



OPEN

SUBJECT AREAS:
DNA METHYLATION
MICROARRAYSReceived
19 November 2013Accepted
17 February 2014Published
12 March 2014Correspondence and
requests for materials
should be addressed to
L.-C.L. (llai@ntu.edu.
tw) or E.Y.C.
(chuangey@ntu.edu.
tw)

Identification of Genes with Consistent Methylation Levels across Different Human Tissues

Tzu-Pin Lu^{1,2}, Kevin T. Chen², Mong-Hsun Tsai^{2,3}, Kuan-Ting Kuo⁴, Chuhsing Kate Hsiao^{2,5},
Liang-Chuan Lai^{2,6} & Eric Y. Chuang^{1,2,7}

¹YongLin Biomedical Engineering Center, National Taiwan University, Taipei, Taiwan, ²Bioinformatics and Biostatistics Core, Center of Genomic Medicine, National Taiwan University, Taipei, Taiwan, ³Institute of Biotechnology, National Taiwan University, Taipei, Taiwan, ⁴Department of Pathology, National Taiwan University Hospital, Taipei, Taiwan, ⁵Department of Public Health, National Taiwan University, Taipei, Taiwan, ⁶Graduate Institute of Physiology, National Taiwan University, Taipei, Taiwan, ⁷Graduate Institute of Biomedical Electronics and Bioinformatics and Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan.

DNA methylation plays an important role in regulating cell growth and disease development. Methylation profiles are examined by bisulfite conversion; however, the lack of markers for bisulfite conversion efficiency and appropriate internal control genes remains a major challenge. To address these issues, we utilized two bioinformatics approaches, coefficients of variances and resampling tests, to identify probes showing stable methylation levels from several independent microarray datasets. Mass spectrometry validated the consistently high methylation levels of the five probes (*N4BP2*, *EGFL8*, *CTRB1*, *TSPAN3*, and *ZNF690*) in 13 human tissue types from 24 cell lines. Linear associations between detected methylation levels and methyl concentrations of DNA samples were further demonstrated in three genes (*N4BP2*, *EGFL8*, and *CTRB1*). To summarize, we identified five genes which may serve as internal controls for methylation studies by analyzing large-scale microarray data, and three of them can be used as markers for evaluating the efficiency of bisulfite conversion.

In recent years, epigenetic changes have been extensively studied, and many studies have demonstrated their association with biological phenomena such as genomic imprinting, immune response regulation, and developmental programming^{1–4}. Epigenetics is the study of the connections between genotype and phenotype and one of its unique revelations is that gene expression patterns can be regulated without altering DNA sequences^{5,6}. Different types of epigenetic changes, such as DNA methylation, microRNA expression, and chromatin modification, have been reported as important players in many physiological functions^{6,7}. Among them, DNA methylation is the most studied mechanism and participates in the pathogenic processes of many diseases, such as cancers, neurodevelopmental disabilities, and allergic diseases^{1,8,9}. Thus, a growing body of research has been devoted to dissecting the methylation profiles in patients and trying to identify potential methylation biomarkers.

In the mammalian genome, DNA methylation usually occurs in a cytosine within a CpG dinucleotide and occasionally is found outside of CpG¹⁰. With the advancement in experimental technologies, several methods, including Illumina Infinium microarray and whole genome shotgun bisulfite sequencing, can be used to investigate genome-wide methylation profiles in tissue samples¹¹. An important feature of these methods is that most of them need to perform bisulfite conversions on DNA samples in order to distinguish methylated and unmethylated nucleotides. Bisulfite conversion transforms cytosine residues into uracil residues but leaves 5-methylcytosine residues unchanged, which allows researchers to quantify the methylation levels. Challenges arise, however, when trying to treat DNA samples with bisulfite. A critical question is how to determine whether input DNA samples are completely converted by bisulfite or not. Although Illumina methylation microarrays do have quality control probes for assessing the efficiency of bisulfite conversion, such information was usually not available in the public datasets. An arbitrary threshold between the intensity ratios of bisulfite-treated and untreated DNAs was used to indicate whether the bisulfite conversion was completed or not, which cannot fully and quantitatively reflect the level of bisulfite conversion. However, incomplete bisulfite conversions lead to overestimation of the methylation levels, since only a portion of cytosine is converted. Alternatively, over-treatment of bisulfite causes degradation of DNA samples and increases the probability of converting a methyl-



lated cytosine to a thymine¹¹. Consequently, identification of gene markers associated with the efficiency of bisulfite conversion may help to overcome this challenge.

High-throughput technologies, such as microarrays and next-generation sequencing, facilitate the identification of genes with altered methylation levels, and other experimental methods are usually performed to validate the results. For example, mass spectrometry has been widely used in methylation analyses^{12–14}. However, few studies have explored genes with consistent methylation levels across different samples. Similar to the concept of “housekeeping” genes showing consistent and stable gene expression levels¹⁵, appropriate internal controls for methylation studies can not only help to reduce the experimental bias from artificial effects, but also provide a better baseline to compare data from distinct biological samples. Therefore, we aimed to perform a large-scale analysis of methylation data in order to identify potential housekeeping genes with stable methylation levels across multiple human tissues.

In this study, we analyzed a total of 682 methylation microarrays generated from Illumina Infinium HumanMethylation27 BeadChips and used a bioinformatics approach to identify 27 genes showing consistent methylation levels across all samples. The top five genes were validated using mass spectrometry in 24 human cell lines, and a linear association between detected methylation levels and methyl concentrations of DNA samples was demonstrated in three genes, suggesting their potential role as markers for the efficiency of bisulfite conversion.

Results

Identification of consistently methylated probes. After the quality checks of samples and probes, a total of 668 samples (Table 1) containing 7,829 probes remained for further analyses. These 668 samples were comprised of more than 10 different cell types from 8 independent experimental batches and ethnicities. The following analysis procedures were all carried out using R version 2.9. For each gene, the coefficient of variance (CV) value and stability score¹⁶ were calculated to estimate the consistency of methylation levels among all samples, and the top 100 probes with the lowest CVs and highest stability scores were recorded as list A and A', respectively. To remove false-positive probes identified by coincidence, resampling tests were performed by randomly splitting the 668 samples into halves with equal sample sizes, i.e., 334 samples each. Similarly, the top 100 probes with lowest CVs were recorded as list B and C, and the top 100 probes with highest stability scores were recorded as list B' and C'. Detailed information about the resampling test is described under Methods. The results of the random trials are summarized in Table S1, which shows that the mean CVs and stability scores of the top 100 probes were generally larger than 80 and even approached or attained 90 among the six lists. Among the 10,000 trials, the number of probes identified in list B at least once was 224, and the number of probes identified at least

once in any of the 6 lists was only 295. Such high concordance suggested that both CV value and stability score approaches were stable and their findings were generally very similar. In addition, these two approaches identified 69 common probes out of the top 100 probes in lists A and A', which further demonstrated the consistency of the results. Therefore, we focused on the intersecting set of probes ($n = 27$) among all six lists for the following analyses.

Methylation levels of the selected 27 probes across different datasets. The 27 candidate probes consistently appearing in all six lists are shown in Table 2. As shown in Figure S1, all of these 27 probes (red dots) displayed high methylation levels and relatively very low CVs. For example, the highest CV value of the 27 probes was only 0.1347, whose rank was 36th among 7,829 probes. To be more specific, we further examined the methylation levels of the 27 probes in all samples from different datasets (Table 2). In general, their β values of methylation were very stable across all 668 samples, independent of different experimental batches, and all of them were higher than 0.8, and even 0.9. For instance, as shown in Figure S2, the M-values of *N4BP2* and *EGFL8* in distinct datasets did not vary much. Therefore, these results suggested that our approach was able to successfully identify probes with consistent methylation levels. The 27 selected probes showed consistent methylation levels across samples with different diseases, tissue types, and ethnicities.

Validation of selected probes using mass spectrometry. To narrow down the target probes for validation, we repeated the same procedures shown in Figure 1, except that only the top 20 probes were tallied. Among the 10,000 resampling trials, only 1.09% of probes ($n = 85$) were identified at least once in all six lists, indicating that our proposed approach to identify probes with stable methylation levels is not sensitive to a change in the number of probes selected. Next, an average number of appearances in the lists B–C' was ranked for experimental validation. The top 5 probes (*N4BP2*, *EGFL8*, *CTRB1*, *TSPAN3*, and *ZNF690*) were selected (Table S2), and all of them were identified more than 9,885 times out of the 10,000 trials, suggesting they had stable methylation levels across different biological samples.

Twenty-four cell lines derived from 13 different cell types were investigated using mass spectrometry (Table 3). After DNA was extracted and bisulfate converted, mass spectrometry experiment was performed according to the standard protocols provided by the manufacturer (Sequenom, San Diego, CA). The results of the mass spectrometry are illustrated in Figure 2, and all of the five genes generally showed consistent and stable methylation levels among all cell lines. *N4BP2* and *EGFL8*, for example, had methylation levels higher than 0.75 in all cell lines, which demonstrated that these two genes were highly methylated independent of tissues type (Figure 2A–B). In addition, *CTRB1*, *ZNF690*, and *TSPAN3* showed high β values (>0.75) in 24 (96%), 23 (92%), and 19 (76%) cell lines.

Table 1 | Characteristics of analyzed Illumina Infinium Human Methylation27 microarray datasets

Data Set ^a	Sample Number	Description
GSE17648	44	Colorectal cancer, tumor vs. adjacent normal
GSE17769	10	Breast cancer, tumor cell lines vs. normal line
GSE20067	195	Irish patients with type 1 diabetes mellitus
GSE20080	48	Normal and preinvasive cervical smear samples
GSE24087	4	HPV(+) and HPV(–) SCC cell lines
GSE26133	160	HapMap Yoruba lymphoblastoid cell lines
GSE27284	10	Primary NSCLC fibroblast and normal cell lines
In-house studies	211	Lung adenocarcinoma; SLE, case vs. control; Cord blood samples with atopic dermatitis
total	682	--

^aAccession number is from Gene Expression Omnibus.

HPV: human papillomavirus; SCC: squamous cell carcinoma; NSCLC: non-small cell lung cancer; SLE: systematic lupus erythematosus.



Table 2 | Information on the 27 probes commonly identified in the six lists

Gene	Description	TargetID	Chromosome	Coordinate	M-value	β value
ADMR	G protein-coupled receptor 182	cg24167841	12	55675308	3.320	0.909
APOL1	Apolipoprotein L, 1	cg08775793	22	34978638	2.874	0.880
C1orf38	Thymocyte selection associated family member 2	cg27573888	1	28078930	4.218	0.949
C2orf18	Solute carrier family 35, member F6	cg24248317	2	26839883	3.046	0.892
CSH1	Chorionic somatomammotropin hormone 1 (placental lactogen)	cg11880211	17	59327200	2.48	0.848
CSH2	Chorionic somatomammotropin hormone 1 (placental lactogen)	cg27178345	17	59304304	3.002	0.889
CTRB1	Chymotrypsinogen B1	cg16863382	16	73810196	3.604	0.924
DAZL	Deleted in azoospermia-like	cg06429195	3	16621586	3.286	0.907
EGFL8	EGF-like-domain, multiple 8	cg16282679	6	32242288	3.235	0.904
F11R	F11 receptor	cg07883333	1	159276377	4.743	0.964
FLJ45684	FLJ45684 locus	cg03410718	19	598423	3.543	0.921
FAM83H	Family with sequence similarity 83, member H	cg20519035	8	144883226	3.624	0.925
GDF11	Growth differentiation factor 11	cg00344358	12	54421960	3.466	0.917
GPS2	G protein pathway suppressor 2	cg11181795	17	7160695	4.075	0.944
HTRA3	HtrA serine peptidase 3	cg24105933	4	8321695	3.604	0.924
IL13	Interleukin 13	cg14523284	5	132021513	2.847	0.878
ITGA2B	Integrin, alpha 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41)	cg17749520	17	39822093	3.202	0.902
N4BP2	NEDD4 binding protein 2	cg13107169	4	39774633	3.219	0.903
PPP2R2A	Protein phosphatase 2, regulatory subunit B, alpha	cg10207787	8	26204112	3.186	0.901
PRH1	Proline-rich protein HaellI subfamily 1	cg13383572	12	10927484	2.888	0.881
RDBP	Negative elongation factor complex member E	cg04710641	6	32036236	3.107	0.896
RPS6KB2	Ribosomal protein S6 kinase, 70 kDa, polypeptide 2	cg23347911	11	66951485	3.286	0.907
SMARCA3	Helicase-like transcription factor	cg21089667	3	150287952	4.585	0.960
TSPAN3	Tetraspanin 3	cg21377793	15	75151574	3.320	0.909
TUBA3D	Tubulin, alpha 3d	cg02774486	2	131949903	3.732	0.930
ZNF142	Zinc finger protein 142	cg04970994	2	219234087	3.524	0.920
ZNF690	Zinc finger and SCAN domain containing 29	cg12784172	15	41449249	3.303	0.908

Thus, the results indicated that our approach can successfully identify genes that are stably and highly methylated across different cell types.

Lastly, we evaluated the sensitivity of detecting methylation levels in the top three genes showing stable methylation levels, including

N4BP2, *EGFL8*, and *CTRB1*, using different concentrations of methylated samples. Two standard DNA samples, which were fully methylated (100%) and unmethylated (0%), were purchased from Qiagen (Valencia, CA) and used to make DNA samples with 0%, 25%, 50%, 75%, and 100% methylation levels. Subsequently, these

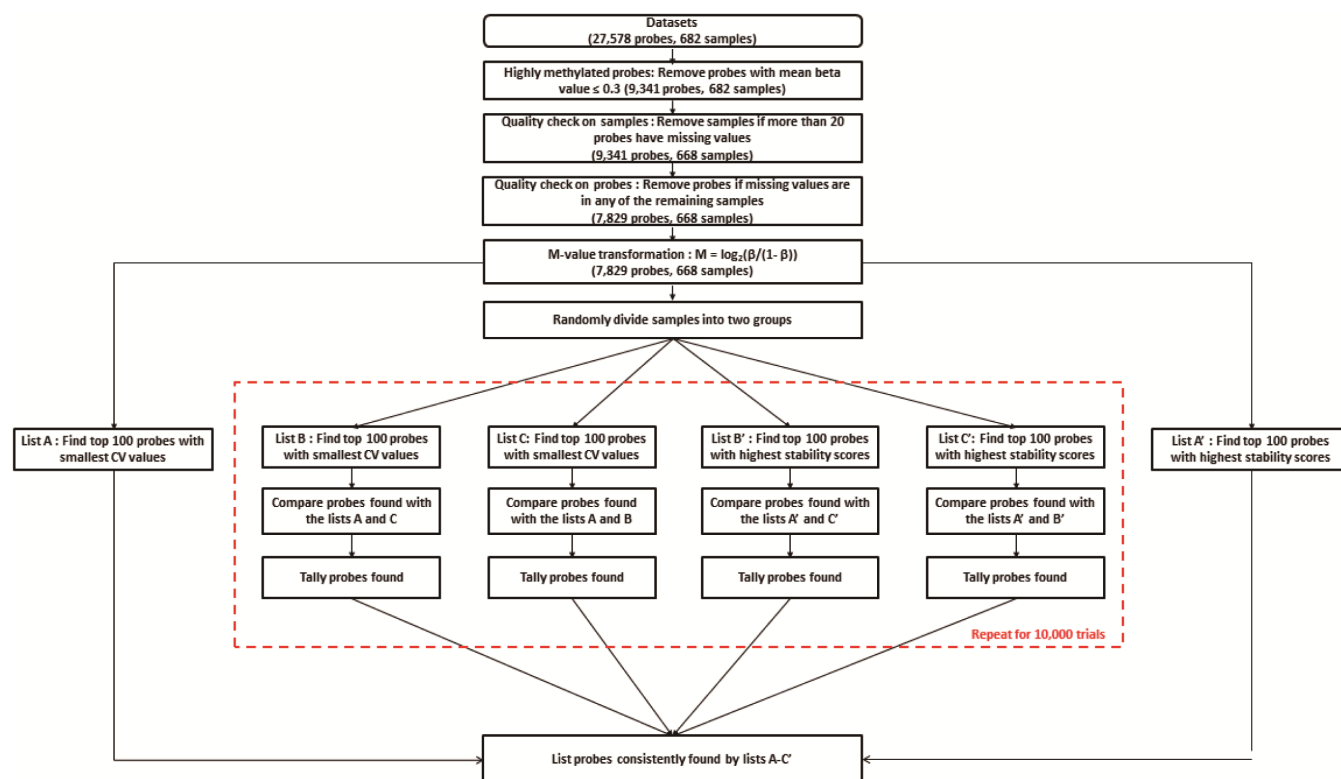


Figure 1 | Flowchart for identification of genes with consistent methylation levels across different samples.



Table 3 | Characteristics of the 24 cell lines investigated using the MassARRAY system

Cell line	Feature
T98	Brain glioblastoma
U87	Brain glioblastoma
SH-SY5Y	Neuroblastoma
GM8680	Fibroblast
GM8402	Fibroblast
TK6	Lymphoblast
WTK1	Lymphoblast
CE81T	Esophageal cancer
MCF7	Breast cancer
CL1-0	Lung cancer
A549	Lung cancer
BEAS2B	Lung normal
PANC-1	Pancreatic cancer
DLD1	Colon cancer
HT-29	Colon cancer
HELA	Cervical cancer
BPH1	Prostate normal
OVCAR3	Ovarian cancer
SKOV3	Ovarian cancer
JH514	Ovarian cancer
OVTOKO	Ovarian cancer
OVTW59_PO	Ovarian cancer
TOV-112D	Ovarian cancer
OVEM	Ovarian cancer

samples were investigated in the MassARRAY system, and the methylation levels of *N4BP2*, *EGFL8* and *CTRB1* are shown in Figure 3. For each gene, a linear relationship ($R^2 \geq 0.98$) was observed between its methylation level and the methyl concentration of DNA samples. In addition, these three genes all showed low (<0.2) and high (>0.8) methylation levels in the 0% and 100% methylated samples, respectively. This suggested that the methylation levels of these three genes were highly associated with the methylated concentrations of DNA samples. Therefore, these genes can serve as potential methylation markers for bisulfate conversion.

Discussion

Changes in methylation have been shown to be an important player in regulating cell growth, normal cellular functions, and even the development of diseases^{17,18}. Thus, how to effectively and accurately measure the methylation levels of multiple genes simultaneously has become a critical issue. Although some experimental technologies, such as enzyme-based gel electrophoresis and affinity-enrichment methods, can be used in methylation studies without performing bisulfite conversions, most of the popular techniques still require treating samples with bisulfite in advance¹¹. However, inappropriate bisulfite conversion may easily introduce systematic errors and lead to incorrect conclusions. A previous study has demonstrated that the rate of cytosine deamination to uracil highly depends on temperature and incubation time¹⁹. Therefore, identification of internal controls for assessing the conversion efficiency of DNA samples is necessary. In addition, internal controls can provide a baseline for comparison of the quality of input DNA samples and provide a stable reference line to normalize methylation data among different samples. For example, the delta-delta cycle threshold (ddCt) method has been widely used in analyzing PCR data for mRNA expression values^{20,21}, and internal controls, such as *ACTB* and 18s rRNA, which have high and stable expression values in different tissues types, are essential for interpreting the results. In this study, we demonstrated that *N4BP2*, *EGFL8*, and *CTRB1* were highly methylated not only in samples detected by microarrays (β values > 0.9 , Table 2), but also in 24 cell lines across 13 tissue types examined by mass spectrometry (β values > 0.75 , Figure 2). Therefore, the results of two independent

techniques both showed that these genes had high methylation levels in several tissue types. In addition, a linear relationship ($R^2 \geq 0.98$) was demonstrated between the methylation levels of three identified genes and the methyl concentration of DNA samples (Figure 3). These data further suggested their capability for serving as internal controls because their methylation levels can be used to reflect the efficiency of bisulfite conversion in input samples. In conclusion, *N4BP2*, *EGFL8*, and *CTRB1* were possible internal controls for methylation studies since their methylation levels were not only consistent in many different human tissues but also proportional to the methyl concentration of DNA samples.

Two approaches, CVs and stability scores, were performed in this study to identify probes showing consistent methylation levels (Figure 1). For a given gene, the CV was used to evaluate consistency across different samples, whereas the stability score approach¹⁶ utilized a rank product method to estimate its suitability in serving as a control in distinct datasets. Interestingly, the results of these two approaches were very similar and identified 69 probes in common out of the top 100 probes, motivating us to use both approaches. Also, moderate to high Pearson correlation coefficients ($r = 0.62\text{--}0.76$) were observed between the rankings of genes obtained from CV and stability score approaches, further suggesting their concordance.

Resampling tests were used to exclude probes identified by random chance, and high similarities were observed in the results (Table S1). In addition, although selecting the top 100 probes is an arbitrary threshold, the results showed minimal variation when the threshold number was changed to 20. To summarize, the results suggest that our procedures were not sensitive to the chosen parameters and were able to reproducibly identify probes by integrating two different approaches.

The expression levels of hypermethylated genes are down-regulated, if these genes are subject to the regulation of DNA methylation¹⁸. Such an epigenetic regulation mechanism is observed in several genes related to embryonic development²². For instance, *DAZL*, one of the top 27 probes, is an important regulator participating in spermatogenesis and oogenesis, and its demethylation is only observed in germ cells but not somatic cells²³. *GDF11* is a growth factor involved in the formation of mesoderm and neurogenesis²⁴, and its gene expression level can be induced by a histone deacetylase (HDAC) inhibitor and inhibited by *HDAC3*²⁵. Accordingly, the results suggest that these identified genes have biological relevance.

Although we used gene symbols to denote the CpG islands showing high methylation, readers should keep in mind that methylation levels are dependent on the specific chromosome coordinates (Table S2), because different methylation statuses of distinct CpG loci in the same gene could be observed. For example, methylation changes were observed in the first exon of *HTRA3* in smoking-related lung cancer, but such alterations were not detected in its promoter region²⁶. However, gene symbols were chosen to represent the CpG islands in this study, since such methylation changes in CpG islands may affect the overall function of the corresponding gene. To date, the literature has rarely reported methylation changes in the top five genes identified by our analysis (*N4BP2*, *EGFL8*, *CTRB1*, *TSPAN3*, and *ZNF690*). A single study has shown that *TSPAN3* was down-regulated in relapsed Wilms tumor; however, such gene expression changes were not controlled by methylation²⁷. Therefore, additional studies of the methylation status of these five genes are required to evaluate their functional roles in relationship to methylation.

In this study, we have demonstrated the consistent methylation levels of *N4BP2*, *EGFL8*, and *CTRB1* in many human tissues and cell lines; however, one caveat is that methylation profiles in each cell line may be affected by *in vitro* cell culture procedures^{28,29}. Epigenetic changes in cells are sensitive to their growth conditions, and thus subtle differences in environment may lead to huge differences in methylation profiles. Two previous studies showed that some varia-

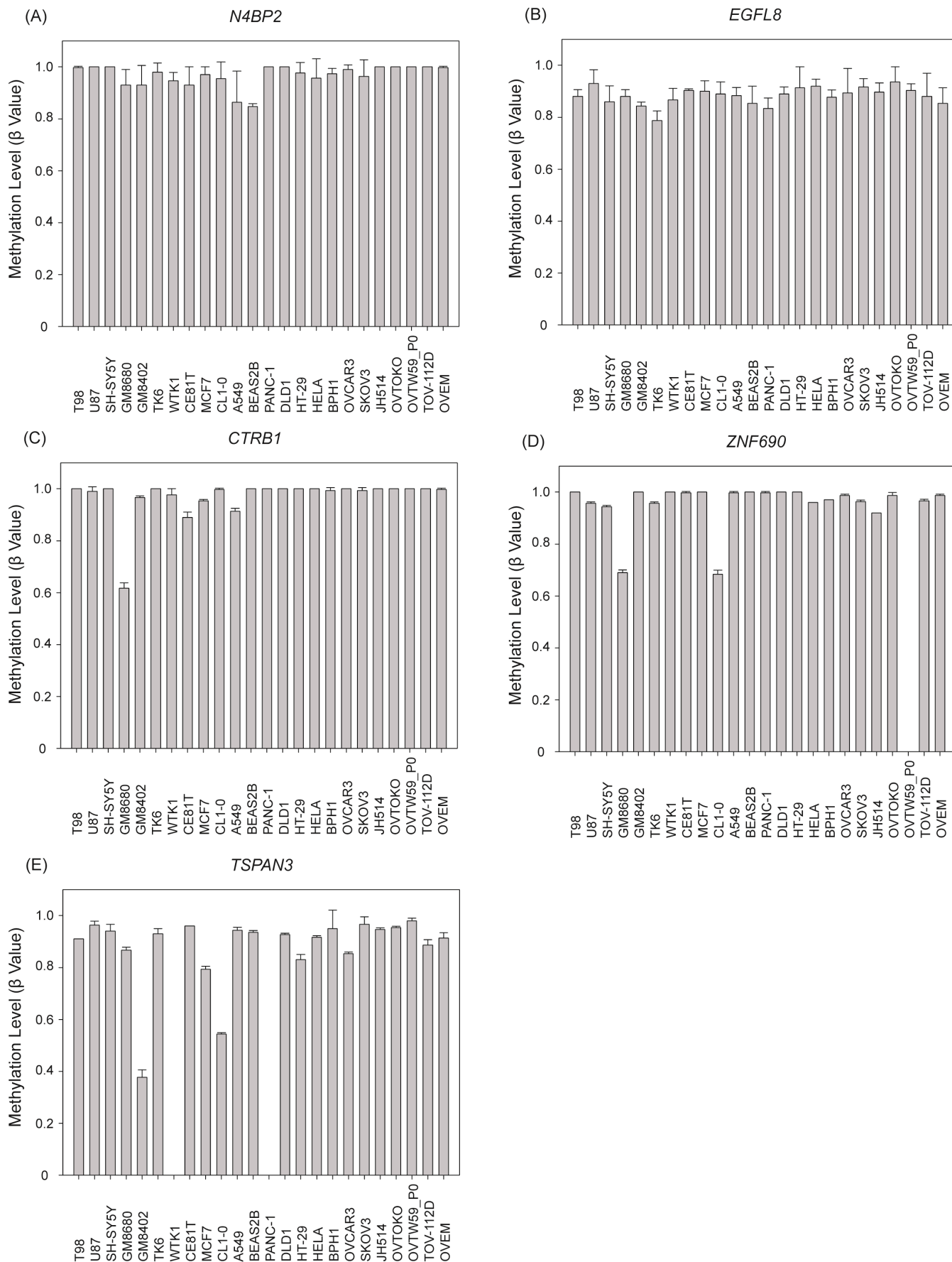


Figure 2 | Methylation levels of the five genes detected by mass spectrometry across 24 cell lines. The X-axis denotes the names of the different cell lines, and the Y-axis represents the average beta value of the methylation level.

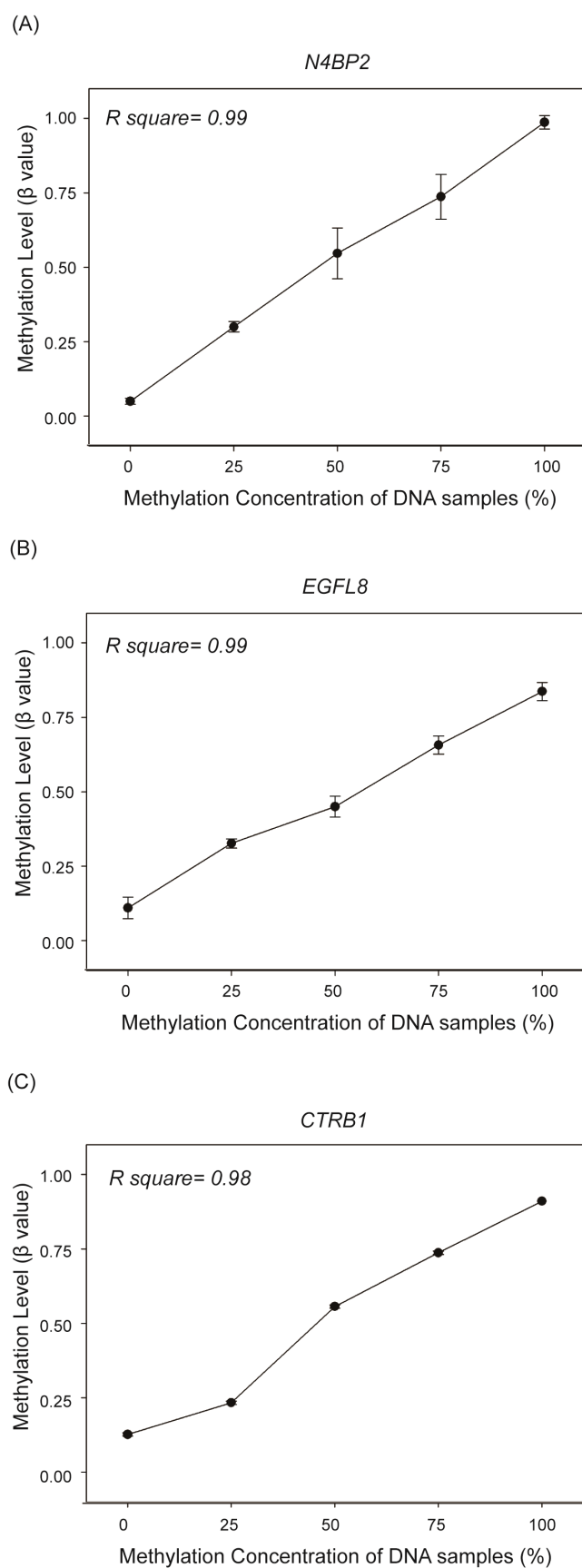


Figure 3 | Correlation between concentration and methylation levels of *EGFL8*, *N4BP2*, and *CTRB1*. Five concentrations, including 0%, 25%, 50%, 75% and 100%, of the methylated DNA samples were investigated by mass spectrometry.

tions in the methylation profiles existed between cell lines and tissues, even if they were from the same organ^{28,29}. Therefore, a preliminary test in different cell lines is prerequisite before utilizing the methylation markers identified in this study.

In conclusion, we have identified five genes with stable hypermethylation across different human tissue types. Among them, *N4BP2*, *EGFL8*, and *CTRB1* not only can serve as internal controls for methylation studies, but also are markers for the efficiency of bisulfite conversion.

Methods

Sample collection. All methylation microarrays analyzed in this study were investigated by using Illumina Infinium HumanMethylation27 BeadChips, containing probes to interrogate 27,578 CpG loci covering more than 14,000 genes. Methylation levels in Illumina methylation assays were quantified by the β value using the ratio of methylated alleles over all alleles for a given CpG locus. Most of the microarray samples were retrieved from the Gene Expression Omnibus website³⁰, with the accession numbers of GSE17648, GSE17769³¹, GSE20067³², GSE20080³², GSE24087, GSE26133³³, GSE27284³⁴, and the other microarrays were collected from our in-house studies. The details of analyzed microarrays are summarized in Table 3.

Processing and filtering of microarray data. The protocol used to identify probes with high and consistent methylation levels is illustrated in Figure 1. First, to remove microarray samples with low quality and intensity, the mean signal of every probe within each slide was calculated in all 682 samples. Samples were excluded for subsequent analyses if the following condition was met: the mean of average β value across all probes was ≤ 0.3 ³⁵. In addition, individual probes were filtered out if they displayed a missing value in any one of the samples. Consequently, 14 samples were excluded and approximately 20,000 probes were filtered out, which resulted in 7,829 probes as potential targets in the following approaches.

Identification of probes with stable methylation levels across different samples. Prior to performing subsequent statistical approaches, the average β values in all microarrays were transformed into “M-values” based on the following equation.

$$M = \log_2(\beta / (1 - \beta))$$

Du et al. reported that this M-value transformation is able to improve the determination of methylation levels in statistical analyses by showing greater consistency and robustness³⁶. After the M-transformation, the coefficient of variances (CV) was utilized to rank the investigated probes for suitability as “housekeeping” probes. Specifically, the CVs of the 7,829 probes were calculated over 668 samples and sorted in ascending order. Based on the results, the top 100 probes having the smallest CV values were reported as possible “housekeeping” candidates (list A). To establish a null baseline for comparison, a resampling test was performed 10,000 times through the following steps. First, the 668 methylation samples were randomly divided in half (334 samples each in lists B and C), and the CV values were calculated. Similar to the approach in identifying list A, the top 100 probes with smallest CV values were recorded and compared with the members of list A. In addition, the top 100 probes identified in list B were compared with the members in list C. Lastly, the matching probes between list A and lists B and C created from 10,000 random trials were recorded, and the common members in list B and C were also tallied for further comparisons.

Verification of possible probes with consistent and stable methylation levels. To evaluate the reliability of identifying housekeeping methylation probes by using CV values, another established algorithm was utilized¹⁶. Briefly, this approach estimated the stability score of each probe based on its methylation level. The formula to calculate the stability score was:

$$s_i = \alpha \log_2(\max\{\mu_i - \beta, 0\}) - \sigma_i.$$

The symbols μ_i and σ_i denote the expression level of gene *i* and the standard deviation across all 668 samples, respectively. The coefficient α was set to its default value, 0.25, as suggested by the authors¹⁶. Similar to the CV value approach, a gene was excluded for further analyses if its mean β value was smaller than 0.3. This criterion was applied in order to yield the same number of probes investigated in both methods to establish a fair baseline for comparison. Moreover, since all samples used in this study were Illumina Infinium HumanMethylation27 BeadChips, the rank product score considering platform-independence, which was outlined by the original authors, was not performed here¹⁶. The scoring scheme in this approach was similar to the previous method implementing CV values, that is, lists of candidate probes over 668 samples were examined and ranked by the stability score in descending order. Likewise, 10,000 random trials were carried out, and three gene lists were obtained for each trial. Meanwhile, the three lists were also compared to each other and the numbers of times that each gene was identified in the lists B' and C' were also tallied. Lastly, the candidate probes with stable methylation levels were narrowed down to those consistently found in all six lists after 10,000 random trials.



Validation of possible gene targets using the MassARRAY system. A total of 24 cell lines were analyzed using the MassARRAY system to validate the methylation levels of selected gene targets. The characteristics of the cell lines are summarized in Table 3. First, genomic DNA was isolated from the cells by proteinase K-phenol/chloroform extraction following standard protocols with 0.5% SDS and 200 µg/ml proteinase K. The DNA concentration of each sample was adjusted to 50 ng/ml and total genomic DNA (500 ng) underwent DNA bisulfate conversion using an EZ DNA Methylation™ kit (ZYMO research, Orange, CA). Among the bisulfate treated DNA products, 200 ng of the bisulfate treated DNA were used for PCR amplification. The primers were designed by using the program EpiDesigner β (<http://www.epidesigner.com/start3.html>). PCR conditions were optimized to preferentially amplify fragments within a size range of 300 to 500 bp. Subsequently, 2 µL of Shrimp Alkaline Phosphatase (SAP) enzyme was added into 5 µL PCR products to dephosphorylate unincorporated dNTPs. Lastly, *in vitro* transcription and RNase A cleavage were carried out, and the mass spectrum was obtained from the PCR reactions. Quantitative methylation analysis software provided by the manufacturer (Sequenom, San Diego, CA) was used to analyze the results.

- Lovinsky-Desir, S. & Miller, R. L. Epigenetics, asthma, and allergic diseases: a review of the latest advancements. *Curr Allergy Asthma Rep* **12**, 211–20 (2012).
- Mohammad, F., Mondal, T. & Kanduri, C. Epigenetics of imprinted long noncoding RNAs. *Epigenetics* **4**, 277–86 (2009).
- Li, C. C., Maloney, C. A., Croyley, J. E. & Suter, C. M. Epigenetic programming by maternal nutrition: shaping future generations. *Epigenomics* **2**, 539–49 (2010).
- Wang, I. J., Chen, S. L., Lu, T. P., Chuang, E. Y. & Chen, P. C. Prenatal smoke exposure, DNA methylation, and childhood atopic dermatitis. *Clin Exp Allergy* **43**, 535–43 (2013).
- Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128**, 635–8 (2007).
- Capell, B. C. & Berger, S. L. Genome-wide epigenetics. *J Invest Dermatol* **133**, e9 (2013).
- Jansson, M. D. & Lund, A. H. MicroRNA and cancer. *Mol Oncol* **6**, 590–610 (2012).
- Shen, H. & Laird, P. W. Interplay between the cancer genome and epigenome. *Cell* **153**, 38–55 (2013).
- Gropman, A. L. & Batshaw, M. L. Epigenetics, copy number variation, and other molecular mechanisms underlying neurodevelopmental disabilities: new insights and diagnostic approaches. *J Dev Behav Pediatr* **31**, 582–91 (2010).
- Fouse, S. D., Nagarajan, R. O. & Costello, J. F. Genome-scale DNA methylation analysis. *Epigenomics* **2**, 105–17 (2010).
- Laird, P. W. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* **11**, 191–203 (2010).
- Ehrlich, M. *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A* **102**, 15785–90 (2005).
- Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–5 (2010).
- Gloss, B. S. *et al.* Integrative genome-wide expression and promoter DNA methylation profiling identifies a potential novel panel of ovarian cancer epigenetic biomarkers. *Cancer Lett* **318**, 76–85 (2012).
- Su, L. J. *et al.* Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. *BMC Genomics* **8**, 140 (2007).
- Popovici, V. *et al.* Selecting control genes for RT-QPCR using public microarray data. *BMC Bioinformatics* **10**, 42 (2009).
- Bock, C. Analysing and interpreting DNA methylation data. *Nat Rev Genet* **13**, 705–19 (2012).
- Robertson, K. D. DNA methylation and human disease. *Nat Rev Genet* **6**, 597–610 (2005).
- Grunau, C., Clark, S. J. & Rosenthal, A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res* **29**, E65–5 (2001).
- Lu, T. P. *et al.* Distinct signaling pathways after higher or lower doses of radiation in three closely related human lymphoblast cell lines. *Int J Radiat Oncol Biol Phys* **76**, 212–9 (2010).
- Lai, L. C. *et al.* Down-regulation of NDRG1 promotes migration of cancer cells during reoxygenation. *PLoS One* **6**, e24375 (2011).
- Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38 (2013).
- Maatouk, D. M. *et al.* DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133**, 3411–8 (2006).
- Mallo, M., Vinagre, T. & Carapuco, M. The road to the vertebral formula. *Int J Dev Biol* **53**, 1469–81 (2009).
- Zhang, X. *et al.* Activation of the growth-differentiation factor 11 gene by the histone deacetylase (HDAC) inhibitor trichostatin A and repression by HDAC3. *Mol Cell Biol* **24**, 5106–18 (2004).
- Beleford, D. *et al.* Methylation induced gene silencing of Htra3 in smoking-related lung cancer. *Clin Cancer Res* **16**, 398–409 (2010).
- Maschietto, M. *et al.* Gene expression analysis of blastemal component reveals genes associated with relapse mechanism in Wilms tumour. *Eur J Cancer* **47**, 2715–22 (2011).
- Paz, M. F. *et al.* A systematic profile of DNA methylation in human cancer cell lines. *Cancer Res* **63**, 1114–21 (2003).
- Varley, K. E. *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res* **23**, 555–67 (2013).
- Barrett, T. *et al.* NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* **37**, D885–90 (2009).
- Chari, R., Coe, B. P., Vucic, E. A., Lockwood, W. W. & Lam, W. L. An integrative multi-dimensional genetic and epigenetic strategy to identify aberrant genes and pathways in cancer. *BMC Syst Biol* **4**, 67 (2010).
- Teschendorf, A. E. *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* **20**, 440–6 (2010).
- Bell, J. T. *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* **12**, R10 (2011).
- Navab, R. *et al.* Prognostic gene-expression signature of carcinoma-associated fibroblasts in non-small cell lung cancer. *Proc Natl Acad Sci U S A* **108**, 7160–5 (2011).
- Nishino, K. *et al.* Defining hypo-methylated regions of stem cell-specific promoters in human iPS cells derived from extra-embryonic amnions and lung fibroblasts. *PLoS One* **5**, e13017 (2010).
- Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).

Acknowledgments

This research was supported in part by a grant from the Department of Health of Taiwan (Grant No. 99-3112-B-002-035). We thank Melissa Stauffer, PhD, of Scientific Editing Solutions, for editing the manuscript, and the YongLin Biomedical Engineering Center of National Taiwan University for additional financial support.

Author contributions

T.P.L., C.K.H. and E.Y.C. conceived and designed the experiments. T.P.L. and K.T.C. performed the experiments. T.P.L., C.K.H. and K.T.C. analyzed the data. M.H.T., K.T.K., L.C.L. and E.Y.C. contributed reagents, materials, and/or analysis tools. T.P.L., K.T.C., L.C.L. and E.Y.C. wrote the paper. All authors reviewed the manuscript.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Lu, T.-P. *et al.* Identification of Genes with Consistent Methylation Levels across Different Human Tissues. *Sci. Rep.* **4**, 4351; DOI:10.1038/srep04351 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>