008; e-pub ahead of print.
29 Attygalle AD, Kyriakou C, Dupuis J, Grogg KL, Diss TC, Wotherspoon AC *et al.* Histologic evolution of angioimmunoblastic T-cell lymphoma in consecutive biopsies: clinical correlation and insights into natural history and disease progression. *Am J Surg Pathol* 2007; **31**: 1077–1088.
30 Jost PJ, Ruland J. Aberrant NF-kappaB signaling in lymphoma: mechanisms, consequences, and therapeutic implications. *Blood* 2007; **109**: 2700–2707.
31 Zhang Z, Swindle CS, Bates JT, Ko R, Cotta CV, Klug CA. Expression of a non-DNA-binding isoform of Helios induces T-cell lymphoma in mice. *Blood* 2007; **109**: 2190–2197.
32 Sun L, Heerema N, Crotty L, Wu X, Navara C, Vassilev A *et al.* Expression of dominant-negative and mutant isoforms of the antileukemic transcription factor Ikaros in infant acute lymphoblastic leukemia. *Proc Natl Acad Sci USA* 1999; **96**: 680–685.
33 Bahram F, von der Lehr N, Cetinkaya C, Larsson LG. c-Myc hot spot mutations in lymphomas result in inefficient ubiquitination and decreased proteasome-mediated turnover. *Blood* 2000; **95**: 2104–2110.
34 Chang SC, Tucker T, Thorogood NP, Brown CJ. Mechanisms of X-chromosome inactivation. *Front Biosci* 2006; **11**: 852–866.
35 Carrington JC, Ambros V. Role of microRNAs in plant and animal development. *Science* 2003; **301**: 336–338.
36 Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N *et al.* The transcriptional landscape of the mammalian genome. *Science* 2005; **309**: 1559–1563.
37 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.

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

Review Article

Non-solid oncogenes in solid tumors: *EML4–ALK* fusion genes in lung cancer

Hiroyuki Mano[1]

Division of Functional Genomics, Jichi Medical University, 3311-1 Yakushiji, Shimotsukeshi, Tochigi 329-0498, Japan

(Received July 22, 2008/Revised August 13, 2008/Accepted August 14, 2008/Online publication November 20, 2008)

It is generally accepted that recurrent chromosome translocations play a major role in the molecular pathogenesis of hematological malignancies but not of solid tumors. However, chromosome translocations involving the e26 transformation-specific sequence transcription factor loci have been demonstrated recently in many prostate cancer cases. Furthermore, through a functional screening with retroviral cDNA expression libraries, we have discovered the fusion-type protein tyrosine kinase echinoderm microtubule-associated protein like-4 (EML4)–anaplastic lymphoma kinase (ALK) in non-small cell lung cancer (NSCLC) specimens. A recurrent chromosome translocation, inv(2)(p21p23), in NSCLC generates fused mRNA encoding the amino-terminal half of EML4 ligated to the intracellular region of the receptor-type protein tyrosine kinase ALK. EML4–ALK oligomerizes constitutively in cells through the coiled coil domain within the EML4 region, and becomes activated to exert a marked oncogenicity both *in vitro* and *in vivo*. Break and fusion points within the *EML4* locus may diverge in NSCLC cells to generate various isoforms of EML4–ALK, which may constitute ~5% of NSCLC cases, at least in the Asian ethnic group. In the present review I summarize how detection of *EML4–ALK* cDNA may become a sensitive diagnostic means for NSCLC cases that are positive for the fusion gene, and discuss whether suppression of ALK enzymatic activity could be an effective treatment strategy against this intractable disorder. (*Cancer Sci* 2008; 99: 2349–2355)

Chromosome translocation is the most prevalent form of somatic changes in the cancer genome, occupying nearly three-quarters of all genetic change analyzed in cancer cells.[1] Such translocations may lead to the generation of novel fusion genes at the ligation points of chromosomes, or may juxtapose growth-promoting genes to aberrant promoter or enhancer fragments, resulting in dysregulated expression of the genes. In either case, such fusion genes or wild-type genes with altered expression may participate directly in the malignant transformation of cells that harbor chromosome translocations.

An archetypal example of such tumor-related translocations is t(9;22), which gives rise to the *breakpoint cluster regio (BCR)-Abelson murine leukemia viral oncogene homolog 1 (ABL1)* fusion gene in chronic myeloid leukemia (CML) and acute lymphoblastic leukemia (ALL).[2] Ligation to BCR constitutively elevates the protein tyrosine kinase (PTK) activity of ABL1, and forced expression of *BCR–ABL1* in the hematopoietic system induces CML and ALL in mice,[3,4] proving that BCR–ABL1 plays a pivotal role in the pathogenesis of such leukemias. Also, molecular detection of *BCR–ABL1* and the development of compounds to suppress BCR–ABL1 enzymatic activity have significantly changed the way we diagnose and treat individuals with CML. Because reverse transcription (RT)–polymerase chain reaction (PCR) can detect *BCR–ABL1* fusion transcripts in almost 100% of individuals with CML, even among those without the characteristic t(9;22) translocation, molecular detection of *BCR–ABL1* has become a standard technique used to diagnose CML. Given the very high sensitivity of PCR, such a strategy is also effective to follow up the tumor burden of leukemia in the patients.[5]

Further, the chemical compound STI571, which suppresses ABL1 kinase activity, has substantially prolonged the survival of patients at the chronic phase of CML, and achieved a higher probability of complete cytogenetic response among them compared to that with the previous treatment regimens.[6,7] Therefore, if translocation-mediated fusion genes encode activated enzymes with direct oncogenic potential, targeting such enzymes could provide a feasible approach to treat individuals harboring the corresponding fusion genes.

However, these fusion-type oncogenes have been reported frequently only in hematological malignancies, and not in solid tumors (especially in epithelial tumors).[8] It has been therefore widely assumed that balanced chromosome cytogenetic aberrations (and the resulting fusion genes) may be rare in the latter conditions. As shown in Table 1, for instance, the incidence of new cases with solid tumors in the USA is ~11 times larger than for those with hematological malignancies.[9] However, the number of recurrent balanced cytogenetic aberrations (RBA) in solid tumors ($n = 125$) is only a quarter of that in hematological malignancies ($n = 495$) worldwide,[10] suggesting that RBA are indeed characteristic of hematological malignancies and that these two tumor types may occur through distinct transformation mechanisms.

However, such a notion has been challenged recently by Mitelman *et al.* who have demonstrated that the number of fusion genes may simply be a function of the number of cases with an abnormal karyotype in both hematological malignancies and solid tumors.[11,12] A correlation between the number of patients with an abnormal karyotype and that of fusion genes is constant throughout all types of cancers ($R^2 = 0.82$, $P < 0.001$). It is therefore possible that infrequent reports of fusion genes in solid tumors (especially in epithelial tumors) may have been attributable to technical difficulties in obtaining clear karyotyping data or to the complex chromosome rearrangements in solid tumors.

If this is the case, many more fusion-type oncogenes may await discovery in solid tumors. Indeed, evidence in support of this prediction has been provided recently both by our discovery of the fusion-type PTK echinoderm microtubule-associated protein like-4 (EML4)–anaplastic lymphoma kinase (ALK), associated with lung cancer,[13–15] and by the detection of recurrent *e26 transformation-specific sequence (ETS)* fusion genes in prostate cancer.[16–18]

[1]E-mail: hmano@jichi.ac.jp

doi: 10.1111/j.1349-7006.2008.00972.x

Table 1. Number of recurrent balanced cytogenetic aberrations (RBA) in cancer

Cancer type	No. RBA†	Annual no. new cases in USA†
Leukemia	336	44 240
Lymphoma	159	71 380
Solid tumors	125	1 329 300

†Calculated from data reported previously.[9,10]

How to screen for oncogenes

Given the marked therapeutic efficacy of STI571 in CML, chemical inhibitors of epidermal growth factor receptor (EGFR) in lung cancer with activated EGFR,[19] and specific antibodies to HER2 in breast cancers with amplification of the *HER2* locus,[20] it is necessary to identify pivotal oncogenes in every cancer and to target these 'Achilles' heels'[21] for developing effective treatment strategies. We therefore tried to establish a functional screening system for transforming genes among a wide variety of cancer specimens.

The focus formation assay with 3T3 or RAT1 fibroblasts[22] has been used extensively to screen for oncogenes from clinical specimens. In such screening, genomic DNA is extracted from samples and transfected into the recipient fibroblasts. Because protein products of oncogenes can interfere with contact inhibition in fibroblasts, cell clones that have received oncogenes may keep growing even after the cells become confluent in culture. Such piled-up foci of cell clones (transformed foci) can be readily identified by visual inspection and subjected to the recovery of incorporated oncogenes. Application of such technologies has indeed succeeded in the isolation of a variety of transforming genes, such as mutated RAS family genes, activated RAF family genes, and a number of activated PTK genes.[23]

However, we have noticed that this type of screening system has a strong tendency to isolate the same sets of genes among different types of cancer (i.e. activated RAS family proteins and guanine nucleotide exchange factors). This could be due to an intrinsic property of the assay system, which isolates genes overriding growth inhibition mediated by cell-to-cell contact in fibroblasts. Another reason may be related to promoter specificity (Fig. 1a). In the screening systems where genomic DNA is used for transfection, any oncogene is controlled transcriptionally in the recipient cells by its own promoter and enhancer fragments. Therefore, if a promoter fragment of a given oncogene is active only in a tissue-specific manner (hematopoietic cell-specific, for instance), that gene would not be transcribed in fibroblasts, and thus could not be captured in the assay. Therefore, a genomic DNA-mediated screening system can only identify oncogenes with promoter fragments that are active in the recipient cells.

To overcome this limitation, it would be desirable to express every oncogene using an exogenous promoter fragment that allows abundant expression in any type of assay cell. For this purpose, we have developed a method to construct retrovirus-based cDNA expression libraries,[13,24–29] which can express any incorporated cDNA using a strong promoter fragment, long-terminal repeat (LTR), of the retroviral genome (Fig. 1b). Our system is so sensitive that we can generate libraries from small quantities of clinical specimens (such as $<1 \times 10^5$ cells). Further, given the high infection efficiency of retrovirus to dividing cells, any type of functional assay can be conducted in any proliferating cell with retroviral libraries.[30,31]

Discovery of a fusion-type PTK, EML4–ALK

Lung cancer remains the leading cause of cancer death, with an estimated ~1.3 million deaths worldwide each year.[32] Although

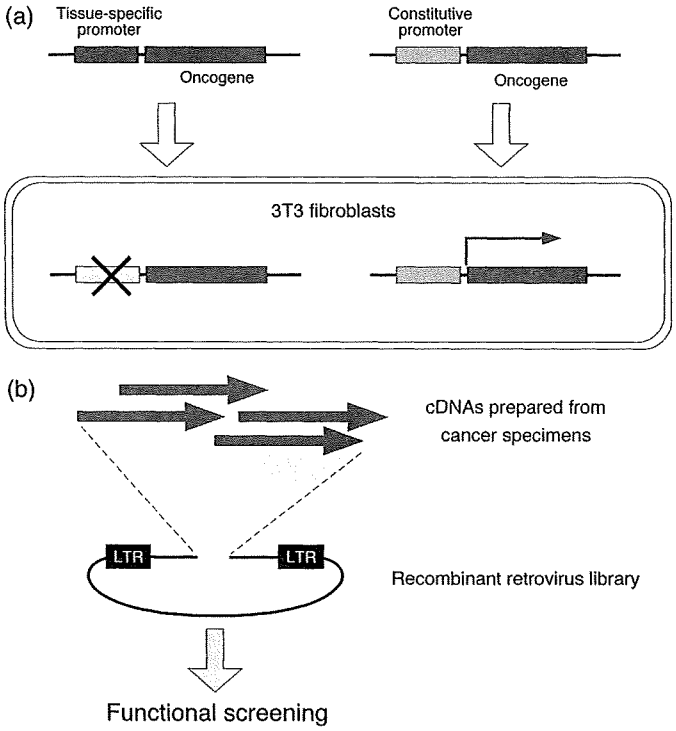

Fig. 1. Development of retroviral cDNA expression libraries for oncogene screening. (a) Although oncogenes controlled by a constitutive promoter or enhancer are expressed when introduced into fibroblasts, those controlled by a tissue-specific promoter-enhancer (e.g. specific to the hematopoietic system) are not transcribed in fibroblasts and will therefore not be detected by such functional screening. (b) To overcome this limitation, we synthesized cDNA from small quantities of clinical specimens, inserted them into a retroviral plasmid, and generated recombinant retroviral libraries. Theoretically, any type of dividing cell can be infected with such libraries, with the incorporated cDNA being expressed at a high level in the recipient cells under the control of the viral long terminal repeat (LTR).

activated EGFR has been identified in non-small cell lung cancer (NSCLC), the major subtype of lung cancer, and specific inhibitors against EGFR provide effective treatment modalities, this type of genetic mutation is found preferentially in non-smokers, young women, and the Asian ethnic group.[33] For other NSCLC patients who are not eligible for the anti-EGFR treatments, there are currently few effective treatments to improve their outcome, unless cancer cells are completely removed by surgery.[34]

We have therefore chosen NSCLC as the target of our retroviral screening system. First, among our consecutive panel of NSCLC specimens, we examined the presence of known transforming genes in lung cancer; that is, mutated *KRAS* and mutated *EGFR*. To raise a retroviral library, from the specimens negative for either mutation we chose a sample of lung adenocarcinoma resected from a 62-year-old man with a smoking history. A total of $>1.4 \times 10^6$ independent retroviral clones (with a mean cDNA size of 1.81 kb) were obtained from the specimen, and were used to infect 3T3 cells for the focus formation assay.

Dozens of transformed foci were readily identified in the assay, from which retroviral insert cDNA was rescued by PCR. Surprisingly, nucleotide sequences of the 5′ and 3′ parts of one cDNA corresponded to two different genes; one for microtubule-associated EML4,[35] and the other for the receptor-type PTK ALK.[36] Nucleotide sequencing of the cDNA revealed that the cDNA was a fusion between exons 1–13 of *EML4* and exons 20–29 of *ALK* (transcript ID ENST00000389048 in the Ensembl database; http://www.ensembl.org/index.html), thus encoding a

doi: 10.1111/j.1349-7006.2008.00972.x

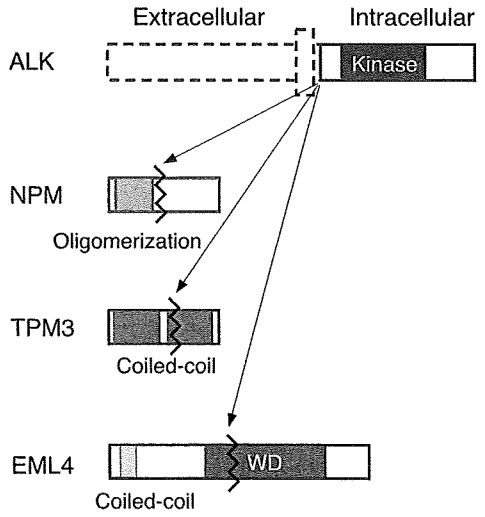

Fig. 2. Anaplastic lymphoma kinase (ALK) fusion proteins. Chromosome rearrangements result in the generation of fusion genes that encode an oligomerization domain (including the coiled coil domain) of nucleophosmin (NPM), tropomyosin (TPM3) or echinoderm microtubule-associated protein like-4 (EML4) fused to the intracellular region of the receptor-type protein tyrosine kinase ALK. WD, WD-repeat domain.

fusion-type PTK between the amino-terminal half of EML4 and the intracellular region of ALK (Fig. 2).[13]

Anaplastic lymphoma kinase was originally identified in anaplastic large cell lymphoma with t(2;5), as a fusion protein to nucleophosmin (NPM).[37,38] NPM–ALK plays an essential role in the lymphomagenesis of this subtype, and is a promising target for therapeutic compounds,[39,40] as is the case for BCR–ABL. In addition to NPM–ALK, ALK kinase may be fused, albeit at a lower frequency, to different partner proteins in the lymphoma through various chromosome translocations, giving rise to TRK-fused gene (TFG)–ALK, 5-aminoimidazole-4-carboximide ribonucleotide formyltransferase/IMP cyclohydrolase (ATIC)–ALK, clathrin, heavy chain (CLTC)–ALK, and others.[41] Subsequently, a non-hematological neoplasm, inflammatory myofibroblastic tumor (IMT) was also shown to harbor ALK fusion proteins such as TPM3–ALK, tropomyosin (TPM)4–ALK, and CLTC–ALK.[41] However, any recurrent translocation involving the *ALK* locus had not been reported for epithelial tumors before our discovery of *EML4–ALK*.

Interestingly, the *ALK* part of our *EML4–ALK* cDNA starts from exon 20 of *ALK*, which is also the fusion point in the vast majority of the other *ALK* fusion cDNA molecules, suggesting the presence of a common fragile locus within intron 19 of the *ALK* gene (Fig. 2). Both the *EML4* and *ALK* genes are mapped closely to the short arm of human chromosome 2 in opposite directions (Fig. 3a). Therefore, a chromosome segment encompassing the *EML4* and *ALK* loci has to become inverted to produce *EML4–ALK* cDNA. We have indeed succeeded to amplify by PCR a genome fragment from a NSCLC specimen that contained the fusion point between the *EML4* and *ALK* genes.[13] In this adenocarcinoma, the *EML4* gene was disrupted at a position 3.6 kb downstream of exon 13, and inverted to become ligated to a position ~300 bp upstream of *ALK* exon 20, proving the presence of inv(2)(p21p23) in the cancer cells.

Therefore, despite the previous notion that epithelial tumors seldom carry fusion-type oncogenes, we discovered an example of a fusion-type PTK with marked oncogenic activity in lung cancer, generated through a chromosome translocation, as is the case for BCR–ABL1, NPM–ALK, and translocation, ETS, leukemia (TEL)–Janus kinase 2 (JAK2) in hematological malignancies.[42] Because mutated EGFR and KRAS have been

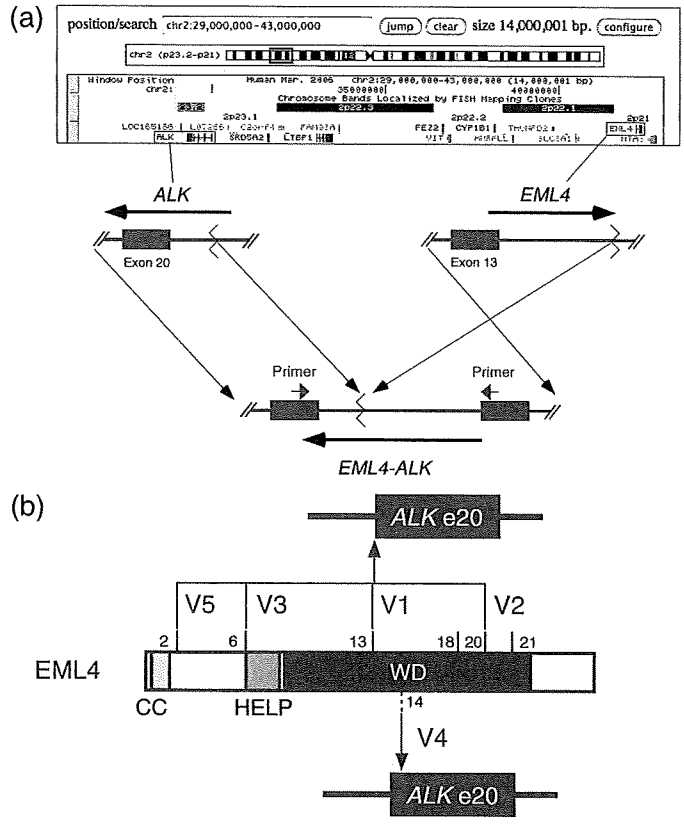

Fig. 3. Diversity in the fusion points between *echinoderm microtubule-associated protein like-4 (EML4)* and *anaplastic lymphoma kinase (ALK)*. (a) *EML4* and *ALK* map in opposite orientations to the short arm of human chromosome 2 (shown in the genome browser of the University of California, Santa Cruz; http://genome.ucsc.edu/cgi-bin/hggateway). Intron 13 of *EML4* is ligated to intron 19 of *ALK* through a chromosome rearrangement, inv(2)(p21p23), generating the *EML4–ALK* (variant 1) fusion gene. Primers flanking the fusion point can be used for the molecular detection of *EML4–ALK*-positive tumors by polymerase chain reaction. Arrows indicate the direction of transcription. (b) Exon boundaries of *EML4* for possible in-frame fusion to exon 20 of *ALK* are shown as vertical bars together with the exon numbers at the corresponding positions in the EML4 protein. Reverse tramscription–polymerase chain reaction screening has identified variants (V) 1, 2, 3, and 5 of *EML4–ALK*, in which exons 13, 20, 6, and 2, respectively, of *EML4* cDNA are fused to exon 20 (e20) of *ALK* cDNA. Unexpectedly, another in-frame fusion was identified in variant 4 cDNA, in which exon 14 of *EML4* was fused via an 11-bp sequence of unknown origin to the nucleotide at position 50 of *ALK* exon 20. CC, coiled coil domain; HELP, hydrophobic EMAL-like protein domain; WD, WD-repeat domain.

found recurrently in NSCLC cells, it is of clinical relevance whether EML4–ALK coexists in cancer cells with active EGFR or KRAS. Interestingly, the presence of EML4–ALK seems to be mutually exclusive to that of EGFR or KRAS mutations in NSCLC,[13,15,43,44] albeit with some exceptions.[45] Therefore, it is likely that EML4–ALK-positive lung cancer forms a subgroup among NSCLC, distinct from that positive for mutated EGFR or KRAS.

Molecular detection of *EML4–ALK*-positive NSCLC

One of the main reasons for the poor prognosis in lung cancer is the lack of sensitive detection methods that can capture tumor cells at early clinical stages (where tumors may be surgically removed). Although pathological examination of sputa and other clinical specimens is used routinely for the diagnosis of lung cancer, reliable detection with such systems usually requires that cancer cells occupy at least a small percentage of the total cells in these specimens. Therefore, patients diagnosed with this

technique as having lung cancer are often at advanced clinical stages already.

In contrast, *EML4–ALK*-positive cells may be detected in a very sensitive way. As *EML4* and *ALK* are mapped to chromosome 2p in opposite directions in normal cells, a set of PCR primers (one at exon 13 of *EML4* and the other at exon 20 of *ALK*; Fig. 3a) will not generate any specific PCR products from cDNA of normal cells or of cancer without inv(2)(p21p23). Therefore, RT-PCR of the cDNA (or PCR of the genome fusion points) should become a highly sensitive yet reliable detection method for EML4–ALK-positive tumors. Given the high sensitivity of PCR, it is even expected that one cancer cell out of 10^5–10^6 normal cells in sputa may be detected, which would significantly help to identify individuals with lung cancer at early resectable stages. Soda *et al.* indeed succeeded in capturing 10 cells/mL of EML4–ALK-positive cells in sputum by RT-PCR.[13] It would therefore be of great importance to test the idea that such RT-PCR-based detection with sputa may be useful as a general screening method for early stages of NSCLC (among individuals with chronic cough or sputa, for instance).

Once detected with such screening systems, individuals positive for *EML4–ALK* may undergo surgical resection of tumors or receive chemotherapies with compounds that specifically suppress ALK activity. Just like the case of *BCR–ABL1* in CML, *EML4–ALK* detection will likely play a pivotal role in the diagnosis of NSCLC positive for the fusion gene. In this regard, it is mandatory that every *EML4–ALK*-positive tumor be identified accurately by the diagnostic PCR system. There is a caveat, however, that the break and fusion points within the *EML4* and *ALK* loci may be more divergent than previously appreciated.

Soda *et al.* first discovered two variants of *EML4–ALK*: exon 13 of *EML4* fused to exon 20 of *ALK* in variant 1, and exon 20 of *ELM4* fused to exon 20 of *ALK* in variant 2.[13] In addition, we and Pasi A. Jänne and colleagues (Dana-Farber Cancer Institute) have recently identified two more variants (variants 3a and 3b, which connect exons 6a and 6b, respectively, of *EML4* to exon 20 of *ALK*) (Fig. 3b).[14,45] Further variants that connect various exons of *EML4* to *ALK* are being identified by a number of groups worldwide.[44–46]

In addition to exons 6 (variant 3), 13 (variant 1), and 20 (variant 2) of *EML4*, an in-frame fusion to exon 20 of *ALK* can occur with exon 2, 18, or 21 of *EML4* (Fig. 3b). Given that the amino-terminal coiled coil domain of EML4 is responsible for the oligomerization of EML4–ALK (see below) and that exon 2 of *EML4* encodes the entire coiled coil domain, all of these possible fusion genes would encode EML4–ALK proteins containing the coiled coil domain and therefore likely produce oncogenic EML4–ALK kinases. To screen for all variants (both known and unknown) of *EML4–ALK* and to estimate the frequency of such oncogenes in human cancers, Takeuchi *et al.* have developed a single-tube multiplex RT-PCR system that captures all possible in-frame fusions between *EML4* and *ALK*.[46] From screening of lung adenocarcinoma specimens ($n = 253$), they have identified a total of 11 samples (4.35%) carrying variants 1, 2, 3, or unknown isoforms (referred to as variants 4 and 5) of *EML4–ALK*.

Unexpectedly, in one of the new isoforms (variant 4), exon 14 of *EML4* is connected to an unknown sequence of 11 bp, and further fused to a nucleotide at position 50 of exon 20 of *ALK*. Although exon 14 of *EML4* is not expected to produce an in-frame fusion to exon 20 of *ALK*, insertion of the unknown 11-bp sequence and its ligation to a position within the *ALK* exon allows an in-frame connection between the two genes.

Additionally, exon 2 of *EML4* is fused to exon 20 of *ALK* or to a nucleotide 117 bp upstream of exon 20 of *ALK*, giving rise to variants 5a and 5b of *EML4–ALK*, respectively. Takeuchi *et al.* further successfully isolated full-length cDNA for variants 4, 5a, and 5b of EML4–ALK, and confirmed the transforming potential of all isoforms.[46] Takeuchi *et al.* have also screened for *EML4–ALK* cDNA, with the same multiplex RT-PCR technique, among other solid tumors ($n = 403$) including squamous cell carcinoma ($n = 71$) and small cell carcinoma ($n = 21$) of the lung. Interestingly, none of these tumors were positive for the fusion cDNA, indicating specificity of the *EML4–ALK* oncogene to lung cancer (especially adenocarcinoma).

Similarly, Wong *et al.* have tried to identify all possible in-frame fusions between *EML4* and *ALK* among a panel of NSCLC specimens ($n = 240$), discovering 13 cases (5.42%) positive for variants 1, 2, 3, and an unknown isoform of *EML4–ALK*.[44] Notably, variant 3 was the most frequent isoform ($n = 8$) in their Chinese cohort. Based on these data, the proportion of EML4–ALK-positive tumors in NSCLC seems to be ~5% in the Asian ethnic group, and may be lower in the others.[44,46,47]

It should be noted that all subtypes of EML4–ALK have not always been assayed in the published screenings and, further, that there may still be other variants not yet discovered. Therefore, to estimate the true prevalence of EML4–ALK-positive tumors within a given ethnic group, it is necessary to examine, in large cohorts, all possible in-frame fusions between *EML4* and *ALK* among the subjects. Additionally, given the increasing number of *EML4–ALK* variants, I personally hope that researchers may, in the near future, develop a more reasonable and uniform nomenclature system for such variants (E13; A20 for variant 1, and E6a; A20 for variant 3a, for instance) than the current one (variants 1, 2, 3, etc).

With regard to other diagnostic tools for EML4–ALK-positive tumors, immunohistochemical detection of EML4–ALK proteins in specimens obtained by biopsy or surgical resection would be a convenient screening system in clinics. In anaplastic large cell lymphoma, such screening with antibodies to the intracellular region of ALK has been used routinely to detect lymphoma positive for NPM–ALK.[48] Unfortunately, however, it is often difficult to stain EML4–ALK with such antibodies in NSCLC that are positive for *EML4–ALK* mRNA (K. Takeuchi and K. Inamura, personal communication). This discrepancy may be due to: (i) the weaker promoter activity of the *EML4* gene (which drives the expression of *EML4–ALK*) compared to that of *NPM* (which drives the expression of *NPM–ALK*), or (ii) a lower stability of the EML4–ALK protein than that of NPM–ALK. Further improvements in the sensitivity of immunohistochemical detection of EML4–ALK would be desirable to apply such systems to routine pathological screenings.

Transforming activity of EML4–ALK

How does fusion to EML4 induce a marked transforming potential in ALK? A number of fusion-type PTK carry an oligomerization motif within the fusion partner regions, which thereby leads to dimerization and autophosphorylation of the corresponding kinase domain.[49,50] Consistent with this notion, the NPM region in the NPM–ALK protein was shown to be essential in the oligomerization and transforming potential of this fusion kinase (Fig. 2).[51] Similarly, TPM3–ALK and TPM4–ALK found in IMT and EML4–ALK in NSCLC all carry a coiled coil domain within the fusion partners, which may act as an oligomerization motif. Indeed, EML4–ALK homodimerizes in cells, but internal deletion of the basic domain of EML4 (which contains the coiled coil domain) severely hampers such physical association. Accordingly, this mutant form of EML4–ALK loses its marked tumor-formation activity *in vivo*, and has decreased tyrosine kinase activity *in vitro*.[13] It should be noted, however, that truncation of subdomains other than the coiled coil domain of EML4–ALK also affects its transforming potential, suggesting that self-oligomerization is not the only mechanism to induce oncogenic potential in EML4–ALK.[13]

As wild-type ALK is a PTK with a single transmembrane region, it is presumed that the *in vivo* function of ALK is that of

doi: 10.1111/j.1349-7006.2008.00972.x

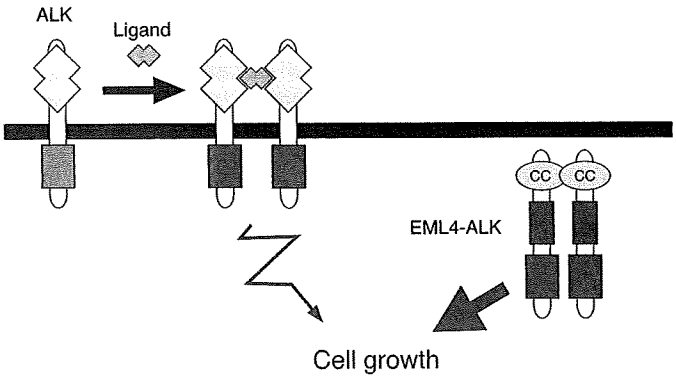

Fig. 4. Activation mechanisms for anaplastic lymphoma kinase (ALK) and echinoderm microtubule-associated protein like-4 (EML4)–ALK. Wild-type ALK is thought to undergo transient homodimerization in response to binding of a specific ligand, resulting in its activation and mitogenic signal transduction. In contrast, EML4–ALK is constitutively oligomerized via the coiled coil domain (CC) of EML4, resulting in persistent mitogenic signaling that eventually leads to malignant transformation.

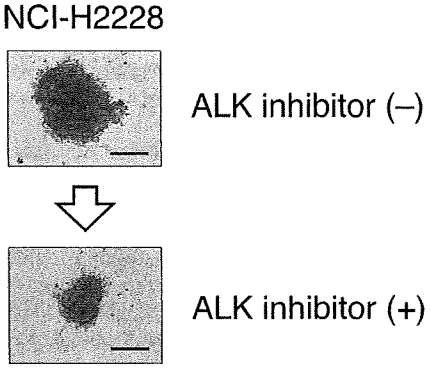

Fig. 5. Suppression of the growth of an echinoderm microtubule-associated protein like-4 (EML4)–anaplastic lymphoma kinase (ALK)-positive cell line with an ALK inhibitor. The non-small cell lung cancer cell line NCI-H2228, which is positive for *EML4–ALK* (variant 3), was maintained in spheroid culture for 2 days and was photographed after an additional 5 days of incubation in the absence (upper panel) or presence (lower panel) of 5 nmol/L 2,4-pyrimidinediamine.[14] Scale bars = 4 mm.

a cell surface receptor for specific ligands (probably growth factors). Unfortunately, however, such ligands have not been isolated in mammals. In *Drosophila melanogaster*, a protein homologous to human and mouse ALK is expressed in visceral mesoderm in the embryo, and malfunctions in Alk lead to visceral mesoderm defects in early embryogenesis, which resembles the phenotype of dysfunction in a secreted protein, jelly belly (Jeb). Upon binding to Jeb, Alk becomes activated to trigger the Ras–mitogen-activated protein kinase cascade and transcriptional activation of a subset of genes,[52] suggesting that Jeb is the ligand of Alk in fly.

In mammals, however, pleiotrophin[53] and midkine[54] are proposed ligands of ALK and are expressed specifically in brain and spinal cord.[55,56] Pleiotrophin may bind ALK at a low dissociation constant and induce tyrosine phosphorylation on ALK as well as putative downstream effector molecules.[53] However, other cellular receptors for pleiotrophin have also been identified and, hence, it is not yet clear if the observed effects of pleiotrophin are mediated mainly through ALK.[57,58]

Although it is theoretically possible that the extracellular region of ALK may act as its own ligand, the Jeb–Alk interaction in fly suggests that ALK likely functions as a cellular receptor for specific ligands in mammalian cells as well. Presumably, upon ligand binding (and only upon such binding), ALK becomes oligomerized and activated to trigger a transient growth signal in cells (Fig. 4). However, EML4–ALK and other ALK fusion proteins are constitutively oligomerized through the binding motif within the corresponding fusion partners, and activated to maintain a persistent mitogenic signal that finally leads to tumor formation.

A pivotal role of ALK fusion proteins in malignant transformation has been demonstrated clearly in the case of NPM–ALK. Retroviral transduction of NPM–ALK mRNA in bone marrow cells induces lymphoma-like disorders in mice,[59,60] and transgenic mice with *Vav* promoter-driven expression of *NPM–ALK* results in the generation of lymphoma.[61] Furthermore, injected lymphoma cells positive for *NPM–ALK* were effectively eradicated from mice by treatment with a specific inhibitor against ALK,[40] suggesting that the activated kinase potential of NPM–ALK is central to the lymphomagenesis.

A few experiments also support such a pivotal role for EML4–ALK in lung cancer. Koivunen *et al.* have found that treatment with a specific inhibitor for ALK (TAE684) induces rapid cell death of one NSCLC cell line (NCI-H3122) in culture, which harbors variant 1 of *EML4–ALK*.[45] Another NSCLC cell line, NCI-H2228, was shown to be positive for variant 3 of *EML4–ALK*, but TAE684 treatment failed to induce drastic effects in this these cells. However, McDermott *et al.* did find partial inhibition of cell viability (66% reduction compared to the control experiments) in NCI-H2228 after treatment with TAE684, as well as in NCI-H3122 (75% reduction).[62] Similarly, we found marked growth suppression of NCI-H2228 with an ALK inhibitor only in a spheroid culture system (Fig. 5),[14] not in a regular *in vitro* culture. These data thus indicate that EML4–ALK may be the principle transforming protein in some NSCLC cells (such as NCI-H3122) that are therefore fully sensitive to ALK inhibitors. However, it is likely that at least one other potential transforming protein is present with EML4–ALK in other NSCLC (e.g. NCI-H2228), and thus inhibition of ALK enzymatic activity provides significant but incomplete growth suppression.

With regard to the *in vivo* role of EML4–ALK, it is important to test if EML4–ALK activity can induce lung cancer *in vivo*. To address this issue, Soda *et al.* have recently generated transgenic mice in which EML4–ALK mRNA is transcribed specifically in lung epithelial cells by the use of a promoter fragment of the surfactant protein C gene.[63] Surprisingly, all independent lines of such mice develop hundreds of adenocarcinoma nodules in both lungs at only a few weeks after birth, proving for the first time the marked transforming activity of EML4–ALK *in vivo* (Soda *et al.*, unpublished data). More importantly, treatment of such mice with a chemical compound that suppresses ALK activity rapidly cleared those nodules from the lungs. Therefore, it is likely that EML4–ALK-positive lung cancer cells are at least partially dependent on their PTK activity for growth, and any means to suppress this activity would be a promising strategy for treating this intractable disorder.

Concluding remarks

The data from *EML4–ALK* transgenic mice clearly show the central role of this fusion kinase in lung cancer, and such mice also provide an efficient *in vivo* screening system for ALK inhibitors. Recently, treatment with an ALK inhibitor was shown to suppress or inhibit the growth of some neuroblastoma cell lines, in addition to NSCLC and anaplastic large cell lymphoma.[62,64] Because the *EML4–ALK* and *NPM–ALK* fusion genes are present in the latter two, such inhibitor-sensitive neuroblastoma cell lines may also possess other *ALK* mutants. These data suggest that tumors of any tissue origin may be treated with the same ALK inhibitors provided that they carry any one of the

oncogenic ALK mutants. Therefore, 'ALKoma'[65,66] may form a novel clinical entity as is the case for v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (KIT) mutant-positive tumors in acute myeloid leukemia, mastocytosis,[67,68] and gastrointestinal stromal tumor.[69]

However, given the marked diversity in the sensitivity of EML4–ALK-positive cell lines to ALK inhibitors,[45,62] identification of coexisting oncogenes in ALK mutant-positive tumors would be valuable to increase the efficacy of treatments with ALK inhibitors.

Furthermore, as various ALK fusion proteins have divergent subcellular localizations (probably dependent on the nature of fusion partner proteins),[46,70] downstream signaling pathways may vary among them. Indeed, although signal transducer and activator of transcription (STAT) proteins likely play a key role in the mitogenic signaling of NPM–ALK, such a role is unlikely for EML4–ALK[46,71] and some other ALK fusions.[70] It is therefore of great clinical relevance to decipher the profiles of oncogenes or tumor-suppressor genes and downstream proteins for each ALK fusion in each cancer subtype.

Acknowledgments

I apologize to all of the authors whose work could not be included in this manuscript owing to space constraints. I thank the members of our laboratory for their support and excellent work, and Yuichi Ishikawa and Kengo Takeuchi for helpful suggestions and discussion. This work was supported in part by a grant for Research on Human Genome Tailor-made from the Ministry of Health, Labor, and Welfare of Japan, and by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

1 Futreal PA, Coin L, Marshall M *et al.* A census of human cancer genes. *Nat Rev Cancer* 2004; **4**: 177–83.
2 Ren R. Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat Rev Cancer* 2005; **5**: 172–83.
3 Li S, Ilaria RL Jr, Million RP, Daley GQ, Van Etten RA. The P190, P210, and P230 forms of the BCR/ABL oncogene induce a similar chronic myeloid leukemia-like syndrome in mice but have different lymphoid leukemogenic activity. *J Exp Med* 1999; **189**: 1399–412.
4 Pear WS, Miller JP, Xu L *et al.* Efficient and rapid induction of a chronic myelogenous leukemia-like myeloproliferative disease in mice receiving P210 bcr/abl-transduced bone marrow. *Blood* 1998; **92**: 3780–92.
5 Oehler VG, Radich JP. Monitoring bcr-abl by polymerase chain reaction in the treatment of chronic myeloid leukemia. *Curr Oncol Rep* 2003; **5**: 426–35.
6 Druker BJ, Talpaz M, Resta DJ *et al.* Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* 2001; **344**: 1031–7.
7 Druker BJ, Guilhot F, O'Brien SG *et al.* Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N Engl J Med* 2006; **355**: 2408–17.
8 Mitelman F. Recurrent chromosome aberrations in cancer. *Mutat Res* 2000; **462**: 247–53.
9 Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. *CA Cancer J Clin* 2007; **57**: 43–66.
10 Mitelman F, Mertens F, Johansson B. Prevalence estimates of recurrent balanced cytogenetic aberrations and gene fusions in unselected patients with neoplastic disorders. *Genes Chromosomes Cancer* 2005; **43**: 350–66.
11 Mitelman F, Johansson B, Mertens F. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nat Genet* 2004; **36**: 331–4.
12 Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 2007; **7**: 233–45.
13 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.
14 Choi YL, Takeuchi K, Soda M *et al.* Identification of novel isoforms of the *EML4–ALK* transforming gene in non-small cell lung cancer. *Cancer Res* 2008; **68**: 4971–6.
15 Inamura K, Takeuchi K, Togashi Y *et al. EML4–ALK* fusion is linked to histological characteristics in a subset of lung cancers. *J Thorac Oncol* 2008; **3**: 13–17.
16 Tomlins SA, Rhodes DR, Perner S *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 2005; **310**: 644–8.
17 Tomlins SA, Laxman B, Dhanasekaran SM *et al.* Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* 2007; **448**: 595–9.
18 Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. Recurrent gene fusions in prostate cancer. *Nat Rev Cancer* 2008; **8**: 497–511.
19 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.
20 Menard S, Pupa SM, Campiglio M, Tagliabue E. Biologic and therapeutic role of HER2 in cancer. *Oncogene* 2003; **22**: 6570–8.
21 Sharma SV, Settleman J. Oncogene addiction: setting the stage for molecularly targeted cancer therapy. *Genes Dev* 2007; **21**: 3214–31.
22 Goldfarb M, Shimizu K, Perucho M, Wigler M. Isolation and preliminary characterization of a human transforming gene from T24 bladder carcinoma cells. *Nature* 1982; **296**: 404–9.
23 Aaronson SA. Growth factors and cancer. *Science* 1991; **254**: 1146–53.
24 Choi YL, Moriuchi R, Osawa M *et al.* Retroviral expression screening of oncogenes in natural killer cell leukemia. *Leuk Res* 2005; **29**: 943–9.
25 Fujiwara S, Yamashita Y, Choi YL *et al.* Transforming activity of the lymphotoxin-β receptor revealed by expression screening. *Biochem Biophys Res Commun* 2005; **338**: 1256–62.
26 Kisanuki H, Choi YL, Wada T *et al.* Retroviral expression screening of oncogenes in pancreatic ductal carcinoma. *Eur J Cancer* 2005; **41**: 2170–5.
27 Choi YL, Kaneda R, Wada T *et al.* Identification of a constitutively active mutant of JAK3 by retroviral expression screening. *Leuk Res* 2007; **31**: 203–9.
28 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.
29 Hatanaka H, Takada S, Choi YL *et al.* Transforming activity of purinergic receptor P2Y, G-protein coupled, 2 revealed by retroviral expression screening. *Biochem Biophys Res Commun* 2007; **356**: 723–6.
30 Kitamura T, Onishi M, Kinoshita S, Shibuya A, Miyajima A, Nolan GP. Efficient screening of retroviral cDNA expression libraries. *Proc Natl Acad Sci USA* 1995; **92**: 9146–50.
31 Yoshizuka N, Moriuchi R, Mori T *et al.* An alternative transcript derived from the trio locus encodes a guanosine nucleotide exchange factor with mouse cell-transforming potential. *J Biol Chem* 2004; **279**: 43 998–4004.
32 American Cancer Society. Global Cancer Facts and Figures 2007. [Cited 18 October, 2008.] Available from URL: http://www.cancer.org/downloads/STT/Global_Facts_and_Figures_2007_rev2.pdf.
33 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.
34 Schiller JH, Harrington D, Belani CP *et al.* Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med* 2002; **346**: 92–8.
35 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.
36 Morris SW, Naeve C, Mathew P *et al.* ALK, the chromosome 2 gene locus altered by the t(2;5) in non-Hodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK). *Oncogene* 1997; **14**: 2175–88.
37 Morris SW, Kirstein MN, Valentine MB *et al.* Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* 1994; **263**: 1281–4.
38 Shiota M, Fujimoto J, Semba T, Satoh H, Yamamoto T, Mori S. Hyperphosphorylation of a novel 80 kDa protein-tyrosine kinase similar to Ltk in a human Ki-1 lymphoma cell line, AMS3. *Oncogene* 1994; **9**: 1567–74.
39 Wan W, Albom MS, Lu L *et al.* Anaplastic lymphoma kinase activity is essential for the proliferation and survival of anaplastic large-cell lymphoma cells. *Blood* 2006; **107**: 1617–23.
40 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 USA* 2007; **104**: 270–5.
41 Pulford K, Morris SW, Turturro F. Anaplastic lymphoma kinase proteins in growth control and cancer. *J Cell Physiol* 2004; **199**: 330–58.
42 Lacronique V, Boureux A, Valle VD *et al.* A TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia. *Science* 1997; **278**: 1309–12.
43 Shinmura K, Kageyama S, Tao H *et al.* EML4–ALK fusion transcripts, but no NPM–, TPM3–, CLTC–, ATIC–, or TFG–ALK fusion transcripts, in non-small cell lung carcinomas. *Lung Cancer* 2008; **61**: 163–9.
44 Wong DD-S, So KK-T, Leung EL-H *et al.* EML4–ALK is a new oncogene in non-small cell lung carcinoma showing wild-type EGFR and K-RAS from

doi: 10.1111/j.1349-7006.2008.00972.x

non-smokers. *American Association of Cancer Research Annual Meeting* 12–16 April, 2008, San Diego, AACR (Philadelphia): (poster #LB-62).
45 Koivunen JP, Mermel C, Zejnullahu K *et al.* EML4–ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer. *Clin Cancer Res* 2008; **14**: 4275–83.
46 Takeuchi K, Choi YL, Soda M *et al.* Multiplex RT-PCR screening for EML4-ALK fusion transcripts. *Clin Cancer Res* 2008; **14**: 6618–24.
47 Perner S, Wagner PL, Demichelis F *et al.* EML4–ALK fusion lung cancer: a rare acquired event. *Neoplasia* 2008; **10**: 298–302.
48 Cataldo KA, Jalal SM, Law ME *et al.* Detection of t(2;5) in anaplastic large cell lymphoma: comparison of immunohistochemical studies, FISH, and RT-PCR in paraffin-embedded tissue. *Am J Surg Pathol* 1999; **23**: 1386–92.
49 Turner SD, Alexander DR. Fusion tyrosine kinase mediated signalling pathways in the transformation of haematopoietic cells. *Leukemia* 2006; **20**: 572–82.
50 Matsumura I, Mizuki M, Kanakura Y. Roles for deregulated receptor tyrosine kinases and their downstream signaling molecules in hematologic malignancies. *Cancer Sci* 2008; **99**: 479–85.
51 Bischof D, Pulford K, Mason DY, Morris SW. Role of the nucleophosmin (NPM) portion of the non-Hodgkin's lymphoma-associated NPM–anaplastic lymphoma kinase fusion protein in oncogenesis. *Mol Cell Biol* 1997; **17**: 2312–25.
52 Lee HH, Norris A, Weiss JB, Frasch M. Jelly belly protein activates the receptor tyrosine kinase Alk to specify visceral muscle pioneers. *Nature* 2003; **425**: 507–12.
53 Stoica GE, Kuo A, Aigner A *et al.* Identification of anaplastic lymphoma kinase as a receptor for the growth factor pleiotrophin. *J Biol Chem* 2001; **276**: 16 772–9.
54 Stoica GE, Kuo A, Powers C *et al.* Midkine binds to anaplastic lymphoma kinase (ALK) and acts as a growth factor for different cell types. *J Biol Chem* 2002; **277**: 35 990–8.
55 Pulford K, Lamant L, Morris SW *et al.* Detection of anaplastic lymphoma kinase (ALK) and nucleolar protein nucleophosmin (NPM)–ALK proteins in normal and neoplastic cells with the monoclonal antibody ALK1. *Blood* 1997; **89**: 1394–404.
56 Iwahara T, Fujimoto J, Wen D *et al.* Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. *Oncogene* 1997; **14**: 439–49.
57 Maeda N, Nishiwaki T, Shintani T, Hamanaka H, Noda M. 6B4 proteoglycan/phosphacan, an extracellular variant of receptor-like protein-tyrosine phosphatase zeta/RPTPbeta, binds pleiotrophin/heparin-binding growth-associated molecule (HB-GAM). *J Biol Chem* 1996; **271**: 21 446–52.
58 Raulo E, Chernousov MA, Carey DJ, Nolo R, Rauvala H. Isolation of a neuronal cell surface receptor of heparin binding growth-associated molecule (HB-GAM). Identification as N-syndecan (syndecan-3). *J Biol Chem* 1994; **269**: 12 999–3004.
59 Kuefer MU, Look AT, Pulford K *et al.* Retrovirus-mediated gene transfer of NPM–ALK causes lymphoid malignancy in mice. *Blood* 1997; **90**: 2901–10.
60 Miething C, Grundler R, Fend F *et al.* The oncogenic fusion protein nucleophosmin–anaplastic lymphoma kinase (NPM–ALK) induces two distinct malignant phenotypes in a murine retroviral transplantation model. *Oncogene* 2003; **22**: 4642–7.
61 Turner SD, Tooze R, Maclennan K, Alexander DR. Vav-promoter regulated oncogenic fusion protein NPM–ALK in transgenic mice causes B-cell lymphomas with hyperactive Jun kinase. *Oncogene* 2003; **22**: 7750–61.
62 McDermott U, Iafrate AJ, Gray NS *et al.* Genomic alterations of anaplastic lymphoma kinase may sensitize tumors to anaplastic lymphoma kinase inhibitors. *Cancer Res* 2008; **68**: 3389–95.
63 Zhao B, Magdaleno S, Chua S *et al.* Transgenic mouse models for lung cancer. *Exp Lung Res* 2000; **26**: 567–79.
64 McDermott U, Sharma SV, Dowell L *et al.* Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling. *Proc Natl Acad Sci USA* 2007; **104**: 19 936–41.
65 Ponzoni M, Terreni MR, Ciceri F *et al.* Primary brain CD30$^+$ ALK1$^+$ anaplastic large cell lymphoma ('ALKoma'): the first case with a combination of 'not common' variants. *Ann Oncol* 2002; **13**: 1827–32.
66 Chiarle R, Voena C, Ambrogio C, Piva R, Inghirami G. The anaplastic lymphoma kinase in the pathogenesis of cancer. *Nat Rev Cancer* 2008; **8**: 11–23.
67 Longley J, Duffy TP, Kohn S. The mast cell and mast cell disease. *J Am Acad Dermatol* 1995; **32**: 545–61.
68 Furitsu T, Tsujimura T, Tono T *et al.* Identification of mutations in the coding sequence of the proto-oncogene c-kit in a human mast cell leukemia cell line causing ligand-independent activation of c-kit product. *J Clin Invest* 1993; **92**: 1736–44.
69 Siehl J, Thiel E. C-kit, GIST, and imatinib. *Recent Results Cancer Res* 2007; **176**: 145–51.
70 Armstrong F, Duplantier MM, Trempat P *et al.* Differential effects of X-ALK fusion proteins on proliferation, transformation, and invasion properties of NIH3T3 cells. *Oncogene* 2004; **23**: 6071–82.
71 Rikova K, Guo A, Zeng Q *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* 2007; **131**: 1190–203.

Genome-wide histone methylation profile for heart failure

Ruri Kaneda[1,2], Shuji Takada[1], Yoshihiro Yamashita[1], Young Lim Choi[1], Mutsuko Nonaka-Sarukawa[2], Manabu Soda[1], Yoshio Misawa[3], Tadashi Isomura[4], Kazuyuki Shimada[2] and Hiroyuki Mano[1,5,*]

Divisions of [1]Functional Genomics, Jichi Medical University, Tochigi 329-0498, Japan
[2]Cardiovascular Medicine, Jichi Medical University, Tochigi 329-0498, Japan
[3]Cardiovascular Surgery, Jichi Medical University, Tochigi 329-0498, Japan
[4]Hayama Heart Center, Kanagawa 240-0116, Japan
[5]CREST, Japan Science and Technology Agency, Saitama 332-0012, Japan

Epigenetic alterations are implicated in the development of cardiac hypertrophy and heart failure, but little is known of which epigenetic changes in which regions of the genome play such a role. We now show that trimethylation of histone H3 on lysine-4 (K4TM) or lysine-9 (K9TM) is markedly affected in cardiomyocytes in association with the development of heart failure in a rat disease model. High-throughput pyrosequencing performed with ChIP products for K4TM or K9TM prepared from human left ventricular tissue with retained or damaged function also revealed that protein-coding genes located in the vicinity of K4TM marks differ between functional and disabled myocytes, yet both sets of genes encode proteins that function in the same signal transduction pathways for cardiac function, indicative of differential K4TM marking during the development of heart failure. However, K9TM mark-profile was less dependent on the disease status compared to that of K4TM. Our data collectively reveal global epigenetic changes in cardiac myocytes associated with heart failure.

Introduction

A variety of conditions, including pressure or volume overload in the cardiovascular system and remodeling of the left ventricle of the heart after ischemic damage, result in heart failure, which is characterized by a reduction in contractile ability and a decrease in the number of viable myocytes in the heart (James *et al.* 2000). Treatment of heart failure remains problematic, and this condition is thus still one of the leading causes of human death (Braunwald 1997).

Epigenetic status has been linked to cardiac hypertrophy and heart failure. The histone acetyltransferase activity of CREB-binding protein (CBP) and p300 is thus required for the induction of hypertrophic changes in cardiac muscle cells by phenylephrine (Gusterson *et al.* 2003). Consistent with this observation, inhibition of histone deacetylase (HDAC) activity results in an increase in the size of cardiac muscle cells (Iezzi *et al.* 2004). Furthermore, HDACs of class II (HDAC4, -5, -7, and -9) suppress cardiac hypertrophy in part by binding to and inhibiting the activity of myocyte enhancer factor 2 (Zhang *et al.* 2002). Induction of the atrial natriuretic peptide gene is associated with acetylation of histones (H3 and H4) located in the 3′ untranslated region of the gene (Kuwahara *et al.* 2001). Histones bound to the β-myosin heavy chain gene have also been shown to be targeted by histone acetyltransferases in cardiomyocytes (Zhang *et al.* 2002). Moreover, dynamic regulation of other histone modifications has been demonstrated in cardiac myocytes (Illi *et al.* 2005; Bingham *et al.* 2007).

It remains to be established, however, (i) which epigenetic marks are dysregulated in association with heart failure *in vivo*, (ii) which regions of the human genome are susceptible to such epigenetic changes, and (iii) how epigenetic dysregulation affects the expression of protein-coding or other genes. To address these issues, we have now studied an animal model of congestive heart failure (CHF), the Dahl salt-sensitive rat (Rapp *et al.* 1989), and found that two histone modifications are markedly affected in cardiac myocytes during the development of CHF. We further confirmed our findings in human left ventricular (LV) myocytes with the use of chromatin immunoprecipitation

Communicated by: Kohei Miyazono
**Correspondence*: hmano@jichi.ac.jp

DOI: 10.1111/j.1365-2443.2008.01252.x

(ChIP) coupled to pyrosequencing. Our results have revealed dynamic histone modifications in the vicinity of a subset of protein-coding genes in the human genome, which directly participate in regulation of the contraction of cardiac myocytes.

Results

Histone modifications in the heart of Dahl rats

We prepared LV myocytes from Dahl salt-sensitive rats, which are genetically intolerant to excessive salt intake (Rapp *et al.* 1989). A high-sodium diet thus induces systemic hypertension and cardiac hypertrophy in Dahl rats within a few weeks. These changes are followed within a few months by the development of CHF and death. We isolated cardiac myocytes from rats with CHF (fed a high-sodium diet) as well as from age-matched animals with a normal heart (fed a low-sodium diet), and we subjected these cells to ChIP with antibodies to acetylated histone H3 (H3Ac), acetylated histone H4 (H4Ac), histone H3 dimethylated on lysine-4 (K4DM), histone H3 trimethylated on lysine-4 (K4TM), histone H3 dimethylated on lysine-9 (K9DM), histone H3 trimethylated on lysine-9 (K9TM), histone H4 trimethylated on lysine-20 (K20TM), or histone H3 dimethylated on lysine-27 (K27DM). The ChIP products as well as cRNA prepared from the normal or failed hearts were then individually subjected to hybridization with high-density oligonucleotide microarrays (Affymetrix Rat Genome 230 2.0 GeneChip) originally developed for expression profiling of rat genes.

Pearson's correlation coefficient for the signal intensity of all probe sets with a "Present" call (by Affymetrix GCOS software) in the normal heart ($n = 13\ 914$) was 0.873 in the cRNA hybridizations for normal and failed hearts (Fig. 1), indicative of a strong correlation in the expression level of most genes between the two samples. Consistent with this observation, the signal intensity for all probe sets with a positive value in the H3Ac ChIP products from the normal heart ($n = 12\ 027$) was highly correlated between these products from normal and failed hearts ($r = 0.724$). A similar strong correlation between the two groups was observed for H4Ac.

Unexpectedly, however, despite the strong correlation ($r = 0.856$) apparent for K4DM, only a weak negative correlation ($r = -0.097$) was detected for the K4TM mark between normal and failed hearts, indicative of marked differences in the associated gene sets. Similarly, although a strong correlation was observed for K9DM ($r = 0.558$), a weak negative correlation ($r = -0.251$) was apparent for K9TM. Hybridization levels were positively correlated between normal and failed hearts for K20TM and K27DM.

Thus, among the epigenetic marks examined, K4TM and K9TM were the histone modifications most affected in heart failure. Although differences in functional roles and genomic distributions between K4DM and K4TM have been described (Santos-Rosa *et al.* 2002; Bernstein *et al.* 2005), little has been known of such differential roles for the methylation level of lysine-9 of histone H3.

K4TM and K9TM profiles in the human heart

We next attempted to identify the genomic regions associated with the K4TM and K9TM marks in human cardiac myocytes. ChIP products for K4TM or K9TM were prepared from a mixture of LV tissue specimens from four individuals with retained pumping function [LV ejection fraction (EF) of 65.5 ± 7.6%, mean ± SD] or from four individuals with CHF (LVEF of 19.8 ± 5.7%) caused by dilated cardiomyopathy (Table 1). The ChIP

Table 1 Clinical characteristics of the subjects who provided specimens for the study

	Sample ID	Disease	Age (years)	Sex	LVEF (%)
HighEF	PM 8	HVD (MSR, ASR)	59	F	65
	PM12	HVD (MSR)	73	F	58
	PM13	HVD (MS)	55	F	76
	PM14	HVD (MS)	62	F	63
CHF	LV13	DCM	52	M	17
	LV14	DCM	55	M	25
	LV18	DCM	57	M	13
	LV20	DCM	64	F	24

HVD, heart valvular disease; MS(R), mitral stenosis (and regurgitation); ASR, aortic stenosis and regurgitation; DCM, dilated cardiomyopathy; F, female; M, male.

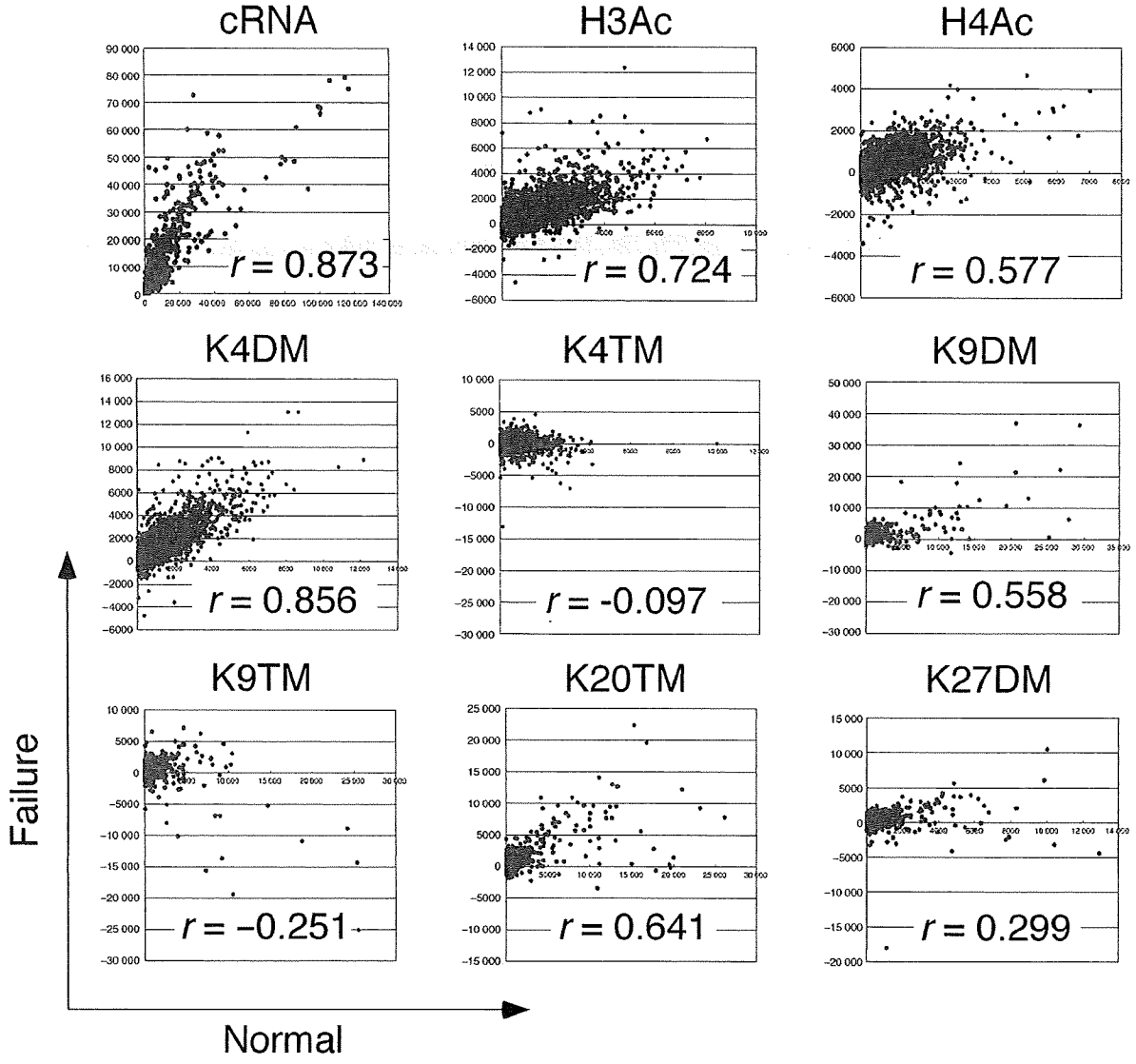

Figure 1 Comparison of epigenetic profiles between normal and failed rat hearts. The expression level of each probe set on oligonucleotide microarrays was compared between total cRNA from normal (x axis) or failed (y axis) hearts by calculation of Pearson's correlation coefficient (r). ChIP-on-chip data for H3Ac, H4Ac, K4DM, K4TM, K9DM, K9TM, K20TM, and K27DM are similarly compared.

products were subjected to pyrosequencing with the Genome Sequencer 20 system (Roche). In this "ChIP-to-seq" experiment, 96 069, 95 596, 116 267, and 96 734 reads were obtained for the K4TM products for specimens with retained LV ejection fraction (HighEF), the K4TM products for CHF, the K9TM products for HighEF, and the K9TM products for CHF, respectively. After quality-filtering, we isolated an average of 36 279 reads per sample, for each of which a single hit with a highest matching score was identified in the human genome sequence (the hg18 assembly of the Genome Bioinformatics Group, University of California at Santa Cruz) (Table S1 in Supporting Information). We thus focused on these reads for further analysis.

Many regions of the genome were identified in which multiple sequence reads mapped closely to each other. We therefore defined a "cluster" as a group of sequence reads localized within a distance of 1 kbp in the human genome (Fig. 2A). A total of 94 202 clusters was identified for all four samples, and 18 725 of these clusters, referred to as "high clusters," contained ≥ 2 sequence reads in ≥ 1 sample (see Table S2 in Supporting Information).

We then examined histone modification at the high clusters for specificity of the epigenetic mark (K4TM or

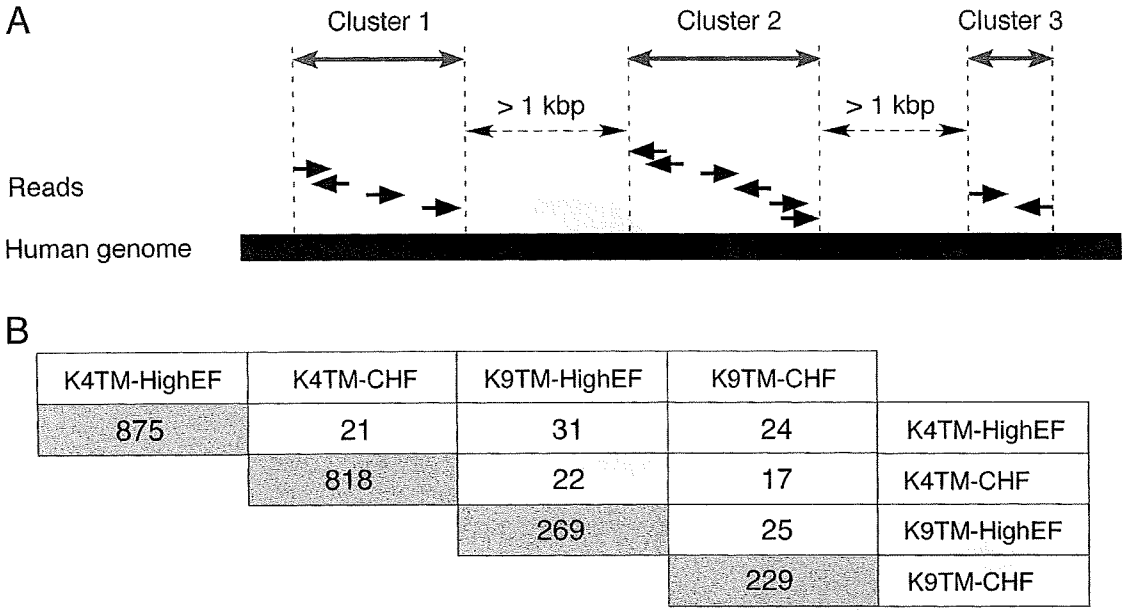

Figure 2 High clusters in K4TM and K9TM ChIP-to-seq data. (A) Groups of sequence reads that map to the human genome within a distance of 1 kbp are defined as "clusters," which are further denoted as "high clusters" when the read number in the cluster is ≥ 2 in ≥ 1 sample. (B) Numbers of high clusters with a read number of ≥ 5 for K4TM or K9TM in HighEF or CHF samples (shaded boxes). The numbers of such clusters shared between any pair of samples is also indicated (open boxes).

Table 2 Disease-specific high clusters

Mark	Characteristics of high clusters	Total number of high clusters	Number of high clusters close to RefSeq genes	Number of high clusters close to CpG islands
K4TM	HighEF ≥ 5, CHF ≤ 1	836	407	129
	HighEF ≤ 1, CHF ≥ 5	786	432	163
K9TM	HighEF ≥ 5, CHF ≤ 1	220	75	18
	HighEF ≤ 1, CHF ≥ 5	196	69	10

K9TM) and disease status (HighEF or CHF). Among the high clusters, 875 had ≥ 5 reads in the K4TM product for HighEF, 818 had ≥ 5 reads in the K4TM product for CHF, 269 had ≥ 5 reads in the K9TM product for HighEF, and 229 had ≥ 5 reads in the K9TM product for CHF (Fig. 2B). Only a few dozen of such high clusters were shared between any pair of samples, indicating the existence of disease-specific as well as methylation site-specific epigenetic profiles. Therefore, despite the heterogeneity in the cause of CHF (sustained systemic hypertension or dilated cardiomyopathy), both the Dahl rat and human data sets revealed a marked difference in the K4TM and K9TM epigenetic profiles between normal and failed hearts. Such specificity is further visualized for human chromosome 1 in Fig. S1 in Supporting Information. In contrast, the profile of read number per cluster was similar among the four groups of human ChIP products (see Fig. S2 in Supporting Information).

Genes mapped closely to disease-dependent clusters

We then isolated disease status-specific high clusters from the data set. A total of 836 high clusters was found to contain ≥ 5 reads in the K4TM products for HighEF but ≤ 1 read in those for CHF (HighEF-specific K4TM clusters); 407 RefSeq genes mapped to within ≤ 5 kbp of these clusters (Table 2). Similarly, 786 high clusters were found to be specific for K4TM and CHF (≤ 1 read in the K4TM products for HighEF but ≥ 5 reads in those for CHF). Smaller numbers of disease-dependent clusters were identified for the K9TM mark (220 HighEF-specific and 196 CHF-specific). These disease-dependent clusters

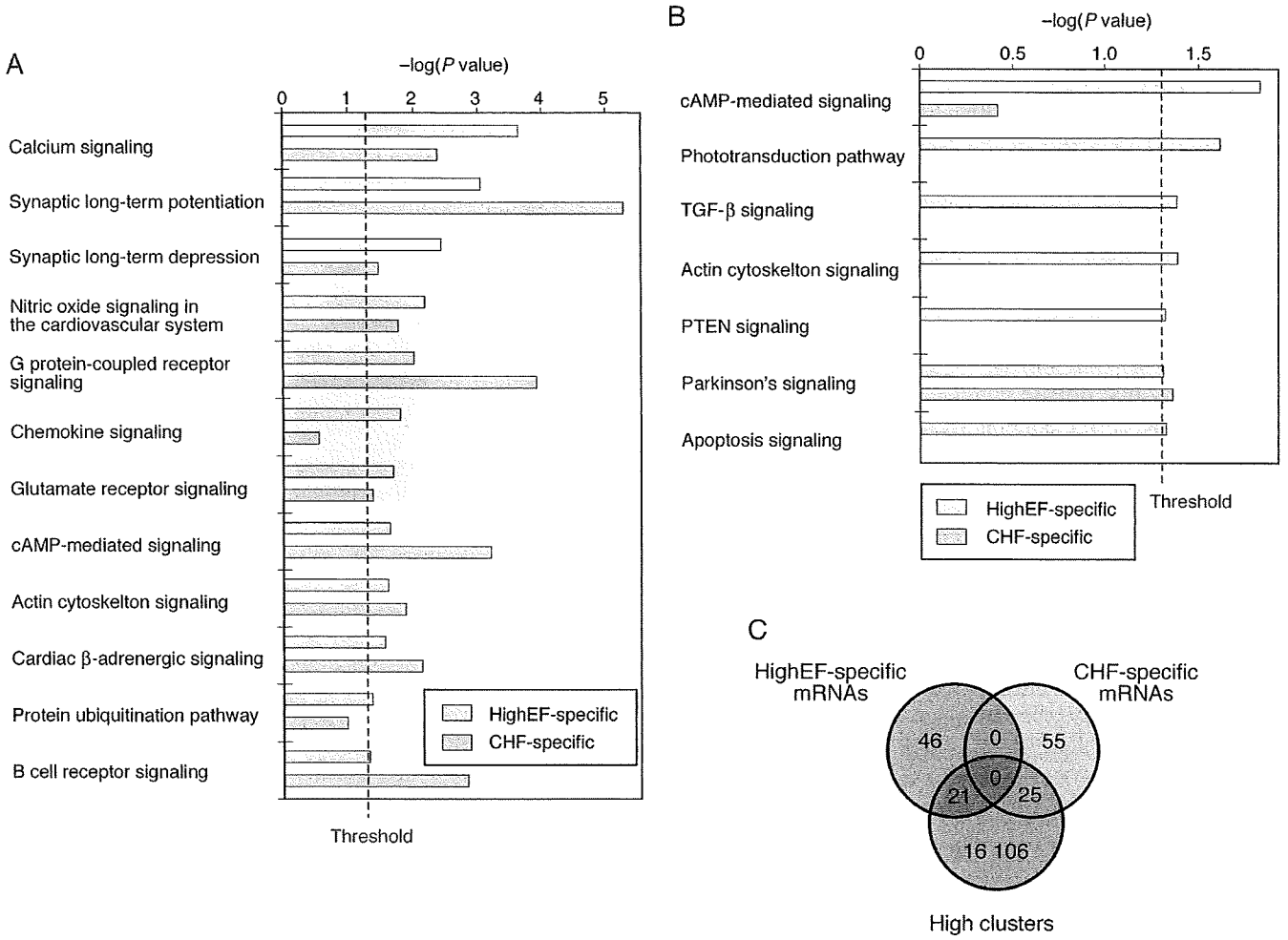

Figure 3 Analysis of genes that map in the vicinity of disease-dependent high clusters. (A) Canonical signaling pathways overrepresented in the HighEF-specific or CHF-specific high clusters for the K4TM ChIP products are listed with the corresponding −log(*P* value) score. (B) Canonical signaling pathways overrepresented in the HighEF-specific or CHF-specific high clusters for the K9TM ChIP products are listed with the corresponding −log(*P* value) score. (C) Venn diagram for comparison of transcripts associated specifically with HighEF or CHF status and those encoded by genes that map within a distance of < 5 kbp relative to a high cluster.

were widely distributed throughout human chromosomes and showed little overlap (see Fig. S3 in Supporting Information).

We examined whether the protein products of RefSeq genes that mapped in the vicinity (a distance of ≤ 5 kbp) of disease-dependent clusters function in canonical intracellular signaling pathways with the use of Ingenuity Pathway Analysis software (Ingenuity Systems; http://www.ingenuity.com). Analysis of the RefSeq genes associated with the disease-dependent K4TM clusters identified 12 canonical pathways that were significantly overrepresented ($P < 0.05$, Fisher's exact test) in HighEF-specific clusters and 20 pathways overrepresented in CHF-specific clusters. Many of the pathways ($n = 10$) were overrepresented in both HighEF-K4TM and CHF-K4TM clusters, almost all of which (including those for calcium signaling, synaptic long-term regulation, and nitric oxide signaling) are related to cardiac function (Fig. 3A).

Consistent with the disease-dependent selection of the clusters, the HighEF-associated and CHF-associated genes were distinct even within the same pathways. The canonical pathway for synaptic long-term potentiation, for example, contains the products of eight HighEF-associated and 12 CHF-associated genes, the interactions among which are shown in Fig. S4 in Supporting Information. Although genes corresponding to the calmodulin complex are present in both gene sets, these genes differ between the HighEF set (*CALM1*) and the CHF set (*CALM3*).

In addition to the proteins of the canonical signaling pathways, many products of the genes in the vicinity of disease-dependent high clusters for K4TM are functionally or physically networked. One such network comprises 34 proteins, 18 of which are encoded by HighEF-associated genes and 16 by CHF-associated genes (Fig. S5 in Supporting Information). Again, the genes for some complexes associated with both gene sets are distinct; those for the ATPase complex, for instance, include that for ATP1B1 in the HighEF-associated set and that for ATP5C1 in the CHF-associated set. Gene products in this network are substantially enriched in those implicated in cardiovascular disease.

In contrast to the K4TM-specific clusters, only a few canonical signaling pathways are linked to the RefSeq genes localized in the vicinity of K9TM-specific clusters. This difference is due in part to the small number of high clusters that contain disease-dependent reads for K9TM. Whereas the numbers of high clusters for HighEF specimens were similar between K4TM and K9TM products ($n = 6547$ and 5594, respectively), the numbers of disease-dependent clusters for the K9TM mark were only approximately 25% of those for the K4TM mark (Table 2). Seven canonical signaling pathways were over-represented ($P < 0.05$, Fisher's exact test) in the genes associated with the HighEF-K9TM clusters, whereas only one such pathway was overrepresented in those associated with the CHF-K9TM clusters (Fig. 3B). The network containing the most disease-dependent K9TM-associated gene products is centered on transforming growth factor β1 (TGFB1) and the tumor suppressor p53 (TP53), implicating K9TM-related regulation in cell death in the heart (see Fig. S6 in Supporting Information).

Our analysis thus revealed differential regulation of K4TM modification for genes related to cardiac function. To examine whether such epigenetic regulation plays a direct role in gene transcription, we performed gene expression profiling with Human Genome U133 Plus 2.0 arrays (Affymetrix) for the individual specimens (four for HighEF and four for CHF) used in the ChIP experiments. From the data obtained for 54 675 probe sets and the eight specimens, we selected HighEF-specific probe sets according to the following criteria: (i) the ratio of the mean expression level between HighEF and CHF was ≥ 3, and (ii) the mean expression level in HighEF was ≥ 10 arbitrary units (U). These criteria resulted in the isolation of 67 probe sets (see Table S3 in Supporting Information). CHF-specific probe sets were also selected on the basis of a CHF/HighEF ratio for mean expression level of ≥ 3 and a mean expression level in CHF of ≥ 10 U, resulting in the identification of 80 probe sets (see Table S4 in Supporting Information). A total of

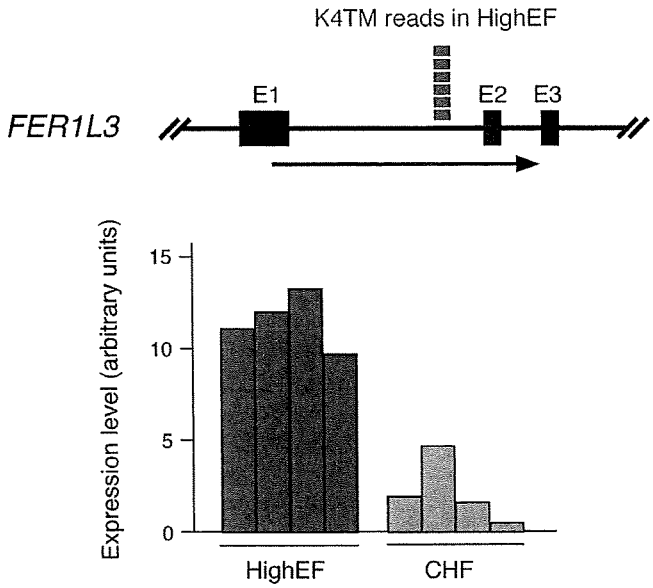

Figure 4 Epigenetic profile and mRNA abundance for *FER1L3*. Six sequence reads were selectively identified in the first intron of the *FER1L3* gene for the K4TM ChIP products of the HighEF sample (upper panel). E, exon. Consistent with this epigenetic profile, the amount of *FER1L3* mRNA was higher in the HighEF specimens than in the CHF specimens, as judged from the microarray data (lower panel).

16 152 of the transcripts measured with the U133 Plus 2.0 arrays mapped within a distance of ≤ 5 kbp relative to the high clusters. A Venn diagram revealed that only 21 probe sets were shared between the HighEF-specific and high cluster–associated transcripts, whereas 25 probe sets were shared between the CHF-specific and high cluster–associated transcripts (Fig. 3C). The K4TM mark has been found to map preferentially to the transcription start sites of active genes (Bernstein *et al.* 2005). Although a typical correlation between the K4TM modification and selective gene expression was apparent for a subset of genes (Fig. 4), our results suggest that this dynamic epigenetic regulation in the heart may not always directly participate in transcriptional regulation of neighboring genes.

Discussion

In the present study, we have revealed heart failure–dependent changes in the epigenetic profiles for K4TM and K9TM marks. The antibodies used in this study have been utilized in other reports for ChIP experiments, with those for K4TM and K9TM being especially employed in a genome-wide epigenetic profiling (Pokholok *et al.* 2005; Vakoc *et al.* 2006). Although it is difficult to

extensively verify our data in this study (because of the lack of knowledge in epigenetic profiles in heart), our ChIP procedure could faithfully confirm the epigenetic data demonstrated in previous studies [You *et al.* have, for instance, revealed that an apicidin treatment decreases the K4TM level while increases the K9TM level in the exon 1 of *DNMT1* in HeLa cells (You *et al.* 2008), and we could observe similar changes in the same experiment (data not shown)], supporting the reliability of our ChIP procedures.

Despite increasing evidence for a role of histone acetylation-deacetylation in the development of cardiac hypertrophy and heart failure, little information has been available for other histone modifications in these conditions (Illi *et al.* 2005; Phan *et al.* 2005; Bingham *et al.* 2007). Given the marked differences between the profiles of dimethylation and trimethylation for both K4 and K9 sites of histone H3, such trimethylation is likely under strict regulation in failed hearts.

The genes positioned close to the K4TM or K9TM marks were highly enriched in those that encode components of signaling pathways related to cardiac function. The HighEF-specific K4TM modification was, for instance, associated with *RYR2*, *CACNA2D1*, and *CACNB2* genes, the products of which directly participate in the regulation of intracellular calcium concentration and in muscle contraction (Cataldi *et al.* 1999; Marx *et al.* 2000). However, such disease-dependent histone methylation was not always linked to the induction or repression of neighboring genes. The expression level of the above three genes thus did not differ significantly between HighEF and CHF specimens (data not shown). Furthermore, only ~30% of HighEF- or CHF-specific transcripts were derived from genes associated with disease-dependent K4TM or K9TM modification (Fig. 3C). Consistent with such observations, the expression ratio for probe sets between normal and failed hearts of Dahl rats was not significantly correlated with the intensity ratio for any of the examined histone modifications, including H3Ac and H4Ac (data not shown). Therefore, despite the marked association between disease status and both transcript abundance and a subset of histone modifications, none of the latter can directly account for the former.

The epigenetic changes associated with heart failure may regulate gene transcription not through a single modification but through a combination of various marks (the "histone code" hypothesis) (Strahl & Allis 2000). The disease-dependent epigenetic changes also may alter the conformation of chromosomes, inducing an open or closed chromatin structure that indirectly affects the targets of subsequent regulation, such as the binding of transcription factors or additional chromatin remodeling. The subsequent regulation step would then play an important role in transcription of neighboring genes. In either case, our epigenetic profiles should facilitate further investigations into the roles of epigenetic changes in the development of heart failure.

Experimental procedures

ChIP-on-chip experiments

Dahl salt-sensitive rats (Japan SLC) at 6 weeks of age were maintained on a low-sodium diet (0.3% NaCl) or switched to a high-sodium diet (8% NaCl); the latter animals developed heart failure, as detected by echocardiography, after 13 weeks, as described previously (Ueno *et al.* 2003). ChIP products were prepared from the LV myocytes of 19-week-old Dahl rats with antibodies specific to H3Ac (Upstate, #17-245), H4Ac (Upstate, #17-229), K4DM (Abcam, #ab7766), K4TM (Abcam, #ab8580), K9DM (Upstate, #07-441), K9TM (Upstate, #07-442), K20TM (Abcam, #ab9053) or K27DM (Upstate, #07-452). The products were amplified by T7 RNA polymerase and subjected to hybridization with Affymetrix Rat Genome 230 2.0 microarrays as described previously (Takayama *et al.* 2007). Total genomic DNA (Pre-ChIP) and cRNA prepared from the LV tissue were also hybridized to the same arrays. The mean expression intensity of all probe sets was set to 500 U in each hybridization, and the fluorescence intensity of each test gene was normalized accordingly. Microarray data for rat and human hearts are available at the Gene Expression Omnibus web site (http://www.ncbi.nlm.nih.gov/geo) under the accession numbers GSE8341 and GSE8331, respectively. For the ChIP data, the signal intensity of each probe set in the Pre-ChIP analysis was then subtracted from that of the corresponding probe set in each ChIP experiment.

ChIP-to-seq experiments

All clinical specimens were obtained with written informed consent, and the study was approved by the ethics committees of Jichi Medical University and Hayama Heart Center. ChIP products were prepared from pooled samples for HighEF or CHF (each derived from four specimens) with antibodies to K4TM or K9TM. The products were converted to cRNA and amplified as described above for ChIP-on-chip experiments. The cRNA was then used to generate double-stranded DNA, which was subjected to pyrosequencing with a Genome Sequencer 20 system (Roche Diagnostics). Keypass wells occupied 82.7% to 87.0% of original Raw wells. Homology searches with the BLAST program were performed against the human genome sequence (the hg18 assembly) for each readout with the following parameter set: –e 2e-19 –v 50 –b 500 –T F –F F –m 8.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas (C) "Medical Genome Science" from the

Ministry of Education, Culture, Sports, Science and Technology of Japan and by a grant (#04C7) from Salt Science Research Foundation (Tokyo, Japan).

References

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J. 3rd, Gingeras, T.R., Schreiber, S.L. & Lander, E.S. (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181.

Bingham, A.J., Ooi, L., Kozera, L., White, E. & Wood, I.C. (2007) The repressor element 1-silencing transcription factor regulates heart-specific gene expression using multiple chromatin-modifying complexes. *Mol. Cell. Biol.* **27**, 4082–4092.

Braunwald, E. (1997) Shattuck lecture—cardiovascular medicine at the turn of the millennium: triumphs, concerns, and opportunities. *N. Engl. J. Med.* **337**, 1360–1369.

Cataldi, M., Secondo, A., D'Alessio, A., Taglialatela, M., Hofmann, F., Klugbauer, N., Di Renzo, G. & Annunziato, L. (1999) Studies on maitotoxin-induced intracellular $Ca^{(2+)}$ elevation in Chinese hamster ovary cells stably transfected with cDNAs encoding for L-type $Ca^{(2+)}$ channel subunits. *J. Pharmacol. Exp. Ther.* **290**, 725–730.

Gusterson, R.J., Jazrawi, E., Adcock, I.M. & Latchman, D.S. (2003) The transcriptional co-activators CREB-binding protein (CBP) and p300 play a critical role in cardiac hypertrophy that is dependent on their histone acetyltransferase activity. *J. Biol. Chem.* **278**, 6838–6847.

Iezzi, S., Di Padova, M., Serra, C., Caretti, G., Simone, C., Maklan, E., Minetti, G., Zhao, P., Hoffman, E.P., Puri, P.L. & Sartorelli, V. (2004) Deacetylase inhibitors increase muscle cell size by promoting myoblast recruitment and fusion through induction of follistatin. *Dev. Cell* **6**, 673–684.

Illi, B., Scopece, A., Nanni, S., Farsetti, A., Morgante, L., Biglioli, P., Capogrossi, M.C. & Gaetano, C. (2005) Epigenetic histone modification and cardiovascular lineage programming in mouse embryonic stem cells exposed to laminar shear stress. *Circ. Res.* **96**, 501–508.

James, M.A., Saadeh, A.M. & Jones, J.V. (2000) Wall stress and hypertension. *J. Cardiovasc. Risk* **7**, 187–190.

Kuwahara, K., Saito, Y., Ogawa, E., Takahashi, N., Nakagawa, Y., Naruse, Y., Harada, M., Hamanaka, I., Izumi, T., Miyamoto, Y., Kishimoto, I., Kawakami, R., Nakanishi, M., Mori, N. & Nakao, K. (2001) The neuron-restrictive silencer element-neuron-restrictive silencer factor system regulates basal and endothelin 1-inducible atrial natriuretic peptide gene expression in ventricular myocytes. *Mol. Cell. Biol.* **21**, 2085–2097.

Marx, S.O., Reiken, S., Hisamatsu, Y., Jayaraman, T., Burkhoff, D., Rosemblit, N. & Marks, A.R. (2000) PKA phosphorylation dissociates FKBP12.6 from the calcium release channel (ryanodine receptor): defective regulation in failing hearts. *Cell* **101**, 365–376.

Phan, D., Rasmussen, T.L., Nakagawa, O., McAnally, J., Gottlieb, P.D., Tucker, P.W., Richardson, J.A., Bassel-Duby, R. & Olson, E.N. (2005) BOP, a regulator of right ventricular heart development, is a direct transcriptional target of MEF2C in the developing heart. *Development* **132**, 2669–2678.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D.K. & Young, R.A. (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517–527.

Rapp, J.P., Wang, S.M. & Dene, H. (1989) A genetic polymorphism in the renin gene of Dahl rats cosegregates with blood pressure. *Science* **243**, 542–544.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J. & Kouzarides, T. (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411.

Strahl, B.D. & Allis, C.D. (2000) The language of covalent histone modifications. *Nature* **403**, 41–45.

Takayama, K., Kaneshiro, K., Tsutsumi, S., Horie-Inoue, K., Ikeda, K., Urano, T., Ijichi, N., Ouchi, Y., Shirahige, K., Aburatani, H. & Inoue, S. (2007) Identification of novel androgen response genes in prostate cancer cells by coupling chromatin immunoprecipitation and genomic microarray analysis. *Oncogene* **26**, 4453–4463.

Ueno, S., Ohki, R., Hashimoto, T., Takizawa, T., Takeuchi, K., Yamashita, Y., Ota, J., Choi, Y.L., Wada, T., Koinuma, K., Yamamoto, K., Ikeda, U., Shimada, K. & Mano, H. (2003) DNA microarray analysis of *in vivo* progression mechanism of heart failure. *Biochem. Biophys. Res. Commun.* **307**, 771–777.

Vakoc, C.R., Sachdeva, M.M., Wang, H. & Blobel, G.A. (2006) Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol. Cell. Biol.* **26**, 9185–9195.

You, J.S., Kang, J.K., Lee, E.K., Lee, J.C., Lee, S.H., Jeon, Y.J., Koh, D.H., Ahn, S.H., Seo, D.W., Lee, H.Y., Cho, E.J. & Han, J.W. (2008) Histone deacetylase inhibitor apicidin downregulates DNA methyltransferase 1 expression and induces repressive histone modifications via recruitment of corepressor complex to promoter region in human cervix cancer cells. *Oncogene* **27**, 1376–1386.

Zhang, C.L., McKinsey, T.A., Chang, S., Antos, C.L., Hill, J.A. & Olson, E.N. (2002) Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell* **110**, 479–488.

Received: 26 June 2008
Accepted: 10 October 2008

Supporting Information/Supplementary materials

The following Supporting Information can be found in the online version of the article:

Figure S1 Distribution of K4TM and K9TM marks on chromosome 1.

Figure S2 Distribution of read number per cluster in ChIP products.

Figure S3 Chromosome distribution of disease-specific high clusters.

Figure S4 Protein complexes in the synaptic long-term potentiation pathway in Fig. 3A are colored red or green on the basis of whether the corresponding genes are associated with HighEF-specific or CHF-specific high clusters for K4TM, respectively.

Figure S5 Interaction map for a protein network that contains the products of 18 and 16 genes associated with the HighEF-specific and CHF-specific high clusters for K4TM, respectively.

Figure S6 Network for the products of genes that mapped in the vicinity of K9TM high clusters.

Table S1 Output of pyrosequencing.

Table S2 High clusters identified in the heart specimens.

Table S3 Expression intensity of HighEF-specific probe sets.

Table S4 Expression intensity of CHF-specific probe sets.

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

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