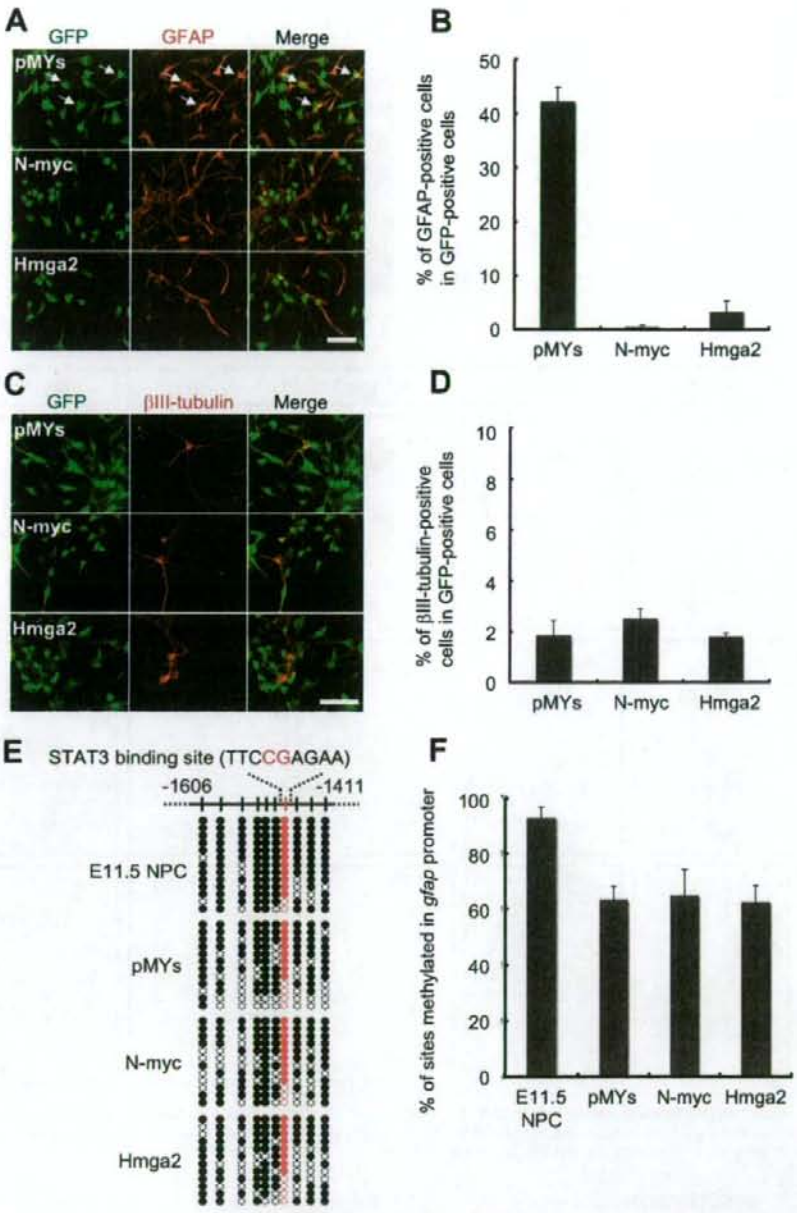

Fig. 3. N-myc and Hmga2 inhibit astrocyte differentiation of NPCs even in the presence of LIF. (A–D) E14.5 NPCs were infected with recombinant retroviruses engineered to express *EGFP* alone (pMYs), and *EGFP* together with either *N-myc* (N-myc) or *Hmga2* (Hmga2), and cultured with LIF (50 ng/ml) for 4 days to induce astrocyte differentiation. The cells were then stained with antibodies against GFP (green) and GFAP (red) (A). Arrows indicate GFP/GFAP double-positive cells. The percentage of GFAP-positive astrocytes in GFP-positive cells was quantified (B). Mean±S.D. The cells were also stained with antibodies against GFP (green) and βIII-tubulin (red) (C). The percentage of βIII-tubulin-positive neurons in GFP-positive cells was quantified (D). Mean±S.D. (E) E11.5 NPCs were infected with recombinant retroviruses engineered to express *EGFP* alone (pMYs), and *EGFP* together with either *N-myc* (N-myc) or *Hmga2* (Hmga2), and subsequently cultured for 4 days. GFP-positive cells were then sorted based on GFP fluorescence, and genomic DNA was extracted for bisulfite sequencing. White and black circles indicate unmethylated and methylated CpG sites, respectively. The E11.5 NPC sample was freshly prepared NPCs from telencephalon at E11.5. The CpG dinucleotide within the STAT3 binding site is indicated in red. (F) Methylation frequency in the *gfap* promoter. Mean±S.D. Scale bars=50 μm (A, C). For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.

been reported to be expressed in a wide range of vertebrate tissues, primarily during embryogenesis (Schreiber-Agus et al., 1993). The mice deficient for functional *N-myc* are embryonic lethal (Stanton et al., 1992). Since N-myc has been shown to be a transcriptional activator, it may inhibit astrocyte differentiation via induction of neurogenic bHLH factors such as *Ngn1* (Sun et al., 2001), which have already been suggested to inhibit astrocyte differentiation in midgestational NPCs. However, this scenario seems unlikely because *N-myc* expression in NPCs did not affect neuronal differentiation, as assessed by monitoring expression of the neuronal marker βIII-tubulin (Fig. 3C, D). On the other hand, Hmga2 possesses an acidic C-terminal tail and three individual DNA-binding domains which bind short stretches of AT-rich DNA with high affinity (Reeves, 2001). *Hmga2* is expressed in pluripotent embryonic stem (ES) cells and in most tissues and organs during embryogenesis, but at very low levels or not at all in adult tissues (Zhou et al., 1995). Its function appears to be critical for cell growth, because mice lacking functional *Hmga2* exhibit a pygmy phenotype (Zhou et al., 1995). Recently, it was reported that Hmga2 specifically accumulates on senescent cell chromatin and that it functions as a structural component of senescence-associated heterochromatin foci and as a repressor of proliferation-associated genes (Narita et al., 2006). We therefore expected that Hmga2 would maintain the hypermethylation status of the astrocyte-specific *gfap* promoter via transcription-repressive heterochromatin formation in E11.5 NPCs. However, our results indicate that this is not the case. The mechanism(s) whereby N-myc and Hmga2 inhibit astrocyte differentiation must await further investigation.

Although DNA methylation is a critical cell-intrinsic determinant for the neurogenic-to-astrogliogenic switch and/or astrocyte differentiation of NPCs, many other spatio-temporally expressed extracellular factors such as CT-1, Notch and Wnt1 (Barnabe-Heider et al., 2005; Hirabayashi and Gotoh, 2005; Nagao et al., 2007) and intracellular factors including Ngn (Sun et al., 2001), N-CoR (Hermanson et al., 2002), N-myc and Hmga2 (this study) complement DNA methylation to ensure the sequential differentiation of NPCs during development. Thus, to better understand the mechanism underlying these processes, this study emphasizes the need to take cell-extrinsic cues, cell-intrinsic programs and factors, and their interaction into consideration.

Acknowledgments—We thank Dr. T. Kitamura (Tokyo University) for pMY vector and Plat-E cells. We appreciate Dr. Y. Bessho and T. Matsui for valuable discussions. We also thank Dr. I. Smith for helpful comments and critical reading of the manuscript. We are very grateful to N. Ueda for excellent secretarial assistance. Many thanks to N. Namihira for technical help. We also thank N. Moriyama for technical help with GeneChip analysis. This work has been supported by a Grant-in-Aid for Science Research on Priority Areas and the NAIST Global COE Program (Frontier Biosciences: Strategies for survival and adaptation in a changing global environment) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

REFERENCES

Abramova N, Charniga C, Goderie SK, Temple S (2005) Stage-specific changes in gene expression in acutely isolated mouse CNS progenitor cells. Dev Biol 283:269–281.

Ajioka I, Maeda T, Nakajima K (2006) Identification of ventricular-side-enriched molecules regulated in a stage-dependent manner during cerebral cortical development. Eur J Neurosci 23:296–308.

Alex R, Sozeri O, Meyer S, Dildrop R (1992) Determination of the DNA sequence recognized by the bHLH-zip domain of the N-Myc protein. Nucleic Acids Res 20:2257–2263.

Barnabe-Heider F, Wasylnka JA, Fernandes KJ, Porsche C, Sendtner M, Kaplan DR, Miller FD (2005) Evidence that embryonic neurons regulate the onset of cortical gliogenesis via cardiotrophin-1. Neuron 48:253–265.

Bonni A, Sun Y, Nadal-Vicens M, Bhatt A, Frank DA, Rozovsky I, Stahl N, Yancopoulos GD, Greenberg ME (1997) Regulation of gliogenesis in the central nervous system by the JAK-STAT signaling pathway. Science 278:477–483.

Brunelli S, Innocenzi A, Cossu G (2003) Bhlhb5 is expressed in the CNS and sensory organs during mouse embryonic development. Gene Expr Patterns 3:755–759.

Bugga L, Gadient RA, Kwan K, Stewart CL, Patterson PH (1998) Analysis of neuronal and glial phenotypes in brains of mice deficient in leukemia inhibitory factor. J Neurobiol 36:509–524.

Cai L, Morrow EM, Cepko CL (2000) Misexpression of basic helix-loop-helix genes in the murine cerebral cortex affects cell fate choices and neuronal survival. Development 127:3021–3030.

Edlund T, Jessell TM (1999) Progression from extrinsic to intrinsic signaling in cell fate specification: a view from the nervous system. Cell 96:211–224.

Graham V, Khudyakov J, Ellis P, Pevny L (2003) SOX2 functions to maintain neural progenitor identity. Neuron 39:749–765.

He F, Ge W, Martinowich K, Becker-Catania S, Coskun V, Zhu W, Wu H, Castro D, Guillemot F, Fan G, de Vellis J, Sun YE (2005) A positive autoregulatory loop of Jak-STAT signaling controls the onset of astrogliogenesis. Nat Neurosci 8:616–625.

Hermanson O, Jepsen K, Rosenfeld MG (2002) N-CoR controls differentiation of neural stem cells into astrocytes. Nature 419: 934–939.

Hirabayashi Y, Gotoh Y (2005) Stage-dependent fate determination of neural precursor cells in mouse forebrain. Neurosci Res 51: 331–336.

Hsieh J, Gage FH (2004) Epigenetic control of neural stem cell fate. Curr Opin Genet Dev 14:461–469.

Kanno J, Aisaki K, Igarashi K, Nakatsu N, Ono A, Kodama Y, Nagao T (2006) "Per cell" normalization method for mRNA measurement by quantitative PCR and microarrays. BMC Genomics 7:64.

Knoepfler PS, Cheng PF, Eisenman RN (2002) N-myc is essential during neurogenesis for the rapid expansion of progenitor cell populations and the inhibition of neuronal differentiation. Genes Dev 16:2699–2712.

Koblar SA, Turnley AM, Classon BJ, Reid KL, Ware CB, Cheema SS, Murphy M, Bartlett PF (1998) Neural precursor differentiation into astrocytes requires signaling through the leukemia inhibitory factor receptor. Proc Natl Acad Sci U S A 95:3178–3181.

Morita S, Kojima T, Kitamura T (2000) Plat-E: an efficient and stable system for transient packaging of retroviruses. Gene Ther 7: 1063–1066.

Nagao M, Sugimori M, Nakafuku M (2007) Cross talk between notch and growth factor/cytokine signaling pathways in neural stem cells. Mol Cell Biol 27:3982–3994.

Nakashima K, Wiese S, Yanagisawa M, Arakawa H, Kimura N, Hisatsune T, Yoshida K, Kishimoto T, Sendtner M, Taga T (1999a) Developmental requirement of gp130 signaling in neuronal survival and astrocyte differentiation. J Neurosci 19:5429–5434.

Nakashima K, Yanagisawa M, Arakawa H, Kimura N, Hisatsune T, Kawabata M, Miyazono K, Taga T (1999b) Synergistic signaling in

fetal brain by STAT3-Smad1 complex bridged by p300. Science 284:479–482.

Narita M, Krizhanovsky V, Nunez S, Chicas A, Hearn SA, Myers MP, Lowe SW (2006) A novel role for high-mobility group a proteins in cellular senescence and heterochromatin formation. Cell 126: 503–514.

Nieto M, Schuurmans C, Britz O, Guillemot F (2001) Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors. Neuron 29:401–413.

Rajan P, McKay RD (1998) Multiple routes to astrocytic differentiation in the CNS. J Neurosci 18:3620–3629.

Ramsay G, Stanton L, Schwab M, Bishop JM (1986) Human proto-oncogene N-myc encodes nuclear proteins that bind DNA. Mol Cell Biol 6:4450–4457.

Reeves R (2001) Molecular biology of HMGA proteins: hubs of nuclear function. Gene 277:63–81.

Saiki Y, Yamazaki Y, Yoshida M, Katoh O, Nakamura T (2000) Human EVI9, a homologue of the mouse myeloid leukemia gene, is expressed in the hematopoietic progenitors and down-regulated during myeloid differentiation of HL60 cells. Genomics 70:387–391.

Sawai S, Kato K, Wakamatsu Y, Kondoh H (1990) Organization and expression of the chicken N-myc gene. Mol Cell Biol 10:2017–2026.

Schreiber-Agus N, Horner J, Torres R, Chiu FC, DePinho RA (1993) Zebra fish myc family and max genes: differential expression and oncogenic activity throughout vertebrate evolution. Mol Cell Biol 13:2765–2775.

Sock E, Rettig SD, Enderich J, Bosl MR, Tamm ER, Wegner M (2004) Gene targeting reveals a widespread role for the high-mobility-group transcription factor Sox11 in tissue remodeling. Mol Cell Biol 24:6635–6644.

Stanton BR, Perkins AS, Tessarollo L, Sassoon DA, Parada LF (1992) Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop. Genes Dev 6:2235–2247.

Sun Y, Nadal-Vicens M, Misono S, Lin MZ, Zubiaga A, Hua X, Fan G, Greenberg ME (2001) Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. Cell 104:365–376.

Takizawa T, Nakashima K, Namihira M, Ochiai W, Uemura A, Yanagisawa M, Fujita N, Nakao M, Taga T (2001) DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain. Dev Cell 1:749–758.

Temple S (2001) The development of neural stem cells. Nature 414:112–117.

Tomita K, Moriyoshi K, Nakanishi S, Guillemot F, Kageyama R (2000) Mammalian achaete-scute and atonal homologs regulate neuronal versus glial fate determination in the central nervous system. EMBO J 19:5460–5472.

Zhou X, Benson KF, Ashar HR, Chada K (1995) Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. Nature 376:771–774.

(Accepted 13 June 2008)
(Available online 21 June 2008)

Development 135, 3511-3519 (2008) doi:10.1242/dev.027144

Functional importance of evolutionally conserved Tbx6 binding sites in the presomitic mesoderm-specific enhancer of *Mesp2*

Yukuto Yasuhiko[1,*], Satoshi Kitajima[1], Yu Takahashi[1], Masayuki Oginuma[2], Harumi Kagiwada[3], Jun Kanno[1] and Yumiko Saga[2,*]

The T-box transcription factor Tbx6 controls the expression of *Mesp2*, which encodes a basic helix-loop-helix transcription factor that has crucial roles in somitogenesis. In cultured cells, Tbx6 binding to the *Mesp2* enhancer region is essential for the activation of *Mesp2* by Notch signaling. However, it is not known whether this binding is required in vivo. Here we report that an *Mesp2* enhancer knockout mouse bearing mutations in two crucial Tbx6 binding sites does not express *Mesp2* in the presomitic mesoderm. This absence leads to impaired skeletal segmentation identical to that reported for *Mesp2*-null mice, indicating that these Tbx6 binding sites are indispensable for *Mesp2* expression. T-box binding to the consensus sequences in the *Mesp2* upstream region was confirmed by chromatin immunoprecipitation assays. Further enhancer analyses indicated that the number and spatial organization of the T-box binding sites are critical for initiating *Mesp2* transcription via Notch signaling. We also generated a knock-in mouse in which the endogenous *Mesp2* enhancer was replaced by the core enhancer of medaka *mespb*, an ortholog of mouse *Mesp2*. The homozygous enhancer knock-in mouse was viable and showed normal skeletal segmentation, indicating that the medaka *mespb* enhancer functionally replaced the mouse *Mesp2* enhancer. These results demonstrate that there is significant evolutionary conservation of Mesp regulatory mechanisms between fish and mice.

KEY WORDS: T-box transcription factor, Enhancer, Targeted disruption, Somitogenesis

INTRODUCTION

Somitogenesis is an important morphogenic process that generates metameric structures in vertebrates, including vertebra, muscles and motoneurons. The segmental boundary of each somite forms at the anterior end of the presomitic mesoderm (PSM) or unsegmented paraxial mesoderm, which are supplied from the primitive streak or tailbud at a later stage of development (Saga and Takeda, 2001). This process proceeds through the interaction of a number of signaling cascades, including Notch, Wnt and Fgf (Delfini et al., 2005; Dunty et al., 2008; Galceran et al., 2004; Hofmann et al., 2004; Moreno and Kintner, 2004; Takahashi et al., 2000). Thus, somitogenesis could be a very useful model system in which to study the interactions among the various signaling cascades that facilitate periodic pattern formation.

The basic helix-loop-helix transcription factor Mesp2 plays a crucial role in both somite segment border formation and in the establishment of the rostrocaudal patterning of each somite (Saga et al., 1997). *Mesp2* shows dynamic and periodic expression in the anterior PSM. This expression pattern defines the positioning of the newly forming somite by suppressing Notch signaling, in part through the activation of lunatic fringe (*Lfng*) (Morimoto et al., 2005). Genetic analyses have revealed that *Mesp2* expression is itself controlled by Notch signaling, indicating the existence of complicated feedback circuitry (Takahashi et al., 2003; Takahashi et al., 2000). We have previously identified the minimal PSM-specific *Mesp2* enhancer (denoted P2PSME) that is sufficient to reproduce the normal *Mesp2* expression pattern in transgenic animals (Haraguchi et al., 2001). We have also demonstrated that the T-box transcriptional regulator Tbx6 directly binds to P2PSME and is essential for P2PSME activity (Yasuhiko et al., 2006). We also showed that Notch signaling strongly enhanced *Mesp2* activation via Tbx6 and we identified the sequences that are important for this enhancement using an in vitro reporter assay (Yasuhiko et al., 2006). However, the question of whether P2PSME is indispensable for *Mesp2* expression during somitogenesis remained to be addressed. Because of differences in the expression patterns of *Mesp2* and Tbx6 – Tbx6 is expressed throughout the PSM and tailbud (Chapman et al., 1996; White and Chapman, 2005) whereas *Mesp2* expression is observed only in the anterior PSM (Saga et al., 1997) – another open question was whether Tbx6 actually binds to P2PSME.

The evolutionary aspect of this system is also noteworthy. We previously identified the *mespb* PSM-specific enhancer in the teleost fish medaka, and reported that the mutation of two T-box binding sites therein diminished its PSM-specific enhancer activity in transgenic embryos (Terasaki et al., 2006). However, definitive evidence as to whether the T-box-factor-dependent regulation is a conserved mechanism among vertebrates remains elusive.

In this study, we established *Mesp2* enhancer knockout mice and confirmed that Tbx6 binding sequences are essential for *Mesp2* expression. The in vivo association of Tbx6 with P2PSME was confirmed in chromatin immunoprecipitation assays, and reporter assays further showed that the number and spatial organization of Tbx6 binding sites are important for P2PSME activity. Furthermore, using a knock-in mouse that harbors the medaka *mespb* enhancer in place of the mouse *Mesp2* enhancer, we show that the T-box-factor-dependent regulation of the Mesp gene is evolutionally conserved between fish and mice.

[1]Division of Cellular and Molecular Toxicology, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan. [2]Division of Mammalian Development, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan. [3]Research Institute for Cell Engineering, National Institute of Advanced Industrial Science and Technology, 3-11-46 Nakoji, Amagasaki, Hyogo 661-0974 Japan.

*Authors for correspondence (e-mails: yasuhiko@nihs.go.jp; ysaga@lab.nig.ac.jp)

Accepted 8 September 2008

MATERIALS AND METHODS

Site-directed mutagenesis

Site-directed mutagenesis of each Tbx6 binding site was performed using previously reported PCR-based procedures (Yasuhiko et al., 2006) with the following primers (mutated nucleotides in lower case): mB1, 5′-CCTTCGAGGGGTCAGAATCgAtAtCTCTGCAAATGGGCCCGCTTT-3′; mB2, 5′-CCTTCGAGaGtaCtGAATCCACACCTCTGCAAATGG-GCCCGCTTT-3′; mD, 5′-AACCTGGCAGGGGACCACCTCgCgaCTT-TAGTCCAGATAAAAGCT-3′; mG, 5′-CTGGGCTCTGTGGGTTTTG-AattCTCTCTGCAACCTGGCA-3′. The mutated Tbx6 binding sites are indicated for each construct, such that P2EmB1D represents a P2PSME containing both mB1 and mD.

Gene targeting

For targeted disruption of P2PSME, a 356-bp DNA fragment containing mutated Site B and Site D was generated by PCR using primers mB1 and mD. As a negative control, the wild-type P2PSME fragment was also generated by PCR. To construct the targeting vectors, a floxed PGK-neoR selection marker cassette was inserted between a 6-kb long arm and the 356-bp DNA fragment with or without mutations (Fig. 1A). The region corresponding to *Mesp2* exon 1, intron 1 and a part of exon 2 served as the short homology arm. The targeting vector was introduced into mouse ES cells (strain TT2) by electroporation. Resulting G418-resistant ES clones were characterized by PCR using primers: Fesneo, 5′-CGCCTTCT-ATCGCCTTCTTGACGAG-3′ and RP213, 5′-CAGGACAGCCACT-GAGCTGCAGGCCTGA-3′. Southern blots were performed to confirm homologous recombination. Positive ES clones were then aggregated with 8-cell stage ICR mouse embryos in order to produce chimeric mice. The ES selection marker PGK-neoR was removed by crossing the chimeric mice with CAG-Cre mice, which express Cre recombinase ubiquitously. The resulting mouse strains, with insertions of either mutated P2PSME or wild-type P2PSME, were designated P2EmB1D or P2EmCont, respectively. Although the knockout mice were established using an ES cell line (TT2) obtained from a C57BL/6 × CBA cross (Yagi et al., 1993), mice were maintained in an ICR background unless otherwise stated.

Skeletal preparation

Embryonic day 17.5 (E17.5) mouse embryos were obtained by crossing the mutants of interest. Embryos were then fixed with 90% ethanol. For genotyping, PCR was performed using a piece of embryonic liver digested with proteinase K (Roche). Alcian Blue and Alizarin Red staining were performed as described (Saga et al., 1997; Takahashi et al., 2000).

Generation of anti-Tbx6 antibody

His-tagged fragments of Tbx6 protein (N-terminal antigen, amino acids 2-78; internal antigen, amino acids 311-408) (White and Chapman, 2005) were produced using the pET system (Novagen) and *Escherichia coli* Rosetta-gamiB (Novagen) as a host strain. The Tbx6 fragments were extracted from bacterial culture using the MagneHis system (Promega), purified by thrombin digestion to remove the His-tag, followed by affinity column purification (Novagen) and dialysis using a semipermeable membrane cassette (Pierce). Rabbits (two animals for each antigen) were immunized with the purified Tbx6 fragments and processed for antibody purification following the standard procedures of Hokudo Bio (Abuta, Hokkaido, Japan).

Protein and mRNA expression analyses

Whole-mount RNA in situ hybridization was performed as described (Saga et al., 1997). Whole-mount immunohistochemistry and simultaneous staining of *Mesp2* mRNA and Tbx6 protein were as previously described (Morimoto et al., 2005; Oginuma et al., 2008).

Chromatin immunoprecipitation (ChIP) assay

Embryonic tails were dissected along the anteroposterior axis into three parts using a tungsten needle. Somite part (s) corresponds to SIV to SII, anterior PSM (ap) is from SI to S-1, and posterior PSM (pp) corresponds to the region posterior to S-2. A total of 120 embryos were dissected, the samples treated with trypsin and dispersed cells counted (around 1×10^6 cells for each sample). Cells were fixed in 1% formaldehyde in PBS for 10 minutes at 37°C. The preparation of cell lysates and ChIP assay were performed using the Chromatin Immunoprecipitation Assay Kit (Upstate biology) according to the manufacturer's protocol. PCR primers used for ChIP assays were: LP286, 5′-AGACATCCAGGTACCTCGAGGTC-3′; LP287, 5′-CGG-GATAGACATCCAGGTACCCA-3′; and RP287, 5′-GGCTGGTGT-GACTCTGGGAAGCT-3′. LP286 and RP287 were used for detection of mutated P2PSME, whereas LP287 and RP287 were used for detection of wild-type P2PSME. As a positive control, the *Dll1* mesoderm (msd) enhancer was amplified using the following primers: LP259, 5′-CCCAACACAGATGATTCTGCCCAGTAACT-3′; and RP255, 5′-GCT-TTGTGTTGAGCATGCCATGAGCTGTA-3′. A sequence 22 kb from P2PSME was amplified by PCR as a negative control, using the following primers: LP285, 5′-GGTCTGTTTGCAGCTGATTCTGAA-3′; and RP286, 5′-CAGTTCTCACCTTGCTTCCATGT-3′.

Electromobility shift assay (EMSA)

The full-length *Tbx6* ORF was obtained from the pACT-Tbx6 construct, which was previously isolated from a yeast one-hybrid screen (Yasuhiko et al., 2006). After ligation to a 3×FLAG tag (Sigma), the tagged *Tbx6* insert was cloned into pCS2+ (Rupp et al., 1994). In vitro transcription/translation was then performed using the TNT In Vitro Translation Kit (Promega) according to the manufacturer's protocol. Oligonucleotide probes were labeled with DIG-11-ddUTP using recombinant TdT (Roche Diagnostics). Five microliters of crude in vitro translated product was subjected to EMSA. As a negative control, reticulocyte lysate without *Tbx6* template was used. The oligonucleotide probes are as follows (mutated nucleotides are indicated in lower case): SiteF, 5′-GCTAAATTACGGGTATATGGACCACAC-CTGTATCAGTCCC-3′; SiteG, 5′-CTGGGCTCTGTGGGTTTTGACA-CCTCTCTGCAACCTGGCA-3′; SiteGmut, 5′-CTGGGCTCTGTGG-GTTTTGAattCTCTCTGCAACCTGGCA-3′; SiteB, 5′-CCTTCGAGG-GGTCAGAATCCACACCTCTGCAAATGGGCCCGCTTT-3′; T1, 5′-CAAGTGCTGGTCTTGGCATCACACCTCTTTATTTGTTTCCATAC-3′; T2, 5′-GCAGAATCTGCAGAGGTGTCACTTCACACCTCTGTGG-CCTGGCT-3′; and T3, 5′-GCTCTCACAGCTGAGGTGTGAAGCG-ACACCTCCAGGCTCATAAG-3′.

EMSA was performed as described (Yasuhiko et al., 2006). Anti-Tbx6 antibody (3.5 μg) was added to the reaction to assess the specificity of the protein-DNA interaction. As a competitor, a 100-fold excess of unlabeled oligonucleotide corresponding to the probe was added to the reaction.

Transgenic assay

DNA fragments with and without mutations in conserved upstream sites were generated from an *Mesp2* genomic fragment using a standard PCR-based protocol. Each transgene comprised the *lacZ* reporter and a 6-kb genomic fragment upstream of the *Mesp2* first ATG, including P2PSME with and without mutated Tbx6 binding sites. The transgenes were injected into the male pronucleus of a fertilized egg as described (Hogan et al., 1994). Embryos recovered at E9.5-10.5 were analyzed for *lacZ* expression by X-Gal staining (Saga et al., 1992) and were subsequently examined for the presence of the transgene by PCR (Sasaki and Hogan, 1996).

Luciferase assay

The *KpnI-NcoI* fragments (356 bp) corresponding to P2PSME, with and without mutations in the Tbx6 binding sites, were subcloned into the pGL3-Basic (Promega) vector to generate luciferase reporter constructs. The expression vectors for the proteins to be assessed were constructed in the same way as those used in the EMSA assays described above. The luciferase assay using COS-7 cells was conducted as described previously (Yasuhiko et al., 2006). Each assay was performed in triplicate and repeated at least twice.

RESULTS

Mutations in the Tbx6 binding site of the *Mesp2* enhancer result in the complete loss of *Mesp2* expression in the presomitic mesoderm

We have shown previously that nucleotide substitutions in two Tbx6 binding motifs in the *Mesp2* PSM enhancer (P2PSME) eliminate Tbx6 binding activity in vitro (Yasuhiko et al., 2006). To establish the function of these Tbx6 binding sites in vivo, we

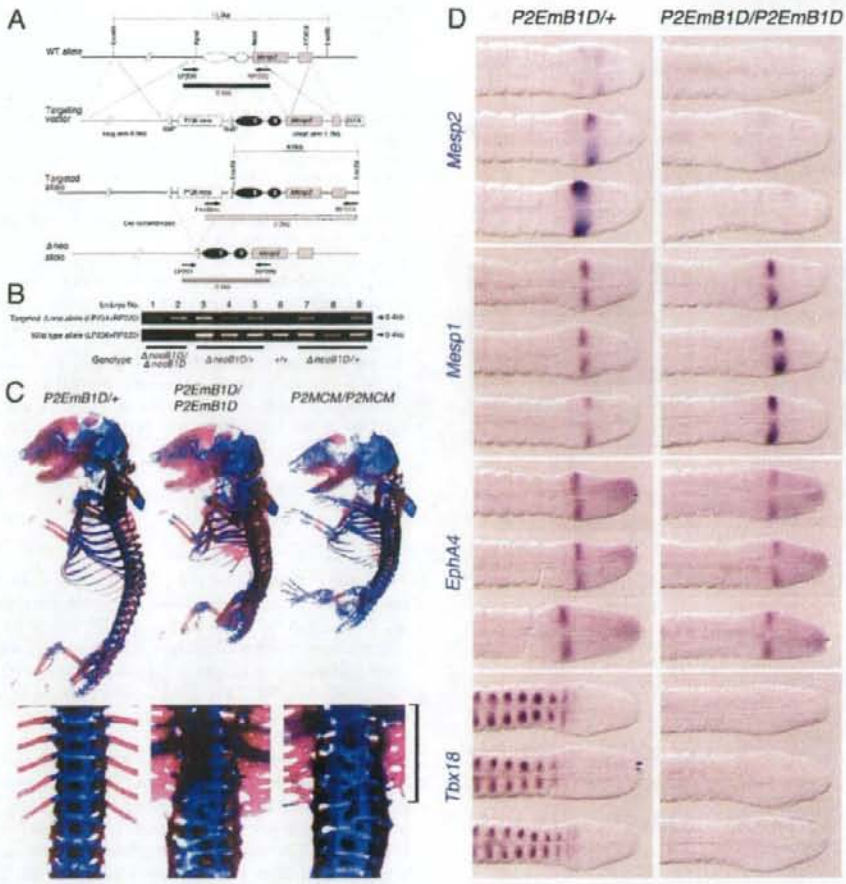

Fig. 1. Disruption of Tbx6 binding sites eliminates *Mesp2* expression. (**A**) Targeting strategy to generate the *Mesp2* enhancer knockout mouse (*P2EmB1D*). A DNA fragment containing mutated Tbx6 binding sites (black ovals with X) was substituted for the wild-type sequence (white ovals) by homologous recombination. The PGK-neoR selection marker was removed by the Cre-loxP system to obtain a Δneo allele. (**B**) PCR detection of homozygotes in the *P2EmB1D* intercross. (**C**) Impaired skeletal segmentation in the *Mesp2* enhancer knockout mouse. The *P2EmB1D/P2EmB1D* mouse exhibits severe skeletal malformation at E17.5 (centre) identical to that of the *Mesp2*-null mouse (*P2MCM/P2MCM*, right). Note the shortened spine with incompletely segmented vertebrae (upper panels) and fused ribs (bracket in lower panels). (**D**) Expression of *Mesp2* and the somite-specific genes *Mesp1*, *Epha4* and *Tbx18* in *P2EmB1D/+* (left column) and *P2EmB1D/P2EmB1D* (right column) embryos. *Mesp2* mRNA expression is eliminated in the *P2EmB1D/P2EmB1D* homozygotes. Wild-type (+/+) and heterozygote (*P2EmB1D/+*) embryos showed varying *Mesp2* expression patterns owing to its cyclic expression. *Mesp1* is upregulated and *Epha4* is not affected, whereas *Tbx18* is completely abolished in *P2EmB1D/P2EmB1D*.

introduced nucleotide substitutions into the mouse genome using a gene-targeting technique. These mutations disrupted two Tbx6 binding sites, denoted Site B and Site D, that were shown to be sufficient to activate *Mesp2* expression in vitro (Yasuhiko et al., 2006) (Fig. 1A). After the establishment of a neo*P2EmB1D* mouse line, the neoR cassette was removed (Δneo) by a cross with the deleter mouse line CAG-Cre. Interbreeding of the Δneo mutants gave rise to homozygotes (*P2EmB1D/P2EmB1D*) that retained a loxP site after neoR removal (Fig. 1B). This residual loxP site appears to have no effect on *Mesp2* expression or somitogenesis because another knock-in mouse, *P2EmCont*, in which wild-type P2PSME is knocked-in using the same strategy, had viable homozygous offspring without any morphological defects (data not shown).

The homozygous *P2EmB1D/P2EmB1D* embryos showed distinct skeletal defects (Fig. 1C) and perinatal lethality, features identical to the previously reported Δneo-type *Mesp2*-null mouse (Fig. 1C, *P2MCM/P2MCM*). As expected from the phenotype, *Mesp2* expression in *P2EmB1D/P2EmB1D* embryos was eliminated (Fig. 1D). Segmental borders were generated during an early stage of

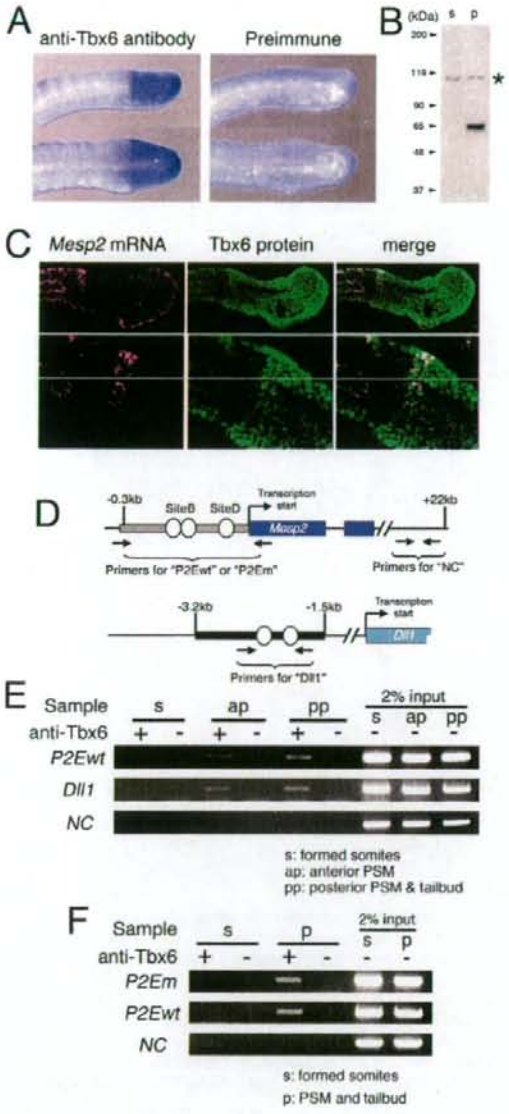

Fig. 2. Tbx6 binds to P2PSME in the PSM and tailbud. (**A**,**B**) Characterization of the anti-Tbx6 antibody produced in this study. (A) Whole-mount immunohistochemistry demonstrating the localization of Tbx6 protein in the mouse PSM and tailbud. (B) Western blot analysis showing that the anti-Tbx6 antibody detected a protein of expected molecular weight (58 kDa) in the PSM and tailbud (p) but not in formed somites (s). The asterisk indicates non-specific binding. (**C**) Double staining of *Mesp2* mRNA (purple) and Tbx6 protein (green) demonstrating the coexistence of both signals in the anterior-most part of the Tbx6-positive region (white in merged image). (**D**) Design of an in vivo technique for detecting Tbx6 binding to the P2PSME by ChIP. Arrows represent primers for the ChIP assay for the *Mesp2* and *Dll1* genes. *Dll1* is known to be downstream of Tbx6 and was therefore used as a positive control. Gray and black boxes represent the P2PSME and *Dll1* mesoderm (msd) enhancers, respectively. White ovals indicate Tbx6 binding sites. P2Em and P2Ewt, mutated and wild-type P2PSME regions, respectively; NC, unrelated sequence as negative control. (**E**) Tbx6 associates with P2PSME in the anterior and posterior PSM. (**F**) The association of Tbx6 with mutated P2PSME as detected by ChIP assay. Mutated and wild-type P2PSME regions were differentially detected by PCR with different sets of primers in the tails of E10.5 embryos obtained from the crossing of *P2EmB1Dl+* and ICR mice.

somitogenesis (Fig. 1D, right-hand panels). However, the borders were unlikely to be maintained because the vertebral bodies were fused along the anteroposterior axis at later developmental stages (Fig. 1C).

To further characterize the phenotypes in these embryos, we first examined the expression of *Mesp1*, which is known to be upregulated and to partially rescue somitogenesis in the absence of *Mesp2* (Morimoto et al., 2006; Takahashi et al., 2007). *Mesp1* expression was upregulated in the *P2EmB1D/P2EmB1D* embryo, but its expression domain was broader than normal (Fig. 1D). *Epha4*, which is required for proper border formation, was expressed normally. However, *Tbx18*, which is implicated in the maintenance of segmental border and somite patterning (Bussen et al., 2004), was not expressed (Fig. 1D). These gene expression patterns were similar to those reported for the Δneo-type *Mesp2*-null mouse (Morimoto et al., 2006; Takahashi et al., 2007). These results confirmed that the Tbx6 binding sites are bona fide enhancer elements required for *Mesp2* expression.

Tbx6 binds the *Mesp2* PSM enhancer in vivo

The Tbx6 protein is normally broadly distributed in the PSM and tailbud (White and Chapman, 2005), whereas *Mesp2* is expressed only in the anterior PSM (Saga et al., 1997). This discrepancy between the Tbx6 and *Mesp2* expression patterns prompted us to investigate whether Tbx6 actually binds to the *Mesp2* enhancer in vivo. We raised an anti-Tbx6 antibody using two different antigens: an N-terminal portion of Tbx6 and an internal portion. The internal antigen yielded an antibody with good specificity and sensitivity. Embryo whole-mount immunohistochemistry confirmed the previously reported distinct Tbx6 staining pattern in the PSM and tailbud (Fig. 2A) (White and Chapman, 2005). Western blot analyses further revealed that this antibody identifies a single band of approximately 58 kDa in cell lysates prepared from the posterior region (PSM and tailbud), but not from the anterior region (formed somite), of E11.5 tails (Fig. 2B). We also performed double staining of *Mesp2* mRNA and Tbx6 protein and confirmed colocalization only in the anterior-most region of the PSM (Fig. 2C).

For ChIP assays, we dissected E11.5 embryo tails into three regions: the tailbud and posterior PSM (pp), the anterior PSM and newly formed somites (ap), and formed somites (s). Protein-DNA complexes were prepared from each pool and used in ChIP assays, which revealed that Tbx6 binds to the *Mesp2* PSM enhancer in the ap and pp regions, but not in the s region, which is consistent with the expression pattern of Tbx6 (Fig. 2E). These results indicate that Tbx6 binds to the P2PSME uniformly in its expression domain, suggesting that Tbx6 alone cannot activate *Mesp2* in the posterior PSM where it binds. *Dll1* is known to be a downstream target of Tbx6 and putative binding sites have been identified in its mesoderm

(msd) enhancer (White and Chapman, 2005). ChIP assays using the putative *Dll1* msd enhancer (Fig. 2E, column Dll1) revealed that Tbx6 also binds to the *Dll1* enhancer in both the ap and pp regions, which is consistent with the expression pattern of *Dll1*. In all cases, the negative control (PCR amplification of an unrelated sequence in the mouse genome) gave no signal in ChIP assays with the anti-Tbx6 antibody (Fig. 2E,F, column NC). Thus, these results confirm our previous finding that Tbx6 binding is required for *Mesp2* expression, but is not sufficient for full transcriptional activation.

Mutated P2PSME contains Tbx6 binding sites that are inactive in vivo

We next applied the ChIP assay system to confirm that the phenotype of our enhancer-specific knockout mouse was due to the lack of Tbx6 binding in the *Mesp2* enhancer region. We performed ChIP assays using the tails of *P2EmB1D* heterozygous embryos and specific primer sets in order to distinguish the mutated DNA fragment from its wild-type counterpart, expecting that Tbx6 would not bind to the mutated enhancer. Surprisingly, mutated P2PSME, which has no PSM-specific transcriptional activity (Fig. 1), gave rise to a band that co-precipitated with the anti-Tbx6 antibody. This indicated that Tbx6 still binds to the mutated PSME in vivo (Fig. 2F). To identify the Tbx6 binding site within the mutated P2PSME, we re-examined this region for a consensus Tbx6 binding sequence (White and Chapman, 2005) and found two additional candidate sites, denoted Site F and Site G, in and upstream of P2PSME (Fig. 3A). EMSA demonstrated that Site G was strongly associated with Tbx6 in vitro (Fig. 3B).

The number and spatial organization of the T-box binding sites are important for initiating *Mesp2* transcription via Notch signaling

We reported previously that the simultaneous mutation of two Tbx6 binding sites, Site B and Site D, eliminates PSM-specific activation of a reporter gene by P2PSME in transgenic embryos (Yasuhiko et al., 2006). To confirm this finding and also investigate the possible involvement of the new Tbx6 binding site, Site G, in enhancer activity, we generated a series of reporter constructs with P2PSME harboring serial mutations in the Tbx6 binding sites. We tested two types of reporter assay: a luciferase assay using cultured cells, and transgenic analyses. In the luciferase assay, the loss of any single Tbx6 binding site among Sites B, D and G, caused a 10-fold reduction in Tbx6-dependent and Tbx6 plus Notch signaling-dependent reporter activation (Fig. 3C, right). Conversely, expression of a *lacZ* reporter in transgenic embryos was not markedly affected by the loss of any individual Tbx6 binding site (Fig. 3C, left). These results suggested that each Tbx6 binding site contributes equally to P2PSME activity, but that the loss of a single site is not sufficient to disrupt the in vivo function of P2PSME.

We next examined the effects of systematically removing multiple Tbx6 binding sites. Removal of two Tbx6 binding sites from P2PSME resulted in a further decrease in luciferase reporter activity (Fig. 3C, lane P2EmDG, and Fig. 3D). *lacZ* expression in transgenic embryos was also diminished, both in intensity and frequency. Out of nine transgene-positive embryos, only one showed weak *lacZ* expression with the P2EmB1 reporter, which has two intact Tbx6 binding sites (Fig. 3D, left). When three out of four Tbx6 binding sites were eliminated, the synergistic effects of Tbx6 and Notch signaling on P2PSME activation were no longer observed and mutants resembled P2EmB1DG, which has lost Tbx6 binding capability at all four sites (Fig. 3D, right). In transgenic embryos, *lacZ* expression was not activated by any single Tbx6 binding site (Fig. 3D, left, P2EmB1D). These results strongly suggest that the PSM-specific expression of *Mesp2* requires at least two Tbx6 binding sites in P2PSME. Notably, the P2PSME reporters with two intact Tbx6 binding sites (P2EmDG, P2EmB1, P2EmB2D, P2EmB2G) showed variable levels of activity in the luciferase assay. This finding contrasts with the uniform reporter activity found with either one or three mutated Tbx6 binding sites (Fig. 3C,D). P2EmDG, with two Tbx6 binding sites in Site B intact, displayed a more than 2-fold stronger activity than P2EmB2G, which harbors single Tbx6 binding sites within Site B and Site D. P2EmB2G activated the luciferase reporter at levels comparable to those of reporters with a single Tbx6 binding site and showed no synergistic activation when Notch signaling was applied (Fig. 3D). Taken together, these data indicate that the four Tbx6 binding sites have equal importance in regulating P2PSME activity, and at least two neighboring sites are required for the Notch signaling-dependent induction of *Mesp2* expression.

The medaka *mespb* PSM enhancer regulates *Mesp2* expression and normal somite formation in the mouse embryo

mespb, the zebrafish homolog of *Mesp2*, shows a similar expression pattern to mouse *Mesp2* during embryogenesis and we speculated that it might exert a similar function in the mouse (Nomura-Kitabayashi et al., 2002). We have previously identified the PSM-specific enhancer of medaka *mespb*, which contains T-box binding sites. Two of these sites, T1 and T2, are important for PSM-specific *mespb* expression (Terasaki et al., 2006) (Fig. 4A). These data suggest that the T-box-protein-dependent expression mechanism is evolutionally conserved between mammals and teleosts (zebrafish, medaka). We demonstrated that zebrafish Tbx24, a T-box protein that is homologous to mouse Tbx6 and is responsible for the *fused somite* (*fss*) mutant phenotype, binds to the medaka *mespb* PSME (Fig. 4B). A sequence comparison revealed three putative T-box binding sites in the medaka *mespb* PSME (Fig. 4A). Two of these had the ability to bind two Tbx24 molecules each, whereas in the mouse P2PSME, only Site B can bind two Tbx6 molecules (Fig. 4B).

To more directly demonstrate the evolutionary conservation of this regulatory mechanism, we generated a knock-in mouse with a medaka *mespb* upstream sequence inserted in place of the endogenous *Mesp2* PSME. For this purpose, we substituted the 356-bp sequence upstream of the *Mesp2* first ATG with 2.8 kb of sequence upstream of the *mespb* first ATG, generating a *medakaP2* mouse (Fig. 4C). Heterozygous mice (*medakaP2/+*) were viable and appeared normal (data not shown). Homozygous mice (*medakaP2/medakaP2*) were also viable and showed no physical malformations (Fig. 4D). In skeletal preparations, we observed that *medakaP2* homozygous fetuses were indistinguishable from heterozygous or wild-type littermates (Fig. 4E), indicating that the PSMEs of medaka *mespb* and mouse *Mesp2* are functionally equivalent, despite some differences in their structural features.

DISCUSSION

The activation of *Mesp2* expression requires at least two Tbx6 binding sites in P2PSME

In our current study, we have shown that Tbx6 binding sites are fundamentally important for P2PSME function and that P2PSME is necessary and sufficient for *Mesp2* expression during somite formation in mouse embryogenesis. However, ChIP assays revealed that Tbx6 binds to P2PSME not only in *Mesp2*-expressing cells, but also in non-expressing cells such as those in the tailbud and posterior

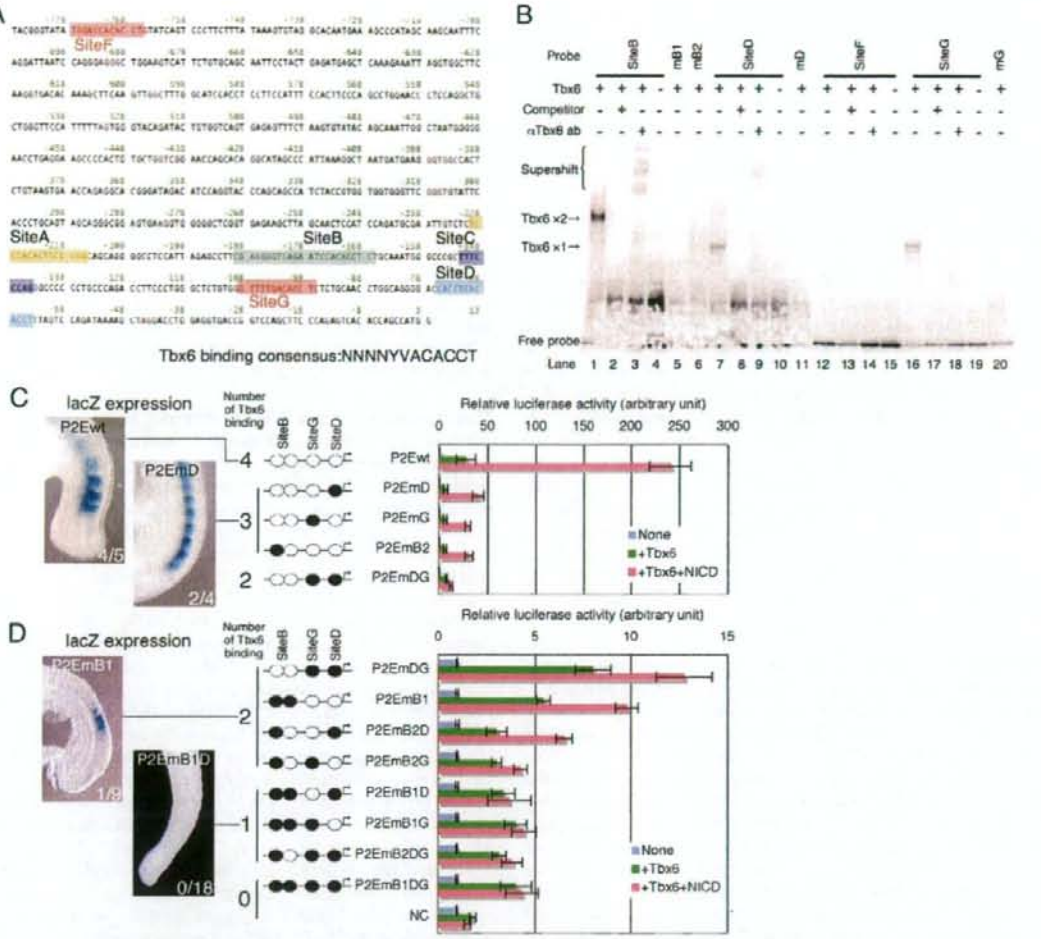

Fig. 3. Multiple Tbx6 binding is required for Mesp2 activation. (**A**) The sequence of the mouse *Mesp2* enhancer region that contains four presumptive Tbx6 binding sites (Sites B, D, F and G). Sites A and C are presumptive RBPJ-κ binding sites (Yasuhiko et al., 2006). The Tbx6 consensus binding site was originally reported by White and Chapman (White and Chapman, 2005). (**B**) Site G, but not Site F, binds to Tbx6 in an electromobility shift assay (EMSA). Site G produced a bandshift indicating a single bound Tbx6 molecule (lane 16), whereas Site B produced two bands (lane 1). Mutated oligonucleotide probes mB1, mD and mG produced no shifted bands (lanes 5, 11 and 20, respectively). The mB2 probe showed a single shifted band, implying the loss of one Tbx6 binding site in Site B (lane 6). (**C**,**D**) Luciferase reporter assays were conducted using several mutated enhancer elements. Luciferase activity was measured after transfection of reporter constructs along with an empty vector (None), Tbx6 expression vector (+Tbx6), or both Tbx6 and Notch intracellular domain expression vectors (+Tbx6 +NICD). Each reporter construct is presented schematically to the left of each graph. Black oval, mutated Tbx6 binding site; white oval, wild-type Tbx6 binding site; arrow, transcription start site. The number of wild-type Tbx6 binding sites is also indicated (number of Tbx6 binding: C, 4 to 2; D, 2 to 0). To the left are representative images of *lacZ* staining in transgenic embryos with P2PSME-*lacZ* reporters bearing the indicated enhancers. The number of *lacZ*-positive/transgene-positive embryos is indicated. The results of a consecutive series of reporter assays, as described in C, are shown in D, but on a different scale owing to the steep declines in activity. The P2EmDG lane represents the same data in both C and D. Each luciferase assay was performed in triplicate in at least three independent experiments. Error bars represent s.d.

PSM (Fig. 2E). This indicates that Tbx6 binding alone is not sufficient to activate *Mesp2* expression. Previously, we showed in vitro that *Mesp2* was activated weakly, if at all, by Tbx6 alone, but rigorously by a coexisting Notch signal (Yasuhiko et al., 2006). Taken together, these data suggest that it is highly likely that the restricted expression pattern of *Mesp2* in the anterior PSM is regulated by a combination of Tbx6 and Notch signaling in vivo. Similarly, Tbx6 activates *Dll1* together with Wnt signaling (Hofmann et al., 2004) and activates *Ripply* in cooperation with Mesp2 (Hitachi et al., 2008). Although the molecular mechanisms

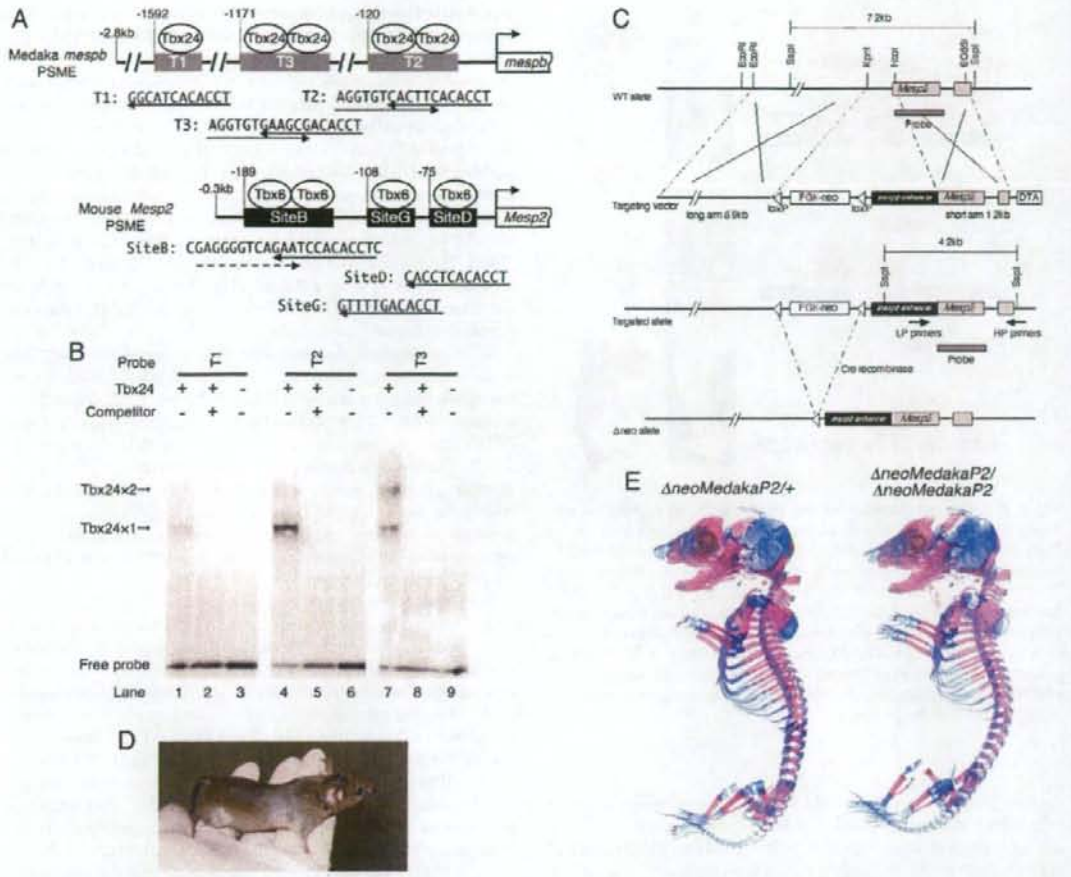

Fig. 4. The medaka *mespb* PSM enhancer is functionally equivalent to its counterpart in the mouse. (**A**) A comparison of the medaka *mespb* and mouse *Mesp2* PSME regions. Black and gray boxes represent presumptive T-box binding sites. The numbers above the boxes represent the nucleotide positions from the first ATG. The nucleotide sequences of the putative T-box binding sequences are shown beneath. Consensus Tbx6 binding sequences and their directions are indicated by arrows. The dashed arrow in Site B of the *Mesp2* PSME depicts an incomplete Tbx6 binding sequence that only binds to Tbx6 if an adjoining complete Tbx6 binding sequence is present. The T-box proteins that might bind to these sequences are indicated. (**B**) EMSA analysis of the T-box binding sites in the medaka *mespb* PSME. T-box binding site T1 associates with a single Tbx24 molecule and T2 and T3 with two Tbx24 molecules, which is consistent with their nucleotide sequences as shown in A. (**C**) The targeting strategy used to generate the medaka *mespb* PSME knock-in mouse (*medakaP2*). A 2.8-kb fragment of *mespb* genomic DNA that is required for PSM-specific *mespb* expression was substituted for *Mesp2* PSME by homologous recombination. The neoR selection marker was removed by recombination using the Cre-loxP system. (**D**) *medakaP2* homozygotes are viable and have normal external features. (**E**) Homozygotes are indistinguishable from heterozygotes and wild-type littermates in skeletal preparations.

by which Tbx6 regulates its target genes together with various partners remain elusive, it is possible that the number and spatial organization of Tbx6 binding sites facilitate the response of P2PSME to Notch signaling.

In total, there are four Tbx6 binding sequences in this region: two palindrome-like sequences in Site B and one each in Sites D and G. Importantly, the P2PSME reporters with only one intact Tbx6 binding site were inactive in both the luciferase assay and the transgenic analyses (Fig. 3D), suggesting that a single P2PSME-bound Tbx6 molecule might not act as a mediator of Notch signaling in the regulatory mechanism controlling *Mesp2* expression. This is consistent with the observation that Site G fails to activate *Mesp2* expression by itself in the *P2EmB1D* mouse.

The loss of two or more of the four Tbx6 binding sites greatly diminishes P2PSME activity in both luciferase and transgenic assays (Fig. 3D). Interestingly, the reporters with two intact Tbx6 binding sites showed varied levels of activity depending upon the position of the intact sites. Two intact Tbx6 binding sites in Site B resulted in the highest reporter activity (Fig. 3D). These data indicate that Site B may be of predominant importance in the function of P2PSME,

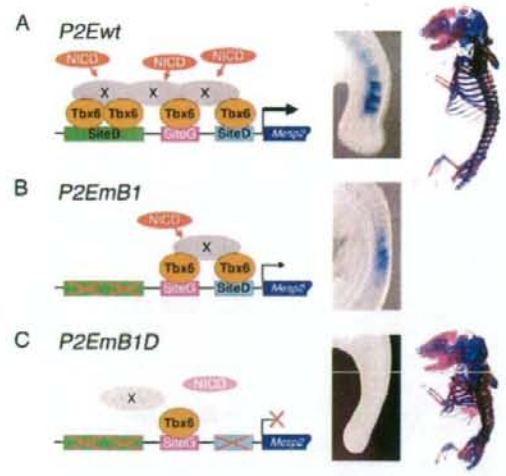

Fig. 5. Putative mechanism for the Tbx6-mediated regulation of Mesp genes. (**A**) Tbx6 activates *Mesp2* expression through multiple Tbx6 binding sites. Notch signaling (NICD, red ovals) activates *Mesp2* expression via factor X (gray ovals), which recognizes two neighboring Tbx6 binding sites in P2PSME. (Left) Schematic description of the Tbx6-dependent activation of *Mesp2*. (Middle column) *lacZ* expression in the P2Ewt transgenic embryo. (Right) Normal skeletal formation in the heterozygous fetus of a *P2EmB1D* mouse. (**B**) Mutation in Site B results in decreased expression of *Mesp2*. (**C**) A single Tbx6 binding site is unable to activate *Mesp2* expression, presumably owing to an inability to respond to Notch signaling.

implying that its two neighboring Tbx6 binding sites play a central role in regulating the activation of *Mesp2*. The binding of Tbx6 to one of the two binding sites in Site B depends on the presence of another Tbx6 molecule binding to this site (Yasuhiko et al., 2006). This property might be related to the unique palindrome-like sequence of this site. Although several T-box binding sites have been identified in the upstream region of other Tbx6-downstream genes, such as *Dll1* (Hofmann et al., 2004) and *Msgn1* (Wittler et al., 2007), the palindrome-like site has thus far been found only in the PSME of *Mesp2* and its medaka ortholog *mespb* (Fig. 4A). It is therefore possible that two neighboring Tbx6 molecules on the palindrome-like site are specifically recognized by as yet unidentified factor(s) (Fig. 5, 'X') that together with Tbx6 constitutes an RBPJ-κ (Rbpj)-independent Notch signaling machinery [disruption of potential RBPJ-κ binding sites does not affect P2PSME activity in transgenic embryos (Yasuhiko et al., 2006)]. Further analyses of *Mesp2* PSME might shed light on these novel regulatory mechanisms that operate during development.

Mutations that removed any one of the Tbx6 binding sites in P2PSME, regardless of which, diminished luciferase reporter activity by the same amount (Fig. 3C), suggesting that each Tbx6 binding site contributes equally to *Mesp2* expression in vitro. In vivo, by contrast, the mutation of a single Tbx6 binding site did not seem to affect PSM-specific gene expression (Fig. 3C). Taken together, these results indicate that the multiple Tbx6 binding sites confer a functional robustness to P2PSME that ensures proper *Mesp2* expression during embryogenesis.

An evolutionally conserved mechanism regulating Mesp expression through multiple T-box binding sites

We previously found that the deletion of two T-box binding sites in the *mespb* PSME greatly reduced its PSM-specific enhancer activity in transgenic medaka embryos (Terasaki et al., 2006), similar to our findings in transgenic mouse embryos. The medaka *mespb* PSME harbors three T-box binding sites (T1-T3), which is similar to the complement of the mouse *Mesp2* PSME (Fig. 4A). However, the total length of the PSME is very different between mouse *Mesp2* and medaka *mespb* (356 bp versus 2.8 kb, respectively) (Terasaki et al., 2006). The number of T-box proteins that bind to the medaka and mouse PSMEs is also different (Fig. 4A,B), and the distance between each element is greater in the *mespb* PSME than in its mouse counterpart.

We have demonstrated, however, that the medaka *mespb* PSME is functionally equivalent to the mouse *Mesp2* PSME. In our transgenic assay, a mutation in the double T-box binding site (Site B in mouse and Site T2 in medaka) had the most profound effect upon PSME activity. Consistent with these results, deletion of medaka Site T1 (harboring a single T-box binding sequence) did not affect reporter gene expression. However, deletion of one of the sites within the double T-box binding sequence (T2) caused a 50% decrease in reporter expression (Terasaki et al., 2006), again demonstrating the importance of the binding to the double T-box site for PSM enhancer function.

In the teleost fish, zebrafish, the T-box transcription factor Tbx24 was identified as responsible for the fused somite (*fss*) mutant phenotype. Tbx24 has a T-box domain that is homologous to that of mouse Tbx6 (Nikaido et al., 2002). The segmentation of somites and expression of *mespb* are eliminated in the *fss* mutant (Sawada et al., 2000), implying that *mespb* is a downstream target of Tbx24, similar to the relationship between *Mesp2* and Tbx6 in mice. However, *fss* mutant fish are viable and fertile (van Eeden et al., 1996), whereas *Tbx6*-null mouse embryos fail to form a mesoderm and die early in development (Chapman and Papaioannou, 1998). This difference might be due to the presence in zebrafish of a *Tbx6* counterpart gene, *spadetail*, which supports paraxial mesoderm formation. Despite this difference, our data clearly demonstrate that the mechanism regulating the PSM-specific expression of *Mesp2* and *mespb* is evolutionarily well conserved between fish and mice.

We thank Hiroyuki Takeda (University of Tokyo) for providing Tbx24 cDNA clones and Mariko Ikumi, Eriko Ikeno and Shunsuke Matsusaka for technical assistance. This work was supported by a grant-in-aid for scientific research from the Ministry of Education, Culture, Sports, Science and Technology, Japan, a grant for Research on Risk on Chemical Substances (H20-004) from the Ministry of Health, Labor and Welfare of Japan, and a grant from the NIG Cooperative Research Program (2007-B66).

References

Bussen, M., Petry, M., Schuster-Gossler, K., Leitges, M., Gossler, A. and Kispert, A. (2004). The T-box transcription factor Tbx18 maintains the separation of anterior and posterior somite compartments. *Genes Dev.* **18**, 1209-1221.

Chapman, D. L. and Papaioannou, V. E. (1998). Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. *Nature* **391**, 695-697.

Chapman, D. L., Agulnik, I., Hancock, S., Silver, L. M. and Papaioannou, V. E. (1996). Tbx6, a mouse T-Box gene implicated in paraxial mesoderm formation at gastrulation. *Dev. Biol.* **180**, 534-542.

Delfini, M. C., Dubrulle, J., Malapert, P., Chal, J. and Pourquie, O. (2005). Control of the segmentation process by graded MAPK/ERK activation in the chick embryo. *Proc. Natl. Acad. Sci. USA* **102**, 11343-11348.

Dunty, W. C., Jr, Biris, K. K., Chalamalasetty, R. B., Taketo, M. M., Lewandoski, M. and Yamaguchi, T. P. (2008). Wnt3a/β-catenin signaling controls posterior body development by coordinating mesoderm formation and segmentation. *Development* **135**, 85-94.

Galceran, J., Sustmann, C., Hsu, S. C., Folberth, S. and Grosschedl, R. (2004). LEF1-mediated regulation of Delta-like1 links Wnt and Notch signaling in somitogenesis. *Genes Dev.* **18**, 2718-2723.
Haraguchi, S., Kitajima, S., Takagi, A., Takeda, H., Inoue, T. and Saga, Y. (2001). Transcriptional regulation of Mesp1 and Mesp2 genes: differential usage of enhancers during development. *Mech. Dev.* **108**, 59-69.
Hitachi, K., Kondow, A., Danno, H., Inui, M., Uchiyama, H. and Asashima, M. (2008). Tbx6, Thylacine1, and E47 synergistically activate bowline expression in Xenopus somitogenesis. *Dev. Biol.* **313**, 816-828.
Hofmann, M., Schuster-Gossler, K., Watabe-Rudolph, M., Aulehla, A., Herrmann, B. G. and Gossler, A. (2004). WNT signaling, in synergy with T/TBX6, controls Notch signaling by regulating Dll1 expression in the presomitic mesoderm of mouse embryos. *Genes Dev.* **18**, 2712-2717.
Hogan, B., Beddington, R., Costantini, F. and Lacy, E. (1994). *Manipulating the Mouse Embryo: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
Moreno, T. A. and Kintner, C. (2004). Regulation of segmental patterning by retinoic acid signaling during Xenopus somitogenesis. *Dev. Cell* **6**, 205-218.
Morimoto, M., Takahashi, Y., Endo, M. and Saga, Y. (2005). The Mesp2 transcription factor establishes segmental borders by suppressing Notch activity. *Nature* **435**, 354-359.
Morimoto, M., Kiso, M., Sasaki, N. and Saga, Y. (2006). Cooperative Mesp activity is required for normal somitogenesis along the anterior-posterior axis. *Dev. Biol.* **300**, 687-698.
Nikaido, M., Kawakami, A., Sawada, A., Furutani-Seiki, M., Takeda, H. and Araki, K. (2002). Tbx24, encoding a T-box protein, is mutated in the zebrafish somite-segmentation mutant fused somites. *Nat. Genet.* **31**, 195-199.
Nomura-Kitabayashi, A., Takahashi, Y., Kitajima, S., Inoue, T., Takeda, H. and Saga, Y. (2002). Hypomorphic Mesp allele distinguishes establishment of rostrocaudal polarity and segment border formation in somitogenesis. *Development* **129**, 2473-2481.
Oginuma, M., Niwa, Y., Chapman, D. L. and Saga, Y. (2008). Mesp2 and Tbx6 cooperatively create periodic patterns coupled with the clock machinery during mouse somitogenesis. *Development* **35**, 2555-2562.
Rupp, R. A., Snider, L. and Weintraub, H. (1994). Xenopus embryos regulate the nuclear localization of XMyoD. *Genes Dev.* **8**, 1311-1323.
Saga, Y. and Takeda, H. (2001). The making of the somite: molecular events in vertebrate segmentation. *Nat. Rev. Genet.* **2**, 835-845.
Saga, Y., Yagi, T., Ikawa, Y., Sakakura, T. and Aizawa, S. (1992). Mice develop normally without tenascin. *Genes Dev.* **6**, 1821-1831.
Saga, Y., Hata, N., Koseki, H. and Taketo, M. M. (1997). Mesp2: a novel mouse gene expressed in the presegmented mesoderm and essential for segmentation initiation. *Genes Dev.* **11**, 1827-1839.
Sasaki, H. and Hogan, B. L. (1996). Enhancer analysis of the mouse HNF-3 beta gene: regulatory elements for node/notochord and floor plate are independent and consist of multiple sub-elements. *Genes Cells* **1**, 59-72.
Sawada, A., Fritz, A., Jiang, Y. J., Yamamoto, A., Yamasu, K., Kuroiwa, A., Saga, Y. and Takeda, H. (2000). Zebrafish Mesp family genes, mesp-a and mesp-b are segmentally expressed in the presomitic mesoderm, and Mesp-b confers the anterior identity to the developing somites. *Development* **127**, 1691-1702.
Takahashi, Y., Koizumi, K., Takagi, A., Kitajima, S., Inoue, T., Koseki, H. and Saga, Y. (2000). Mesp2 initiates somite segmentation through the Notch signalling pathway. *Nat. Genet.* **25**, 390-396.
Takahashi, Y., Inoue, T., Gossler, A. and Saga, Y. (2003). Feedback loops comprising Dll1, Dll3 and Mesp2, and differential involvement of Psen1 are essential for rostrocaudal patterning of somites. *Development* **130**, 4259-4268.
Takahashi, Y., Yasuhiko, Y., Kitajima, S., Kanno, J. and Saga, Y. (2007). Appropriate suppression of Notch signaling by Mesp factors is essential for stripe pattern formation leading to segment boundary formation. *Dev. Biol.* **304**, 593-603.
Terasaki, H., Murakami, R., Yasuhiko, Y., Shin, I. T., Kohara, Y., Saga, Y. and Takeda, H. (2006). Transgenic analysis of the medaka mesp-b enhancer in somitogenesis. *Dev. Growth Differ.* **48**, 153-168.
van Eeden, F. J., Granato, M., Schach, U., Brand, M., Furutani-Seiki, M., Haffter, P., Hammerschmidt, M., Heisenberg, C. P., Jiang, Y. J., Kane, D. A. et al. (1996). Mutations affecting somite formation and patterning in the zebrafish, Danio rerio. *Development* **123**, 153-164.
White, P. H. and Chapman, D. L. (2005). Dll1 is a downstream target of Tbx6 in the paraxial mesoderm. *Genesis* **42**, 193-202.
Wittler, L., Shin, E. H., Grote, P., Kispert, A., Beckers, A., Gossler, A., Werber, M. and Herrmann, B. G. (2007). Expression of Msgn1 in the presomitic mesoderm is controlled by synergism of WNT signalling and Tbx6. *EMBO Rep.* **8**, 784-789.
Yagi, T., Tokunaga, T., Furuta, Y., Nada, S., Yoshida, M., Tsukada, T., Saga, Y., Takeda, N., Ikawa, Y. and Aizawa, S. (1993). A novel ES cell line, TT2, with high germline-differentiating potency. *Anal. Biochem.* **214**, 70-76.
Yasuhiko, Y., Haraguchi, S., Kitajima, S., Takahashi, Y., Kanno, J. and Saga, Y. (2006). Tbx6-mediated Notch signaling controls somite-specific Mesp2 expression. *Proc. Natl. Acad. Sci. USA* **103**, 3651-3656.

ELSEVIER

Experimental Hematology

Experimental Hematology 2009;37:195–205

Comparison of murine gene expression profiles between spontaneous and radiation-induced myelogenous leukemias: Stochastic and probabilistic expression variances in the former vs radiation-specific expression commonalities in the latter

Yoko Hirabayashi[a], Isao Tsuboi[a,b], Kunio Kitada[c], Katsuhide Igarashi[a], Yukio Kodama[a], Jun Kanno[a], Kazuko Yoshida[d], Nicholas Dainiak[e], and Tohru Inoue[a]

[a]*Division of Cellular and Molecular Toxicology, National Center for Biological Safety and Research, National Institute of Health Sciences, Tokyo, Japan;* [b]*Department of Functional Morphology, Nihon University School of Medicine, Tokyo, Japan;* [c]*Kamakura Research Laboratories, Chugai Pharmaceutical Co. Ltd., Kamakura, Japan;* [d]*Research Center for Radiation Emergency Medicine, National Institute of Radiological Sciences, Chiba, Japan;* [e]*Department of Medicine, Bridgeport Hospital, Yale University School of Medicine, Bridgeport, Conn., USA*

(Received 25 July 2008; revised 29 September 2008; accepted 6 October 2008)

Objective. **To elucidate the common characteristics of murine radiation-induced myelogenous leukemias, global gene-chip expression profiles were compared with age-matched steady-state bone marrow tissue profiles and spontaneous myelogenous leukemia profiles.**

Materials and Methods. **Six each of C3H/He mice-derived radiation-induced and spontaneously developed myelogenous leukemias were analyzed. Bone marrow cells from five each of 2- and 21-month-old mice were used to subtract nonleukemic information in the analysis. mRNAs from individual mice were analyzed separately using 45,101 gene chips followed by computational biological analysis.**

Results. **First, principal component analysis (PCA) was performed to discriminate the gene expression profiles of individual mice with radiation-induced myelogenous leukemia from those of bone marrow cells from 2- or 21-month-old mice. Discriminant union genes for individual leukemias were then selected, which finally yielded 242 genes, among which six are radiation-related genes including *Hus-1*, *Edf1a2*, and *Vegf-c*; 16 are apoptosis/cell-death – related genes, 13 are cell-cycle/cell-growth – related genes, and 50 are suppressor/promoter genes. PCA of these 242 genes consistently enabled the discrimination of the radiation-induced leukemias from the spontaneous leukemias. Second, the other components of the same PCA provided four different eigenvector clusters in an unsupervised manner representing four histopathological findings, with which the differential diagnosis in molecular taxonomy was significant as determined by analysis of variance of the global gene expression profiles.**

Conclusion. **Discriminant union genes in radiation-induced myelogenous leukemias against spontaneous myelogenous leukemias and age-matched nonleukemic bone marrow profilings generated by unsupervised computational analysis essentially represent probabilistic biomarkers for radiation-induced myelogenous leukemias, which may contribute to developing a model for risk of secondary carcinogenesis in patients treated by whole-body irradiation.**

Currently, extensive DNA microchip analyses are utilized to elucidate the xenobiotic effects of chemical compounds and to determine toxicological and pathological characteristic gene expression profiles. In this study, the original aim is to elucidate the gene expression profiles of radiation-induced myelogenous leukemias using DNA microchips [1]. Several studies have elucidated radiation-specific gene expression profiles; however, no clear radiation-specific profile was obtained. This may be due to induction of unique expression profiles by radiation exposure, compared to profiles resulting from chemical exposure; namely, radiation exposure does not provide a homologous specific adduct formation, but rather provides stochastic and probabilistic modifications of gene expression.

Offprint requests to: Tohru Inoue, M.D., National Center for Biological Safety and Research, National Institute of Health Sciences, 1-18-1 Kamiyohga, Setagayaku, Tokyo 158-8501, Japan; E-mail: tohru@nihs.go.jp

0301-472X/09 $–see front matter.
doi: 10.1016/j.exphem.2008.10.006

C3H/He mice are prone to develop myelogenous leukemias with up to 50% probability after a single graded increased dose of whole-body irradiation from 1 to 5 Gy; unirradiated mice can also spontaneously develop myelogenous leukemias, but with only ~1% probability [2,3]. The former and the latter myelogenous leukemias are histologically undefined. Although spontaneous myelogenous leukemias show a high degree of stochasticity of gene expression, radiation-induced myelogenous leukemias show, homogenous characteristics. Homogeneity may be due to the presence of a limited number of radiation-induced fragile sites that may initiate radiation damage. On the other hand, stochastic characteristics may be the result of random leukemogenic processes that are induced by epigenetic process after radiation damage [4,5]. To determine whether the gene expression profile of radiation-induced myelogenous leukemias is based on any biomarker genes related to radiation activation of fragile sites, if any, and, whether that of spontaneous myelogenous leukemias is based on differences in gene expression, the gene expression profiles of spontaneous and radiation-induced myelogenous leukemias were analytically compared [6–9].

It seems difficult to obtain possible discriminant profiles specific to radiation-induced myelogenous leukemias. When one attempts to compare sample gene expressions from radiation-induced myelogenous leukemia profiles with those from spontaneous myelogenous leukemia profiles using a dendrogram, those myelogenous leukemia profiles of two different origins cannot be identified readily. This may be due to similarities in the expression patterns that characterize the pathological diagnosis. When one compares the average expression profiles with that from animals of radiation-induced myelogenous leukemias, from a mathematical operations required for consensus percentages, it makes generalization of Procrustes analysis unavoidable; thus, no radiation-specific characteristic profiles can be identified [6,10,11].

Use of senescence-linked gene expression profiles to subtract and discriminate between the profiles of spontaneous and radiation-induced myelogenous leukemias was also considered. Namely, the development of not only spontaneous leukemias, but also radiation-induced myelogenous leukemia is prolonged over the life-span of the mouse [12]. Thus, the gene expression profiles of both types of leukemia were analyzed, relative to profiles of senescent animals. To determine whether senescence-linked expression profiling is also associated with myelogenous leukemias, the profiles of both types of myelogenous leukemia were compared with those of steady-state bone marrow cells in senescent mice. Senescence-related genes are simultaneously expressed and observed in the gene expression profiles of both spontaneous and radiation-induced myelogenous leukemias.

Radiation-specific discriminant profiles of myelogenous leukemias were obtained by maintaining inhomogeneity throughout the computation. We identified 242 relevant genes that composed a discriminant gene repertoires.

Materials and methods

Mice

C3H/He male mice purchased from Japan SLC (Hamamatsu, Japan) or C3H/HeNirsMs male mice were maintained in the board-approved laboratory animal facility of the National Institute of Health Sciences (NIHS), Japan or the National Institute of Radiological Sciences, Japan. All experimental protocols involving the laboratory mice used in this study were reviewed by the Interdisciplinary Monitoring Committee for the Proper Animal Use and Welfare of Experimental Animals, a peer review panel established at NIHS, and approved by the Committee for Animal Care and Use of the NIHS with the experimental code #9639415-2007. All animal studies were conducted using humane protocols approved by the Committee for Animal Care and Use of the NIHS, Japan.

Spontaneous myelogenous leukemias

C3H/He mice are prone to develop myelogenous leukemias but generate spontaneous leukemias with only 1% probability [2,3], the histological types of which range from myelogenous leukemias with marked differentiation to blast-type leukemias with little differentiation. Mice with spontaneous myelogenous leukemias were obtained from 5-week-old C3H/He male mice purchased from Japan SLC at 188 mice per week for 3 weeks. All the mice were monitored throughout their lifetime, except for those showing symptoms of advanced hematopoietic neoplasms, such as anemia and palpable splenomegaly, which were euthanized at the agonal period and then examined hematopathologically. Pieces of leukemic spleens were stored in an RNA stabilization reagent (RNAlater; QIAGEN GmbH, Hilden, Germany) until RNA extraction. The remaining mice were also examined histopathologically.

Radiation-induced myelogenous leukemias

The method used here to induce myelogenous leukemias by irradiation is described elsewhere [2]. Briefly, 10-week-old C3H/HeNirsMs male mice were exposed to 3 Gy of whole-body x-ray irradiation at a 200-kV/20-mA pulse using a therapeutic x-ray irradiator (Simadzu, Kyoto, Japan) with 0.5-mm Al and 0.5-mm Cu filters at a dose rate of 0.614 Gy/minute and a 56-cm focus surface distance. Mice were monitored at least twice daily until they showed symptoms of advanced hematopoietic neoplasms. They were then euthanized at the agonal period for hematopathological examination. Pieces of leukemic spleens as well as tubes of cell suspension in 10% dimethyl sulfoxide with 10% fetal calf serum from radiation-induced leukemias were similarly stored with or without RNA stabilization reagent, as described in the case of spontaneous myelogenous leukemia cases. When analyzing the gene expression profiles of leukemia cases, frozen leukemic cells were thawed and injected into C3H/HeNirsMs mice.

Histopathological and cytological examination

For the evaluation of spontaneous hematopoietic malignancies, mice were sacrificed under ethyl ether anesthesia for autopsy. Whenever possible, cytological examinations for spleen and bone marrow tissues were examined. For histopathological examination, all the visceral organs including the thymus, spleen, sternum, and femoral bone marrow were fixed in 4% neutral buffered formalin for 24 hours. The sternum and femoral bone marrow were decalcified in 7.5% formic acid for 72 hours. After routine processing, paraffin-embedded sections were stained

with hematoxylin and eosin and then examined histopathologically under a light microscope [13,14].

Bone marrow cell collection for RNA extraction

Two-month-old C3H/He mice or 21-month-old C3H/HeNirsMs mice from which bone marrow cells were collected for RNA extraction were carefully chosen by evaluating peripheral blood cell number and bone marrow cellularity using a blood cell counter (Sysmex M-4500; Sysmex Co., Kobe, Japan) and by comparing the results with those reported previously [15,16]. All the 21-month-old mice used here were examined histopathologically and confirmed to have neither tumors nor life-threatening fatal diseases.

Bone marrow cells were harvested from the femur of each mouse after the animals were sacrificed by cervical dislocation under deep anesthesia with ethyl ether. A 26-gauge needle was inserted into the femoral bone cavity through the proximal and distal ends of the bone shafts, and bone marrow cells were flushed out under pressure by injecting 2 mL α−minimum essential medium with ribonucleosides and deoxyribonucleosides (Invitrogen, Carlsbad, CA, USA). After resuspending, the bone marrow cell pellet in guanidine isothiocyanate−containing lysis buffer (Buffer RLT) (QIAGEN GmbH) with 2-mercaptoethanol (Sigma, St Louis, MO, USA), the bone marrow cells were immediately frozen in liquid nitrogen and stored in a −80°C deep freezer until RNA extraction.

Analysis of mRNA expression level by microarray

Total RNA was extracted from bone marrow cells collected from each individual mouse or leukemic cells and applied on a GeneChip Mouse Genome 430 2.0 Array (Affymetrix, Santa Clara, CA, USA), as described elsewhere [17] and briefly here as follows.

Total RNA preparation. Total RNA was extracted from bone marrow cells collected from each individual mouse or leukemic cells using columns of an RNeasy kit (QIAGEN GmbH) in accordance with the manufacturer's instructions. The extracted total RNA yielded OD ratios ($OD_{260/280}$) of 1.7 to 2.1; its purity and concentration were confirmed by capillary electrophoresis using an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA, USA).

Target preparation from total mRNA. First-strand cDNA was synthesized by incubating 40 μg total RNA with SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) and a T7-$(dT)_{24}$ primer. Second-strand cDNA was synthesized by incubating the first-strand cDNA with DNA polymerase (Invitrogen), and residual RNA was digested using RNase H (Invitrogen). The purified products of double-strand cDNA were used to synthesize biotin-labeled cRNAs by in vitro transcription using a BioArray High-Yield RNA transcript labeling kit (Enzo Diagnostics, Farmingdale, NY, USA). The labeled cRNA was then purified using an RNeasy minikit (QIAGEN GmbH) in accordance with the manufacturer's instruction. The purified cRNA was then fragmented in 1× fragmentation buffer (40 mM Tris-acetate, 100 mM KOAc, and 30 mM MgOAc) at 94°C for 35 minutes.

Hybridization, washing, staining, and antibody amplification. For hybridization to the GeneChip Mouse Genome 430 2.0 Array (Affymetrix), the fragmented cRNA (15 μg) was diluted with control oligonucleotide B2, Eukaryotic Hybridization Controls (Affymetrix) and injected into a prehybridized GeneChip cartridge. The GeneChip cartridge was incubated in a 45°C rotisserie oven (Hybridization Oven 640; Affymetrix) for 16 hours. After hybridization, the cartridge was washed and stained with a GeneChip Fluidic Station (Affymetrix) using appropriate antibody amplification washing and staining protocols.

Probe array scanning. The phycoerythrin-stained array was scanned using GeneChip Scanner 7G (Affymetrix) as a digital image file and analyzed using Microarray Suite version 5.0 software (Affymetrix). All the procedures such as experimental design, array design, sampling, hybridization, signal measurements, and normalization control were in accordance with the MIAME guidelines [18].

Analysis of mRNA expression level by quantitative real-time polymerase chain reaction

Total RNA extracted from each leukemic mouse was also used for quantitative reverse transcriptase polymerase chain reaction (RT-PCR) assay as described elsewhere [19] and briefly here as follows. The isolated mRNA was reverse-transcribed using Superscript (Life Technologies, Grand Island, NY, USA) and Oligo-dT (Promega Corp, Madison, WI, USA). Next, triplicates of the transcribed cDNA per sample were analyzed by quantitative RT-PCR using TaqMan Universal PCR master mix (Applied Biosystems, Foster City, CA, USA) and specific primers and probes using the Applied Biosystems 7900 Sequence Detection System. The specific primers and probes for each of the six spontaneous and six radiation-induced leukemia samples, and ribosomal RNA (rRNA) genes selected on the basis of their expression intensity were purchased from Applied Biosystems (TaqMan Gene Expression Assays; *Sh2d1a*, Mm00448717_m1; *Trib3*, Mm00454879_m1; *Mmp9*, Mm00600163_m1; *Gml/Hemt1*, Mm00468837_m1; *Evi1*, Mm00514814_m1; *Sox4*, Mm00486320_s1; *Gngt1*, Mm00802677_m1; *Lect1*, Mm00495289_g1; *Sgip1*, Mm00513890_m1; *Igll1*, Mm01716143_m1; *Tmem178*, Mm00509628_m1; *Mov10l1*, Mm00473179_m1; rRNA, P/N 4308329). PCR conditions and data analysis were in accordance with the instructions provided with the Sequence Detection System, version 2.0. All the reactions were performed in triplicate. In accordance with manufacturer's instruction, the signals examined were normalized with rRNA signals using the formula $2^{-\delta Ct} = 2^{-(Ct\ rRNA\ -\ Ct\ Signal)}$.

Statistical analysis

Computational data processing. Line configuration analysis, dendrographic analysis, principal component analysis (PCA) with contributing scoring and Pearson correlation calculation for each gene, and gene ontology analysis with likelihood ratio calculation were performed using GeneSpring GX 7.3.1 (Agilent Technologies Inc.). PCA was also applied using SPSS 14.1 (SPSS Inc., Chicago, IL, USA). Student's *t*-test and fold-change calculation were carried out using Microsoft Office Excel 2003 (Microsoft, Redmond, WA, USA). Differences were considered significant at $p < 0.05$.

Data normalization and statistical analysis. The obtained microarray data were normalized and analyzed using GeneSpring GX 7.3.1 (Agilent Technologies Inc.), SPSS 14.1 (SPSS Inc.) and Microsoft Office Excel 2003 (Microsoft). In this normalization, all the expression intensities measured were converted into Euclidian distance, which is based on the 50th percentile of 1.0. Namely,

for normalization, the following steps were applied to the data: 1) data transformations: values <0.01 were set at 0.01; 2) per chip: each measurement was divided by the 50th percentile of all measurements in that sample; and 3) per gene: the measurement of each gene was divided by the median of its measurements in all samples. When the median of the raw values was <10, each measurement for that gene was divided by 10 when the numerator was >10; otherwise, the measurement was disregarded.

Generation of union gene list. To generate a gene list that can discriminate between radiation-induced and spontaneous myelogenous leukemias, maintenance of the stochasticity of expression characteristics was attempted during PCA. The gene expression profiles of the age-matched group were compared separately with each individual expression profile of radiation-induced myelogenous leukemias by PCA followed by the selection of genes with high contribution scores from certain PCA components. A union gene list can be generated from each PCA combination corresponding to individual radiation-induced myelogenous leukemias. Other union gene lists are also incorporated, for example, those obtained by the combination of the profiles of spontaneous myelogenous leukemias and each individual expression profile of radiation-induced myelogenous leukemias.

Results

Linear configurations of spontaneous and radiation-induced myelogenous leukemias

Figure 1 shows linear configurations of each expressed gene that have been connected for >12 cases of leukemias, including six cases per group from the leftmost end, nos. 01 to 06, and another six cases, from number 07 to the rightmost end, number 12. The former six cases on the left are from nonirradiated spontaneous myelogenous leukemias, and the other six cases on the right are from radiation-induced leukemias that developed after 3-Gy x-ray exposure. When one focuses on a case, number 12 for example, and designates genes in the order of expression intensity from red (the highest) to green (the lowest), then the order from red genes to green genes in each case, as observed in number 12, is essentially comparable to those on the six individual cases on the right (no. 07 to no. 12), namely, the orders of gene expression intensities of the six cases on the right seem to be comparable among the groups (i.e., upregulated genes in the case of number 12 designated by red are nearly identical to those in the cases from nos. 07 to 11.) However, the order of the gene expression profiles in the case of spontaneous myelogenous leukemias shown from number 01 to 06 is the opposite across the cases as compared with those in number 12; namely, red genes, which are highly expressed in number 12, tend to appear in the low-expression zone in the case of spontaneous myelogenous leukemias, from number 01 to 06. Furthermore, the red linear configurations, i.e., genes that are highly expressed in the case with radiation-induced myelogenous leukemias, are largely located below 1.0 for the spontaneous myelogenous leukemia cases. The green linear configurations, i.e., relatively low-expression level genes associated with radiation-induced leukemias, are reversed and show a high-expression intensity.

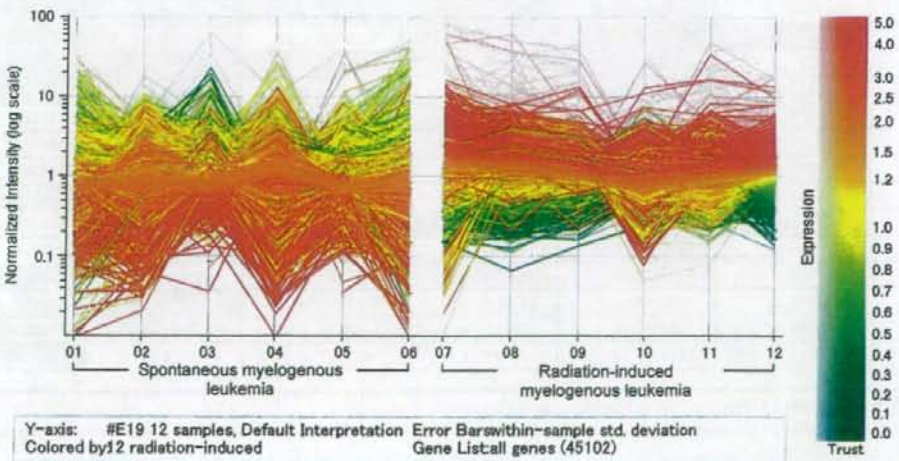

Figure 1. Linear configurations of spontaneous and radiation—induced myelogenous leukemias: microarray profiling differentiates radiation—induced myelogenous leukemias from spontaneous leukemias. The six individual data points on the left panel are from spontaneously developed myelogenous leukemias in C3H/He mice. The other six individual cases on the right panel are from 3—Gy—radiation—induced myelogenous leukemias in C3H/He mice [2,3]. Along the gene expression intensity scale from the highest (red) to the lowest (green) of number 12 case, the same gene is connected and indicated by the same color. Accordingly, genes overexpressed in the radiation—induced myelogenous leukemia groups are largely repressed in the spontaneous myelogenous leukemia groups. See text.

Table 1. Top 20 genes in terms of expression intensity

A. Spontaneous myelogenous leukemia

ID[a]	1	2	3	4	5	6
1	1449393_at	1433438_x_at	1448291_at	1422400_a_at	1438325_at	1425502_x_at
2	1425500_x_at	1456182_x_at	1449366_at	1425145_at	1425583_at	1449393_at
3	1449835_at	1428864_at	1416298_at	1436155_at	1450545_a_at	1426215_at
4	1440156_s_at	1426065_a_at	1442339_at	1427357_at	1433438_x_at	1419156_at
5	1436598_at	1423580_at	1419764_at	1428816_a_at	1449757_x_at	1452565_x_at
6	1433924_at	1441054_at	1425451_s_at	1431751_a_at	1456182_x_at	1433438_x_at
7	1422340_a_at	1456022_at	1456182_x_at	1417837_at	1419764_at	1425452_s_at
8	1437542_at	1456456_x_at	1450009_at	1435285_at	1419538_at	1419157_at
9	1425502_x_at	1425288_at	1452732_at	1430182_a_at	1455626_at	1424367_a_at
10	1421930_at	1438833_at	1435761_at[b]	1428359_s_at	1425584_x_at	1422632_at
11	1452565_x_at	1419411_at	1436598_at	1426243_at	1420805_at	1424547_at
12	1421931_at	1421277_at	1451537_at	1419271_at	1433939_at	1449903_at
13	1427656_at	1419421_at	1422953_at	1451596_a_at	1452855_at	1427608_a_at
14	1426208_x_at	1421009_at	1419334_at	1420712_a_at	1427608_a_at	1426168_a_at
15	1442544_at	1438189_s_at	1449393_at	1434137_x_at	1421628_at	1423608_at
16	1419684_at	1452740_at	1439902_at	1428221_at	1452565_x_at	1455279_at
17	1456956_at	1434553_at	1419709_at[b]	1428209_at	1417155_at	1433471_at
18	1452127_a_at	1421008_at	1422190_at	1450333_a_at	1425451_s_at	1456182_x_at
19	1449888_at	1447655_x_at	1421930_at	1426887_at	1454867_at	1440782_at
20	1452539_a_at	1431194_at	1425324_x_at	1456527_at	1425471_x_at	1443937_at

B. Radiation-induced myelogenous leukemia

ID[a]	7	8	9	10	11	12
1	1451633_a_at	1460258_at	1423327_at	1452730_at	1429175_at	1451633_a_at
2	1429175_at	1451633_a_at	1451633_a_at	1420176_x_at	1420176_x_at	1419340_at
3	1460258_at	1425219_x_at	1429175_at	1423915_at	1449388_at	1427170_at
4	1425219_x_at	1429175_at	1460258_at	1453203_at	1449254_at	1449388_at
5	1425167_a_at	1425167_a_at	1425180_at	1448470_at	1449869_at	1453203_at
6	1425180_at	1428685_at	1425219_x_at	1421681_at	1425180_at	1435761_at
7	1417649_at	1437060_at	1425167_a_at	1449388_at	1419340_at	1421681_at
8	1425295_at	1425180_at	1425181_at	1421594_a_at	1453203_at	1429175_at
9	1416271_at	1425295_at	1425295_at	1457123_at	1425181_at	1425219_x_at
10	1438452_at	1441259_s_at	1417673_at	1417074_at	1421681_at	1425167_a_at
11	1427239_at	1427239_at	1441259_s_at	1451679_at	1451633_a_at	1417745_at
12	1435761_at[b]	1425181_at	1449922_at	1441258_at	1456777_at	1416713_at
13	1441259_s_at	1418796_at	1419340_at	1434848_at	1448696_at	1422837_at
14	1421681_at	1449388_at	1452388_at	1449074_at	1453357_at	1425295_at
15	1425181_at	1437932_a_at	1449388_at	1450513_at	1441258_at	1418796_at
16	1416776_at	1419321_at	1449831_at	1448949_at	1419321_at	1429348_at
17	1419709_at[b]	1427170_at	1427239_at	1427277_at	1425843_at	1460258_at
18	1455901_at	1452986_at	1435887_at	1419748_at	1425219_x_at	1453357_at
19	1449965_at	1423582_at	1429512_at	1448573_a_at	1457123_at	1449254_at
20	1437060_at	1418395_at	1448990_a_at	1437467_at	1455771_at	1457123_at

Each column indicates gene names designated as Affymetrix systematic name.
The color of each column indicates appearance frequency of each gene among 6 cases each of spontaneous and radiation-induced myelogenous leukemias.

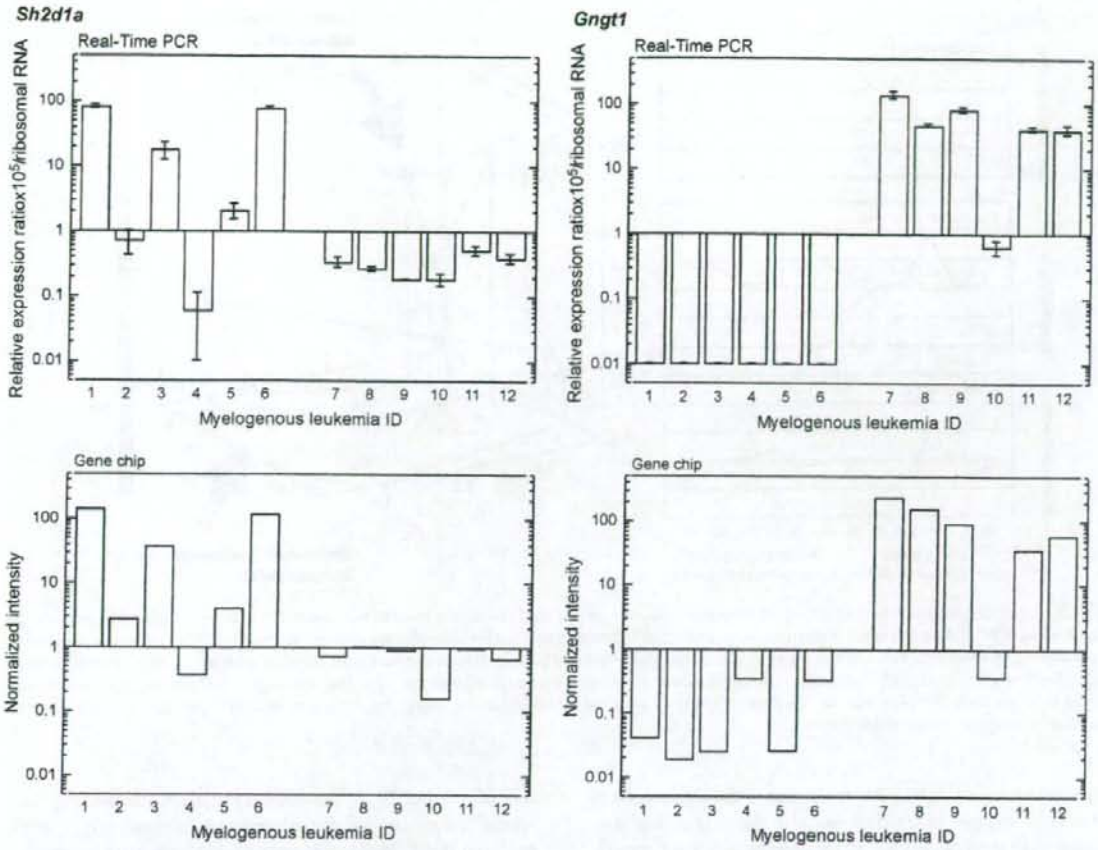

Figure 2. Comparison between the expression ratio of the gene relative to that of ribosomal RNA (top) and relative expression of the gene in the gene chip analysis (bottom) of two genes; *Sh2d1a* and *Gngt1*. The left shows data for *Sh2d1a*, which showed the highest expression intensity in the spontaneous myelogenous leukemia number 1. The right shows those for *Gngt1*, which showed the highest expression intensity in the radiation—induced myelogenous leukemia no. 7 (and no. 12). Similar comparisons between reverse transcriptase polymerase chain reaction (RT—PCR) data and normalized expression obtained by gene chip analysis for the other genes with the highest expression intensity in each of the spontaneous and radiation—induced myelogenous leukemias are shown in Supplementary Figures 1—1 and 1—2, respectively (because of overlap genes, the second highest genes were added). Expression intensity of each gene in each case of myelogenous leukemias (nos. 1 to 12) is shown by a pair of graphs, RT—PCR (top) and gene chip analysis (bottom). The ordinates of the RT—PCR data (top, left and right) indicate relative expression ratio multiplied by 10^5 divided by ribosomal RNA expression, and the ordinates of the gene chip data indicate normalized gene expression intensities in the gene chip analyses.

Comparison of gene expression intensities between genes from mice with spontaneous and radiation-induced myelogenous leukemias

The gene expression profiles of radiation-induced leukemias were relatively homogeneous compared with those of spontaneous leukemias, as shown by the linear configuration (Fig. 1). Thus, because genes with high-expression intensities in each mouse with radiation-induced leukemia were considered to show similar repertoires with similar high-expression intensities, the top 20 genes with the highest expression intensity in each mouse with leukemia were focused on and selected (Table 1A and B; spontaneous and radiation-induced myelogenous leukemias, respectively; also, see Suppl. Tables 1-1 and 1-2 for detail). In Table 1A and B, note that the 20 most expressed genes differ across samples. Of the 120 genes (6 samples × 20 genes) shown in Table 1, 24 genes overlapped 77 times (64%) in two or more samples in the radiation-induced myelogenous leukemia cohort, whereas only 10 genes overlapped 25 times (i.e., as low as 21%) in up to two samples in the six mice with spontaneous myelogenous leukemia cohort. From Tables 1A and B, genes with the highest expression intensity in each of the six mice with spontaneous and six mice with radiation-induced myelogenous leukemias were separately selected and subjected to quantitative RT-PCR analysis (in the case that overlapping genes show the highest expression intensities in other leukemias, genes with the second or third highest expression intensities were

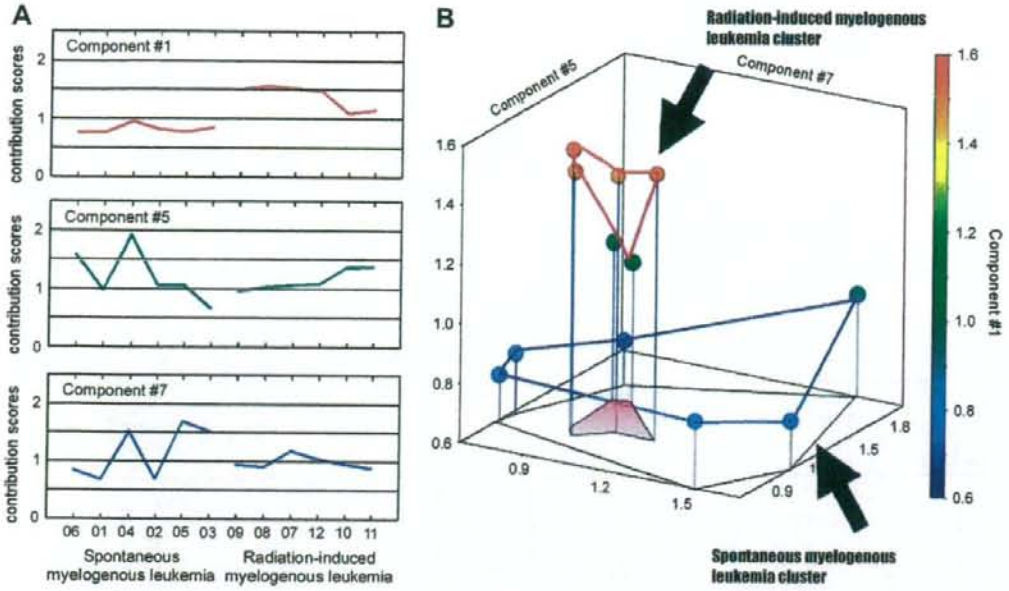

Figure 3. Principal component analysis (PCA) differentiates radiation—induced leukemias from spontaneous leukemias: three—dimensional expression clustering. Results of PCA of six cases each of spontaneous and radiation—induced myelogenous leukemias are shown in the three—dimensional contribution scores for component numbers 1, 5, and 7, which discriminate the radiation—induced myelogenous leukemia cluster from the spontaneous myelogenous leukemia cluster. The line graph on the left shows actual contribution scores converted from each eigenvector value (**A**), which were used for the three—dimensional expression on the right (**B**). Note that the contribution scores of the spontaneous leukemias, except for component number 1, are comparatively divergent from those of radiation—induced leukemias.

subjected to the analysis). Figure 2 shows the expression of two sets of sample genes. For *Sh2d1a*, the expression ratio relative to that of ribosomal RNA determined from spontaneous leukemia number 1 (top left; out of six cases, nos. 1 through 6) is compared with the relative expression intensity of the same gene in the gene chip in the spontaneous myelogenous leukemias (six cases in the bottom left, number 1 through 6). For *Gngt1*, the expression ratio relative to that of ribosomal RNA determined from the radiation-induced myelogenous leukemia number 7 (top right; out of six cases, nos. 7 through 12) is compared with the relative expression intensity of the same gene in the gene chip in the radiation-induced myelogenous leukemias (six cases in the bottom right, nos. 7 through 12).

Sh2d1a expression level was evaluated in not only spontaneous leukemias but also radiation-induced leukemias, and *Gngt1* expression level was evaluated in not only radiation-induced leukemias, but also spontaneous leukemias. Results show that the *Sh2d1a* expression level highly vary in the spontaneous group, whereas it is relatively low, reversed, and homogenous in the radiation-induced group (nos. 7–12); these trends are comparable between the quantitative RT-PCR data (top) and the relative expression intensities in the gene chip analysis (bottom). On the other hand, as shown on the right panels of Figure 2 on the right, the *Gngt1* expression levels are generally high and homogenous, except for that in number 10, in the case of radiation-induced myelogenous leukemias, on the other hand, interestingly, the expression levels in spontaneous myelogenous leukemias are nearly very low. These trends are similar not only between the quantitative RT-PCR data and the relative expression intensities in the gene chip analyses, but also in the expression intensities of the other genes except *Sgip1*, as shown in the radiation-induced myelogenous leukemias examined (see Suppl. Fig. 1). The 24 genes that overlapped in the radiation-induced myelogenous leukemias are shown in Supplementary Table 1-2, which included *G proteins*, *thrombospondin 4*, and *stem cell growth factor*. None of the 24 genes associated with radiation-induced myelogenous leukemias overlapped with the 10 genes associated with the spontaneous myelogenous leukemia (Suppl. Table 1-1). To confirm the nominal discriminant power of these genes associated with the spontaneous and the radiation-induced myelogenous leukemias, a reference dendrogram based on these 240 genes in total in both leukemias (actually, 170 genes in total after subtraction of the overlapping genes) is shown in Supplementary Figure 2. Based on gene expression intensity, despite the clear discrimination between both myelogenous leukemias, as shown in Supplementary Figure 2, none of the highly expressed genes show any mechanistic discrimination between the gene expressions in the myelogenous leukemia cases.

Table 2. Gene list obtained according to the contribution score of PCA component #1 up to 0.999

Affymetrix systematic name	GenBank ID	Description
1439612_at	AW060892	calcium channel, voltage-dependent, N type, alpha 1B subunit (Cacna1b)
1449476_at	NM_011973	renal tumor antigen (Rage)
1450492_at	NM_013927	cyclic nucleotide gated channel beta 3 (Cngb3)
1422990_at	NM_008591	met proto-oncogene (Met)
1422931_at	NM_008037	fos-like antigen 2 (Fosl2)
1426175_a_at	U42405	mast cell protease 7 (Mcpt7)
1419463_at	AF108501	chloride channel calcium activated 2 (Clca1)
1440134_at	AI647584	Transcribed sequences (Cyp4a10)
1442092_at	BG063268	Fanconi anemia, complementation group D2 (Fancd2)
1427581_at	BI737649	hypermethylated in cancer 2 (Hic2)
1422116_at	NM_008032	fragile X mental retardation 2 homolog (Fmr2)

Spontaneous and radiation-induced myelogenous leukemias are differentiated by PCA

Evaluations by PCA of both myelogenous leukemias are largely defined by component number 1 (Fig. 3A) and further discriminated by the combination of the additional component numbers 5 and 7, as shown in Figure 3B. The entire PCA data of 12 cases are shown in Supplementary Table 2-1. When one analyzes the discriminant components numbers 1, 5, and 7, the radiation-induced myelogenous leukemias are found in one particular cluster in the three-dimensional representation; on the other hand, the spontaneous myelogenous leukemias are separated from the cluster of radiation-induced myelogenous leukemias, but are scattered between the axes of numbers 5 and 7, and the lower axis of number 1.

Whether responsible discriminant genes associated with the two myelogenous leukemia groups form a possible gene cluster that may be relevant to radiation-induced myelogenous leukemogenesis is of interest. Among the genes that contribute to PCA component number 1, 11 genes with contribution scores of >0.99 based on the PCA component number 1 are shown in Table 2. These genes include met proto-oncogene (*Met*), fos-like antigen2 (*Fosl2*), Fanconi anemia, complementation group D2 (*Fancd2*), and fragile X mental retardation 2 homolog (*Fmr2*).

Generation of union gene profiles of each individual gene expression associated with radiation-induced myelogenous leukemias while maintaining inhomogeneity of each gene expression profile

Regardless of the difference between spontaneous and radiation-induced myelogenous leukemias, it takes nearly a lifetime for both myelogenous leukemias to become fatal; i.e., the former ranges from 465 to 636 days, and the latter ranges, in general, from 300 to 800 days, depending on radiation dose. Thus, the profile of each type of myelogenous leukemia may be associated with age-related expression profiles. The extent by which such age-related gene expression profiles overlap with the profiles of myelogenous leukemias is of particular interest, because knowledge of this may help in elucidating more specific gene expressions associated with spontaneous myelogenous leukemia or gene expressions associated with radiation-induced myelogenous leukemias. In Supplementary Figure 3, the linear configurations of genes from the bone marrow of 2-month-old untreated mice (a), 21-month-old -untreated mice (b), mice with spontaneous myelogenous leukemias (c), and mice with radiation-induced myelogenous leukemias (d) are shown from left to right. Strongly expressed genes in untreated 2-month-old mice are expressed less strongly in the mice with radiation-induced leukemias. On the other hand, genes that are expressed less strongly in the steady state in 2-month-old mice are relatively expressed more strongly in mice with radiation-induced leukemias.

To exclude possible age-related factors and to maintain the stochastic and probabilistic expression characteristics of each individual case (rather than an averaged grouping analysis), the profiles of each case of radiation-induced myelogenous leukemia were compared with those of age-related groups as well as with those of spontaneous myelogenous leukemias (data not shown for spontaneous myelogenous leukemias). The procedure for generating union genes is described in the Materials and Methods section. Namely, the gene expression profiles of the 2-month-old and 21-month-old groups were compared separately with each individual expression profile of radiation-induced myelogenous leukemias by PCA, followed by the selection of genes with a contribution score of >1.0 from component numbers 2 or 3 and 6, respectively, based on the results of PCA. Forty-five genes that overlapped with another 249 union genes from the PCA combination between the gene expression profiles of the 21-month-old bone marrow group and each individual expression profile of the spontaneous myelogenous leukemia group were subtracted from the obtained 287 union genes. The union gene list of discriminant gene profiles associated with radiation-induced myelogenous leukemias was then obtained (Suppl. Table 3), which consequently generated a total of 242 genes associated with radiation-induced myelogenous leukemias and included six