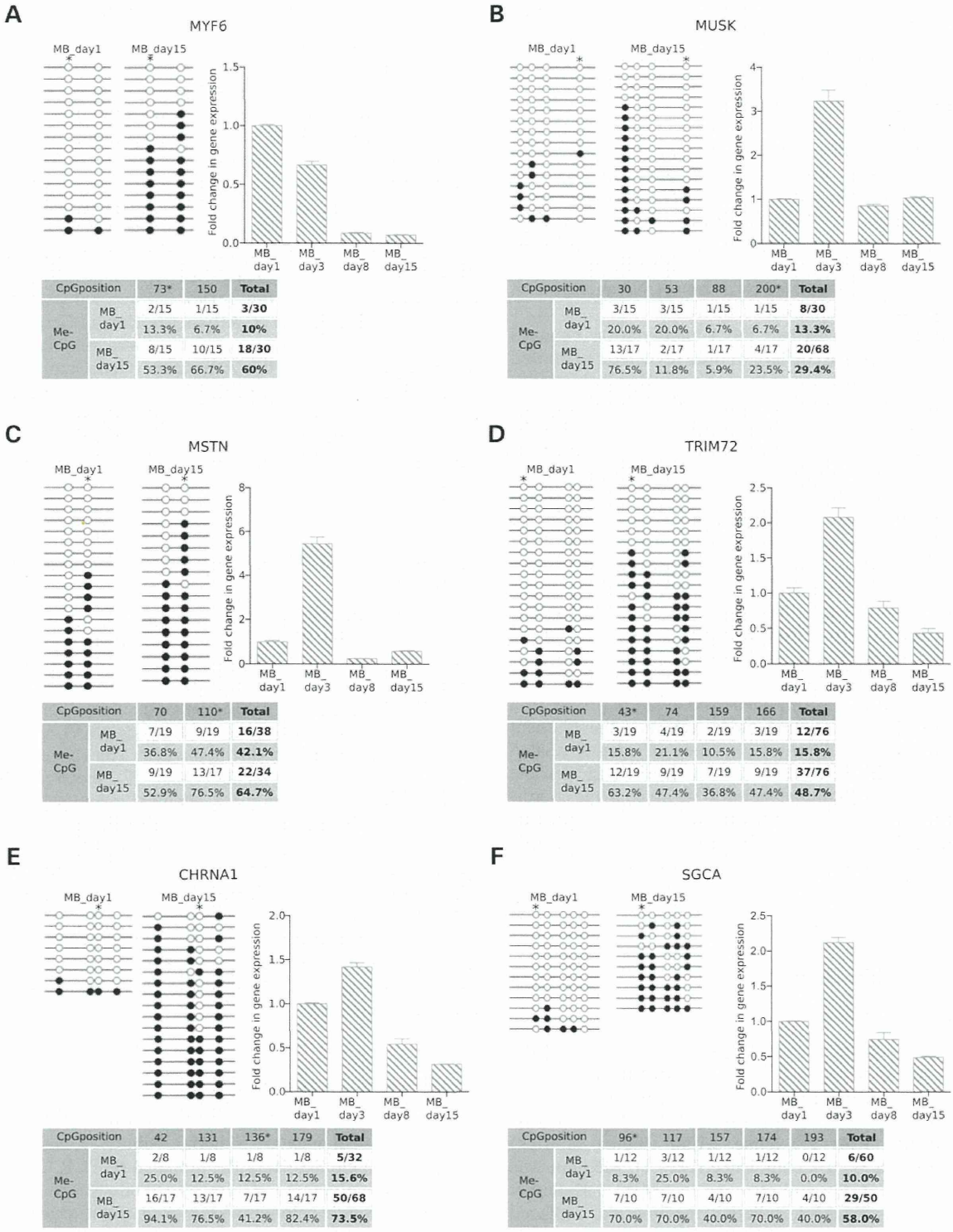

A	CpGposition	73*	150	Total
Me-CpG	MB_ day1	2/15	1/15	**3/30**
		13.3%	6.7%	**10%**
	MB_ day15	8/15	10/15	**18/30**
		53.3%	66.7%	**60%**

B	CpGposition	30	53	88	200*	Total
Me-CpG	MB_ day1	3/15	3/15	1/15	1/15	**8/30**
		20.0%	20.0%	6.7%	6.7%	**13.3%**
	MB_ day15	13/17	2/17	1/17	4/17	**20/68**
		76.5%	11.8%	5.9%	23.5%	**29.4%**

C	CpGposition	70	110*	Total
Me-CpG	MB_ day1	7/19	9/19	**16/38**
		36.8%	47.4%	**42.1%**
	MB_ day15	9/19	13/17	**22/34**
		52.9%	76.5%	**64.7%**

D	CpGposition	43*	74	159	166	Total
Me-CpG	MB_ day1	3/19	4/19	2/19	3/19	**12/76**
		15.8%	21.1%	10.5%	15.8%	**15.8%**
	MB_ day15	12/19	9/19	7/19	9/19	**37/76**
		63.2%	47.4%	36.8%	47.4%	**48.7%**

E	CpGposition	42	131	136*	179	Total
Me-CpG	MB_ day1	2/8	1/8	1/8	1/8	**5/32**
		25.0%	12.5%	12.5%	12.5%	**15.6%**
	MB_ day15	16/17	13/17	7/17	14/17	**50/68**
		94.1%	76.5%	41.2%	82.4%	**73.5%**

F	CpGposition	96*	117	157	174	193	Total
Me-CpG	MB_ day1	1/12	3/12	1/12	1/12	0/12	**6/60**
		8.3%	25.0%	8.3%	8.3%	0.0%	**10.0%**
	MB_ day15	7/10	7/10	4/10	7/10	4/10	**29/50**
		70.0%	70.0%	40.0%	70.0%	40.0%	**58.0%**

Figure 8. Bisulfite sequencing and quantitative RT–PCR analyses to validate alterations of DNA methylation and gene expression levels for six muscle-related genes, *MYF6* (**A**), *MUSK* (**B**), *MSTN* (**C**), *TRIM72* (**D**), *CHRNA1* (**E**) and *SGCA* (**F**), during myogenic differentiation. Open and closed circles represent unmethylated and methylated CpG sites, respectively. Asterisks indicate the position of the CpG site that was found to be hypermethylated in MB_day15 compared with MB_day1 in the analysis of Infinium HumanMethylation450 BeadChip data [the corresponding name (IlmnID) of the probe is listed in Supplementary Material, Table S12]. Each row of circles corresponds to an individual clone sequenced. For each gene, the methylation ratios (%) of the CpG sites examined are shown in the table underneath. Gene expression levels in MB_day3, MB_day8 and MB_day15 relative to MB_day1 were displayed as bar charts. Each bar represents the mean fold change and the standard error of technical replicates ($n = 3$).

ubiquitously expressed housekeeping genes. In contrast, the promoters of tissue-specific genes tend to be non-CGIs (28,29). In our integrated analysis of DNA methylation and gene expression data for MB_day1 and MB_day15, *de novo* DNA methylation of non-CGI promoters was more frequently accompanied with the suppression of gene expression than *de novo* DNA methylation of CGI promoters. Furthermore, GO analysis for the 27 genes with non-CGI promoters that were down-regulated (expression fold change ≤ 0.5) and hypermethylated ($\Delta\beta \geq 0.2$) upon myogenic differentiation revealed significant enrichment ($P = 8.567E-06$) for muscle-related genes. These findings support the hypothesis that muscle-related genes, especially those with non-CGI promoter, are regulated by *de novo* DNA methylation during myogenic differentiation.

In this study, we described DNA methylation changes occurring during myoblast differentiation revealed by detailed and comprehensive analyses of the Infinium HumanMethylation 450 BeadChip array data. This method uses bisulfite-converted DNA and does not distinguish 5-methylcytosine and 5-hydroxymethylcytosine (5 hmC), the latter of which is generated by oxidation of the former by ten eleven translocation (TET) enzymes. It has been suggested that the TET-medicated 5 hmC not only triggers DNA demethylation but also participates in epigenetic regulation *per se* in essential biological processes, such as cell pluripotency, embryonic development and cellular differentiation (30,31). Examination of genome-wide distribution and alteration of 5 hmC during myoblast differentiation is a promising future study to broaden our understanding of epigenetic mechanisms regulating myoblast differentiation.

In summary, analysis of the DNA methylation profile in a human myoblast differentiation model revealed the frequent occurrence of *de novo* DNA methylation of muscle-related genes with non-CGI promoter. Furthermore, it suggested the existence of key transcriptional factor(s) involved in the coordinated regulation of such muscle-related genes during myogenic differentiation. Our findings provide a basis to better understand the epigenetic principles that control myogenesis.

MATERIALS AND METHODS

Cell culture and samples

Human primary skeletal muscle myoblasts (Lonza, CC-2580) were cultured in SkGM-2 BulletKit medium (Lonza, CC-3160). Differentiation into myotubes was stimulated by switching the culture medium to DMEM-F12 (Thermo scientific) supplemented with 2% horse serum (Sigma–Aldrich) when cells reached ~50% confluence (Day 1). Medium was replaced every other day and changed to SkGM-2 BulletKit medium at Day 8. SkGM-2 BulletKit medium was replaced every other day until Day 15. Cells were collected on Days 1, 3, 8 and 15 for subsequent analyses. Human normal skeletal muscle tissues from two individuals were obtained from Tissue Solutions (SK052009, SKm1; SKM102210A, SKm2). hMSCs were grown in DMEM (Thermo Scientific) supplemented with 10% fetal bovine serum (Thermo Scientific). All cell lines and human tissues were derived from male individuals.

RNA isolation, RT–PCR and expression array analysis

Total RNA was isolated from myoblasts (Days 1, 3, 8 and 15, $n = 1$ for each time point) using ISOGEN (Nippon Gene) and reverse-transcribed using SuperScript Reverse Transcriptase II (Invitrogen). Quantitative real-time PCR was performed with SYBR Green PCR Master Mix (Applied Biosystems). Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) was used as an internal control for mRNA expression. Gene expression changes were quantified using the delta–delta CT method. Primer sequences for real-time PCR are listed in Supplementary Material, Table S11. For gene expression array analysis, 200 ng of RNA isolated from myoblasts (Day 1, 3, 8, or 15) was labeled with biotin using GeneChip 3′ IVT Express Kit (Affymetrix). Samples were hybridized on Affymetrix Human Genome U133 Plus 2.0 arrays. Expression array data were summarized with MAS5 algorithm using the Affymetrix GeneChip Command Console Software. Probes recognized as 'Absent' in myoblasts on Day 1, 3, 8 or 15 were excluded. Signal intensities of multiple probes for one gene were averaged, and fold changes in gene expression were calculated. Statistical analysis was performed using Mann–Whitney *U*-test.

DNA isolation and Illumina HumanMethylateion450 BeadChip

Myoblasts, hMSCs and human SKm_1 and SKm_2 tissues were harvested and eluted in 0.1% sodium dodecyl sulfate/Tris–EDTA buffer (pH 8.0). Cell lysates were incubated with Proteinase K (Sigma–Aldrich) and RNase A (Invitrogen) at 50°C for 2 h. Phenol/chloroform extraction and ethanol precipitation were performed to purify genomic DNA (gDNA), and precipitates were dissolved with distilled water. The concentration of gDNA was measured with the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies). Bisulfite reactions were performed using the Epitect Plus DNA Bisulfite Kit (QIAGEN) with an input of 1.5 μg of gDNA. After bisulfite conversion, 300 ng of each sample was whole-genome-amplified, enzymatically fragmented and hybridized to the Illumina HumanMethylateion450 BeadChip array, which contains probes to determine the DNA methylation levels of >480 000 CpG sites in a quantitative manner. After hybridization, the BeadChip array was processed for the single-base extension reaction, stained and imaged on an Illumina iScan.

Cloned-based bisulfite sequencing

The sequences of the primers used for bisulfite-PCR reactions are listed in Supplementary Material, Table S12. Bisulfite-PCR products were cloned using the StrataClone PCR Cloning Kit (Agilent Technologies) and transformed into StrataClone Competent Cells. Single colonies were picked up and used as starting material to amplify plasmid DNA within them using the TempliPhi DNA Amplification Kit (GE Healthcare). Sequencing reactions for individual amplified clones were conducted using the BigDye Terminator version 3.1 Cycle Sequencing kit (Applied Biosystems) with the M13 Rev primer. Sequence data were obtained using ABI3130xl Genetic Analyzer and analyzed using the QUMA website (http://quma.cdb.riken.jp/).

DNA methylation data analysis

The methylation level of each of the >480 000 CpG sites was calculated using the GenomeStudio Methylation Module Software ver. 1.0 as methylation β-value ranging from 0 if completely unmethylated to 1 if completely methylated [β-value = intensity of the methylated allele/(intensity of the unmethylated allele + intensity of the methylated allele + 100)]. We excluded probes with a detection P-value of >0.05 or blank β-value from further analyses. Differences in the β-values (delta-beta, $\Delta\beta$) between target and control samples were interpreted as follows: $\Delta\beta \geq 0.2$ and $\Delta\beta \leq -0.2$ were regarded as hyper- and hypo-methylated, respectively. This is based on the fact that a $\Delta\beta$ detection sensitivity of 0.2 (95% confidence level) was previously estimated for >90% of 27 000 infinium I assay probes (32). CpG sites were categorized into seven gene feature groups or into three groups according to the probe annotation provided by Illumina. When a single CpG site was assigned to multiple gene symbols or gene features, the corresponding β-value was used for multiple times. R packages were used for data processing, statistical analysis and graphic visualization.

CGI annotation for RefSeq transcriptional start sites

Chromosomal and nucleotide positions of RefSeq transcriptional start sites were retrieved from http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz. For each of the RefSeq transcripts, 100-bp DNA sequences flanking both sides of the TSS were compiled, and the resultant 201-bp sequence was assessed whether it fulfills CGI criteria of a GC content of ≥50% and an observed-to-expected ratio of CpG of ≥0.6.

GO analysis and motif discovery

Genomic regions identified as hyper- or hypo-methylated regions were tested for enrichment of GO terms using the Genomic Regions Enrichment of Annotations Tool (GREAT, http://bejerano.stanford.edu/great/public/html/) version 2.0.2 (19). Parameters were set to 'Single nearest gene, within 1000.0kb'. Binomial P-values were displayed as bar charts. Gene ontology term analyses using gene symbols were performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID; http://david.abcc.ncifcrf.gov) (33). As a cut-off for functional categories, we chose a P-value of 0.05. The motif analysis was conducted using the MEME algorithm (http://meme.nbcr.net/meme/) on the super-computing resource provided by Human Genome Center, Institute of Medical Science, University of Tokyo. Flanking sequences (from −50 to +50 bp) of hyper- or hypo-methylated CpG sites were obtained from the reference human genome sequence (hg19) using custom scripts and used as input sequences in multi-FASTA format. Parameters used were 'zero or one per sequence' for distributed, 'minimum 6 and maximum 15' for the width of the motif, and '10' for the maximum number of motifs to find (34). Motifs identified by MEME were subsequently subjected to a motif comparison tool, Tomtom (http://meme.nbcr.net/meme/cgi-bin/tomtom.cgi), for comparison with known motifs in databases. Significant thresholds were set to 'E-value < 10' and 'Human and Mouse (Jolma2013)' was selected as a database in the Tomtom analysis (35,36).

Data deposition

The data used in publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE55571 for DNA methylation array and GSE55034 for gene expression array.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Hiromi Kamura for her assistance to obtain HumanMethylation450 BeadChip data.

Conflict of Interest statement. None declared.

FUNDING

This work was supported, in part, by The Grant of National Center for Child Health and Development, (#24-3 to K.N. and #25-1 to H.A.); JSPS KAKENHI (Grant Numbers 23249071, 24115707 and 24659669 to H.A.); Japan Science and Technology Agency (Core Research for Evolutional Science and Technology to H.A.); Takeda Science Foundation, Bristol-Myers RA Research Fund (to H.A.); Mochida Memorial Foundation for Medical and Pharmaceutical Research (to H.A.); National Institute of Biomedical Innovation (to H.A.) and National Institutes of Health (AR050631 to H.A.).

REFERENCES

1. Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, **9**, 465–476.
2. Ghosh, S., Yates, A.J., Frühwald, M.C., Miecznikowski, J.C., Plass, C. and Smiraglia, D. (2010) Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. *Epigenetics*, **5**, 527–538.
3. Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R. and Bird, A.P. (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.*, **6**, e1001134.
4. Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209–213.
5. Li, E., Bestorm, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.
6. Okano, M., Bellm, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
7. Chen, J.C. and Goldhamer, D.J. (2003) Skeletal muscle stem cells. *Reprod. Biol. Endocrinol.*, **1**, 101.
8. Péault, B., Rudnicki, M., Torrente, Y., Cossu, G., Tremblay, J.P., Partridge, T., Gussoni, E., Kunkel, L.M. and Huard, J. (2007) Stem and progenitor cells in skeletal muscle development, maintenance, and therapy. *Mol. Ther.*, **15**, 867–877.
9. Dilworth, F.J. and Blais, A. (2011) Epigenetic regulation of satellite cell activation during muscle regeneration. *Stem Cell Res. Ther.*, **2**, 18.
10. Taylor, S.M. and Jones, P.A. (1979) Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine. *Cell*, **17**, 771–779.

11. Davis, R.L., Weintraub, H. and Lassar, A.B. (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, **51**, 987–1000.
12. Palacios, D., Summerbell, D., Rigby, P.W. and Boyes, J. (2010) Interplay between DNA methylation and transcription factor availability: implications for developmental activation of the mouse Myogenin gene. *Mol. Cell Biol.*, **30**, 3805–3815.
13. Hupkes, M., Jonsson, M.K., Scheenen, W.J., van Rotterdam, W., Sotoca, A.M., van Someren, E.P., van der Heyden, M.A., van Veen, T.A., van Ravestein-van Os, R.I., Bauerschmidt, S. *et al.* (2011) Epigenetics: DNA demethylation promotes skeletal myotube maturation. *FASEB J.*, **25**, 3861–3872.
14. Tsumagari, K., Baribault, C., Terragni, J., Varley, K.E., Gertz, J., Pradhan, S., Badoo, M., Crain, C.M., Song, L., Crawford, G.E. *et al.* (2013) Early de novo DNA methylation and prolonged demethylation in the muscle lineage. *Epigenetics*, **8**, 317–332.
15. Bird, A. (1992) The essentials of DNA methylation. *Cell*, **70**, 5–8.
16. Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M. and Schübeler, D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.*, **39**, 457–466.
17. Mofarrahi, M. and Hussain, S.N. (2011) Expression and functional roles of angiopoietin-2 in skeletal muscles. *PLoS One*, **6**, e22882.
18. Deaton, A.M. and Bird, A. (2011) CpG islands and the regulation of transcription. *Genes Dev.*, **25**, 1010–1022.
19. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M. and Bejerano, G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
20. Jen, Y., Weintraub, H. and Benezra, R. (1992) Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. *Genes Dev.*, **6**, 1466–1479.
21. Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L. and Weintraub, H. (1990) The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell*, **61**, 49–59.
22. Langlands, K., Yin, X., Anand, G. and Prochownik, E.V. (1997) Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. *J. Biol. Chem.*, **272**, 19785–19793.
23. Yokoyama, S., Ito, Y., Ueno-Kudoh, H., Shimizu, H., Uchibe, K., Albini, S., Mitsuoka, K., Miyaki, S., Kiso, M., Nagai, A. *et al.* (2009) A systems approach reveals that the myogenesis genome network is regulated by the transcriptional repressor RP58. *Dev. Cell*, **17**, 836–848.
24. Hsieh, C.L. (1994) Dependence of transcriptional repression on CpG methylation density. *Mol. Cell Biol.*, **14**, 5487–5494.
25. Fuks, F., Burgers, W.A., Godin, N., Kasai, M. and Kouzarides, T. (2001) Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J.*, **20**, 2536–2544.
26. Mastroyiannopoulos, N.P., Nicolaou, P., Anayasa, M., Uney, J.B. and Phylactou, L.A. (2012) Down-regulation of myogenin can reverse terminal muscle cell differentiation. *PLoS One*, **7**, e29896.
27. Oikawa, Y., Omori, R., Nishii, T., Ishida, Y., Kawaichi, M. and Matsuda, E. (2011) The methyl-CpG-binding protein CIBZ suppresses myogenic differentiation by directly inhibiting myogenin expression. *Cell Res.*, **21**, 1578–1590.
28. Vinson, C. and Chatterjee, R. (2012) CG methylation. *Epigenomics*, **4**, 655–663.
29. Slieker, R.C., Bos, S.D., Goeman, J.J., Bovée, J.V., Talens, R.P., van der Breggen, R., Suchiman, H.E., Lameijer, E.W., Putter, H., van den Akker, E.B. *et al.* (2013) Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin*, **6**, 26.
30. Delatte, B. and Fuks, F. (2013) TET proteins: on the frenetic hunt for new cytosine modifications. *Brief Funct. Genomics.*, **12**, 191–204.
31. Delatte, B., Deplus, R. and Fuks, F. (2014) Playing TETris with DNA modifications. *EMBO J.*, **33**, 1198–1211.
32. Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R. and Gunderson, K.L. (2009) Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics*, **1**, 177–200.
33. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
34. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
35. Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G. *et al.* (2013) DNA-binding specificities of human transcription factors. *Cell*, **152**, 327–339.
36. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. and Noble, W.S. (2007) Quantifying similarity between motifs. *Genome Biol.*, **8**, R24.

RESEARCH PAPER

Epigenetics 9:5, 1–8; May 2014; © 2014 Landes Bioscience

Variable maternal methylation overlapping the *nc886/vtRNA2-1* locus is locked between hypermethylated repeats and is frequently altered in cancer

Valeria Romanelli[1], Kazuhiko Nakabayashi[2], Miguel Vizoso[3], Sebastian Moran[3], Isabel Iglesias-Platas[4], Naoko Sugahara[2], Carlos Simón[5], Kenichiro Hata[2], Manel Esteller[3,6,7], Franck Court[1], and David Monk[1,*]

[1]Imprinting and Cancer Group; Cancer Epigenetic and Biology Program; Institut d'Investigació Biomedica de Bellvitge; Hospital Duran i Reynals; Barcelona, Spain; [2]Department of Maternal-Fetal Biology and Department of Molecular Endocrinology; National Research Institute for Child Health and Development; Tokyo, Japan; [3]Cancer Epigenetics Group; Cancer Epigenetic and Biology Program; Institut d'Investigació Biomedica de Bellvitge; Hospital Duran i Reynals; Barcelona, Spain; [4]Servicio de Neonatología; Hospital Sant Joan de Déu; Fundació Sant Joan de Déu; Barcelona, Spain; [5]Fundación IVI; Instituto Universitario IVI; Universidad de Valencia; INCLIVA; Valencia, Spain; [6]Department of Physiological Sciences II; School of Medicine; University of Barcelona; Barcelona, Spain; [7]Institucio Catalana de Recerca i Estudis Avançats (ICREA); Barcelona, Spain

Keywords: imprinting, DNA methylation, miRNAs, vault RNAs, vtRNA2-1, nc886

Abbreviations: Nc, non-coding; vtRNA, vault complex RNA; miRNA, microRNA; pri-miRNA, precursor microRNA; dsRNA, double stranded RNA; CRC, colorectal carcinomas; LOH, loss-of-heterozygosity; UPD, uniparental disomy; DMR, differentially methylated region; ICM, inner cell mass; hES, human embryonic stem cells; pa-oocytes, parthenogenetically-activated metaphase II oocytes; AML, acute myeloid leukemia

Cancer is as much an epigenetic disease as a genetic one; however, the interplay between these two processes is unclear. Recently, it has been shown that a large proportion of DNA methylation variability can be explained by allele-specific methylation (ASM), either at classical imprinted loci or those regulated by underlying genetic variants. During a recent screen for imprinted differentially methylated regions, we identified the genomic interval overlapping the non-coding *nc886* RNA (previously known as *vtRNA2-1*) as an atypical ASM that shows variable levels of methylation, predominantly on the maternal allele in many tissues. Here we show that the *nc886* interval is the first example of a polymorphic imprinted DMR in humans. Further analysis of the region suggests that the interval subjected to ASM is approximately 2 kb in size and somatically acquired. An in depth analysis of this region in primary cancer samples with matching normal adjacent tissue from the Cancer Genome Atlas revealed that aberrant methylation in bladder, breast, colon and lung tumors occurred in approximately 27% of cases. Hypermethylation occurred more frequently than hypomethylation. Using additional normal-tumor paired samples we show that on rare occasions the aberrant methylation profile is due to loss-of-heterozygosity. This work therefore suggests that the *nc886* locus is subject to variable allelic methylation that undergoes cancer-associated epigenetic changes in solid tumors.

Introduction

DNA methylation is assumed to be complementary on both alleles, except in the cases of parent-of-origin-dependent DNA methylation, which regulates imprinted gene clusters and the random silencing of one of the two X-chromosomes in females to ensure dosage compensation with males. A third type of allelic specific methylation (ASM) has recently been reported, which is associated with DNA sequences in *cis* (ie, single nucleotide polymorphisms, SNPs), and is termed methylation quantitative trait loci (mQTL).[1] This form of allele-specific methylation has been reported to show tissue and population specificity, since it is dictated by underlying DNA sequence and acquired somatically.[2] In contrast, ASM associated with imprinted differentially methylated regions (DMRs) are *cis*-acting regulatory elements that influence the allelic expression of neighboring genes in a parent-of-origin manner. Most imprinted DMRs acquire their differential methylation status during gametogenesis, when the two parental genomes are separated, resulting from the cooperation of the de novo methyltransferase DNMT3A and its co-factor DNMT3L.[3,4] These primary, or germline, imprinted DMRs are stably maintained throughout somatic development by the

*Correspondence to: David Monk; Email: dmonk@iconcologia.net
Submitted: 01/09/2014; Revised: 02/18/2014; Accepted: 02/22/2014; Published Online: 03/03/2014
http://dx.doi.org/10.4161/epi.28323

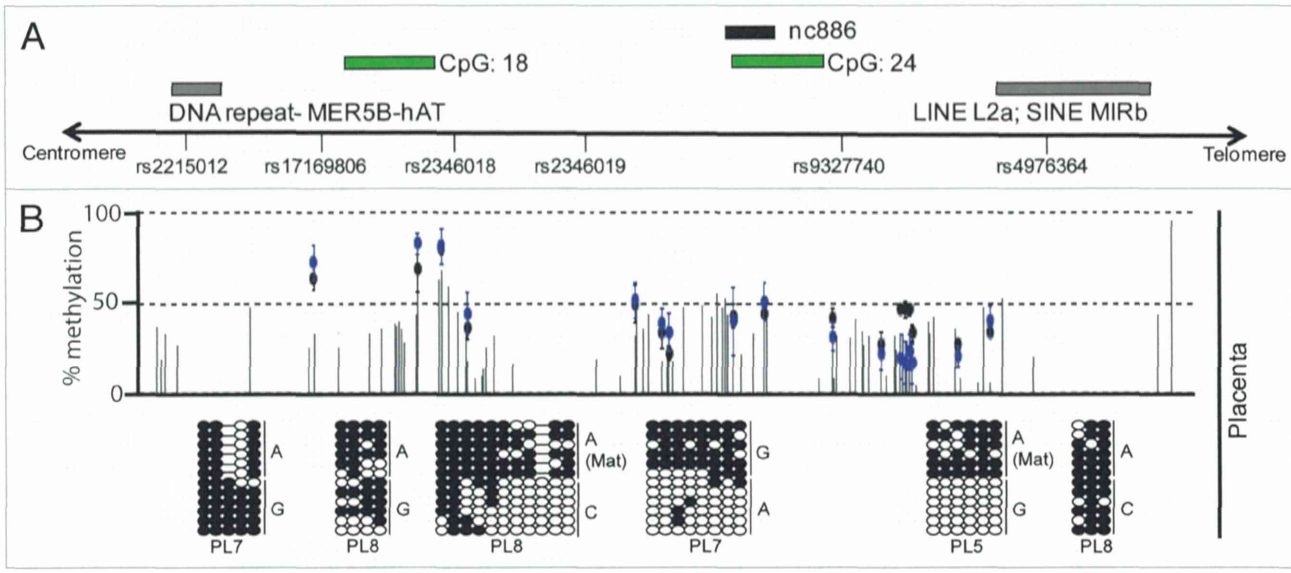

Figure 1A and B. Mapping the extent of allelic methylation at the *nc886* interval in 4 human tissues. (**A**) Map of the *nc886* locus, showing the location of the transcript, CpG islands, genetic variants and DNA repeat-elements. (**B**) Detailed methylation map of the *nc886* interval for placenta as determined by WGBS and Infinium HumanMethylation BeadChip array. Vertical dark gray lines in the WGBS tracks represent the methylation value for individual CpG dinucleotides and each dot representing the methylation of a single array probe: normal placenta (black dots) and hydatidiform mole (blue dots). The pattern of methylation was confirmed using bisulphite PCR in heterozygous samples so that allelic origin could be ascertained. Each circle represents a single CpG dinucleotide on a DNA strand, a methylated cytosine (●) or an unmethylated cytosine (○).

UHRF1-DNMT1 complex,[5] surviving the epigenetic reprogramming at the oocyte-to-embryo transition.[6,7]

Both types of ASM can shed light on disease susceptibility, including multifactorial phenotypes and cancer. For example, a constitutional epimutation of the MutL homolog 1 (*MLH1*) tumor suppressor gene in hereditary nonpolyposis colon cancer is associated with a specific polymorphic variant that modifies gene expression,[8] whereas most tumor types present with frequent loss-of-imprinting (For a review see ref. 9).

During a genome-wide screen for regions of ~50% methylation in multiple tissues that could manifest as novel imprinted DMRs, we identified a region of partial methylation overlapping the non-coding RNA (ncRNA) *nc886* region located on chromosome 5q31. The 102 bp *nc886* transcript is a novel class of RNA. It was originally proposed to be a precursor microRNA (pri-miRNA) or a RNA component of the vault complex RNA (vtRNAs); however, based on the size of the RNA and its dicer-independent processing, it has been shown to be neither.[10]

The *nc886* transcript was recently shown to be a tumor-suppressing ncRNA and key for tumor surveillance through its ability to bind and repress the double-stranded RNA (dsRNA)-dependent kinase (PKR).[10,11] Tumor-associated repression of *nc886* activates PKR, the phosphorylated form being the kinase that phosphorylates EIF2α, which in turn inhibits global protein synthesis and ultimately induces cell death. At least in acute myeloid leukemia (AML), hypermethylation of the *nc886* promoter results in abolished expression, causing the inhibitory effect to disappear.[12] Stress/chemotherapy-induced PKR-phosphorylation then switches to favor the activation of NFK-β and cancer cell survival at the expense of EIF2α-induced apoptosis.[13]

Here, we have characterized the methylation profile of the *nc886* region in numerous normal tissues and observed variable methylation profile in which 76% of the normal samples analyzed were predominantly methylated on one allele and, when informative, this was the maternal chromosome. While this manuscript was in preparation, the region overlapping *nc886* was shown to be polymorphic maternally methylated region in 83% of tissues, corroborating our observations.[14] To ascertain whether this region is associated with aberrant methylation in solid tumors and could potentially be a marker for disease, we profiled tumor and normal paired samples from the Cancer Genome Atlas. Abnormal methylation within this interval occurred in 25% of bladder, 38% of breast, 21.5% of colorectal and 32% of lung primary tumor samples. Using additional paired samples sets, we showed that, occasionally, the aberrant epigenetic profile was explained by cancer-associated loss-of-heterozygosity (LOH), suggesting that deletion of 5q31 also occurs in solid tumors.

Results

Methylation profiling of the *nc886* region by WGBS and Infinium Methylation Arrays

We have previously used whole genome bisulphite sequencing (WGBS) and the Illumina Infinium HumanMethylation450 BeadChip array to identify novel human imprinted loci.[15] During this screen, we identified a region overlapping the *nc886* transcript that was consistently partially methylated in DNA samples derived from leukocytes, liver, brain and placenta tissues. Twenty probes present on the Infinium array platform confirm

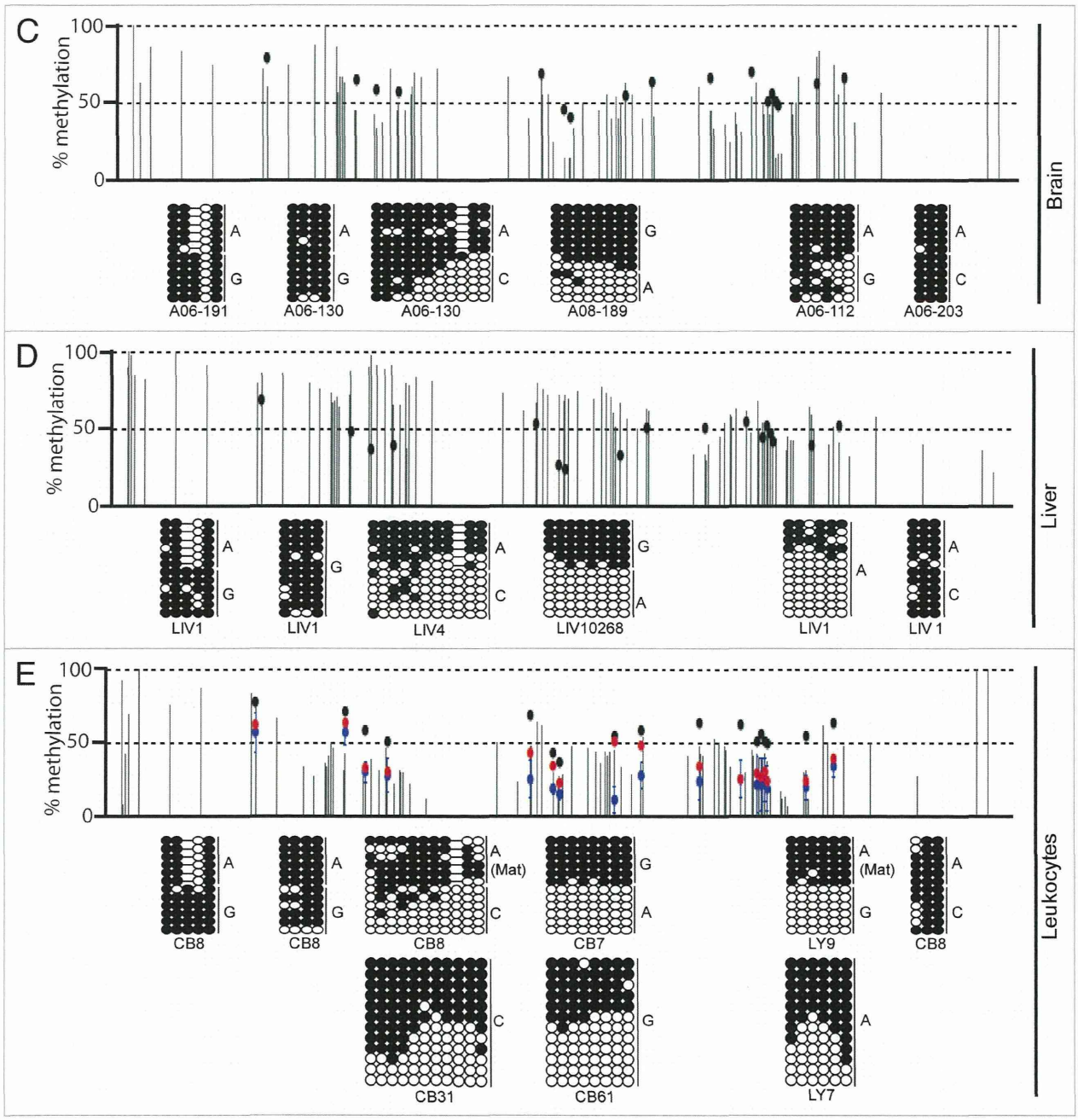

Figure 1C–E. Mapping the extent of allelic methylation at the *nc886* interval in 4 human tissues. The methylation profiles for brain (**C**), liver (**D**) and leukocytes (**E**) also reveal partial methylation over the *nc886* transcript. The lower panel also shows the DNA methylation profile for normal leukocytes (black), genome-wide pUPD (blue) and mUPD (red) samples as determined by the Infinium methylation array.

this partially methylated profile (**Fig. 1**) in tissues and immune-selected blood lineages (**Fig. S1**). Despite appearing as a promising candidate for a novel imprinted DMR, this region did not show the correct parent-of-origin methylation profile in reciprocal genome-wide uniparental disomy (UPD) samples or hydatidiform moles (containing only paternal chromosomes), which would be expected for a bona fide imprinted DMR (**Fig. 1**).

Allelic methylation of the *nc886* promoter

Despite the contradictory evidence for parent-of-origin methylation within the *nc886* locus, we decided to pursue the mechanism leading to the partially methylated profile. We utilized bisulphite PCR and subsequent sequencing of sub-cloned DNA strands incorporating the SNP rs9327740, which is located ~200 bp upstream of CpG island 24 in the promoter of *nc886*. This semi-quantitative technique was used rather than quantitative pyrosequencing, as the latter does not allow for allelic

discrimination or assessment of PCR clonality. We observed that in 87% of heterozygous samples (n = 7), irrespective of tissue type analyzed, the A allele was more methylated (**Fig. 1**; **Table S1**). In informative samples, the methylation was always on the maternal chromosome (**Fig. 2**). Using this genetic variant to assay for allelic methylation, we could not distinguish between polymorphic imprinted and SNP-associated methylation. To exclude that the methylation pattern we observe was directly due to this SNP variant, we analyzed samples that were homozygous for rs9327740. Most homozygous tissue samples (9/16) had both methylated and unmethylated strands consistent with imprinting (**Fig. 1**).

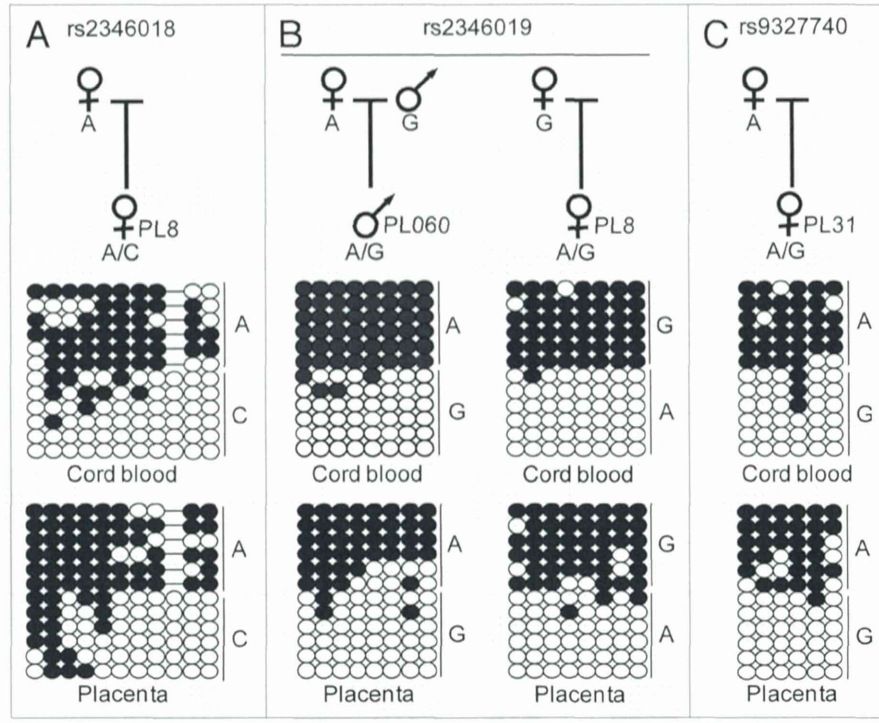

Figure 2. Confirmation that *nc886* is a maternally methylated DMR. Examples of bisulphite PCRs confirming the presence of methylation on the maternal allele in cord blood and placental-derived DNA samples heterozygous for the SNPs (**A**) rs2346018, (**B**) rs2346019 and (**C**) rs9327740.

Variable maternal methylation across the interval *nc886*

To determine the extent of the allelic methylation we performed a similar bisulphite PCR strategy for 5 additional regions that contained highly informative SNPs. The genomic interval overlapping the rs2346019 polymorphism also presented with allelic methylation in 83% of samples, with the G allele being methylated in 9/10 heterozygous samples (**Table S1**). When informative, the maternal chromosome was more methylated (**Figs. 1 and 2**). For only one placenta and cord blood pair (sample set 060) the A allele was inherited maternally and was methylated, confirming that the methylation in this domain is inherited in a parent-of-origin manner and not dictated by SNP genotype (**Fig. 2**). The next PCR amplicon targeted SNP rs2346018 near CpG island 18 and presented with a gradient methylation profile. The maternal allele was heavily methylated, whereas the paternal allele gained methylation in the centromeric direction in 63% of samples, confirming the WGBS data that indicated this was the boundary of the allelic methylation. To determine the extent of the ASM we performed allelic bisulphite PCR on the regions surrounding rs2215012 and rs4976364. These SNPs are located within MER5B-hAT sequences and a LINE L2a-SINE MIRb element, respectively. Both of these repeats were hypermethylated on both chromosomes in all analyzed somatic tissues (**Fig. 1**; **Table S1**). These observations are concordant with the discrete boundaries of the DMR identified by Paliwal et al. using high throughput bar-code bisulfite PCR Fluidigm AccessArrays.[14]

ASM associated with *nc886* is somatically acquired

To address whether the DNA methylation within the *nc886* domain is germline- or somatically-acquired, we performed bisulphite analysis on DNA extracted from sperm, stem cells derived from parthenogenetically-activated metaphase-2 oocyte blastocysts (phES)[16,17] and human embryonic stem (hES) cell lines generated from both seven-cell blastomeres and the inner cell mass (ICM) of blastocysts. We have previously shown that low passage phES cells are good surrogates for primary oocytes since they maintain the correct methylation profile at imprinted loci.[15]

The bisulphite analysis of a sperm sample heterozygous for rs2346018 revealed that spermatozoa with either chromosome 5 homolog were completely unmethylated. This interval was also devoid of methylation in the phES and hES cells generated from seven-cell blastomere, but allelic in hES cells derived from the ICM. This strongly suggests that the methylation at *nc886* is a post-fertilization event, occurring early in development at a time point between post-fertilization days 4–6 (**Fig. 3**).

Tumor samples show aberrant methylation changes at *nc886* with occasional loss-of-heterozygosity

Recent studies have shown that *nc886* has tumor suppressing function, with hypermethylation and concomitant decreased expression predicting outcome in AML patients.[12] To determine if the methylation of this interval has prognostic value in solid tumors, we determined the methylation index for a range of tumors and adjacent normal tissues using the Infinium HumanMethylation450 BeadChip array. This analysis revealed that 65, 71, 66.5 and 56.5% of normal bladder, breast, colon and lung tissues, respectively, were ~50% methylated and that the corresponding tumors samples maintained this profile. In some instances, we observed samples with ~50% in the normal tissues but either hyper- or hypo-methylation in the matched tumor sample (**Fig. 4**; **Table S2**). We confirmed these observations

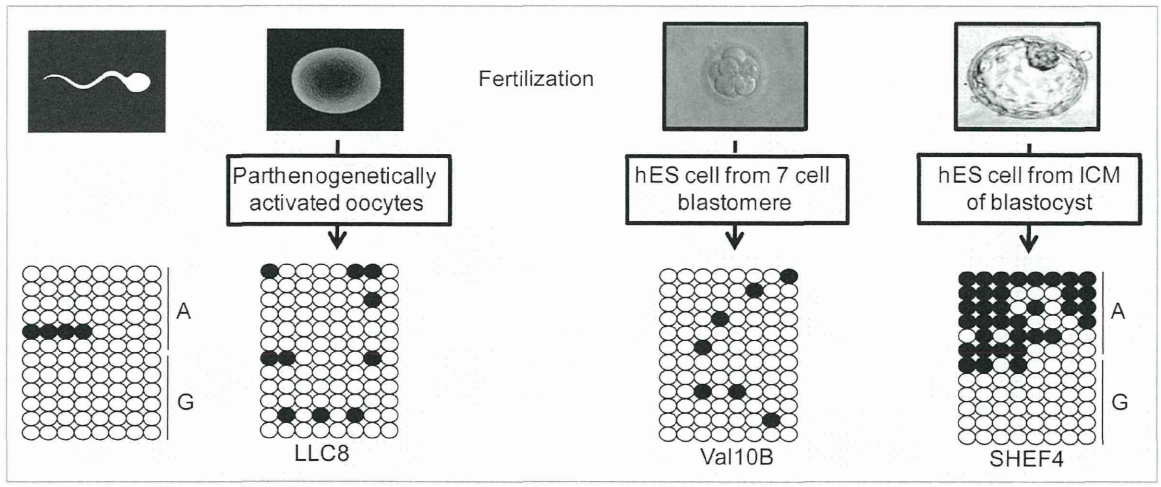

Figure 3. Timing of the acquisition of DNA methylation at *nc886*. The methylation profiles for the bisulphite PCR incorporating the SNP rs2346019 in DNA extracted from mature sperm, phES and hES cells derived from different aged pre-implantation embryos.

using bisulphite PCR and sub-cloning in samples for which we had DNA. Three colorectal sample sets were heterozygous for the rs2346019 variant. Interestingly, one clearly presented with hypermethylation on the normally unmethylated allele (colon 1184), whereas the remaining two sample pairs (colon 326 and 973) were associated with loss of the unmethylated allele, consistent with a cancer-associated heterozygous deletion (**Fig. 4**). In addition, we observed tumor-associated hypermethylation when the normal sample was constitutionally unmethylated. For the colon sample set 1404, three different tumor sections presented with various degrees of hypermethylation.

Discussion

Through the combined use of genome-wide methylation screening, we have identified a region of partial methylation overlapping the *nc886* region. Allelic methylation analysis using various SNPs throughout the interval revealed that the methylation is predominantly on one allele in ~73% of all samples, suggesting it is an atypical imprinted DMR. To date, all known human imprinted DMRs are stably maintained in somatic tissues, irrespective of gene expression levels, by the DNA methyltransferase activity of DNMT1 during semi-conservative DNA replication.[5] Polymorphic imprinting has been observed in human placental samples at the *IGF2R* locus.[18,19] In-depth studies trying to delineate the mechanism for polymorphic imprinting at this locus failed to identify an epigenetic signature that differentiates imprinted and biallelic expressing samples.[18] Both categories maintained a maternally methylated DMR within intron 2 of *IGF2R*, overlapping the promoter of the ncRNA *AIRN*. Therefore, our observations, substantiated by those noted by Paliwai et al.,[14] are the first evidence for a polymorphic imprinted DMR. Unfortunately, the ~102 bp *nc886* transcript does not contain any genetic variants that would allow us to directly determine whether the presence of maternal methylation within this interval directly influences allele-specific expression.

The *nc886* gene is located at chromosome 5q31. This region is frequently associated with LOH in many tumor types. Deletions of chromosome 5q are associated with poor outcomes in AML.[12] Heterozygous deletions encompassing 5q, identified by microsatellite polymorphisms and CGH array analyses, have also been identified in colorectal and breast tumors. While no significant association between LOH and recurrence or survival was observed in colorectal cancer,[20,21] LOH of this interval often occurred with *BRCA2* mutations in sporadic breast tumors.[22] Furthermore, 5q31 LOH was described as a hallmark for basal-like breast tumors and was associated with estrogen-receptor status.[23] Taken together, these observations suggest the presence of one or more tumor-suppressor genes at this locus. The *nc886* RNA is a confirmed tumor suppressor ncRNA that acts by downregulating the cancer survival factor pPKR.[12,24] Our results suggest that abolished expression of *nc886* can be achieved by either promoter hypermethylation or LOH, as seen in 25, 36, 16.5 and 29% of bladder, breast, colon and lung tumors, respectively. However, in addition to hypermethylation, ~5% tumors, respective of tissue of origin, presented with decreased methylation and, presumably, overexpression of the *nc886* transcript, suggesting that survival of some tumors is not governed by the *nc886*-pPKR pathway.

Conclusions

Our results strongly suggest that the ~2 kb region overlapping the *nc886* RNA locus somatically acquires DNA methylation of the maternal allele in the majority of individuals, with approximately 27% of the population not associated with ASM. In solid tumors, this region is subject to cancer-associated methylation changes as well as LOH. Further work is now required to determine if, as in AML, aberrant *nc886* expression disrupts the

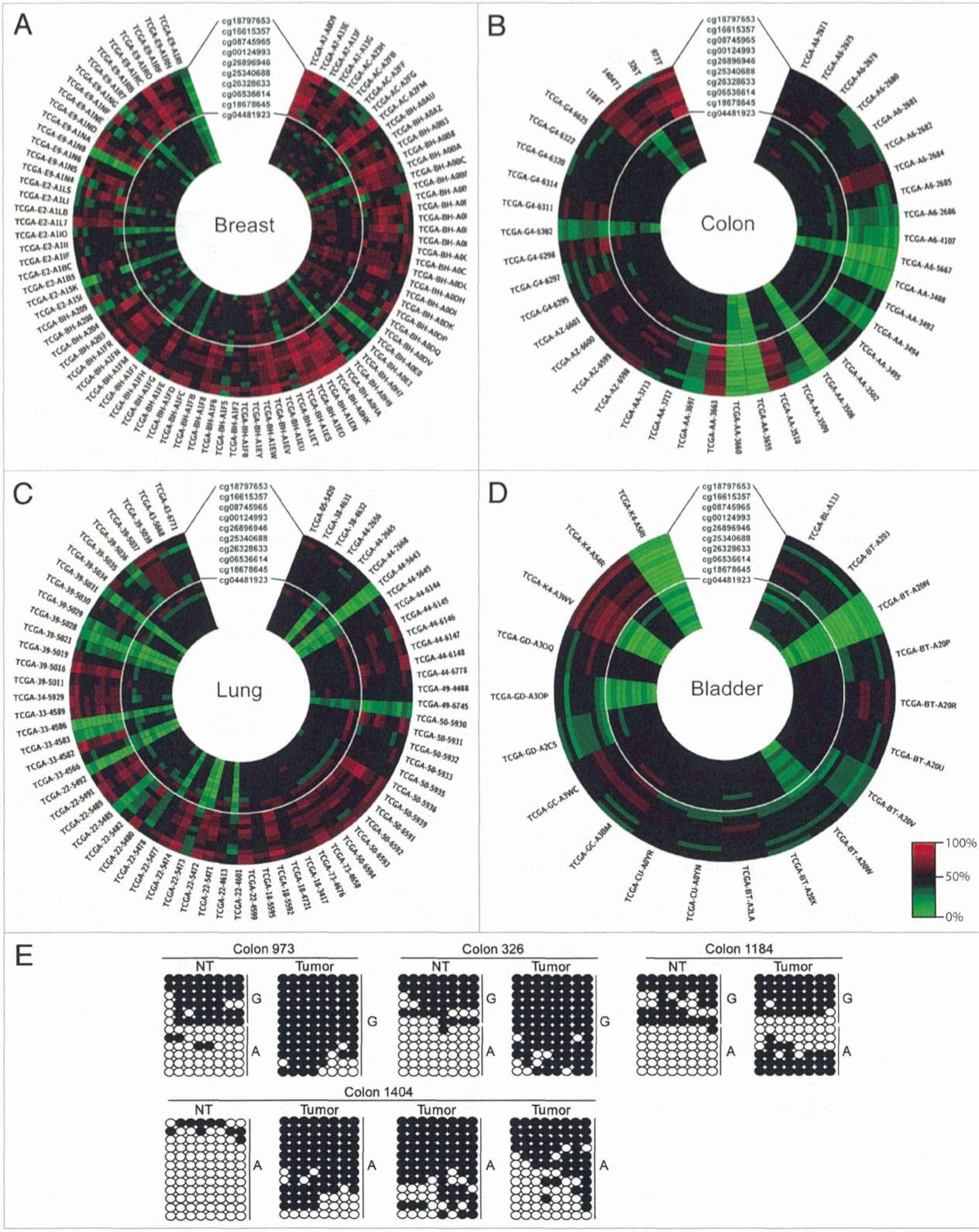

Figure 4. Profiling the methylation status of the *nc886* locus in human tumor samples. Circular heat map of Infinium array probes mapping to the *nc886* locus in (**A**) breast, (**B**) colorectal, (**C**) lung, and (**D**) bladder tumors. In each case the inner circle represents the methylation values of the normal tissue and the outer circle the corresponding tumor. (**E**) Confirmation of the methylation profile overlapping the rs2346019 interval in various CRC samples sets. Each circle represents a single CpG dinucleotide on a DNA strand, a methylated cytosine (●) or an unmethylated cytosine (○).