a single spot at an isoelectric point (p*I*) of 5.0 (Fig. 2B, left panel), slightly more acidic than actin (Fig. 2B, right panel). NCR-G3 cells expressed exactly the same Raft.2-positive proteins as F9 cells (Fig. 2C). The other two anti-SSEA Mabs, MC-813-70 and MC-631, bound none of the SDS-denatured proteins on a nitrocellulose sheet (data not shown).

In order to identify the spot bound by Mab Raft.2, we carried out MALDI-TOF MS/MS analysis of this spot in a 2-D gel (Fig. 2B). Eleven possible peptide signals with masses (*m/z*) of 912.6, 1200, 1203.6, 1291.7, 1419.8, 1615, 1698, 1740.9, 2082.1, 2172.2, and 2298.2, respectively, were detected (Fig. 3, underlined) and matched to the fragments of 40S ribosomal protein SA, known as 34/67 laminin receptor or LBP. The MS/MS spectra of five peptides (*) were obtained and the fragments were matched to the sequence from Lys_{42} to Arg_{128} of LBP (Table 1).

In order to confirm that LBP has a sialylated carbohydrate chain, we examined whether [^{14}C]galactose is incorporated into LBP. Membrane proteins prepared from F9 cells, which were metabolically labeled with [^{14}C]galactose, were separated by 2-D PAGE and subjected to autoradiography. The LBP spot was detected on the autoradiogram obtained by BAS 2000 (Fig. 4A). Next, we examined the effect of sialidase digestion on Raft.2 binding. Two nitrocellulose sheets, on which membrane proteins of NCR-G3 cells had been separated by 2-D PAGE followed by transfer, were stained with Ponceau 3R Stain (Fig. 4B). Small rectangles containing LBP were excised from the nitrocellulose sheets. The excised blot was incubated in 50 mM sodium acetate buffer, pH 4.5, with or without *C. perfringens* sialidase to release terminal sialic acids from glycoconjugates. This treatment reduced the Raft.2 binding to LBP (Fig. 4C), confirming that LBP contains sialic acid.

Next, we fractionated the organelles of F9 cells by sucrose density gradient centrifugation to identify the

Table 1
Mass fragments fitted for murine LBP by Mascot search

m/z	Start	End	Peptide sequence
912.6*	121	128	LLVVTDPR
1200	54	63	TWEKLLLAAR
1203.6*	90	102	FAAATGATPIAGR
1291.7	43	53	SDGIYIINLKR
1419.8	42	53	KSDGIYIINLKR
1615*	86	102	AVLKFAAATGATPIAGR
1698	103	117	FTPGTFTNQIQAAFR
1740.9*	64	80	AIVAIENPADVSVISSR
2082.1*	103	120	FTPGTFTNQIQAAFREPR
2172.2	81	102	NTGORAVLKFAAATGATPIAGR

The five fragments (*) were subjected to MS/MS analysis.

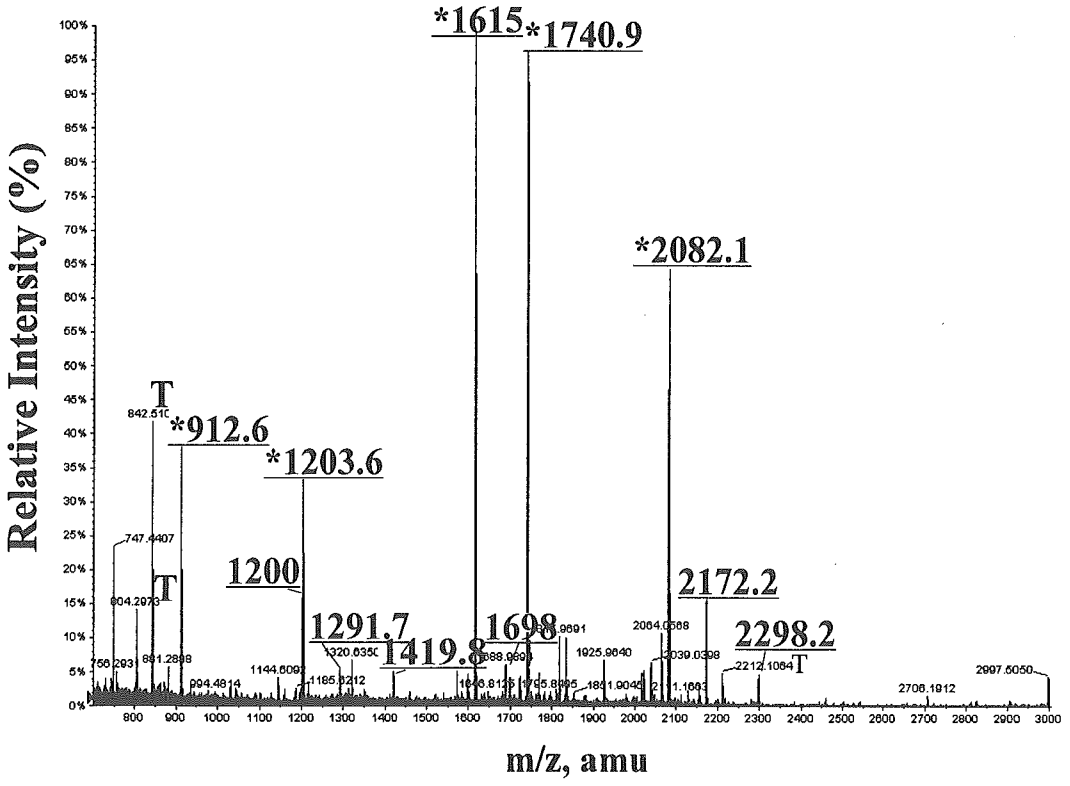


Fig. 3. MALDI-TOF MS spectra of the mw 44K molecule recognized by Raft.2. The Raft.2-binding molecule indicated by the arrow in Fig. 2C was in-gel digested with trypsin and subjected to oMALDI-Qq-TOF MS/MS QSTAR Pulser *i* spectrometry. The masses of the underlined peaks matched those of LBP. The peaks marked "T" were derived from trypsin.

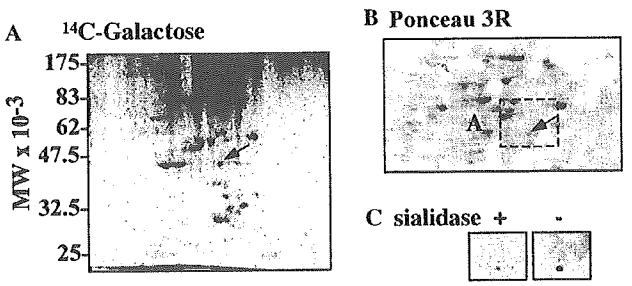


Fig. 4. Detection of galactose and sialic acid residues in LBP. (A) Autoradiogram of membrane proteins prepared from [^{14}C]galactose-labeled F9 cells and separated by 2-D PAGE. (B) The membrane proteins separated by 2-D PAGE were transferred to a nitrocellulose sheet and stained with Ponceau 3R. "A" indicates actin and the arrows point to LBP. (C) The blot containing LBP was incubated with (+) and without (−) sialidase and probed with Raft.2.

organelles in which LBP was localized. Triton X-100 lysates of each layer were 2-D separated and stained with CBB (Fig. 5). Since Src kinase Yes was found only in layer 1 (data not shown), the plasma membrane appears to be recovered from layer 1. Nuclear fragments were precipitated to the bottom. LBP was found only in layer 1, indicating that LBP is localized in plasma membranes.

fsConfocal microscopic observation confirmed Raft.2 binding on the cell surface. Cell suspensions of F9 and NCR-G3 were stained with Raft.2 and observed with a confocal microscope. Since Raft.1 recognizes Gβ located within cells [10], no F9 cells were stained with Raft.1 (Fig. 6A). As shown in Fig. 6B, some cells showed uneven, dot-like staining with Raft.2, while others did not. This indicates that some, not all, F9 cells express Raft.2 binding molecules on their cell surfaces. NCR-G3 cells were evenly stained with Raft.2 (Fig. 6C).

Discussion

We previously established Mab Raft.2, which specifically binds to sialylGb5 on TLC and cell surfaces. SialylGb5 carries the SSEA-4 epitope, which is naturally found in glycolipids. Interestingly, although mouse EC cell line F9 cells do not express sialylGb5, they are stained with Raft.2. Western analysis and the subsequent mass spectrometric analysis revealed the Raft.2 binding molecule to be LBP. The human EC cell line NCR-G3, which synthesizes sialylGb5, was also found to express Raft.2-binding LBP. Weak acid treatment or sialidase digestion reduced Raft.2 binding to LBP and [^{14}C]galactose was incorporated into LBP. These results confirm that LBP carries the SSEA-4 epitope. The precise carbohydrate structure of LBP must be obtained to provide further confirmation. Mass spectrometric analysis of the sugar moiety of LBP is currently underway, employing microLC-ESI MS/MS.

Raft.2 is different from another anti-SSEA-4 Mab, MC-813-70 in reactivity with SSEA-4 epitope. Raft.2 did not react with GM1b purified from mouse spleen [11] (data not shown), but MP-813-70 does [5]. These two Mabs appear to be different from each other in reactivity with glycoproteins.

Many developmentally regulated antigens, including SSEAs, ABH, Forssman, globoside, and Ii, are carbohydrates in nature [12]. Such antigenic determinants are carried by lipids and/or by protein molecules. It was previously reported that SSEA-1, i.e., Lewis x antigen (Le^x), is carried by both proteins [13] and lipids [4] and that SSEA-3 is carried by both membrane glycolipids and glycoproteins [1]. By contrast, it is generally accepted that the SSEA-4 is carried only by globoseries glycosphingolipids [5]. Herein, however, we present

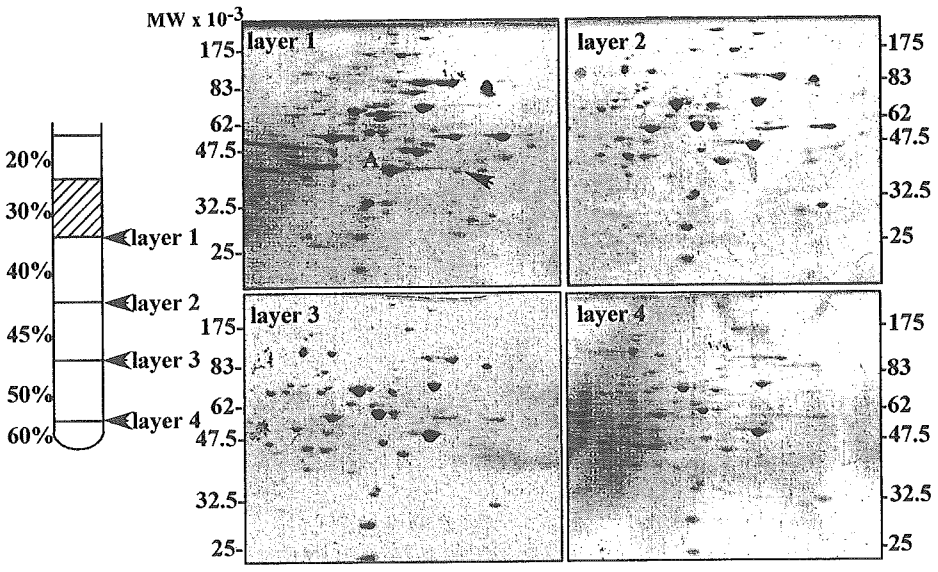


Fig. 5. Subcellular localization of LBP. The membrane proteins recovered from layers 1, 2, 3, and 4 after sucrose density-gradient centrifugation were analyzed by 2-D PAGE. The 2-D gels were stained with CBB. The arrow in the layer 1 gel points to LBP.

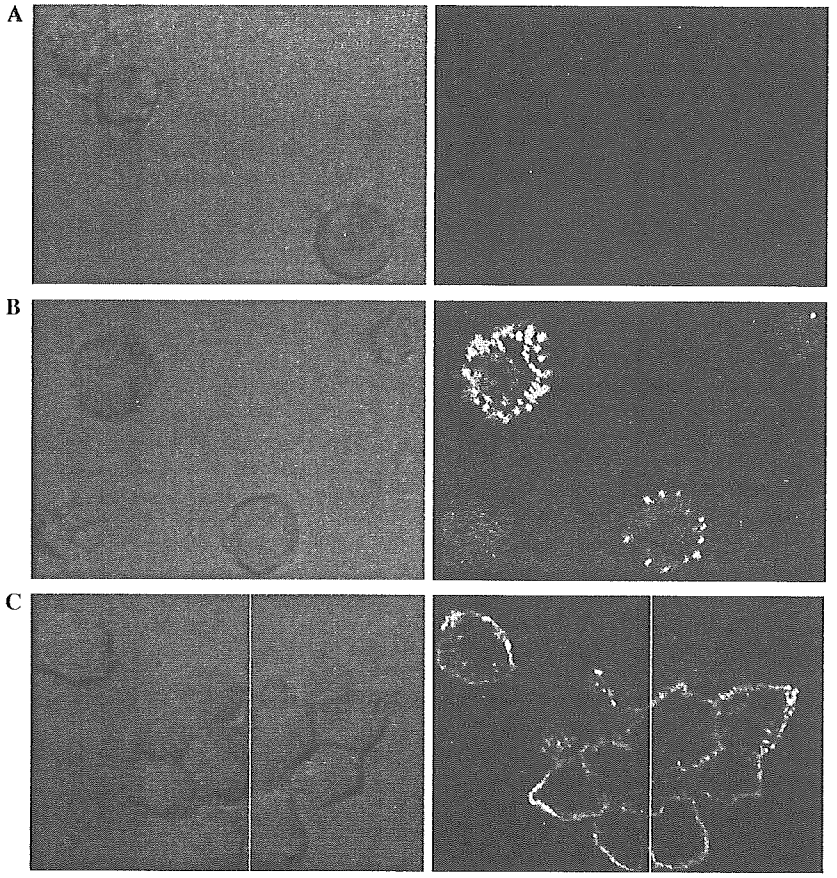
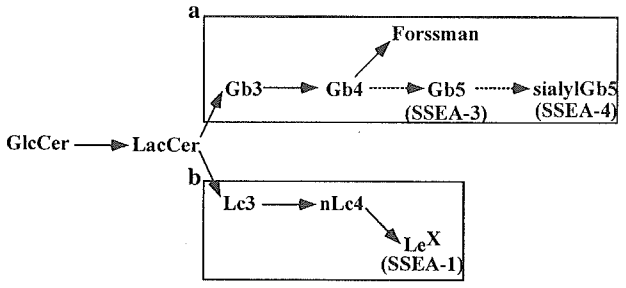


Fig. 6. Immunostaining of mouse and human EC cells with SSEA-4. Phase-contrast micrograph (left) and confocal micrograph (right). (A) Isotype-matched negative control Mab (Raft.1)/mouse EC F9—there is no staining. (B) Raft.2/F9—dot-like surface-staining. (C) Raft.2/human EC NCR-G3—even surface-staining.



GlcCer: Glcβ1→Cer
LacCer: Galβ1,4Glcβ1→Cer
Gb3: Galα1,4Galβ1,4Glcβ1→Cer
Gb4: GalNAcβ1,3Galα1,4Galβ1,4Glcβ1→Cer
Forssman: GalNAcα1,3GalNAcβ1,3Galα1,4Galβ1,4Glcβ1→Cer
Gb5: Galβ1,3GalNAcβ1,3Galα1,4Galβ1,4Glcβ1→Cer
sialylGb5: NeuAcα2,3Galβ1,3GalNAcβ1,3Galα1,4Galβ1,4Glcβ1→Cer
Lc3: GlcNAcβ1,3Galβ1,4Glcβ1→Cer
nLc4: Galβ1,4GlcNAcβ1,3Galβ1,4Glcβ1→Cer
Le^X: Galβ1,4(Fucα1,3)GlcNAcβ1,3Galβ1,4Glcβ1→Cer

Fig. 7. Scheme summarizing the major glycosylation pathways in mouse EC F9 cells. The globoseries glycolipid synthesis pathway (a) and the neolactoseries glycolipid synthesis pathway (b) lead to the synthesis of SSEA-3 and -4 and SSEA-1 active glycolipids, respectively. The cells exhibit a distinct pattern of globo- and neolacto-series oligosaccharide chain elongation (solid arrows). Retinoic acid treatment converts the synthesis of Forssman to SSEA-3 and -4 (dotted arrows).

evidence that the SSEA-4 epitope is also carried by a protein, 34/67 laminin receptor, also called LBP.

In culture, the differentiation of murine EC or ES cells into endoderm-like cells is typically characterized by the loss of SSEA-1 expression and the appearance of SSEA-3 and SSEA-4 [14], and may be accompanied by up-regulation of laminin synthesis [15]. Laminins are a family of extracellular matrix proteins that constitute the major non-collagenous glycoproteins found in the basement membrane and are involved in multiple important biological activities, such as assembly of the basement membrane, cell attachment, migration, neurite outgrowth, and angiogenesis [16,17]. In addition, laminins have well demonstrated roles in diverse developmental processes, from the pre-implantation period onwards [18]. Laminin receptors are divided into two major groups: integrin and non-integrin receptors. One of the non-integrin receptors is the mw 67K laminin receptor, LBP [19]. A highly conserved multifunctional mw 37K laminin receptor protein is the precursor of the 67K laminin receptor but the exact manner by which it forms a mature laminin receptor is not clear [20]. Endo and co-workers demonstrated that the glycans of α-dystroglycan include *O*-mannosyl oligosaccharides,

and that a sialyl *O*-mannosyl glycan, Siaα 2,3Galβ1,4GlcNAcβ1,2Man, of α-dystroglycan is a laminin-binding ligand of α-dystroglycan [21,22]. Furthermore, α-dystroglycan from sheep brain has an SSEA-1 structure O-linked to Man [23]. We found that another non-integrin laminin receptor, LBP of human and mouse EC cells, has the SSEA-4 epitope. Since laminin is essential to autocrine- or paracrine-signaling throughout mammalian development and differentiation, the up-regulation of laminin production might have some relationship with SSEA-4 expression on the laminin receptor. However, Raft.2 had no significant effects on F9 cell adhesion to a coverslip precoated with the murine Engelbreth–Holm–Swam tumor-derived laminin (data not shown).

The F9 cell exhibits a distinct pattern of globo- and neolacto-series oligosaccharide chain elongation (Fig. 7). In the globoseries pathway, globotetraosylceramide (Gb4) is not elongated to globopentaosylceramide (Gb5) by the addition of Gal, but rather to the Forssman antigen by the addition of GalNAc [24]. The β1,3-galactosyltransferase-V (β3GalT-V) can catalyze the transfer of Gal not only to GlcNAc-based acceptors with a preference for the core3 O-linked glycan GlcNAc(β1,3)-GalNAc structure, but also to the terminal GalNAc unit of Gb4, thereby leading to the synthesis of Gb5 [25]. These investigators further confirmed that Gb5 synthesis, or SSEA-3 expression in F9 cells is due to β3GalT-V. SSEA-3 synthase, i.e., β3GalT-V, can be said to catalyze Gal addition to both the acceptor of glycolipid and protein. SSEA-4 synthase, which transfers sialic acid to Gb5, was recently cloned from a human renal cancer cell line, ACHN, and the cloned cDNA was found to have a sequence identical to previously cloned α2,3-sialyltransferase (ST3Gal II) [26]. It catalyzes the transfer of sialic acid to the Galβ1,3GalNAc epitope of Gb5 in addition to asialo-GM1 and GM1a [27]. Whether ST3Gal II acts on a proteinous acceptor, such as LBP, is unknown. These transferases, which can glycosylate LBP, should be characterized to elucidate the role of the sugar moiety in differentiation.

Acknowledgments

We thank Ms. S. Yamauchi for her excellent secretarial work, and Drs. Susumu Watanabe and Tomomi Kayamori of Hitachi Science Systems, for mass spectrometric analysis of the Raft.2-binding protein. This work was supported in part by MEXT. KAKENHI 16017321, a grant from the Japan Health Sciences Foundation for Research on Health Sciences Focusing on Drug Innovation KH21014 and also by the CREST grant from the Japan Science and Technology.

References

[1] L.H. Shevinsky, B.B. Knowles, I. Damjanov, D. Solter, Monoclonal antibody to murine embryos defines a stage-specific embryonic antigen expressed on mouse embryos and human teratocarcinoma cells, Cell 30 (1982) 697–705.

[2] N.W. Fox, I. Damjanov, B.B. Knowles, D. Solter, Stage-specific embryonic antigen 3 as a marker of visceral extraembryonic endoderm, Dev. Biol. 103 (1984) 263–266.

[3] P.W. Andrews, Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro, Dev. Biol. 103 (1984) 285–293.

[4] R. Kannagi, E. Nudelman, S.B. Levery, S. Hakomori, A series of human erythrocyte glycosphingolipids reacting to the monoclonal antibody directed to a developmentally regulated antigen SSEA-1, J. Biol. Chem. 257 (1982) 14865–14874.

[5] R. Kannagi, N.A. Cochran, F. Ishigami, S. Hakomori, P.W. Andrews, B.B. Knowles, D. Solter, Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells, EMBO J. 2 (1983) 2355–23561.

[6] Y.U. Katagiri, K. Ohmi, C. Katagiri, T. Sekino, H. Nakajima, T. Ebata, N. Kiyokawa, J. Fujimoto, Prominent immunogenicity of monosialosyl galactosylgloboside, carrying a stage-specific embryonic antigen-4 (SSEA-4) epitope in the ACHN human renal tubular cell line-a simple method for producing monoclonal antibodies against detergent-insoluble microdomains/raft, Glycoconj. J. 18 (2001) 347–353.

[7] T. Maruyama, A. Umezawa, S. Kusakari, H. Kikuchi, M. Nozaki, J. Hata, Heat shock induces differentiation of human embryonal carcinoma cells into trophectoderm lineages, Exp. Cell Res. 224 (1996) 123–127.

[8] K. Nakamura, M. Suzuki, C. Taya, F. Inagaki, T. Yamakawa, A. Suzuki, A sialidase-susceptible ganglioside, IV3 alpha(NeuGc alpha 2-8NeuGc)-Gg4Cer, is a major disialoganglioside in WHT/Ht mouse thymoma and thymocytes, J. Biochem. 110 (1991) 832–841.

[9] H. Nakajima, Y.U. Katagiri, N. Kiyokawa, T. Taguchi, T. Suzuki, T. Sekine, K. Mimori, H. Nakao, T. Takeda, J. Fujimoto, Single-step method for purification of Shiga toxin-1 B subunit using receptor-mediated affinity chromatography by globotriaosylceramide-conjugated octyl sepharose CL-4B, Protein Exp. Purif. 22 (2001) 267–275.

[10] Y.U. Katagiri, K. Ohmi, W. Tang, H. Takenouchi, T. Taguchi, N. Kiyokawa, J. Fujimoto, Raft.1, a monoclonal antibody raised against the raft microdomain, recognizes G-protein β1 and 2, which assemble near nucleus after Shiga toxin binding to human renal cell line, Lab. Invest. 82 (2002) 1735–1745.

[11] K. Nakamura, Y. Hashimoto, M. Suzuki, A. Suzuki, T. Yamakawa, Characterization of GM1b in mouse spleen, J. Biochem. 96 (1984) 949–957.

[12] R. Kannagi, S.B. Levery, F. Ishigami, S. Hakomori, L.H. Shevinsky, B.B. Knowles, D. Solter, New globoseries glycosphingolipids in human teratocarcinoma reactive with the monoclonal antibody directed to a developmentally regulated antigen, stage-specific embryonic antigen 3, J. Biol. Chem. 258 (1983) 8934–8942.

[13] M. Ozawa, T. Muramatsu, D. Solter, SSEA-1, a stage-specific embryonic antigen of the mouse, is carried by the glycoprotein-bound large carbohydrate in embryonal carcinoma cells, Cell Differ. 16 (1985) 169–173.

[14] J.K. Henderson, J.S. Draper, H.S. Baillie, S. Fishel, J.A. Thomson, H. Moore, P.W. Andrews, Preimplantation human embryos and embryonic stem cells show comparable expression of stage-specific embryonic antigens, Stem Cells 20 (2002) 329–337.

[15] A. van de Stolpe, M. Karperien, C.W. Lowik, H. Juppner, G.V. Segre, A.B. Abou-Samra, S.W. de Laat, L.H. Defize, Parathyroid hormone-related peptide as an endogenous inducer of parietal endoderm differentiation, J. Cell Biol. 120 (1993) 235–243.
[16] K.M. Malinda, K.H. Kleinman, The laminins, Int. J. Biochem. Cell Biol. 28 (1996) 957–959.
[17] K.M. Malinda, M. Nomizu, M. Chung, M. Delgado, Y. Kuratomi, Y. Yamada, H.K. Kleinman, M.L. Ponce, Identification of laminin alpha1 and beta1 chain peptides active for endothelial cell adhesion, tube formation, and aortic sprouting, FASEB J. 13 (1999) 53–62.
[18] J.H. Miner, P.D. Yurchenco, Laminin functions in tissue morphogenesis, Annu. Rev. Cell Dev. Biol. 20 (2004) 255–284.
[19] E.E. Simon, J.A. McDonald, Extracellular matrix receptors in the kidney cortex, Am. J. Physiol. 259 (1990) F783–F792.
[20] S. Buto, E. Tagliabue, E. Ardini, A. Magnifico, C. Ghirelli, F. van den Brule, V. Castronovo, M.L. Colnaghi, M.E. Sobel, S. Menard, Formation of the 67-kDa laminin receptor by acylation of the precursor, J. Cell. Biochem. 69 (1998) 244–251.
[21] A. Chiba, K. Matsumura, H. Yamada, T. Inazu, T. Shimizu, S. Kusunoki, I. Kanazawa, A. Kobata, T. Endo, Structures of sialylated O-linked oligosaccharides of bovine peripheral nerve alpha-dystroglycan. The role of a novel O-mannosyl-type oligosaccharide in the binding of alpha-dystroglycan with laminin, J. Biol. Chem. 272 (1997) 2156–2162.
[22] T. Sasaki, H. Yamada, K. Matsumura, T. Shimizu, A. Kobata, T. Endo, Detection of O-mannosyl glycans in rabbit skeletal muscle alpha-dystroglycan, Biochim. Biophys. Acta 1425 (1998) 599–606.
[23] N.R. Smalheiser, S.M. Haslam, M. Sutton-Smith, H.R. Morris, H.R.A. Dell, Structural analysis of sequences O-linked to mannose reveals a novel Lewis X structure in cranin (dystroglycan) purified from sheep brain, J. Biol. Chem. 273 (1998) 23698–23703.
[24] J.G. Krupnick, I. Damjanov, A. Damjanov, Z.M. Zhu, B.A. Fenderson, Globo-series carbohydrate antigens are expressed in different forms on human and murine teratocarcinoma-derived cells, Int. J. Cancer 59 (1994) 692–698.
[25] D. Zhou, T.R. Henion, F.B. Jungalwala, E.G. Berger, T. Hennet, The beta 1,3-galactosyltransferase beta 3GalT-V is a stage-specific embryonic antigen-3 (SSEA-3) synthase, J. Biol. Chem. 275 (2000) 22631–22634.
[26] S. Saito, H. Aoki, A. Ito, S. Ueno, T. Wada, K. Mitsuzuka, M. Satoh, Y. Arai, T. Miyagi, Human alpha2,3-sialyltransferase (ST3Gal II) is a stage-specific embryonic antigen-4 synthase, J. Biol. Chem. 278 (2003) 26474–26479.
[27] Y.J. Kim, K.S. Kim, S.H. Kim, C.H. Kim, J.H. Ko, I.S. Choe, S. Tsuji, Y.C. Lee, Molecular cloning and expression of human Gal beta 1,3GalNAc alpha 2,3-sialytransferase (hST3Gal II), Biochem. Biophys. Res. Commun. 228 (1996) 324–327.

Human Molecular Genetics, 2005, Vol. 14, No. 22 *3379–3388*
doi:10.1093/hmg/ddi369
Advance Access published on October 3, 2005

Detecting tissue-specific alternative splicing and disease-associated aberrant splicing of the *PTCH* gene with exon junction microarrays

Kazuaki Nagao[1], Naoyuki Togawa[2], Katsunori Fujii[3], Hideki Uchikawa[1,3], Yoichi Kohno[3], Masao Yamada[1] and Toshiyuki Miyashita[1,*]

[1]Department of Genetics, National Research Institute for Child Health and Development, Tokyo 157-8535, Japan, [2]Yokohama Research Laboratories, Mitsubishi Rayon Co., Ltd, Yokohama 230-0053, Japan and [3]Department of Pediatrics, Graduate School of Medicine, Chiba University, Chiba 260-8670, Japan

Received June 8, 2005; Revised and Accepted September 23, 2005 GenBank accession numbers[†]

Mutations in the human ortholog of *Drosophila patched* (*PTCH*) have been identified in patients with autosomal dominant nevoid basal cell carcinoma syndrome (NBCCS), characterized by minor developmental anomalies and an increased incidence of cancers such as medulloblastoma and basal cell carcinoma. We identified many isoforms of *PTCH* mRNA involving exons 1–5, exon 10 and a novel exon, 12b, generated by alternative splicing (AS), most of which have not been deposited in GenBank nor discussed earlier. To monitor splicing events of the *PTCH* gene, we designed oligonucleotide arrays on which exon probes and exon–exon junction probes as well as a couple of intron probes for the *PTCH* gene were placed in duplicate. Probe intensities were normalized on the basis of the total expression of *PTCH* and probe sensitivity. Tissue-specific regulation of AS identified with the microarrays closely correlated with the results obtained by RT–PCR. Of note, the novel exon, exon 12b, was specifically expressed in the brain and heart, especially in the cerebellum. Additionally, using these microarrays, we were able to detect disease-associated aberrant splicings of the *PTCH* gene in two patients with NBCCS. In both cases, cryptic splice donor sites located either in an exon or in an intron were activated because of the partial disruption of the consensus sequence for the authentic splice donor sites due to point mutations. Taken together, oligonucleotide microarrays containing exon junction probes are demonstrated to be a powerful tool to investigate tissue-specific regulation of AS and aberrant splicing taking place in genetic disorders.

INTRODUCTION

Alternative splicing (AS) is one of the major mechanisms by which humans produce the complexity of the proteome. It has been estimated that greater than 55% of all genes and at least 74% of multi-exon genes are alternatively spliced in humans (1,2). Protein isoforms produced by AS can have antagonistic functions, such as anti-apoptotic Bcl-x_L versus pro-apoptotic Bcl-x_S (3), or completely different amino-acid compositions due to different reading frames, such as p16(INK4a) versus p19(ARF) (4). In addition, AS is also implicated in pathophysiological processes and it has been estimated that at least 15% of point mutations that cause human genetic diseases affect splicing (5).

Nevoid basal cell carcinoma syndrome (NBCCS), also called Gorlin's syndrome, is an autosomal dominant neurocutaneous disorder characterized by large body size, developmental and skeletal abnormalities, radiation sensitivity, basal

*To whom correspondence should be addressed at: Department of Genetics, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan. Tel: +81 334160181; Fax: +81 354947035; Email: tmiyashita@nch.go.jp

[†]The nucleotide sequence data of human and mouse isoforms, +12b, have been deposited with the GenBank Library under Accession Nos AB214500 and AB214501, respectively. Human isoforms, −3, 4.5, +4′ and −10, have been deposited with the GenBank Library under Accession Nos AB233423, AB233424 and AB233422, respectively.

cell carcinoma (BCC) and an increased incidence of medulloblastoma (6). NBCCS is caused by inactivating mutations in the *patched* (*PTCH*) gene (7,8). The human *PTCH* gene contains 23 exons spanning ~ 65 kb and is predicted to encode a protein of 1447 amino-acid residues containing 12 transmembrane-spanning domains and two large extracellular loops (7). Heterozygous loss of *PTCH* found in certain sporadic and familial cases of BCC indicates that *PTCH* is also a tumor suppressor gene (9,10). In vertebrates, a second *patched* gene (*PTCH2* in humans) was identified (11,12). So far, no mutations in *PTCH2* have been reported in NBCCS although a limited number of mutations were found in BCC and medulloblastoma (11).

Recently, others and we have identified that *PTCH* undergoes complex AS between multiple first exons and a second exon (13–16). In addition, we also have identified additional mRNA isoforms downstream of exon 2. Therefore, to survey tissue-specific AS and disease-associated aberrant splicing of *PTCH*, we have developed oligonucleotide microarrays designed for profiling AS. Current conventional microarray technologies are limited in their ability to distinguish and analyze mRNA isoforms. Several groups have reported a survey of human AS using exon junction microarrays that can circumvent this problem (2,17–20). In this study, considering the volume and complexity of the data produced by genome-wide studies and the cost–benefit ratio of these arrays, we have developed microarrays focused on the *PTCH* gene in which probes were designed for individual exons and splice junctions including junctions derived from rare AS. Using these arrays, we demonstrate a detailed evaluation of the data on tissue-specific regulation of AS. In addition, we describe the use of DNA microarrays to identify the aberrant splicing taking place in a genetic disorder.

RESULTS

Detection and validation of AS by oligonucleotide microarrays

So far, at least five alternatively used first exons have been reported in the *PTCH* gene (13–16). Recently, using RNA from various human tissues, we have identified additional mRNA isoforms generated by AS involving downstream exons (exons 2, 3, 4, 5 and 10) and an alternative exon named 12b by RT–PCR and sequencing (Fig. 1A and B). The numbering of exons is according to Johnson *et al.* (7) with GenBank accession no. U59464. Among these isoforms, the one skipping exon 10 has been briefly described previously (21) and the one skipping exons 4 and 5 has been deposited in Genbank (accession no. AB209495) but not discussed in a prior publication. The rest of the isoforms are novel. Wicking *et al.* (22) reported our exon 13 as exon 12b. However, as this is a constitutive exon, it is named as exon 13 in most publications and databases and the following exons are numbered by counting from 5′ to 3′. No AS involving exon 13 or exons further downstream could be identified in any tissues examined. If alternative exons can be spliced independently, then the *PTCH* locus potentially encodes 483 isoforms. To monitor splicing events of the *PTCH* gene, we designed oligonucleotide arrays on which 17 exon probes and 23 exon–exon junction probes as well as a couple of intron probes were placed in duplicate (Fig. 1A). Before analyzing the data obtained from various tissues and NBCCS patients, we validated each probe using plasmids encoding various *PTCH* isoforms. For this purpose, each cDNA sequence encoding six *PTCH* isoforms was amplified by PCR followed by *in vitro* transcription and labeling with Cy-5 and then applied onto the microarrays. The intensity data prior to normalization represented as an intensity heatmap is shown in Figure 1B. Each probe's intensity was then adjusted in two ways: first, for the mean probe intensity for each construct and then for each probe's sensitivity. Finally, mean probe intensities that should be positive according to the plasmid and probe sequences were adjusted to 1, and those that should be negative were adjusted to zero. As depicted in Figure 1C, all constructs showed profiles expected from exon composition and could be clearly discriminated from one another. No intensities in-between were observed. However, when placing probes at exon–exon junctions, there is little choice as to the underlying nucleotide composition. So, two junction probes, exon 9–11 and exon 12–13, were non-informative, e.g. saturated, constant intensity and could lead to erroneous predictions, therefore excluded from the figure.

Tissue-specific regulation of AS

To investigate tissue-specific regulation of AS using array data, we normalized probe intensities as described in Materials and Methods and in Figure 4. If no AS takes place at the exons or exon–exon junctions for which probes are designed, then the normalized relative probe intensities are near 1.0 in all tissues. In contrast, probes involved in AS should give intensities with large standard deviations. As expected from mRNA isoforms we have identified so far, a marked variation in normalized probe intensities was observed in exons 2–5, 10 and 12b (Fig. 2A). The tissue-specific accuracy of these array data was confirmed by RT–PCR. For example, when we focused on exon 12b, in tissues in which elevated probe intensities for exons 12–12b, 12b and 12b–13 were observed (i.e. the brain and heart), RT–PCR products containing exon 12b were also evident (Fig. 2B, lanes 1 and 8). We next addressed the question of where in the brain exon 12b was highly expressed. The subsequent investigation demonstrated that exon 12b was particularly expressed in the cerebellum among various brain tissues (Fig. 2B, lane 2). Additionally, two independent methods, microarray analysis and RT–PCR, demonstrated a strong correlation (Fig. 2C), validating the method of array data normalization. Array data as to inclusion or exclusion of exon 10 was similarly compared with the RT–PCR results. As shown in Figure 2D, these two methods again correlated well with each other. Isoforms skipping exon 10 were the most highly expressed in the thymus and the lung based on both these methods.

As shown in Figure 1A, AS involving exons 2–5 is relatively complicated and cannot be explained simply by inclusion or exclusion of a single exon. When we focused on isoforms skipping exons 4 and 5, the data from the two methods still correlated (Fig. 2E). However, the degree of correlation represented by R^2 was weaker than that obtained with

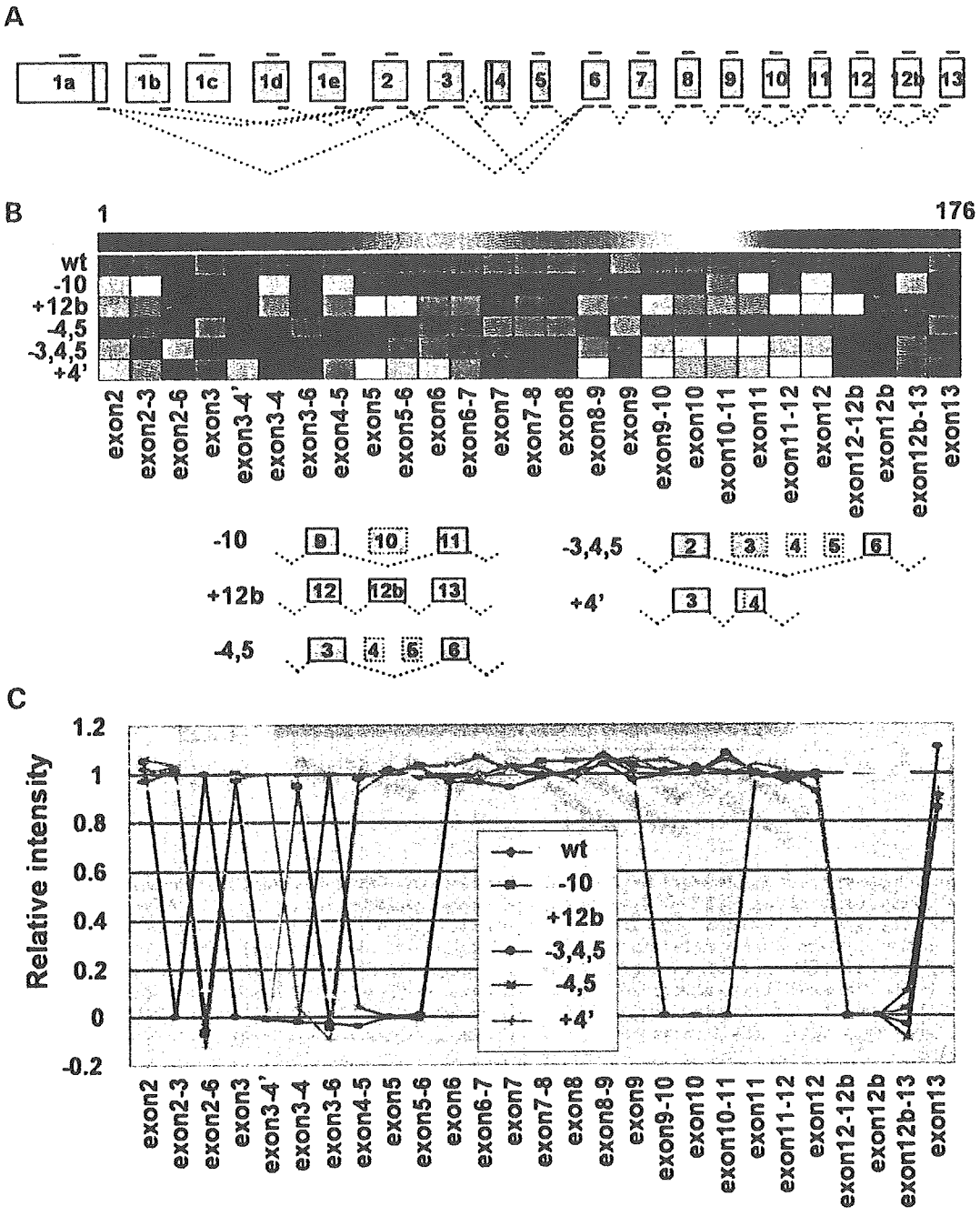


Figure 1. Detection and validation of the known isoforms of *PTCH*. (**A**) Exon structure of the *PTCH* gene. Positions of the exon probes and the exon junction probes are indicated. (**B**) Probe intensities obtained with Cy-5-labeled RNA from constructs encoding various isoforms of *PTCH*. The labeling reaction was performed with T7 RNA polymerase by using the PCR product as a template obtained from the constructs. Each matrix point shows the S/N ratio of one exon or exon junction probe in one construct. Probes are ordered horizontally, 5′–3′, together with some probes for infrequently spliced junctions such as exon 2–6. Exon 3–4′ is an exon junction probe between exon 3 and the alternative splice acceptor site located in intron 3. Probe sequences for most 5′ exons were not included in some of the constructs and were excluded from this figure. Hybridization samples form the vertical axis of each matrix. Exon compositions of the expression plasmids are briefly depicted at the bottom. (**C**) Normalized probe intensities. The mean positive intensity and the mean negative intensity are adjusted to 1 and 0, respectively.

probes for exon 12b or 9 and 10, because AS skipping exons 4 and 5 is a rare event [up to 3% of the authentic isoform (shown by *X*-axis in Fig. 2E)].

Next, we evaluated the microarray data regarding the usage of the alternative first exons. RT–PCR was performed using the same forward primers for each alternative exon 1, as those used for the microarray analysis. As shown in Figure 2F and G, the two methods showed a good correlation regarding the usage of exon 1b. However, isoforms starting from exons 1a, 1d and 1e failed to demonstrate significant

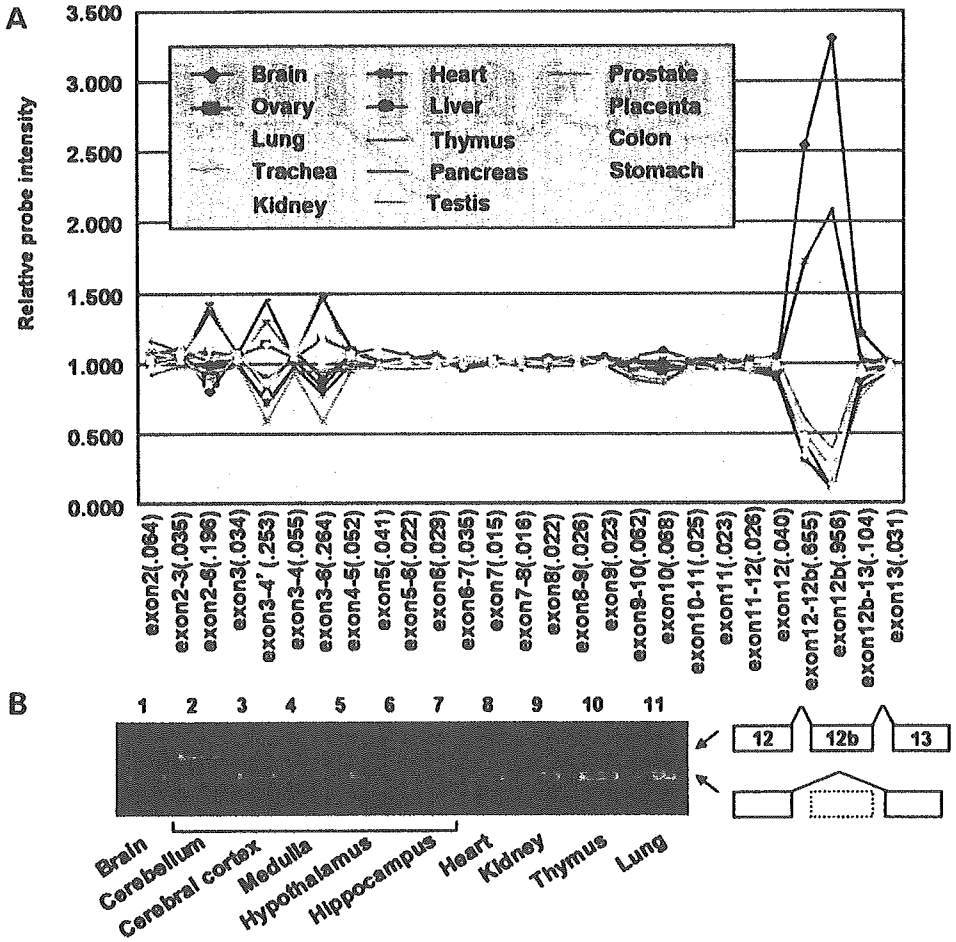


Figure 2. Detection of tissue-specific regulation of AS. (**A**) AS profiles in various tissues. Probes were ordered horizontally as in Figure 1B. Standard deviations for each probe are indicated in parentheses. (**B**) RT–PCR results from a primer pair hybridizing to exons 11 and 14. AS events corresponding to each RT–PCR product are depicted at the right. (**C**) Comparison of the data obtained by the two methods. The *Y*-axis represents normalized relative probe intensity shown in (A). RT–PCR products were applied onto Agilent Bioanalyzer and the percentage of the isoform containing exon 12b is presented by the *X*-axis. (**D**) Two data obtained by microarray and RT–PCR were compared as in (C). A primer pair was constructed on exons 8 and 13. *X*-axis represents the percentage of the isoform skipping exon 10. (**E**) The data obtained by microarray analysis and RT–PCR were compared as in (C). A primer pair was constructed on exons 1b and 6. The *X*-axis represents the percentage of the isoform skipping exons 4 and 5. (**F, G**) The data obtained by microarray analysis [probe 1b in (F) and junction probe 1b-2 in (G)] and RT–PCR were compared as in (C). A primer pair was constructed on exons 1b and 2. The *X*-axis represents the estimated molarity of the isoform starting from exon 1b.

correlations (data not shown). This is probably because the expression of isoforms starting from exons 1a and 1e is much lower than that starting from exon 1b (average expression of exons 1a and 1e is 3.4 and 13.8% of exon 1b, respectively, when pooled S/N ratios are compared). In addition, the PCR efficiency of the exon 1d primer is significantly lower than that of the exon 1b primer (37% of the exon 1b primer) (Supplementary Material, Fig. S1).

Detection of aberrant splicing in NBCCS patients

We have been investigating *PTCH* mutations in NBCCS patients (23) and have detected mutations in 13 out of 17 cases analyzed so far. A list of all cases is presented in Supplementary Material, Table S1. Among 13 cases, G17 had a mutation, c.584G>A (as per GenBank entry NM_000264.2: the A of the ATG of the initiator Met codon is counted as nucleotide +1), on exon 3. This raised two possibilities that explain the effects on the coding for the PTCH protein. One is a missense mutation, p.R195K (as per GenBank entry NP_000255.1). However, as this point mutation was located at the 3′ end of exon 3, and potentially disrupts a splice donor site, we sought the second possibility that the mutation may affect splicing. The data of the microarray analysis showed a significant decreased intensity for the junctional probe exon 3–4, but otherwise the data looked normal indicating an abnormal splicing between exons 3 and 4 (Fig. 3A). The electrophoretogram of the RT–PCR product revealed an additional band which had a larger molecular weight indicating the presence of an aberrant splicing (Fig. 3B, left panel). Sequencing of the additional product demonstrated that abnormal splicing was indeed taking place in which a cryptic splice donor site located in intron 3 was activated, resulting in the insertion of a 37-bp

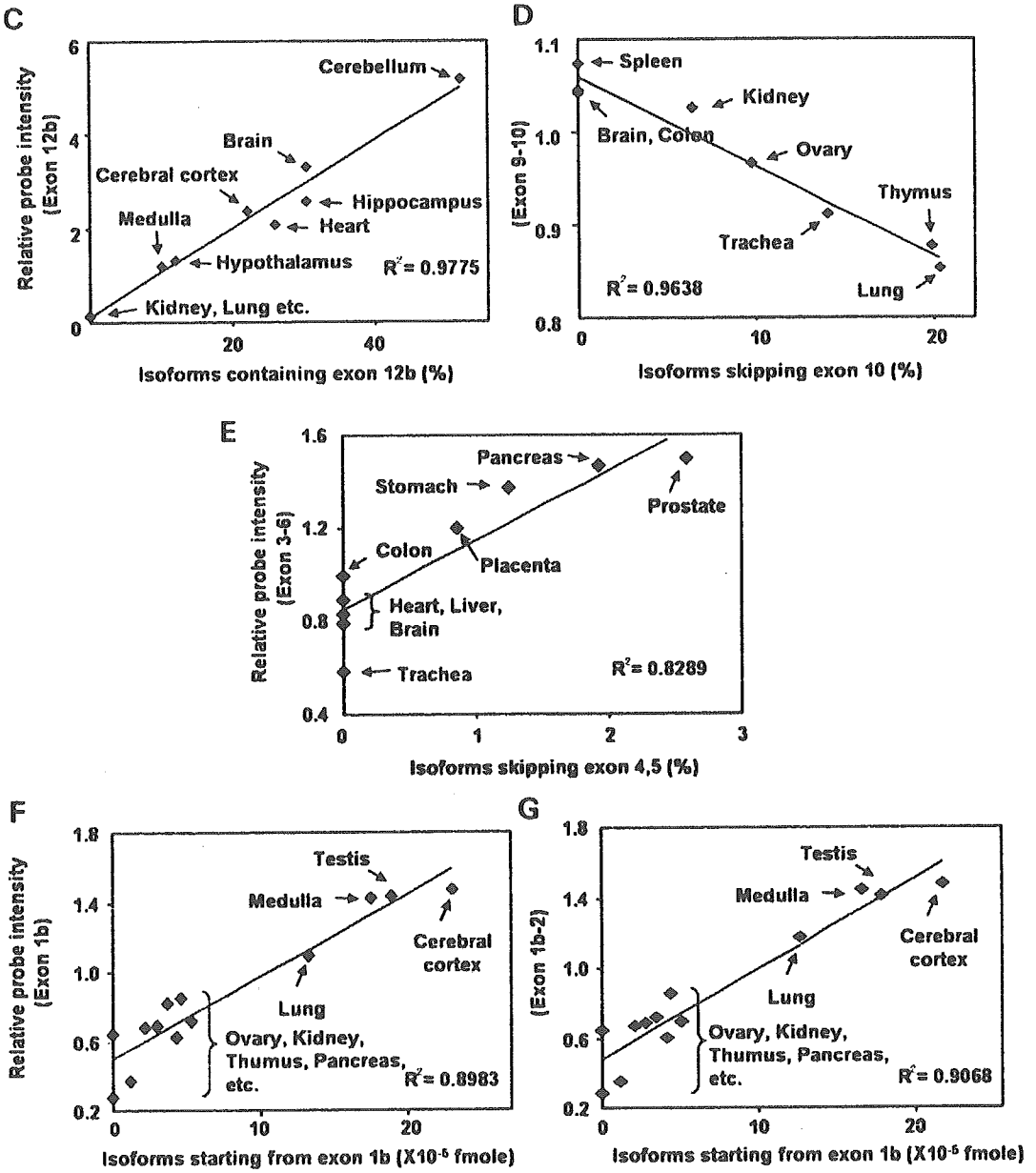


Figure 2. Continued.

intronic sequence between exons 3 and 4 causing a premature termination of the PTCH protein (Fig. 3C). We therefore concluded that it was the aberrant splicing rather than the missense mutation that caused the disease phenotype in this patient.

Another case, G9, did not contain mutations in any of the coding exons of *PTCH*. However, the array data demonstrated a markedly decreased intensity for the junctional probe exon 6–7 (Fig. 3A). The detection of the smaller RT–PCR product in G9 (Fig. 3B, right panel) prompted us to sequence this RT–PCR product, which again revealed the presence of an aberrant splicing. But in this case, a cryptic splice donor site located in exon 6 was activated generating the deletion of an exonic sequence of 87 bp (Fig. 3C). As a result, 29 amino-acid residues located in the first extracellular loop important for Shh binding were deleted. Mutational analysis in intron 6 identified a point mutation, c.945+5G>T, that partially disrupts the consensus sequence for the splice donor site (Fig. 3C).

The third case, G8, had a mutation, c.1526G>A, on exon 11. This mutation presumably results in a missense mutation, p.G509D. However, a growing body of evidence indicates that some single base changes may be capable of switching regulation from positive to negative or vice versa by disrupting or creating exonic splicing enhancers or exonic splicing silencers that are just beginning to be understood (24). Therefore, we investigated such possibilities using microarrays. As shown in Figure 3A, the splicing profile of G8 was similar

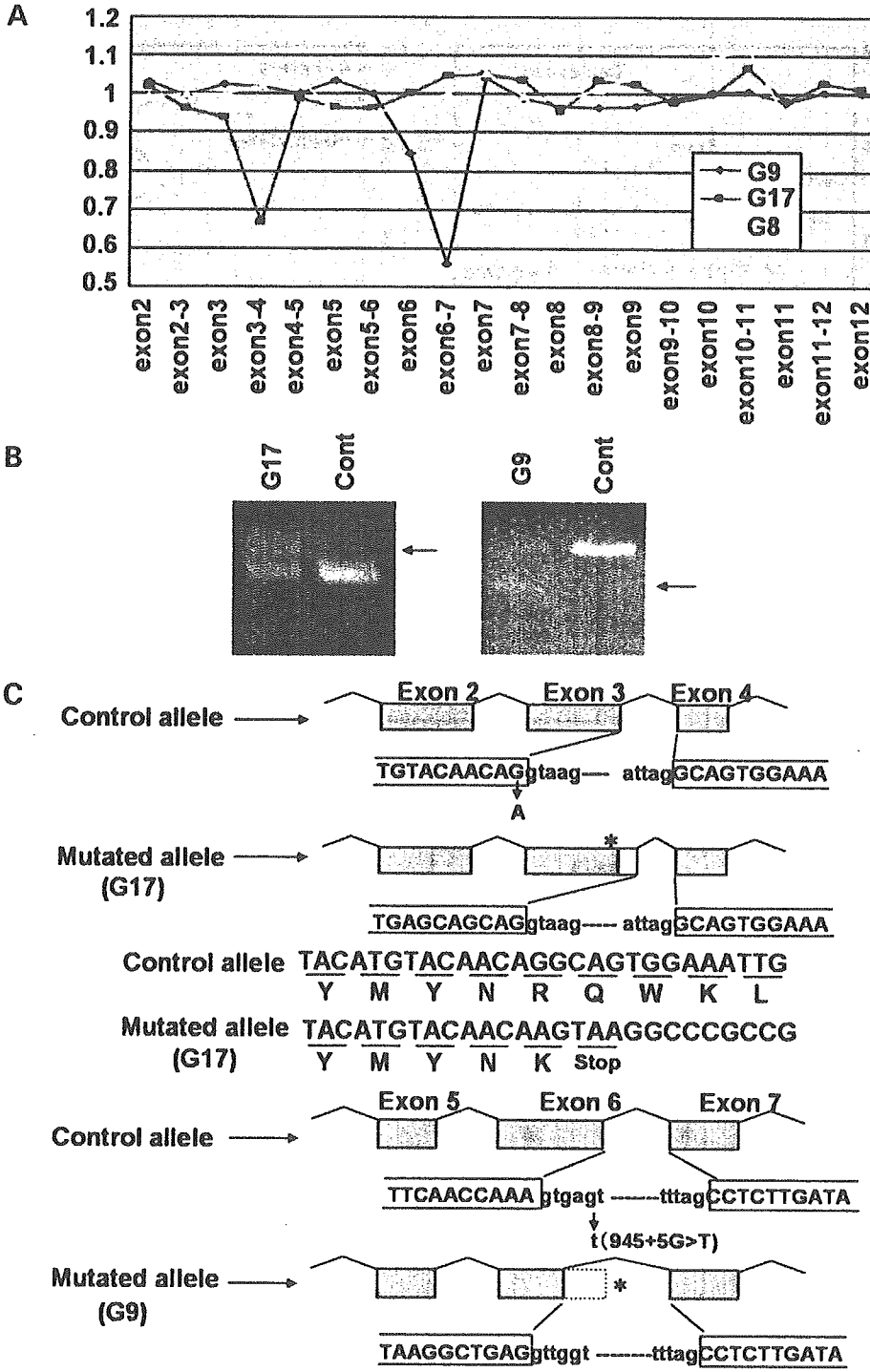


Figure 3. Detection of aberrant splicing in NBCCS patients. (**A**) Normalized probe intensities obtained from NBCCS patients, G8, G9 and G17. Data were analyzed as described in Figure 2A. (**B**) RT–PCR analysis of AS in the *PTCH* gene. Exon 3 (G17) and exon 6 (G9) and flanking exons were amplified by RT–PCR and subjected to agarose gel electrophoresis. The positions of extra bands not observed in a control healthy individual (Cont) are indicated by arrows. (**C**) Schematic representation of abnormal splicing identified in G17 and G9. The positions of the point mutations are indicated by asterisks.

to that of the control and no significant changes in probe intensity were observed. So, we concluded that the mutation found in G8 was a bona fide missense mutation. Interestingly, substitutions of the same amino-acid residue have been reported in NBCCS (p.G509R and p.G509V) (25,26) and these mutations are predicted to disrupt the sterol-sensing domain of the PTCH protein (27). It has also been observed that the mutation p.G509V results in dominant negative activity *in vivo* in *Drosophila* (28). Two out of 17 patients, G7 and G13, did not have a mutation in the region we

have sequenced, nor have any abnormal splicing profile by microarray analysis (Supplementary Material, Table S1). They may have a mutation in non-coding region important for the transcription of *PTCH*, or have a mutation in a gene other than *PTCH*. Altogether, these results demonstrate that exon junction microarrays can be used for detecting the mutations which result in aberrant splicing.

DISCUSSION

In this paper, we described the use of oligonucleotide microarrays containing exon junction probes to investigate tissue-specific AS and to detect disease-associated aberrant splicing. Our microarrays have been demonstrated to be a particularly powerful tool to detect the exon skipping/inclusion type of AS, the most common type (38%) of AS (29), in which even a rare AS event (no more than 3%) can be quantitatively detected using microarrays. Theoretically, previously unrecorded mRNA isoforms can be predicted with this method. If unexpectedly high variation is detected in certain probes, then AS is suggested in this region.

In this study, we found a novel alternative exon named exon 12b. High expression of exon 12b in the brain, particularly in the cerebellum, is intriguing, because proliferative effects of sonic hedgehog (Shh), the ligand of the PTCH protein, in the external granular layer of the cerebellum are well characterized (30) and patients with NBCCS are prone to develop the cerebellar tumor medulloblastoma (6). As exon 12b has an in-frame stop codon, the mRNA isoform containing this exon encodes a truncated PTCH protein that is presumably non-functional. However, this protein may have a dominant negative effect on the wild-type protein and such a possibility needs to be ruled out. Importantly, exon 12b was also found in mice and was also expressed in a brain- and heart-specific manner (data not shown). Recently, the splicing-regulatory element, UGCAUG, was reported to be evolutionarily conserved in introns that flank brain-specific alternative exons (31). Such an element was indeed identified in the intron between exons 12b and 13 both in humans and in mice (data not shown). In contrast, skipping of exon 10 leads to a 52 amino-acid in-frame deletion in the second and third transmembrane domains. The effect on the function of the PTCH protein is currently unknown and this PTCH isoform may be functional in some context.

Detection of AS using exon junction microarrays is limited in several ways. First, as the detection requires probes that match specific exons and splice junctions, probe selection is tightly constrained and some probes are non-informative due to cross-hybridization with somewhere else in human genes. In our study, two out of 42 probes, exon 9–11 and exon 12–13, gave constitutive high intensities and therefore were excluded from further study. Secondly, detection is based on differential expression. Therefore, if two isoforms are present in the same proportion in every tissue, no prediction will result, because normalized probe intensities will behave similarly to the pool. Third, precaution should be taken when interpreting the data on alternative usage of 5′-terminal exons, because not only expression levels but also other factors such as the sequences of forward primers constructed for 5′-terminal exons can have an effect on probe sensitivities.

In this study, we monitored a relatively small number of AS events of one gene in a relatively small set of samples. It should be noted, however, that this is the first report of microarrays used for the identification of aberrant splicings in a genetic disorder. In NBCCS, a considerable number of the patients do not have mutations within the coding region of *PTCH* (23,32). In such patients, exon junction microarrays will be a valuable tool for detecting mutations affecting the splicing event. According to the *PTCH* mutation database (http:/www.cybergene.se/cgi-bin/w3-msql/ptchbase/index.html), at least 20 mutations have been reported to potentially result in abnormal splicing in *PTCH* and some of them have been proven experimentally (21,26). Thus, mutations having an effect on splicing events do not seem to be uncommon. Not only NBCCS, but also an increasing number of genetic diseases are known to be caused by mutations that alter splicing in *cis* (at least 15% of point mutations) (5 and reviewed in 24). Some of these mutations weaken or activate *cis*-acting element such as intronic splicing enhancers or intronic splicing silencers that are sometimes located in intronic sequences distant from exons. In such cases, mutations are not identified by the sequencing of coding regions, warranting the use of exon junction microarrays like the ones described in this study to rapidly search the candidate regions to be sequenced.

Apart from *PTCH*, germ-line mutations in the genes encoding Shh signaling components such as Shh or Gli3 are responsible for a variety of genetic disorders, most of which are accompanied by developmental anomalies in the central nervous system (33). In addition, somatic mutations of the genes involved in this signaling pathway, such as *PTCH2* or *Smoothened*, are associated with various sporadic cancers (33). Moreover, distinct roles of *PTCH2* splice variants in Shh signaling have recently been proposed (34). Therefore, oligonucleotide microarrays containing exon and exon junction probes widely covering these genes would be an attractive tool to study the pathogenesis of these disorders. In addition, these microarrays may also be useful to detect mutations in the dystrophin gene, because the removal of exon(s) is found in 60–65% of patients with Duchenne and Becker muscular dystrophies (35).

MATERIALS AND METHODS

Plasmids

The plasmid encoding myc-tagged full-length PTCH protein (exons 1b–23) (pMyc-Ptc1) was kindly provided by Dr Jeffrey Ming (15). The plasmids encoding other isoforms of *PTCH* were created by PCR-mediated mutagenesis as described previously (36) using pMyc-Ptc1 as a template. The details of construction are available on request.

Oligonucleotide microarray construction

All probes 34–76 bp in length were designed to have approximately the same annealing temperature (T_m). Splice junction probes between two exons were designed so that the T_m for the first exon sequence is the same as that for the second

$$P=\begin{pmatrix} P_{11} & \cdots & \cdots & \cdots & P_{1n} \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ P_{m1} & \cdots & \cdots & \cdots & P_{mn} \end{pmatrix} \quad M=\begin{pmatrix} 1/\overline{P}_1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1/\overline{P}_n \end{pmatrix} \quad \overline{P}_j=\left(\sum_i P_{ij}\right)/m$$

$$PM=\begin{pmatrix} P_{11}/\overline{P}_1 & P_{12}/\overline{P}_2 & \cdots & \cdots & P_{1n}/\overline{P}_n \\ P_{21}/\overline{P}_1 & P_{22}/\overline{P}_2 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ P_{m1}/\overline{P}_1 & \cdots & \cdots & \cdots & P_{mn}/\overline{P}_n \end{pmatrix}$$

$$N=\begin{pmatrix} 1/\overline{p}_1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1/\overline{p}_m \end{pmatrix} \quad \overline{p}_i=\left(\sum_j P_{ij}/\overline{P}_j\right)/n$$

$$NPM=\begin{pmatrix} P_{11}/\overline{p}_1\overline{P}_1 & P_{12}/\overline{p}_1\overline{P}_2 & \cdots & \cdots & P_{1n}/\overline{p}_1\overline{P}_n \\ P_{21}/\overline{p}_2\overline{P}_1 & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ P_{m1}/\overline{p}_m\overline{P}_1 & \cdots & \cdots & \cdots & P_{mn}/\overline{p}_m\overline{P}_n \end{pmatrix}$$

Figure 4. Matrix representation. The relationship between probes and tissues can be represented by an $m \times n$ matrix P. The total number of probes is m, and n is the total number of tissues. Let M be an $n \times n$ diagonal matrix where $\overline{P}j$ represents the mean P in tissue j. The probe intensities normalized by total *PTCH* expression in each tissue can then be expressed as PM. Let N be an $m \times m$ diagonal matrix where $\overline{p}i$ represents the mean normalized P in probe i. The matrix NPM therefore represents the probe intensities normalized by both factors described above and $P_{ij}/\overline{p}i\overline{P}j$ is used in Figures 2 and 3.

exon sequence (Supplementary Material, Table S2). The oligonucleotide microarray, GenoPal™ (Mitsubishi Rayon Co., Ltd), was made in the following manner. Plastic hollow fibers were bundled in an orderly arrangement, and hardened with resin to form a block (Supplementary Material, Fig. S2). Oligonucleotide-capture probes were chemically bonded inside each hollow fiber with hydrophilic gel. The block was then sliced to make thin chips, each of which was set into a holder (http://www.mrc.co.jp/genome/e/index.html for details).

Preparation of labeled probe

Total RNA from a panel of human tissues was purchased from Ambion. Lymphoblastoid cell lines from NBCCS patients immortalized by Epstein-Barr virus were also used to obtain total RNA. All studies using patient samples were approved by the local ethic committee. The template for *in vitro* transcription was generated through a modified procedure using a SuperScript One-Step RT–PCR System with Platinum Taq (Invitrogen). In brief, 2.5 μg of total RNA was reverse-transcribed with SuperScriptIII RT/Platinum Taq Mix and T7-exon15 reverse primer, 5′-TAATACGACTCACTATAGGGGTCATATTCTCTGGTTTCCCGAGGTACAATGTC-3′, for 30 min at 50°C. To evaluate the quality of RNA, T7-GAPDH reverse primer, 5′-TAATACGACTCACTATAGGGAGGAGGGGAGATTCAGTGTGGT-3′ was also added in the reaction. The RNA was degraded with the addition of RNaseH (Invitrogen) for 15 min at 37°C. After the addition of forward primers specific for each first exon (5′-AGCGCCTGTTTACCCAGGAG-3′ for exon 1a, 5′-GGACCGGGACTATCTGCACC-3′ for exon 1b, 5′-AAATGCCGCGCCGGGGAGCAGCCT-3′ for exon 1d and 5′-TTCTCGGCGGGGGTCCAGTT-3′ for exon 1e) and T7-exon15 reverse primer (final concentration 0.5 μM each) and the activation of Platinum Taq for 10 min at 94°C, PCR was run for 30 cycles of denaturation at 94°C for 30 s, annealing at 56°C for 30 s, and extension at 72°C for 2 min. Plasmid DNA encoding

various isoforms of *PTCH1* was also subjected to PCR using the exon 2 forward primer, 5′-GCTGAGAGCGAAGTTTC AGA-3′, and T7-exon 15 reverse primer. PCR products generated from the reverse-transcribed cDNA and the plasmid DNA were purified with a PCR Purification kit (QIAGEN) according to the manufacturer's instructions. Then, a Cy-5-labeled probe was generated by using the PCR product as a template with a MEGAscript T7 kit (Ambion). The reaction contained a 1:1.5 mixture of uridine triphosphate (UTP) and Cy-5-UTP (Amersham Biosciences). The product was purified with an RNeasy mini kit (QIAGEN) following the manufacturer's instructions. One microgram of labeled probe was fragmented with RNA Fragmentation Reagents (Ambion) at 70°C for 3 min.

Hybridization and detection

Hybridization was carried out in a final volume of 0.1 ml injected into a hybridization chamber at 65°C for 16–24 h in 0.5× SSC with 0.2% SDS. The microarray chip was then washed twice in 0.5× SSC with 0.2% SDS at 55°C for 20 min and once in 0.5× SSC at 55°C for 10 min, before being slowly cooled to room temperature. GenoPal was then scanned and the image was captured with a cooled CCD-type Microarray Image Analyzer (Mitsubishi Rayon Co., Ltd). Fluorescent intensity was analyzed with software developed by Mitsubishi Rayon Co., Ltd. Fluorescence throughout the three-dimensional structure of each array feature can be efficiently captured due to the long focal depth of the optical system of the image analyzer (http://www.mrc.co.jp/genome/e/index.html).

Microarray data analysis

To analyze changes in the AS of a gene, changes in the total expression of a gene and differences in probe sensitivity should be separated and excluded. We used a simple and generalized pooling strategy presented by Le *et al.* (20) with modifications. The probe response P_{ij} (represented as an S/N ratio) for a specific probe i to a specific tissue sample j was normalized using the total expression of *PTCH* and probe sensitivity as described in Figure 4.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We thank Kaori Takeuchi-Inoue and Mayu Yamazaki-Inoue for technical support, and Kayoko Saito for preparing the manuscript. This work was supported by the Naito Foundation and Grants for Cancer Research and Child Health and Development from the Ministry of Health, Labour and Welfare; a Grant-in-Aid for Scientific Research and the Budget for Nuclear Research from the Ministry of Education, Culture, Sports, Science and Technology.

Conflict of Interest statement. None declared.

REFERENCES

1. Kan, Z., Rouchka, E.C., Gish, W.R. and States, D.J. (2001) Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res.*, **11**, 889–900.
2. Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R. and Shoemaker, D.D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, **302**, 2141–2144.
3. Boise, L.H., Gonzalez-Garcia, M., Postema, C.E., Ding, L., Lindsten, T., Turka, L.A., Mao, X., Nunez, G. and Thompson, C.B. (1993) bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell*, **74**, 597–608.
4. Quelle, D.E., Zindy, F., Ashmun, R.A. and Sherr, C.J. (1995) Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. *Cell*, **83**, 993–1000.
5. Krawczak, M., Reiss, J. and Cooper, D.N. (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.*, **90**, 41–54.
6. Gorlin, R.J. (1987) Nevoid basal-cell carcinoma syndrome. *Medicine*, **66**, 98–113.
7. Johnson, R.L., Rothman, A.L., Xie, J., Goodrich, L.V., Bare, J.W., Bonifas, J.M., Quinn, A.G., Myers, R.M., Cox, D.R., Epstein, E.H., Jr *et al.* (1996) Human homolog of *patched*, a candidate gene for the basal cell nevus syndrome. *Science*, **272**, 1668–1671.
8. Hahn, H., Wicking, C., Zaphiropoulous, P.G., Gailani, M.R., Shanley, S., Chidambaram, A., Vorechovsky, I., Holmberg, E., Unden, A.B., Gillies, S. *et al.* (1996) Mutations of the human homolog of *Drosophila patched* in the nevoid basal cell carcinoma syndrome. *Cell*, **85**, 841–851.
9. Gailani, M.R., Stahle-Backdahl, M., Leffell, D.J., Glynn, M., Zaphiropoulos, P.G., Pressman, C., Unden, A.B., Dean, M., Brash, D.E., Bale, A.E. *et al.* (1996) The role of the human homologue of *Drosophila patched* in sporadic basal cell carcinomas. *Nat. Genet.*, **14**, 78–81.
10. Unden, A.B., Holmberg, E., Lundh-Rozell, B., Stahle-Backdahl, M., Zaphiropoulos, P.G., Toftgard, R. and Vorechovsky, I. (1996) Mutations in the human homologue of *Drosophila patched* (*PTCH*) in basal cell carcinomas and the Gorlin syndrome: different *in vivo* mechanisms of PTCH inactivation. *Cancer Res.*, **56**, 4562–4565.
11. Smyth, I., Narang, M.A., Evans, T., Heimann, C., Nakamura, Y., Chenevix-Trench, G., Pietsch, T., Wicking, C. and Wainwright, B.J. (1999) Isolation and characterization of human *patched 2* (*PTCH2*), a putative tumour suppressor gene in basal cell carcinoma and medulloblastoma on chromosome 1p32. *Hum. Mol. Genet.*, **8**, 291–297.
12. Zaphiropoulos, P.G., Unden, A.B., Rahnama, F., Hollingsworth, R.E. and Toftgård, R. (1999) *PTCH2*, a novel human patched gene, undergoing alternative splicing and up-regulated in basal cell carcinomas. *Cancer Res.*, **59**, 787–792.
13. Kogerman, P., Krause, D., Rahnama, F., Kogerman, L., Unden, A.B., Zaphiropoulos, P.G. and Toftgård, R. (2002) Alternative first exons of *PTCH1* are differentially regulated *in vivo* and may confer different functions to the PTCH1 protein. *Oncogene*, **21**, 6007–6016.
14. Ågren, M., Kogerman, P., Kleman, M.I., Wessling, M. and Toftgård, R. (2004) Expression of the *PTCH1* tumor suppressor gene is regulated by alternative promoters and a single functional Gli-binding site. *Gene*, **330**, 101–114.
15. Nagao, K., Toyoda, M., Takeuchi-Inoue, K., Fujii, K., Yamada, M. and Miyashita, T. (2005) Identification and characterization of multiple isoforms of a murine and human tumor suppressor, *Patched*, having distinct first exons. *Genomics*, **85**, 462–471.
16. Shimokawa, T., Rahnama, F. and Zaphiropoulos, P.G. (2004) A novel first exon of the *Patched1* gene is upregulated by Hedgehog signaling resulting in a protein with pathway inhibitory functions. *FEBS Lett.*, **578**, 157–162.
17. Castle, J., Garrett-Engele, P., Armour, C.D., Duenwald, S.J., Loerch, P.M., Meyer, M.R., Schadt, E.E., Stoughton, R., Parrish, M.L., Shoemaker, D.D. *et al.* (2003) Optimization of oligonucleotide arrays and RNA amplification protocols for analysis of transcript structure and alternative splicing. *Genome Biol.*, **4**, R66.
18. Wang, H., Hubbell, E., Hu, J.S., Mei, G., Cline, M., Lu, G., Clark, T., Siani-Rose, M.A., Ares, M., Kulp, D.C. *et al.* (2003) Gene structure-based splice variant deconvolution using a microarray platform. *Bioinformatics*, **19** (Suppl. 1), i315–i322.

19. Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S. and Fu, X.D. (2002) Profiling alternative splicing on fiber-optic arrays. *Nat. Biotechnol.*, **20**, 353–358.
20. Le, K., Mitsouras, K., Roy, M., Wang, Q., Xu, Q., Nelson, S.F. and Lee, C. (2004) Detecting tissue-specific regulation of alternative splicing as a qualitative change in microarray data. *Nucleic Acids Res.*, **32**, e180.
21. Smyth, I., Wicking, C., Wainwright, B. and Chenevix-Trench, G. (1998) The effects of splice site mutations in patients with naevoid basal cell carcinoma syndrome. *Hum. Genet.*, **102**, 598–601.
22. Wicking, C., Gillies, S., Smyth, I., Shanley, S., Fowles, L., Ratcliffe, J., Wainwright, B. and Chenevix-Trench, G. (1997) *De novo* mutations of the *Patched* gene in nevoid basal cell carcinoma syndrome help to define the clinical phenotype. *Am. J. Med. Genet.*, **73**, 304–307.
23. Fujii, K., Kohno, Y., Sugita, K., Nakamura, M., Moroi, Y., Urabe, K., Furue, M., Yamada, M. and Miyashita, T. (2003) Mutations in the human homologue of *Drosophila patched* in Japanese nevoid basal cell carcinoma syndrome patients. *Hum. Mutat.*, **21**, 451–452.
24. Garcia-Blanco, M.A., Baraniak, A.P. and Lasda, E.L. (2004) Alternative splicing in disease and therapy. *Nat. Biotechnol.*, **22**, 535–546.
25. Chidambaram, A., Goldstein, A.M., Gailani, M.R., Gerrard, B., Bale, S.J., DiGiovanna, J.J., Bale, A.E. and Dean, M. (1996) Mutations in the human homologue of the *Drosophila patched* gene in Caucasian and African-American nevoid basal cell carcinoma syndrome patients. *Cancer Res.*, **56**, 4599–4601.
26. Pastorino, L., Cusano, R., Nasti, S., Faravelli, F., Forzano, F., Baldo, C., Barile, M., Gliori, S., Muggianu, M., Ghigliotti, G. *et al.* (2005) Molecular characterization of Italian nevoid basal cell carcinoma syndrome patients. *Hum. Mutat.*, **25**, 322–323.
27. Kuwabara, P.E. and Labouesse, M. (2002) The sterol-sensing domain: multiple families, a unique role? *Trends Genet.*, **18**, 193–201.
28. Hime, G.R., Lada, H., Fietz, M.J., Gillies, S., Passmore, A., Wicking, C. and Wainwright, B.J. (2004) Functional analysis in *Drosophila* indicates that the NBCCS/PTCH1 mutation G509V results in activation of smoothened through a dominant-negative mechanism. *Dev. Dyn.*, **229**, 780–790.
29. Sugnet, C.W., Kent, W.J., Ares, M., Jr and Haussler, D. (2004) Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput.*, 66–77.
30. Dahmane, N. and Altaba, A. (1999) Sonic hedgehog regulates the growth and patterning of the cerebellum. *Development*, **126**, 3089–3100.
31. Minovitsky, S., Gee, S.L., Schokrpur, S., Dubchak, I. and Conboy, J.G. (2005) The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. *Nucleic Acids Res.*, **33**, 714–724.
32. Wicking, C., Shanley, S., Smyth, I., Gillies, S., Negus, K., Graham, S., Suthers, G., Haites, N., Edwards, M., Wainwright, B. *et al.* (1997) Most germ-line mutations in the nevoid basal cell carcinoma syndrome lead to a premature termination of the PATCHED protein, and no genotype–phenotype correlations are evident. *Am. J. Hum. Genet.*, **60**, 21–26.
33. Cohen, M.M., Jr (2003) The hedgehog signaling network. *Am. J. Med. Genet.*, **123A**, 5–28.
34. Rahnama, F., Toftgård, R. and Zaphiropoulos, P.G. (2004) Distinct roles of PTCH2 splice variants in Hedgehog signalling. *Biochem. J.*, **378**, 325–334.
35. Muntoni, F., Torelli, S. and Ferlini, A. (2003) Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurol.*, **2**, 731–740.
36. Imai, Y., Matsushima, Y., Sugimura, T. and Terada, M. (1991) A simple and rapid method for generating a deletion by PCR. *Nucleic Acids Res.*, **19**, 2785.

A Comprehensive Analysis of Allelic Methylation Status of CpG Islands on Human Chromosome 21q

Yoichi Yamada,[1,2] Hidemi Watanabe,[3,4] Fumihito Miura,[1,2] Hidenobu Soejima,[5] Michiko Uchiyama,[6] Tsuyoshi Iwasaka,[6] Tsunehiro Mukai,[5] Yoshiyuki Sakaki,[2,4] and Takashi Ito[1,7,8]

[1]Division of Genome Biology, Cancer Research Institute, Kanazawa University, Kanazawa 920-0934, Japan; [2]Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan; [3]Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan; [4]Genomic Science Center, RIKEN Yokohama Institute, Yokohama 230-0045, Japan; [5]Division of Molecular Biology and Genetics, Department of Biomolecular Sciences, and [6]Department of Obstetrics and Gynecology, Saga University Faculty of Medicine, Saga 849-8501, Japan; [7]Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa 277-8562, Japan

Approximately half of all human genes have CpG islands (CGIs) around their promoter regions. Although CGIs usually escape methylation, those on Chromosome X in females and those in the vicinity of imprinted genes are exceptions: They have both methylated and unmethylated alleles to display a "composite" pattern in methylation analysis. In addition, aberrant methylation of CGIs is known to often occur in cancer cells. Here we developed a simple HpaII–McrBC PCR method for discrimination of full, null, incomplete, and composite methylation patterns, and applied it to all computationally identified CGIs on human Chromosome 21q. This comprehensive analysis revealed that, although most CGIs (103 out of 149) escape methylation, a sizable fraction (31 out of 149) are fully methylated even in normal peripheral blood cells. Furthermore, we identified seven CGIs showing the composite methylation, and demonstrated that three of them are indeed methylated monoallelically. Further analyses using informative pedigrees revealed that two of the three are subject to maternal allele-specific methylation. Intriguingly, the other CGI is methylated in an allele-specific but parental-origin-independent manner. Thus, the cell seems to have a broader repertoire of methylating CGIs than previously thought, and our approach may contribute to uncover novel modes of allelic methylation.

[Supplemental material is available online at www.genome.org.]

Mammalian genomes contain CpG dinucleotides much less frequently than expected from their GC contents (i.e., CpG suppression), and most of them are modified by methylation at the 5-position of cytosine (Ponger et al. 2001). However, CpG suppression is not observed or much less evident in characteristic regions termed CpG islands (CGIs) despite their high GC contents (Gardiner-Garden and Frommer 1987; Antequera and Bird 1993). CGIs are generally found near promoter regions of genes, including most housekeeping and many tissue-specific ones, and intriguingly escape methylation, often regardless of the expression of flanking genes (Macleod et al. 1998; Grunau et al. 2000; Ioshikhes and Zhang 2000).

Although aberrant methylation of CGIs is frequently observed in cancer cells, some exceptional CGIs are physiologically methylated in an allele-specific manner. It is well known that one of the two X-chromosomes in females is inactivated. The CGIs on the inactivated X-chromosome are heavily methylated, similar to other regions on this chromosome (Norris et al. 1991). On autosomes, a small number of imprinted genes that display exclusive or highly skewed expression of specific allele depending on their parental origins (Morison and Reeve 1998) have been demonstrated to accompany regions subject to parental-origin-dependent methylation. These regions are termed allelic differentially methylated regions (DMRs), and have been demonstrated to play pivotal roles in genomic imprinting (Wutz et al. 1997; Yoon et al. 2002). Although allelic DMRs show base composition similar to CGIs and often contain tandem repeat sequences (Neumann et al. 1995), they share no apparent sequence similarity.

Allelic DMRs have been extensively searched around imprinted genes but not in other regions. In other words, their distribution has not been analyzed in an unbiased, hypothesis-free manner. Although several methods have been developed for the purpose, they are not truly comprehensive and have missed many DMRs (Plass et al. 1996). We thus intended to thoroughly examine the methylation status of CGIs based on the established genome sequence data, which allows one to identify all CGIs in silico. The experimental method to be used for the evaluation of methylation status should not only be rapid and simple but also be capable of detecting the coexistence of methylated and unmethylated alleles (i.e., composite methylation).

As a method to fulfill the requirement, we developed a simple method called HpaII–McrBC PCR, which is based on the complementary sensitivity of the two enzymes HpaII and McrBC to DNA methylation. We applied it for the analysis of 149 CGIs computationally identified on human Chromosome 21q, one of the most completely sequenced chromosomes. The analysis, which is the very first thorough analysis of CGIs on a chromosome-wide scale, revealed an unexpectedly high incidence of normally methylated CGIs and, furthermore, three

[8]Corresponding author.
E-MAIL titolab@kenroku.kanazawa-u.ac.jp; FAX 81-76-234-4508.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.1351604.

allelic DMRs, including one subject to a novel mode of allelic methylation.

RESULTS

HpaII–McrBC PCR for Rapid Evaluation of Allelic Methylation Status

A comprehensive methylation analysis requires a rapid and simple method to examine methylation status. Although the so-called HpaII-PCR has been widely used, it cannot distinguish between fully methylated and compositely methylated sequences, the latter of which include CGIs on X-chromosomes in female and allelic DMRs in the vicinity of imprinted genes. To overcome this drawback, we developed a novel method termed HpaII–McrBC PCR by exploiting two enzymes with complementary methylation sensitivity. The method can readily distinguish regions subject to full, null, composite, and incomplete methylation.

In HpaII–McrBC PCR, genomic DNA is divided into two portions, each of which is subsequently digested with HpaII (or other methylation-sensitive enzymes such as HhaI) or McrBC, and used as templates for PCR (Fig. 1). Whereas HpaII cuts unmethylated alleles at CCGG sites, McrBC digests methylated alleles at $R^mCN_{40\ 80}R^mC$ (Fig. 1; Sutherland et al. 1992; Stewart and Raleigh 1998). In the case of a fully methylated sequence, HpaII totally fails to digest the target, whereas McrBC cuts it completely (Fig. 1A). Amplification would be thus achieved only from the HpaII-digested template. On the other hand, an unmethylated region is digested only with HpaII but not with McrBC, and hence amplification would be successful only from the McrBC-digested DNA. Accordingly, amplification from both HpaII- and McrBC-digested DNAs indicates the presence of both methylated and unmethylated alleles in the sample or the "composite" methylation. If the target region is incompletely methylated, amplification will be obtained from neither HpaII- nor McrBC-digested DNAs, because both enzymes digest the template. Therefore, HpaII–McrBC PCR can, in principle, distinguish four different statuses of allelic methylation (Fig. 1A).

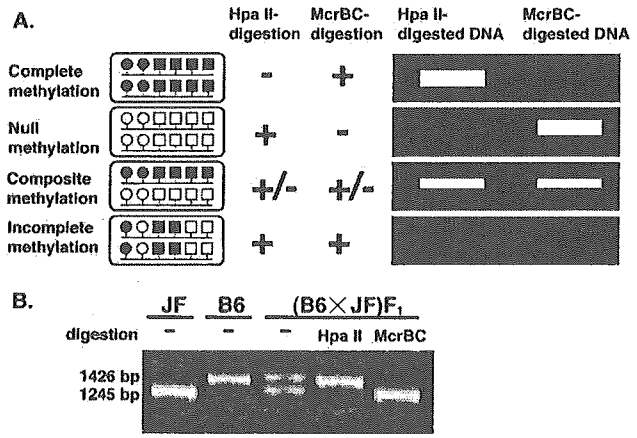


Figure 1 HpaII–McrBC PCR. (*A*) Principle of HpaII–McrBC PCR that distinguishes four different patterns in allelic methylation. (*Left*) The open and closed circles indicate unmethylated CCGG and methylated C^mCGG sites, respectively. Similarly, the open and closed squares indicate unmethylated RC and methylated R^mC sites, respectively. Each line with two circles and four squares indicates each allele of genomic DNA. (*Middle*) The +, +/ , and mean complete, incomplete, and no digestion with HpaII or McrBC, respectively. (Right) A schematic gel pattern of HpaII–McrBC PCR products in individual cases. (*B*) Proof of principle for HpaII–McrBC PCR. HpaII–McrBC PCR was applied to the intronic DMR of mouse *Impact*, a paternally expressed gene. JF, B6, and (B6 JF) F_1 indicate *Mus musculus molossinus* JF1, *Mus musculus domesticus* C57BL/6, and F_1 hybrid generated between JF and B6, respectively. The PCR products from mock-treated ( ), HpaII-digested, or McrBC-digested DNAs from JF, B6, or (B6 JF) F_1 were electrophoresed, stained with ethidium bromide, and visualized by UV illumination.

As a proof-of-principle experiment, we applied the method to the allelic DMR of mouse *Impact*, which we identified as a gene expressed exclusively from the paternal allele bearing a maternally methylated CGI or allelic DMR in its first intron (Hagiwara et al. 1997; Okamura et al. 2000). We analyzed an F_1 hybrid between *Mus musculus domesticus* C57BL/6 (B6) and *Mus musculus molossinus* JF1 (JF), because the CGI displays an obvious length polymorphism between the two species, allowing one to distinguish the two alleles by simple gel electrophoresis (Okamura et al. 2000).

As shown in Figure 1B, both the maternally derived B6 allele (1426 bp) and the paternally derived JF one (1245 bp) were detected from the mock-treated DNA prepared from a (B6 JF) F_1 mouse. Note that only the maternally derived B6 allele was amplified from the HpaII-digested DNA, which serves as a template for methylated portions. On the other hand, only the paternally derived JF allele was amplified from the McrBC-treated DNA, in which methylated alleles were digested. These results unequivocally indicate maternal methylation of this CGI, consistent with the previous observation (Okamura et al 2000).

We also examined the CGI spanning the promoter region of human *IMPACT*, which we had previously shown to escape methylation biallelically and hence represent a conventional nonmethylated CGI (Okamura et al. 2000). As expected, amplification was achieved only from the McrBC-digested DNA but not from the HpaII-digested one, indicative of unmethylated pattern (data not shown).

These results demonstrated that HpaII–McrBC PCR serves as a rapid and simple method to evaluate allelic methylation status.

Strategy for a Comprehensive HpaII–McrBC PCR Analysis of CGIs on Human Chromosome 21q

Having a versatile method in hand to examine allelic methylation status, we planned a comprehensive methylation analysis of CGIs on human Chromosome 21q. Because this chromosome provides the most complete and accurate sequence data ever generated, one can identify CGIs most thoroughly in silico and examine their methylation status in a truly comprehensive manner.

According to the original definition of CGI, it should be longer than 200 bp, have a GC content higher than 50%, and display an expected CpG frequency (ECF) larger than 0.6 (Gardiner-Garden and Frommer 1987). This definition has, however, turned out to allow contamination of repetitive DNA elements as well as exons. Thus, more stringent criteria are generally used in recent studies. For instance, in the initial annotation of human Chromosome 21q, we used a criterion requiring that the length, GC content, and ECF be larger than 400 bp, 55%, and 0.6, respectively, to identify 137 CGIs, most of which were found linked to the 5 -portions of genes (Hattori et al. 2000). Here we used a slightly relaxed condition (length >400 bp, GC content >50%, ECF >0.6) with the masking of Alu and LINE-1 sequences to extract 149 CGIs in total (Table 1).

For these CGIs, we designed PCR primers using a free program called prima. We selected each primer pair so that the amplicon keeps its GC content as low as possible and contains more than two recognition sites for HpaII or HhaI, thereby avoiding the difficulty in amplification and minimizing the effect of incomplete digestion, and at least one site that would be recognized by McrBC when methylated. The program designed 101 primer pairs that work in amplification, and 47 other pairs were

Table 1. Methylation and Other Features of CpG Islands on Human Chromosome 21q

CpG islands[a]	Methylation	Repeat[b]	Locus	GC %[c]	Obs/exp[d]	Size (bp)	Nucleotide position[e]		Location in the linked gene[f]	CGI-linked genes
							Start	End		
#1 (NT_002836.4 740746-742525)	Complete methylation		21q11.1	72.0	0.88	780	13998888	13999667	CDS	Similarity to feminization 1 homolog a (*C. elegans*)
#2 (NT_002836.4 798428-799837)	Complete methylation		21q11.1	72.2	0.80	410	14056570	14056979	CDS	Similarity to feminization 1 homolog a (*C. elegans*)
#3 (NT_002836.4 1099894-1101335)	Complete methylation		21q11.1	68.1	0.80	442	14358036	14358477	5 -UTR and CDS	Similarity to hypothetical protein DKFZp434A171; KIAA0565; ankyrin repeat-containing proteins
#4 (NT_002836.4 2100365-2102278)	Unmethylation		21q11.2	76.2	0.98	914	15358453	15359366	5 -UTR	Nuclear receptor interacting protein 1 (NRIP1)
#5 (NT_002836.4 2765740-2767861)	Unmethylation		21q11.2	74.7	0.88	1320	16023802	16025122	5 -UTR and CDS	Ubiquitin specific protease 25 (USP25)
#6 (NT_002836.4 4548804-4550994)	Unmethylation		21q21.1	71.1	0.83	1229	17806771	17807999	5 -UTR and CDS	Coxsackie virus and adenovirus receptor (CXADR)
#7 (NT_002836.4 4648389-4650557)	Unmethylation		21q21.1	73.4	0.96	1169	17906356	17907524	5 -UTR	BTG family member 3 (BTG3)
#8 (NT_002836.4 4854926-4857042)	Unmethylation		21q21.1	68.0	0.97	1117	18112871	18113987	5 -UTR and CDS	Chromosome 21 open reading frame 91 (C21 orf91), hypothetical protein LOC54149 (YG81)
#9 (NT_002836.4 12511562-12513060)	Unmethylation		21q21.2	70.3	0.88	499	25856202	25856700		—
#10 (NT_002836.4 12684911-12686522)	Unmethylation		21q21.2	71.7	0.79	612	26029550	26030161	5 -UTR	ATP synthase H+ transporting mitochondrial F0 complex subunit F6
#11 (NT_002836.4 13119202-13121651)	Unmethylation		21q21.3	70.8	0.83	1450	26463778	26465227	5 -UTR and CDS	Amyloid beta precursor protein (APP)
#12 (NT_002836.4 13793868-13795756)	Unmethylation		21q21.3	67.2	0.98	889	27138374	27139262	5 -UTR and CDS	Matrix metalloprotease (ADAMTS1)

(continued)

Table 1. *Continued*

CpG islands[a]	Methylation	Repeat[b]	Locus	GC %[c]	Obs/exp[d]	Size (bp)	Nucleotide position[e] Start	End	Location in the linked gene[f]	CGI-linked genes
#13 (NT_002836.4 13915193-13916938)	Unmethylation		21q21.3	70.1	0.87	746	27259699	27260444	5 -UTR and CDS	A disintegrin-like and metalloprotease with thrombospondin type 1 motif5 (ADAMTS5)
#14 (NT_002836.4 13917154-13918663)	Unmethylation		21q21.3	61.7	0.83	510	27261660	27262169		—
#15 (NT_002836.4 15834745-15836237)	Unmethylation		21q22.11	62.4	0.93	493	29179194	29179686	5 -UTR and CDS	Putative N6-DNA-methyltransferase (N6AMT1)
#16 (NT_002836.4 16246692-16248676)	Unmethylation		21q22.11	76.5	0.90	985	29591133	29592117	5 -UTR	BTB and CNC homology 1, basic leucine zipper transcription factor 1 (BACH1)
#17 (NT_002836.4 18506100-18509038)	Incomplete methylation		21q22.11	71.7	0.94	1939	31850340	31852278	5 -UTR	Similarity to T-lymphoma invasion and metastasis-inducing protein 1 (TIAM1)
#18 (NT_002836.4 18607855-18609856)	Unmethylation		21q22.11	69.0	0.94	1002	31952095	31953096	5 -UTR and CDS	Superoxide dismutase (SOD-1)
#19 (NT_002836.4 18679791-18682057)	Unmethylation		21q22.11	74.3	0.90	1267	32024044	32025310		Similarity to CTD-binding SR-like protein rA4
#20 (NT_002836.4 18821198-18823391)	Unmethylation		21q22.11	72.7	0.87	1374	32165461	32166834	5 -UTR and CDS	Hormonally up-regulated Neu-associated kinase (HUNK)
#21 (NT_002836.4 19227134-19228707)	Unmethylation		21q22.11	65.1	0.85	574	32571397	32571970	5 -UTR and CDS	FAPP1-associated protein 1 (FASP1); chromosome 21 open reading frame 45 (C21orf45)
#22 (NT_002836.4 19248734-19250388)	Complete methylation	+	21q22.11	54.0	1.21	655	32592997	32593651		—
#23 (NT_002836.4 19359982-19362161)	Unmethylation		21q22.11	70.2	0.86	1180	32704245	32705424		—
#24 (NT_002836.4 19561116-19562518)	Unmethylation		21q22.11	62.0	0.82	403	32905379	32905781	5 -UTR	Chromosome 21 open reading frame 59 (C21orf59)
#25 (NT_002836.4 19675812-19677620)	Unmethylation		21q22.11	71.4	0.83	809	33020075	33020883	5 -UTR and CDS	Synaptojanin 1 (SYNJ1)

(continued)

Table 1. *Continued*

CpG islands[a]	Methylation	Repeat[b]	Locus	GC %[c]	Obs/exp[d]	Size (bp)	Nucleotide position[e] Start	End	Location in the linked gene[f]	CGI-linked genes
#26 (NT_002836.4 19719511-19721351)	Unmethylation		21q22.11	70.2	0.98	841	33063774	33064614	5 -UTR and CDS	Chromosome 21 open reading frame 66 (C21orf66); Putative transcription factor (ORF1)
#27 (NT_002836.4 19972011-19973417)	Unmethylation		21q22.11	62.9	0.83	407	33316261	33316667		—
#28 (NT_002836.4 19975198-19977216)	Unmethylation		21q22.11	68.6	0.90	1019	33319448	33320466	5 -UTR, CDS and 3 -UTR	Oligodendrocyte lineage transcription factor 2 (OLIG2)
#29 (NT_002836.4 20018303-20020533)	Unmethylation		21q22.11	73.9	0.84	1283	33362553	33363835	5 -UTR, CDS and 3 -UTR	Oligodendrocyte transcription factor 1 (Olig1)
#30 (NT_002836.4 20178164-20179883)	Unmethylation		21q22.11	72.0	0.99	720	33522413	33523132	5 -UTR	Interferon (alpha, beta and omega) receptor 2(IFNAR2)
#31 (NT_002836.4 20273122-20274794)	Unmethylation		21q22.11	64.7	0.93	673	33617371	33618043	5 -UTR and CDS	Interferon (alpha, beta and omega) receptor 1(IFNAR1)
#32 (NT_002836.4 20351550-20353202)	Unmethylation		21q22.11	79.1	0.87	653	33695798	33696450	5 -UTR and CDS	Interferon gamma receptor 2 (interferon gamma transducer 1) (IFNGR2)
#33 (NT_002836.4 20427755-20429718)	Unmethylation		21q22.11	73.0	0.93	964	33772003	33772966	5 -UTR	Chromosome 21 open reading frame 4 (C21orf4)
#34 (NT_002836.4 20536702-20538471)	Unmethylation		21q22.11	72.7	0.92	770	33880950	33881719	5 -UTR and CDS; 3 -UTR	Downstream neighbor of SON (DONSON); Crystallin, zeta (quinone reductase)-like 1 (CRYZL1)
#35 (NT_002836.4 20590829-20592722)	Unmethylation		21q22.12	75.8	0.88	894	33935077	33935970	5 -UTR	Intersectin 1 (SH3 domain protein) (ITSN1)
#36 (NT_002836.4 21021661-21023784)	Unmethylation		21q22.12	77.3	1.02	1124	34365908	34367031	5 -UTR and CDS	Mitochondrial ribosomal protein S6 (MRPS6)
#37 (NT_002836.4 21323771-21325344)	Unmethylation		21q22.12	68.1	0.84	574	34668018	34668591		—
#38 (NT_002836.4 21562903-21564940)	Unmethylation		21q22.12	75.3	0.95	1038	34907150	34908187	5 -UTR and CDS	Down syndrome critical region gene 1 (DSCR1)

(continued)