

RESEARCH

Open Access

DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes

Dvir Aran^{1,2}, Sivan Sabato² and Asaf Hellman^{1*}

Abstract

Background: Abnormal epigenetic marking is well documented in gene promoters of cancer cells, but the study of distal regulatory sites has lagged behind. We performed a systematic analysis of DNA methylation sites connected with gene expression profiles across normal and cancerous human genomes.

Results: Utilizing methylation and expression data in 58 cell types, we developed a model for methylation-expression relationships in gene promoters and extrapolated it to the genome. We mapped numerous sites at which DNA methylation was associated with expression of distal genes. These sites bind transcription factors in a methylation-dependent manner, and carry the chromatin marks of a particular class of transcriptional enhancers. In contrast to the traditional model of one enhancer site per cell type, we found that single enhancer sites may define gradients of expression levels across many different cell types. Strikingly, the identified sites were drastically altered in cancers: hypomethylated enhancer sites associated with upregulation of cancer-related genes and hypermethylated sites with downregulation. Moreover, the association between enhancer methylation and gene deregulation in cancer was significantly stronger than the association of promoter methylation with gene deregulation.

Conclusions: Methylation of distal regulatory sites is closely related to gene expression levels across the genome. Single enhancers may modulate ranges of cell-specific transcription levels, from constantly open promoters. In contrast to the remote relationships between promoter methylation and gene dysregulation in cancer, altered methylation of enhancer sites is closely related to gene expression profiles of transformed cells.

Keywords: Cancer, DNA methylation, distal control elements, enhancers, epigenomics, gene-enhancer pairing, gene regulation, machine-learning

Background

DNA methylation is a key determinant of regulatory chromatin complexes transmitted through cell divisions. Relationships between DNA methylation and gene expression levels were first recognized at gene promoters and CpG islands [1], but were recently observed across the genome [2-6]. In gene promoters, DNA methylation mediates silencing by altering the binding of transcription modulators to their DNA targets [7], but the cause and function of expression-associated methylation away from gene promoters has been elusive. The relationships between aberrant DNA methylation

and the altered expression profiles of cancer cells are also not well understood: the predominant pattern in gene promoters is *denovo* methylation of *polycomb*-repressed genes [8-11]. Since these genes are already inactive in the normal tissue and generally remain inactive in the cancer [12,13], it is hard to establish direct contribution to the cancer process [14]. Away from gene promoters, perturbed methylation has been associated with reduced genomic stability [15,16] and global silencing of large chromatin domains [17,18], but the effect on transcription of particular genes remains unknown.

Transcriptional enhancers support tissue-specific expression profiles through physical interactions with gene promoters. Based on analyses of chromatin structures, hundreds of thousands of transcriptional enhancers were predicted in mammalian genomes (reviewed in [19]). These sites bind chromatin-modulating factors,

* Correspondence: asafh@ekmd.huji.ac.il

¹The Department of Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada, Faculty of Medicine, Hebrew University-Hadassah Medical School, Ein Kerem Campus, Jerusalem 91120, Israel
Full list of author information is available at the end of the article

interact with distal promoters through DNA loops, and demonstrate a unique pattern of DNA methylation [20]. In some particular examples, enhancer sites turned out to be less methylated in cells expressing the controlled genes than in non-expressing cells [21-24]. Taking a more systematic approach, Wienchet *al.* recently revealed activity-dependent methylation in a group of distal enhancers in the mouse [25], and Bock *et al.* [26] demonstrated correlations between enhancer methylation and expression of particular developmental genes during mouse tissue differentiation.

We conducted a methodical analysis of distal DNA methylation sites associated with gene expression in normal and cancerous human cells. The results suggest that enhancer methylation corresponds closely with expression profiles of cancer genes in transformed cells.

Results and discussion

To explore relationships between DNA methylation and gene expression levels across the genome, we took the following strategy. First, we developed a model for the relationships between methylation in gene promoters and gene expression using a machine-learning algorithm. Then, we applied the derived model to methylation sites away from gene promoters and characterized the chromatin structure and binding profile of the discovered sites. Finally, we explored how the identified sites were altered in cancer.

We analyzed available DNA methylation and gene expression data for 58 human cell types (Table S1 in Additional file 1). The methylation data were produced by two different assays: reduced representation bisulfite sequencing (RRBS) and Infinium HumanMethylation450 Bead-Chip (Illumina Inc., San Diego, CA, USA). Following verification of high agreement between the assays (Figures S1A and S2A, Bin Additional file 1), we combined the two datasets. We then analyzed variation in methylation levels across the panel of cell types (see Materials and methods for detailed description). Sites with methylation levels that did not change across the cell types were eliminated, and the remaining 670,906 variable methylation sites (VMSs) were used in the study (Figure S1 Bin Additional file 1). VMSs in gene promoters were negatively correlated with expression whereas VMSs in gene bodies were positively correlated with expression (Figure S2 in Additional file 1), as previously reported [6]. The connection with expression levels was most evident at -500 to +2,000 base pairs relative to transcription start sites (TSSs) (Figure S2A-C in Additional file 1). Therefore, we used the CpG methylation sites in this range for model development.

Promoter-based model of regulatory methylation sites

As a first step in our study, we tested whether we could couple promoters with the appropriate genes based

solely on DNA methylation and gene expression data. We identified 15,205 VMSs in promoters of 3,434 genes. These gene-CpG pairs provided the 'true' methylation-expression sample inputs (Figure 1A). We then trained a machine-learning algorithm to discriminate these true gene-CpG pairs out of an excess of false (randomized) pairs, produced by matching promoter CpGs with genes from other chromosomes. Through rounds of training and test sets, the algorithm optimizes parameters of linear and monotonic correlations between methylation and expression, allowing the best discrimination of the true pairs (Figure 1B). The output was a general model for the methylation-expression relationship in gene promoters (including promoters with or without CpG islands). Based on the learned model, we further produced a score for each of the gene-CpG pairs. At score ≥ 0.85 , the model successfully paired 87.2% of the genes to their actual promoters, compared with a null expectation of 50% (Figure 1C). The sensitivity of the test under these conditions was 2.63% (that is, predictions were made for 2.63% of the promoter VMSs), and the false discovery rate was 12.8%. Thus, the developed model successfully paired promoter methylation sites with their genes based on methylation and expression data alone.

Relating genes to distal regulatory sites using DNA methylation

We then extended the analysis to the VMSs residing from one megabases (Mb) upstream of the TSSs through one Mb downstream of the transcription end sites of 17,862 human genes (Figure 2A). We chose to exclude more distant sites to decrease complexity and false positives, and because it was previously shown that distal regulatory elements tend to concentrate in the global vicinities (usually within a few hundred kilobases (kb)) of the targeted genes [27,28]. Out of 14,702,075 analyzed CpG-gene pairs, 2,824 pairs obtained high scores (score ≥ 0.9) according to the model. The distribution of methylation-expression relationships of the high-scoring sites over the cell types is shown in Figure 2B. High-scoring sites were significantly more frequent ($P < 10^{-5}$) in actual gene intervals than in random intervals (produced by shuffling the expression data of the actual gene with expression data of foreign genes taken from other chromosomes). The entire list of CpG-gene pairs at score > 0.75 is given in Additional file 2.

We then sorted out the sites within 5 kb of promoters or alternative promoters of the associated genes. The remaining 1,911 pairs included 486 unique genes corresponding with 1,041 unique distal methylation regions (we arbitrarily clustered CpGs with < 3 kb between them into a single region). On average, we obtained 2.14 distal regions and 3.93 methylation sites per associated gene. The following

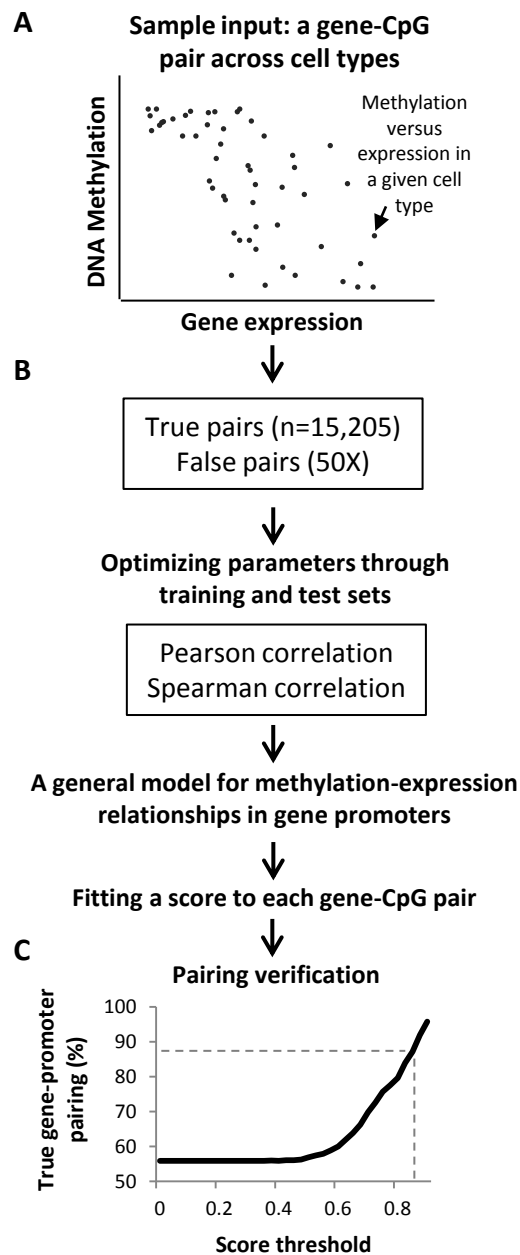
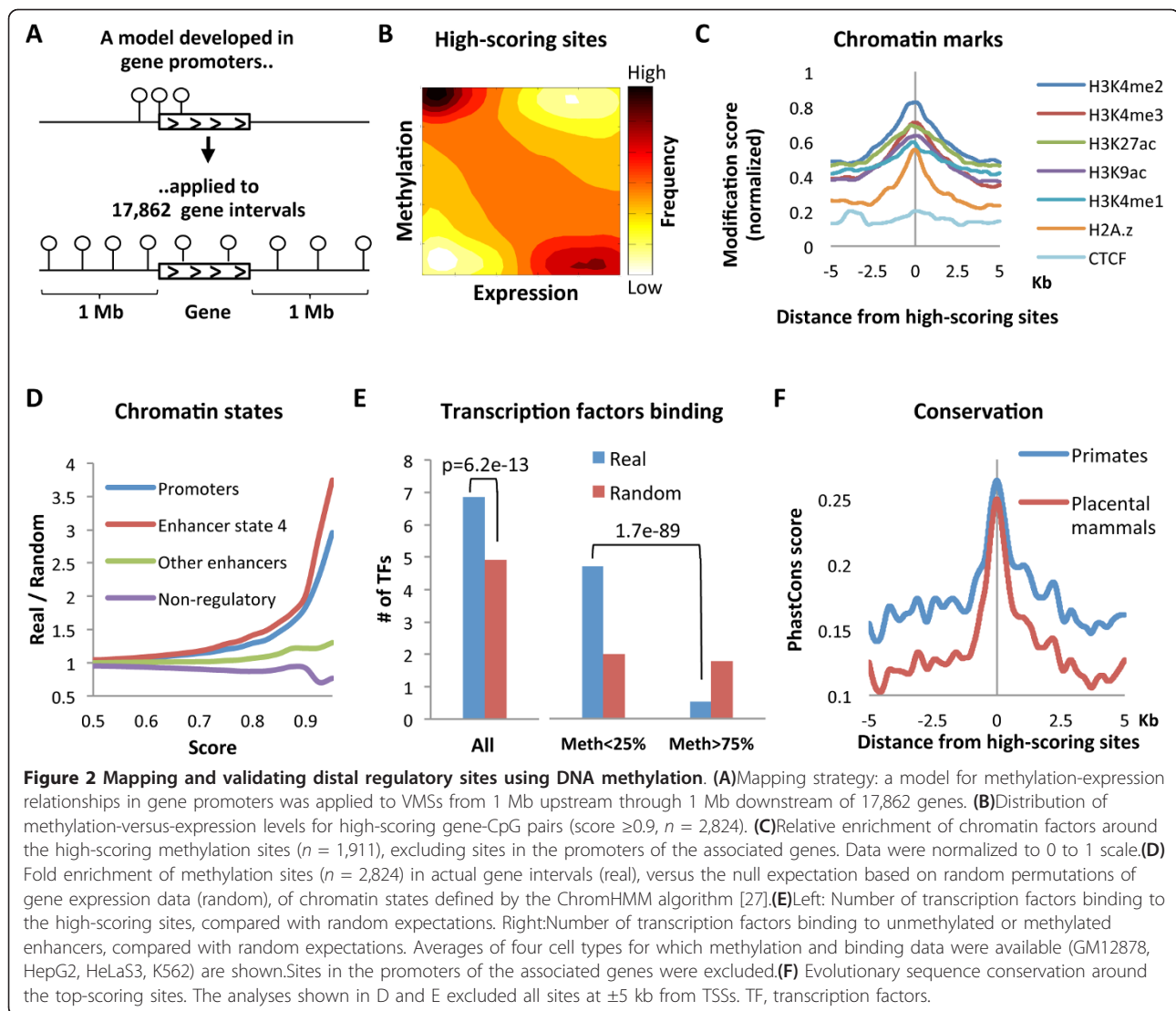


Figure 1 Promoter-based model of methylation-expression relationships at regulatory sites. **(A)** A typical gene-CpG sample pair, conveying the methylation of a promoter-based CpGsite in relation to the expression levels of its linked gene across cell types. **(B)** A machine-learning algorithm (SVM-MAP) was trained to distinguish true gene-CpG pairs out of 50-fold excess of false (randomized) pairs. Through rounds of training and test sets, the algorithm optimized parameters of linear (Pearson coefficient) and monotonic (Spearman) correlations to provide the best discrimination between true and false pairs, producing a general model for methylation-transcription relationships in gene promoters. Based on fitting with the learned model, a score was assigned to each gene-CpG pair. **(C)** Rates of successful gene-promoter pairing as a function of thresholds on the scores (null expectation = 50%). At score ≥ 0.85 , the model successfully paired 87.2% of the genes to their actual promoters (dashed lines).

analyses were done on these 1,911 high-scoring CpG-gene pairs.

To confirm that the high-scoring methylation sites were indeed regulatory sites, we analyzed their chromatin signatures. High-scoring sites were enriched by histone

modifications that have been previously associated with transcription activation, including histone 3 lysine 4 methylations, lysine 27 acetylation, lysine 9 acetylation and histone variant H2A.Z, but not with *CTCF* binding (Figure 2C). Further analysis of chromatin states defined



by the ChromHMM algorithm (based on various chromatin factors [27]) revealed significant enrichment of the high-scoring sites within promoters and within a particular class (ChromHMM state 4) of strong enhancers ($P = 2.4e^{-15}$; Figure 2D). In agreement with this, VMSs were hypomethylated when included in enhancer chromatin, compared to the methylation of the same sites in cells where they were included in non-regulatory chromatin (Figure S3 in Additional file 1). In addition, the highest frequency of hypomethylated sites was observed in state 4 enhancers (Figure S3C in Additional file 1). We also analyzed the protein-binding capacities of the identified distal sites. The mapped methylation sites bind a larger number of transcription factors than expected. Moreover, unmethylated sites bind more factors than methylated ones (Figure 2E). Finally, we found that the high-scoring methylation sites are evolutionarily

conserved (Figure 2F). Together, these findings suggest that distal expression-related methylation sites populate a particular class of transcriptional enhancers, which bind transcription factors in a methylation-dependent manner.

Validation of enhancer-promoter pairing

To validate distal enhancer-promoter interactions, we compared our predictions with a recent comprehensive study of long-range DNA interactions, measured by the chromosome conformation capture carbon copy (5C) technique in three cell types [29]. We identified enhancer sites ($n = 318$) assessed by both methylation score and 5C probes, and determined whether the methylation-based gene-enhancer pairs also reveal physical DNA interactions. Between-assay agreements were markedly increased with methylation scores: at scores

≥ 0.85 , 5C interactions were found in 53% of the cases, compared with 29% of lower score interactions ($P = 0.021$). A representative example of a long-range enhancer-promoter interaction captured by the both assay is shown in Figure S4 in Additional file 1. We also analyzed the frequency of DNA interactions in 21 enhancer-TSS pairs predicted by our assay, versus the interactions of the same enhancers with the other 264 TSSs in the regions (10 kb to 1 Mb from the enhancer site). The frequency of 5C interactions was significantly higher with the TSS predicted by our assay than with the other TSSs (308.5 ± 738.2 junction sequence reads in the high-scoring pairs, versus 90.7 ± 277.7 reads in the other pairs, $P = 0.009$). Thus, the large-scale looping analysis affirms the fidelity of enhancer-promoter pairings based on DNA methylation signals. Given that our study examined many more cell types than the three analyzed in the 5C study, and thus could identify additional enhancers not active in these three cell types, this level of between-assay agreement is probably an underestimation of the actual degree of valid pairs.

Since certain epigenetic profiles may not be fully preserved in cell lines [6], we also confirmed the ability to map enhancer-promoter interactions in freshly obtained (uncultured) tissue samples. Analysis of seven tissue biopsies from two donors [30] ([GSE:30654]) revealed significant intersection ($P < 0.01$) with the enhancer-gene pairs observed in the original set of 58 cell types (see Materials and methods for detailed description). Thus, a significant number of the gene-enhancer interactions predicted in cell cultures were replicated in uncultured tissue samples. This analysis also indicates a significant degree of overlapping between gene-enhancer pairing in diverse tissue collections.

Enhancer methylation defines gradients of cell-type transcription levels from permissive promoters

Interestingly, many of the promoters for which we identified high-scoring enhancers were constantly unmethylated across all cell types, regardless of expression levels (Figure 3A,B and Figures S4 and S5 in Additional file 1). This finding does not contradict the bulk of evidence suggesting that methylation sites in and around promoters tend to correlate with expression levels (our model was in fact built based on this general phenomena). Instead, it suggests that in cases of permissive - but not necessarily active - promoters, enhancer methylation may serve as a main determinant of gene transcription levels. Moreover, enhancers tend to show gradual, rather than distinct, methylation levels across cell types, in tight correlation with gradients of expression (note that our model does not enforce these gradients). Enhancers have previously been connected with cell-type-specific

transcription patterns. However, it was generally assumed that sets of enhancers, each of them active in a particular tissue or tissues, are needed [31]. According to this model, we would expect that a given enhancer site would be fully methylated in most of the cell types, and fully unmethylated in some. The tendency of the identified sites to show gradual methylation differences between cell types supports an alternative model, in which even a single enhancer site can mediate distinct transcription levels over wide ranges of tissues and cell types (Figure 3C).

Altered enhancer methylation predicts changes in the expression profiles of cancer genes

Altered histone modifications have recently been demonstrated in enhancers in cancer cells [32]. We therefore explored whether methylation of enhancer DNA is also altered in cancers. We analyzed the methylation levels of the 1,911 distal methylation sites associated with the expression of 486 genes in normal human mammary epithelial cells versus mammary gland adenocarcinoma (MCF7) cells. A subset of the predicted enhancers was hypo- or hypermethylated in the cancer cells compared to the normal cells. The genes associated with hypomethylated enhancers tended to be upregulated in the cancer, and the genes associated with the hypermethylated enhancers tended to be downregulated (Figure 4A, upper row). Moreover, altered gene expression was better correlated with altered enhancer methylation than with altered promoter methylation ($R = -0.37$ versus $R = -0.16$, respectively), and the effect of methylation differences on gene expression was significantly higher in enhancer sites than in promoter sites ($P = 3.55e^{-15}$). Thus, enhancer methylation is remarkably altered in cancer and, moreover, enhancer methylation is more closely related to changes in gene expression than promoter methylation.

To examine whether these patterns are exclusive to mammary cells, we repeated the analysis in lung epithelia. The analysis of normal human bronchial epithelial cells versus epithelial lung carcinoma (A549) cells revealed similar relationships between enhancer and promoter methylation and gene expression (Figure 4A, middle row). We have further analyzed a set of 18 normal cell types of various tissue origins versus 18 cancers of various tissue origins (Table S2 in Additional file 1). Once again, the methylation of enhancers was drastically altered and was tightly connected with alteration of gene expression, whereas promoters showed smaller alterations and weaker correlation with expression ($R = -0.51$ versus $R = -0.24$); $P < 1e^{-20}$; Figure 4A, lower row). Thus, alteration of enhancer methylation is common to many cancer types, and is associated with substantial changes in gene expression.

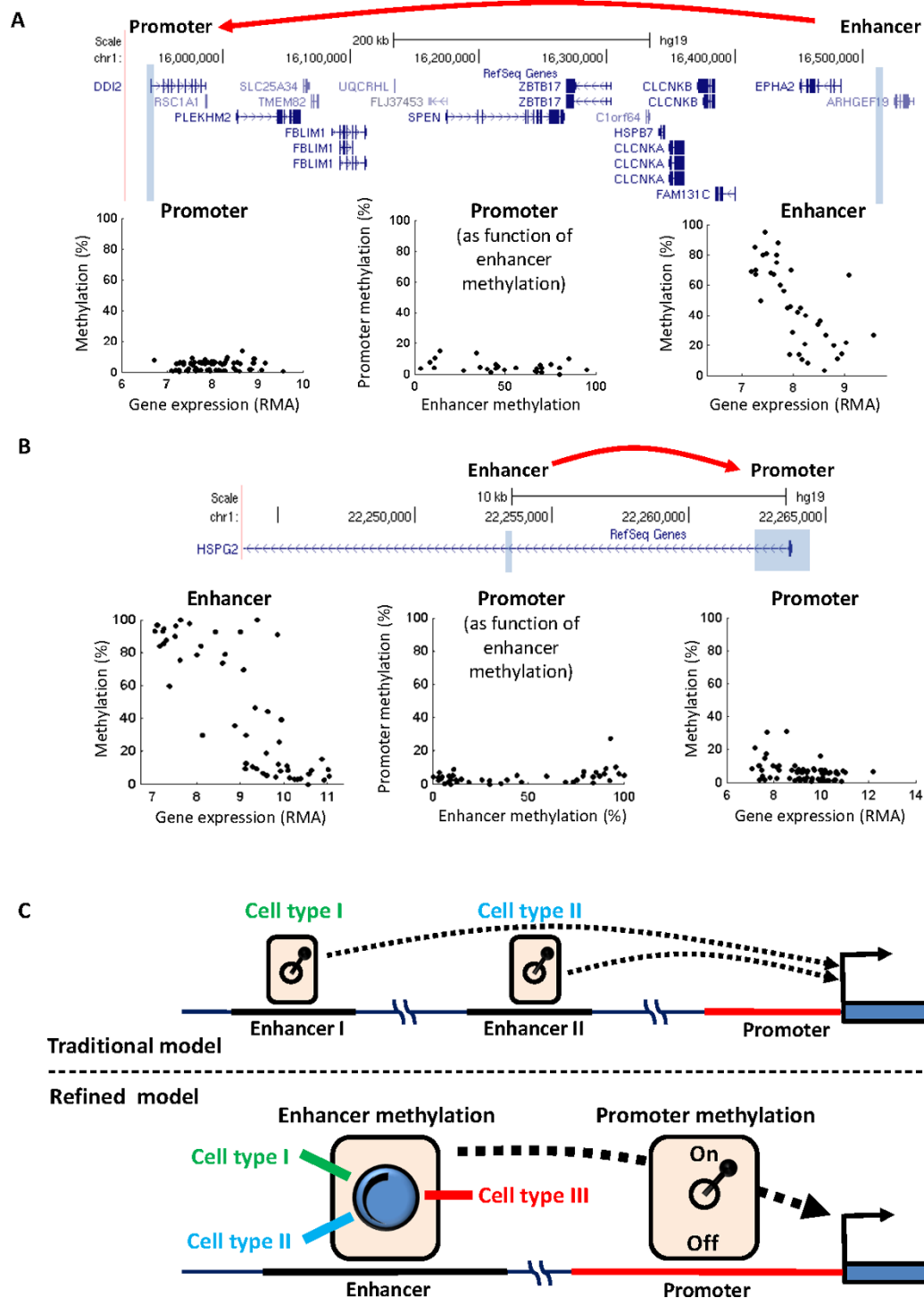
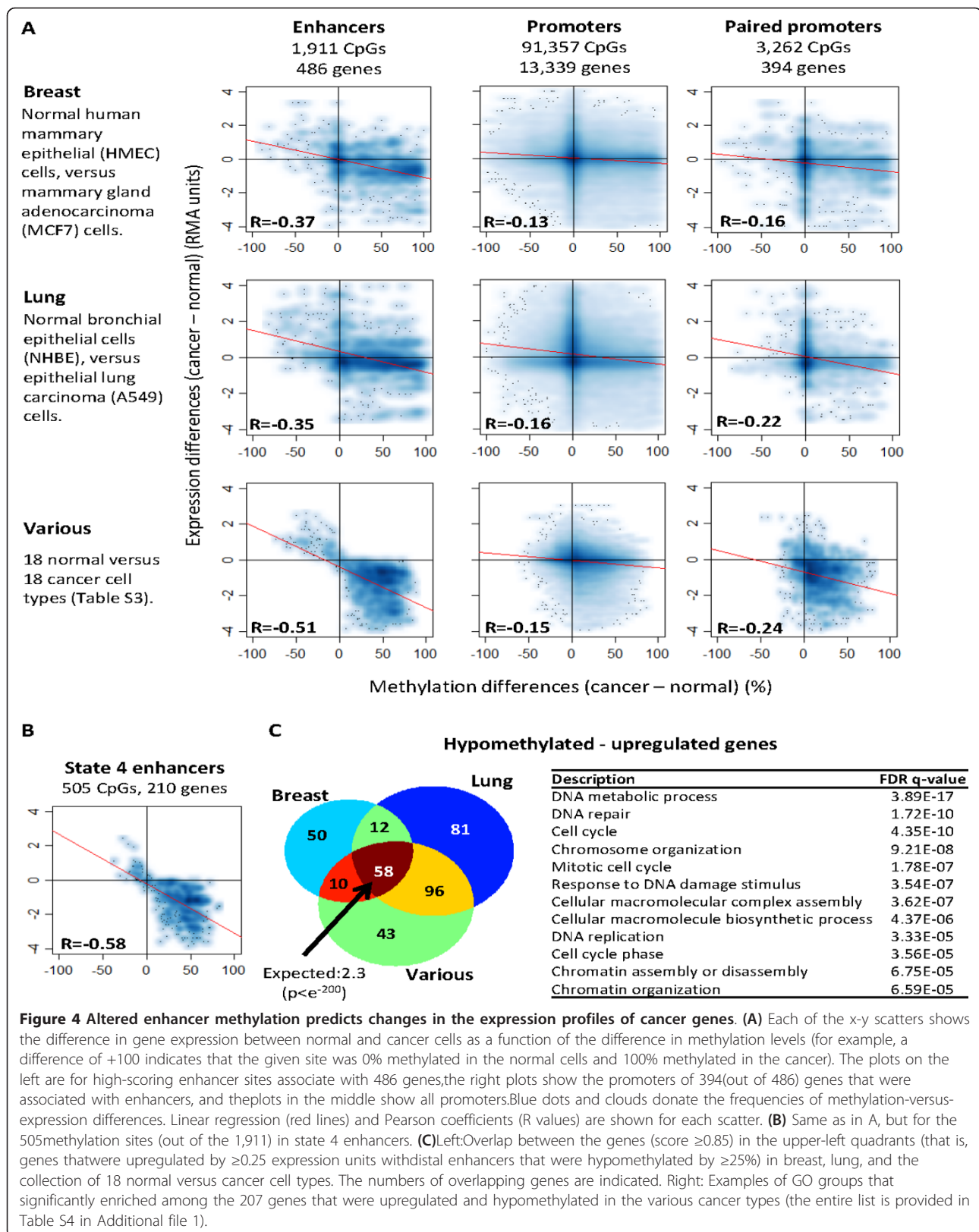


Figure 3 Enhancer methylation defines gradients of cell-type transcription levels. **(A)** An example of gene-enhancer pair associated with expression of the *RSC1A1* gene region. Gray boxes mark an enhancer methylation site associated with expression of the *RSC1A1* gene, and the promoter of this gene. The x-y scatters below the map show methylation versus expression across the cell types for the promoter methylation sites (left), and for the enhancer site (right). The middle scatter shows the methylation of the promoter sites versus methylation of the enhancer across the cell types. **(B)** Similar to A, but for an enhancer within the first intron of the *HSPG2* gene, interacting with *HSPG2* expression. **(C)** Models for enhancer control of cell-type expression levels. Upper panel: The traditional model suggesting that enhancers act like cell-type-specific switches of gene transcription, each of them supports expression in a given cell type (or types). Bottom: A refined model suggesting that even a single enhancer site may function as a dimmer switch, mediating gradients of transcription levels across many cell types as long as the promoter is unmethylated and thus permissive for transcription. RMA, (Robust Multi-chip Average).



Since our enhancers were specifically enriched in the chromatin class defined as ChromHMM state 4 (Figure 2C), and since enhancers in this state showed the highest methylation differences compared to non-regulatory chromatin (Figure S3 in Additional file 1), we hypothesized that state-4 enhancers would show the closest association with changes in gene expression in cancers. Indeed, this group of enhancers showed an even tighter correlation with gene expression levels compared with the entire group of enhancers ($R = -0.58$ versus $R = -0.51$; Figure 4B).

We also sought to analyze what characteristics the altered genes might share in addition to association with enhancer methylation. For this, we first analyzed whether the same genes tended to be upregulated or downregulated across cancer types, and then asked whether they belonged to defined gene-ontology (GO) categories. As shown in Figure 4, a majority of the genes were hypermethylated and downregulated, and a smaller fraction of the genes showed the opposite pattern. We found no significant overlap between the genes that were hypermethylated and downregulated in mammary, lung or in the sets of various cell types. These genes also showed no significant GO clustering. Thus, the hypermethylated genes might be cancer-type-specific (for example, cell-type-specific genes that are downregulated in the derived cancer). By contrast, the hypomethylated, upregulated genes significantly repeated between cancer types (Figure 4C, left), and 58 of them appeared in all cancer types examined ($P < e^{-20}$, compared with the null expectation of 2.3 overlapping genes; Table S3 in Additional file 1). GO analysis showed that the hypomethylated and upregulated genes were frequently involved in cell growing-related functions (Figure 4C right; an extended list of significant GO clusters is given in Table S4 in Additional file 1). Thus, the hypomethylated enhancers associate with genes that are active in many cancer types and support cell proliferation.

An interesting question raised from the above is whether enhancer methylation may also distinguish between cancer and normally-dividing cells. To examine this, we compared Epstein-Barr virus-immortalized lymphoblastoid (GM12878) cells, which are rapidly growing in culture but otherwise have normal karyotype and no particular characteristics of cancer cells, with lymphoblastoid (Jukat) cells derived from an acute leukemia. We identified 74 genes that were hypomethylated and upregulated in the leukemia compared to the normal lymphoblastoids. As expected, these were not proliferative genes - the proliferative genes were already active in the dividing normal cells, so they were not expected to show differences compared to the cancer. Instead, this analysis captured genes that were involved in cancer-related processes other than cell proliferation, such as the *PRAME* gene, which is believed to inhibit myeloid differentiation in

certain myeloid leukemias[33] (Table S5 in Additional file 1). Of the 74 upregulated genes, 31 were also hypomethylated and upregulated in the sets of normal versus cancer cell types (Figure 4), a significantly higher number than expected by chance ($P = 7e^{-14}$). Thus, enhancer hypomethylation may associate with upregulation of genes involved in a variety of cancer-related pathways.

Enhancer methylation maybe specifically altered in cancer

Finally, we asked whether the observed perturbations of enhancer methylation in cancers (Figure 4) are fully explained by a global perturbation of the DNA methylation blueprint in cancer cells, or are the outcome of enhancer-specific processes. For this, we first analyzed the overall differences between the methylation of normal and cancer cells. As previously reported, methylation levels along the genomes of normal cells from a given tissue (left scatter in Figure 5A), or even from different tissues (left scatter in Figure 5B), generally resemble each other. Cancer cells, however, are grossly differentiated from normal cells: sites that are normally unmethylated tended to gain methylation, sites that were normally methylated tended to lose methylation, and partially-methylated regions shifted to both directions, accumulating at the extremities of the methylation scale (since larger and smaller shifts must converge at the borders). Due to this effect, cancers are overall more similar to each other than to the normal tissues (Figure 5B). We also analyzed the overall shift in the methylation of particular genomic sections. As expected, promoters, which are generally unmethylated in normal cells, tended to gain methylation in the cancer samples (Figure 5C), while the rest of the genome was slightly hypomethylated (Figure 5D). Surprisingly, high-scoring enhancer sites demonstrated the most dramatic shift between normal and cancer cells: they were partly methylated in the normal cells, and become highly methylated in the cancer (Figure 5E).

With these baseline analyses in hand, we then asked whether any particular group of methylation sites behaved as expected from the global trends, or stand alone. Figure 5F,G shows the average alteration of a given genomic site in cancer, as a function of its methylation level in the normal cells. For example, if a given site was 10% methylated in the normal cells, its average methylation shift in the cancer, based on the analyses shown in Figure 5A,B, was about plus 30% methylation. We found that promoters (green lines in Figure 5E,G) were behaving more or less as expected from the global trend of the genome (black lines). In other words, the alteration of promoter methylation in cancer was similar to the alteration of non-promoter sites with similar normal methylation levels. Looking more carefully into the promoter classes, it appeared that active promoters tended to maintain low methylation against the global

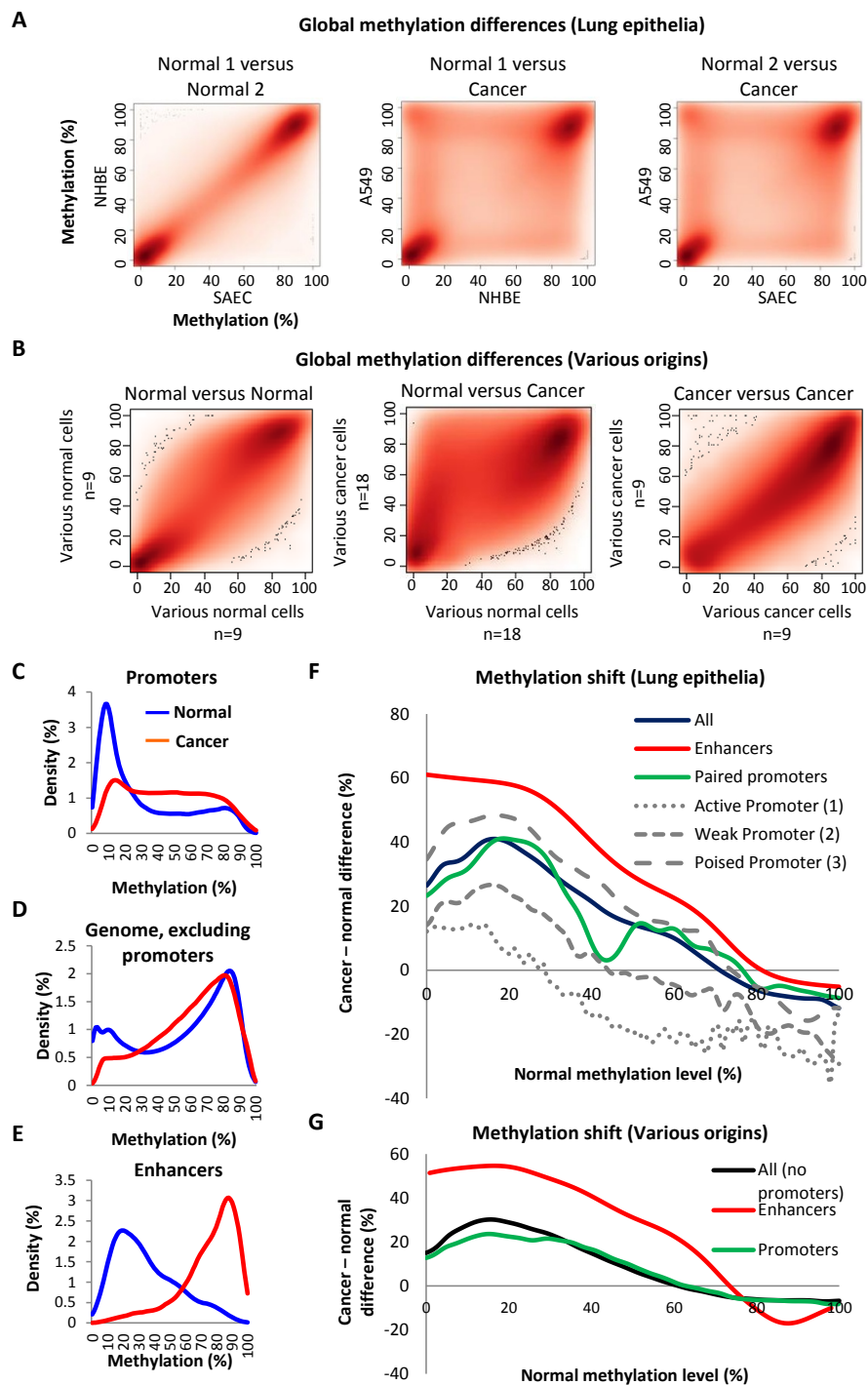


Figure 5 Enhancer methylation may be specifically altered in cancer. (A,B) x-y scatters showing the methylation of sites across the genome in a given cell type (or a collection of cell types) versus another cell type (or types). (A) Normal and cancerous lung epithelia: genome-wide methylation of normal versus normal (left) or of cancer versus normal (middle and right) cell types. The given cell types are indicated, as listed in table S2. (B) Normal and cancerous cell types of various tissue origins: genome-wide methylation of nine normal cell types of various tissue origins, versus another nine cell types (left), of 18 cancer versus 18 normal cell types (middle), or of nine cancers versus other cancers (right). (C-E) Distributions of methylation levels in the collection of normal and cancer samples shown in B, in promoter sites, in all sites excluding promoters, or in the high-scoring enhancer sites. (F) Average gain or loss of methylation in cancer, as a function of the methylation level in normal cells: for any given level of methylation in the normal cells, the graph shows the average change in the cancer, according to the global trend shown in B. Data are shown for all sites; for sites in the high-scoring enhancers; for sites in the promoters paired with the high-scoring enhancers; or for sites in strong, weak or poised promoters. The analysis is based on methylation levels and chromatin states in human mammary epithelial cells and MCF7 cells shown in A. (G) Same analysis as in F, but for 18 normal and 18 cancer cell types shown in B.

trend, whereas poised (*polycomb*-repressed) promoters gained more methylation than expected from the global trend. Yet, the overall alteration of promoter methylation in cancer was relatively close to that expected from the global trend. The shift in the methylation of the high-scoring enhancers, however, was clearly beyond what could be expected from the global trend (Figure 5F,G). Thus, the alteration of the high-scoring enhancer sites in cancer is probably not simply the outcome of the global loss of methylation control in cancer cells. Rather, these sites may be under the effect of a more specific mechanism, such as targeted methylation or demethylation and/or cell selection.

Conclusions

The mapping of associations between distal regulatory sites and the genes they control is a challenging task, which only recently began to be confronted on the genome-wide scale. Attempts to predict gene-enhancer pairs were based on the profiling of chromatin states or transcription factor binding [27,34], or of long-range DNA looping [28,29]. Here, we show that enhancers can be also associated with genes using DNA methylation. In contrast to the above mapping approaches, methylation data are readily available and are highly quantitative, and thus may enhance mapping of gene-enhancer pairing.

We found that distal expression-related methylation sites are abundant in the human genome, co-localizing with enhancer chromatin marks, and are more predictive of expression levels than promoter methylations. While not all distal regulatory sites in the genome must exhibit promoter-like methylation, we showed that a large number of enhancer sites demonstrate reverse correlation between methylation and expression, as in gene promoters.

We have further shown that hypomethylation state is directly related to enhancer activity across cell types (Figure S3 in Additional file 1). The observation that hypomethylated enhancers bind more transcription factors than methylated ones (Figure 2E) suggests a possible mechanism underlying the connection between DNA methylation and enhancer activity. Consistent with this possibility, high-scoring enhancers are particularly enriched within a defined chromatin state (Figures 2D), which is particularly hypomethylated compared to non-regulatory chromatin (Figure S3C in Additional file 1). Whether this chromatin state holds particularly active enhancers, or perhaps a unique class of methylation-related enhancers, remains to be elucidated.

The range of cell types we analyzed in this study was determined by the availability of methylation and expression data. In addition, the RRBS and Infinium Human-Methylation450 BeadChip methylation data we used provide limited genomic coverage and are strongly biased towards promoters and certain other portions of the

genome, while enhancers are not efficiently targeted. Because of this, it is likely that more complete methylomic coverage will expose many additional enhancer-gene pairs. Whole-genome bisulfite sequencing approaches have recently become popular and whole methylome analyses of human tissues are rapidly accumulating. Utilizing our approach, these additional data should allow the production of more comprehensive maps of enhancer-gene pairing across tissues, cell types and conditions.

Unmethylated promoters are permissive for, but do not necessarily determine, transcription initiation. We showed that enhancer methylation associates with cell-type-specific expression levels, even when the promoter is constantly unmethylated (Figure 3). Moreover, enhancer methylation characterizes small (and larger) expression differences. Thus, enhancers are not just on-off switches of cell-type transcription levels, as previously suggested, but may also mediate ranges of expression levels across multiple cell types. In contrast to the traditional model of one enhancer site per cell type, we suggest that a gradient of methylation states at a single enhancer site may direct distinct expression levels in many different cell types (Figure 3C).

In occasional examples, enhancer methylation level has been suggested to be associated with the control of cancer-related genes [35-37]. However, to our knowledge this is the first report on a global association between perturbed enhancer methylation and aberrant expression of cancer genes. We have shown that hypomethylated enhancers associated with the upregulation of many cancer genes controlling various cellular functions (Figure 4C), some of them involved in cell proliferation and some in other cancer-related processes. Moreover, many of these hypomethylated genes were found in most cancer types examined, suggesting a pan-cancer mechanism. However, the larger group of hypermethylated enhancers seemed to target cancer-type-specific genes. Given the limited genomic coverage of this study, many additional cancer-related enhancers are expected in the genome.

To date, almost all studies of cancer-related methylation have focused on gene promoters and CpG islands. Among these, the predominant event in cancers is hypermethylation of *polycomb*-repressed promoters [9-11]. This hypermethylation does not directly affect expression levels, as the associated genes are inactive in the normal tissue and generally remain inactive in the cancer (although it may limit the potential for re-activation of silenced genes in the cancer). Here, we established a very different occurrence in the other large group of regulatory sites - the transcriptional enhancers. These sites are drastically altered in cancers, to both hypo- and hypermethylation, and are closely related to substantial modifications in the expression levels of cancer-related genes

(Figure 4). Moreover, their aberrant methylation in cancers might derive from targeted methylation or demethylation or from selection of the altered cells (Figure 5). Whether targeted or selected, aberrant enhancer methylation may be involved in important events during cancer development.

We have uncovered a class of distal methylation sites that closely describe cell-type transcription levels. These sites reside in a particular subclass of transcriptional enhancers and are associated with cell-type-specific enhancer activity, possibly through communication with the binding of transcription factors. Methylation levels of these enhancers associate with gradual expression differences across cell types, even when the linked promoters are consistently unmethylated across the cell types. The radical changes in methylation of these sites in cancer is beyond that expected from the general profile of the cancer methylome, and may reflect specific targeting of the methylation and demethylation machinery to these sites, and/or functional contribution to tumor development. Further analyses of these sites may provide crucial information about paradigms of gene expression control in normal and cancerous cells.

Materials and methods

Imported datasets

The following datasets were downloaded from the ENCODE website [38]:

1. DNA methylation data for 58 cell types, for which expression data were also available. Methylation data were the average of two RRBS experiments (52 of the cell types, as detailed in Table S1 in Additional file 1) or Infinium HumanMethylation450 BeadChip data (36 of the RRBS cell types plus additional six cell types for which only BeadChip data were available).

2. Expression data for the above cell types produced by the Affymetrix Human Exon array. Gene-level expression data for 17,862 genes was extracted from the raw data following Robust Multi-chip Average normalization.

3. Histone chromatin immunoprecipitation (ChIP)-sequencing peaks and ChromHMM annotation data available for six of the above cell types, including K562, human mammary epithelial cells, human skeletal muscle myoblasts, normal human lung fibroblasts, GM12878 and HepG2 cells.

4. ChIP-sequencing peaks data for 122 DNA binding factors in various cell types.

Human whole-genome methylation data were from Lister *et al.* [39]. Genomic locations are according to human genome version hg19 of the human genome. Gene sizes, start and end sites, promoters and alternative promoters are according to the RefSeq database (National Center for Biotechnology Information, Bethesda, MD, USA). Methylation and expression data

for somatic tissues of two individuals are from [30] ([GSE:30654] patients #1 and 2).

Data filtering and categorization

VMSs were CpGs with standard deviation >5 after omitting outliers utilizing the Thomson's Tau method ($\alpha = 0.1\%$). A 'sample' (gene-CpG pair) was the two-dimensional matrix describing the methylation levels of a given CpG site as a function of the expression levels of a given gene across the cell types. For the RRBS data, only methylation sites that were sequenced at least 10 times in at least 60% (80% for the Support Vector Machine (SVM) algorithm) of the cell types were included.

Machine learning

The following four features were extracted for each CpG-gene sample: positive Pearson coefficient (0 if negative); negative Pearson coefficient (0 if positive); positive Spearman coefficient (0 if negative); and negative Spearman coefficient (0 if positive). Based on these features, we learned a model distinguishing false from true samples, applying a machine-learning algorithm. The SVMmap algorithm is a SVM algorithm for predicting rankings. It performs supervised learning using binary labeled training examples, with the goal of optimizing mean average precision (MAP) [40]. We used SVM-MAP for ranking the CpG-gene samples in 3,434 gene queries (RRBS data in 52 cell types) containing 15,205 true samples and a 50-fold excess of randomly selected false samples, where the goal was fitting a model that ranks the true samples above the false samples. The regularization parameter was optimized by dividing the gene queries into training and test sets (4:1 ratio) and performing 10-fold cross-validation. Query P -values (the probability of obtaining the top ranked true sample by chance) were calculated for the test set queries as follows:

We denote $X = \min_T(\text{rank})$, where T represents the true samples. The event that $X = x$ is equivalent to selecting $|T| - 1$ samples from $\{x + 1, x + 2, \dots, n\}$. Therefore the probability of this event can be obtained by the hypergeometric distribution:

$$\Pr(X = x) = \frac{\binom{n-x}{|T|-1}}{\binom{n}{|T|}}$$

And the probability that $X \leq \min_T(\text{rank})$ is:

$$\Pr(X \leq x) = \sum_{i=1}^x \frac{\binom{n-i}{|T|-1}}{\binom{n}{|T|}}$$

The resulting model for the methylation-expression relationship consisted of the following learned features: positive Pearson coefficient = -0.0058, negative Pearson coefficient = 0.0104, positive Spearman coefficient = -0.0038, and negative Spearman coefficient = 0.0088.

Applying the learned model to the datasets

The model learned in gene promoters was applied to all CpGs in 17,862 gene intervals (1 Mb upstream, gene-body, 1 Mb downstream to each gene), resulting in 14,702,075 CpG-gene pairs. The scores were normalized to a 0 to 1 range ($c \in C$ a CpG site, $g \in G$ a gene):

$$\text{Score}(c, g) = \frac{\text{model, features}(c, g) - \min(s(C, G))}{\max(s(C, G)) - \min(s(C, G))}$$

Data analyses

Frequency map of high-scoring sites

The distribution of methylation-expression relationships of the high-scoring sites over the cell types is a contour representation of the model learned to distinguish true from random pairs. The frequency map shown in Figure 2B is a representation of the corresponding scatter plot (SVM-MAP for 25×25 grids of high-scoring pairs), with smoothing of each point around a sphere (radius = 3).

Enrichment of chromatin marks around the high-scoring sites

ChIP-seq data ('modification score') were downloaded from the ENCODE website, averaged across the analyzed cell types, normalized to a 0 to 1 scale, and averaged across the high-scoring methylation sites. Loess smoothing with a span of 10% was applied to the presented data.

Enrichment of chromatin states around the high-scoring sites

ChromHMM states were identified for the CpG sites across the cell types. (A state was called when a given site was found in a given state in at least one of the cell types. Sites may be related to more than one state). The actual number of sites in a given chromatin category and a given score level ('real'), was divided by the number obtained from shuffling the expression data between genes 10 times ('random'). P -values were calculated based on the normal distribution of the shuffled data.

Binding of transcription factors to the high-scoring sites

ChIP-sequencing peaks data for 122 DNA binding factors in four cell types (GM12878, HepG2, HeLaS3, K562) were downloaded. For every high-scoring site we counted the number of factors that bind, and averaged across the cell types. Random expectation and real versus random P -values were obtained as above. The P -value for the difference between methylated and unmethylated sites was calculated using the Wilcoxon rank-sum test for difference between averages.

Conserved sequences around the high-scoring sites

PhastConsdata (phastCons46wayPrimates and phastCons46wayPlacental) were downloaded from the University of California, Santa Cruz genome browser and plotted around the high-scoring sites.

Agreement with gene-enhancer pairing by long-range chromatin interactions

From 5C data produced at the University of Massachusetts in three cell lines (Gm12878, HeLa-S3, K562) we sorted out the enhancer sites that were probed by the 5C, for which we also have at least one CpG site available to our analysis at ± 500 base pairs from the probed enhancer. We then located the TSS that obtained the highest score in relation to the enhancer CpG. This yielded 318 interactions, at distances of 10 kb to 1 Mb between the CpG side and the TSS side. We then counted the number of cases in which the maximum number of interaction sequence reads (average of two repeats over three cell types) was higher than the average interaction read number (98.23 reads). P -values were obtained by comparing the binomial probability of the agreement in high versus low scoring sites. We also compared the number of 5C interaction reads in high-scoring (score ≥ 0.85) enhancer-TSS pairs, versus the number interactions between these enhancers and other TSSs at 10 kb to 1 Mb from the enhancer sites. The P -value of the difference was calculated by the Wilcoxon Rank-Sum test.

Replication in uncultured samples

Methylation and expression data for bladder, lymph node, ureter, lung, stomach, skeletal muscles and adipose tissue biopsies of two individuals were downloaded from a recently-published dataset [GSE:30654]. Out of the enhancer-promoter interactions that were predicted in the original set of the 58 cell types, 876 were also assessable in the new dataset (that is, methylation data were available for the relevant CpG sites, and expression data were available for the interacting genes). We re-scored these gene-CpG pairs, using our model, in the new dataset. If the interactions that were predicted in the 58 cell types were not relevant to interactions in the uncultured tissue samples, than we expected that 20.5 ± 4.5 out of the 876 would (by chance) obtain high-scoring (score > 0.9) interactions in individual #1, and 19.9 ± 3.5 in individual #2. Instead, we observed 29 in individual #1 ($P = 0.0145$) and 28 ($P = 0.0084$) in individual #2. The P -value (normal distribution, two-sided) indicated in the result and discussion section is for the average of both individuals (expected = 20.21 ± 3.206 , observed = 28.5 interactions, $P = 0.0098$).

Alteration in cancer

Differences between normal and cancer cells were measured by Robust Multi-chip Average expression units and by methylation percentages. The distributions of methylation and expression differences between normal

and cancer samples (Figure 4) were smoothed (color density representation) using a kernel density estimate (transformation function = x^3). Linear regressions were performed (red lines and Pearson coefficients in Figure 4). The *P*-value of the frequencies of signals in the upper-left and lower-right quarters, in enhancers versus promoters, were calculated using normal distribution of two-proportion *z* test. *P*-values of the overlap between genes in the cancer types were calculated based on the expectation from random intersections between the groups.

Gene ontology analysis

Analysis of GO terms of the hypomethylated upregulated genes was done with the GOrilla tool [41], using the list of high-scoring genes as a background.

Additional material

Additional file 1: Supplementary tables and figures. Table S1.

Human cell types and DNA methylation data that used in the development of the promoter-based model. **Table S2.** Normal and cancer cell types used in the study. **Table S3.** Genes that are hypomethylated and upregulated in the examined cancer types, compared with normal cells. **Table S4.** Enriched GO groups among the genes that were hypomethylated and upregulated in various cancers. **Table S5.** Hypomethylated upregulated genes in acute leukemia. **Figure S1.** The overall structure of the DNA methylation data. **Figure S2.** Methylation levels around TSSs as function of gene expression levels. **Figure S3.** Genomic sites are hypomethylated when marked as enhancer chromatin, compared with the methylation of the same sites in non-regulatory chromatin. **Figure S4.** An example of a long-range enhancer-promoter interaction captured by both methylation-based gene-enhancer pairing and long-range chromatin interactions assessed by the 5C technique. **Figure S5.** Representative examples of methylation-based gene-enhancer pairing.

Additional file 2: Data file. A CSV file (3.3 MB) denoting CpG position and gene symbols for the 118,417 CpG-gene pairs that obtained score >0.75.

Abbreviations

5C: chromosome conformation capture carbon copy technique; ChIP: chromatin immunoprecipitation; GO: gene ontology; MAP: mean average precision; kb: kilobase pairs; Mb: megabase pairs; RRBS: reduced representation bisulfite sequencing; SVM: Support Vector Machine; TSS: transcription start site; VMS: variable methylation site.

Authors' contributions

DA participated in the design of the study, collected and prepared the data, developed the promoter-based model, performed the analyses, and helped to draft the manuscript. SS aided in the development of the promoter-based model. AH conceived and supervised the study and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We are grateful to Shai Shalev-Shwartz for his helpful advice regarding the development of the machine-learning algorithm. Thanks are also due to Yehudit Bergman, Gadi Glaser, Amir Eden, Benjamin Glaser, Reuven Laskov and Howard Cedar for helpful discussions and comments. This work was supported by the Adams Fellowship Program of the Israel Academy of

Sciences and Humanities (to SS), by the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation (Grant No. 41/11 to AH), and by the Israel Science Foundation (Grant No. 887/08 to AH).

Author details

¹The Department of Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada, Faculty of Medicine, Hebrew University-Hadassah Medical School, Ein Kerem Campus, Jerusalem 91120, Israel. ²The Rachel and Selim Benin School of Computer Science and Engineering, Hebrew University of Jerusalem, Edmond J. Safra Campus, Jerusalem 91904, Israel.

Received: 23 August 2012 Revised: 21 February 2013

Accepted: 12 March 2013 Published: 12 March 2013

References

1. Klose RJ, Bird AP: Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 2006, **31**(2):89-97.
2. Hellman A, Chess A: Gene body-specific methylation on the active X chromosome. *Science* 2007, **315**(5815):1141-1143.
3. Suzuki MM, Bird A: DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 2008, **9**(6):465-476.
4. Zemach A, McDaniel IE, Silva P, Zilberman D: Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 2010, **328**(5980):916-919.
5. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, Ukomadu C, Sadler KC, Pradhan S, Pellegrini M, Jacobsen SE: Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci USA* 2010, **107**(19):8689-8694.
6. Aran D, Toperoff G, Rosenberg M, Hellman A: Replication timing-related and gene body-specific methylation of active human genes. *Hum Mol Genet* 2011, **20**(4):670-680.
7. Cedar H, Bergman Y: Programming of DNA methylation patterns. *Annu Rev Biochem* 2012, **81**:97-117.
8. Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, Lin JC, Liang G, Jones PA, Tanay A: Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc Natl Acad Sci USA* 2008, **105**(35):12979-12984.
9. Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, Bergman Y, Simon I, Cedar H: Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* 2007, **39**(2):232-236.
10. Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, Berman DM, Jenuwein T, Pruitt K, Sharkis SJ, Watkins DN, Herman JG, Baylin SB: A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* 2007, **39**(2):237-242.
11. Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, Laird PW: Epigenetic stem cell signature in cancer. *Nat Genet* 2007, **39**(2):157-158.
12. Easwaran H, Johnstone SE, Van Neste L, Ohm J, Mosbrugger T, Wang Q, Aryee MJ, Joyce P, Ahuja N, Weisenberger D, Collisson E, Zhu J, Yegnashubramanian S, Matsui W, Baylin SB: A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res* 2012, **22**(5):837-849.
13. Sproul D, Nestor C, Culley J, Dickson JH, Dixon JM, Harrison DJ, Meehan RR, Sims AH, Ramsahoye BH: Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. *Proc Natl Acad Sci USA* 2011, **108**(11):4364-4369.
14. De Carvalho DD, Sharma S, You JS, Su SF, Taberlay PC, Kelly TK, Yang X, Liang G, Jones PA: DNA methylation screening identifies driver epigenetic events of cancer cell survival. *Cancer Cell* 2012, **21**(5):655-667.
15. Lengauer C, Kinzler KW, Vogelstein B: DNA methylation and genetic instability in colorectal cancer cells. *Proc Natl Acad Sci USA* 1997, **94**(6):2545-2550.
16. Eden A, Gaudet F, Waghmare A, Jaenisch R: Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* 2003, **300**(5618):455.
17. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, Valsesia A, Ye Z, Kuan S, Edsall LE, Camargo AA, Stevenson BJ, Ecker JR, Bafna V, Strausberg RL, Simpson AJ, Ren B: Global DNA hypomethylation coupled

- to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res* 2012, **22**(2):246-258.
18. Raddatz G, Gao Q, Bender S, Jaenisch R, Lyko F: **Dnmt3a protects active chromosome domains against cancer-associated hypomethylation.** *PLoS Genet* 2012, **8**(12):e1003146.
 19. Bulger M, Groudine M: **Functional and mechanistic diversity of distal transcription enhancers.** *Cell* 2011, **144**(3):327-339.
 20. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, Schubeler D: **DNA-binding factors shape the mouse methylome at distal regulatory regions.** *Nature* 2011, **480**(7378):490-495.
 21. Xu J, Watts JA, Pope SD, Gadue P, Kamps M, Plath K, Zaret KS, Smale ST: **Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells.** *Genes Dev* 2009, **23**(24):2824-2838.
 22. Sandovici I, Smith NH, Nitert MD, Ackers-Johnson M, Uribe-Lewis S, Ito Y, Jones RH, Marquez VE, Cairns W, Tadayyon M, O'Neill LP, Murrell A, Ling C, Constanca M, Ozanne SE: **Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the Hnf4a gene in rat pancreatic islets.** *Proc Natl Acad Sci USA* 2011, **108**(13):5449-5454.
 23. Carone BR, Fauquier L, Habib N, Shea JM, Hart CE, Li R, Bock C, Li C, Gu H, Zamore PD, Meissner A, Weng Z, Hofmann HA, Friedman N, Rando OJ: **Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals.** *Cell* 2010, **143**(7):1084-1096.
 24. Taberlay PC, Kelly TK, Liu CC, You JS, De Carvalho DD, Miranda TB, Zhou XJ, Liang G, Jones PA: **Polycomb-repressed genes have permissive enhancers that initiate reprogramming.** *Cell* 2011, **147**(6):1283-1294.
 25. Wiench M, John S, Baek S, Johnson TA, Sung MH, Escobar T, Simmons CA, Pearce KH, Biddie SC, Sabo PJ, Thurman RE, Stamatoyannopoulos JA, Hager GL: **DNA methylation status predicts cell type-specific enhancer activity.** *EMBO J* 2011, **30**(15):3028-3039.
 26. Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Boyle P, Gnirke A, Fuchs E, Rossi DJ, Meissner A: **DNA methylation dynamics during in vivo differentiation of blood and skin stem cells.** *Mol Cell* 2012, **47**:633-647.
 27. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE: **Mapping and analysis of chromatin state dynamics in nine human cell types.** *Nature* 2011, **473**(7345):43-49.
 28. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, et al: **Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation.** *Cell* 2012, **148**(12):84-98.
 29. Sanyal A, Lajoie BR, Jain G, Dekker J: **The long-range interaction landscape of gene promoters.** *Nature* 2012, **489**(7414):109-113.
 30. Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, Garitaonandia I, Muller FJ, Wang YC, Boscolo FS, Fakunle E, Dumevska B, Lee S, Park HS, Olee T, D'Lima DD, Semechkin R, Parast MM, Galat V, Laslett AL, Schmidt U, Keirstead HS, Loring JF, Laurent LC: **Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives.** *Cell Stem Cell* 2012, **10**(5):620-634.
 31. Visel A, Rubin EM, Pennacchio LA: **Genomic views of distant-acting enhancers.** *Nature* 2009, **461**(7261):199-205.
 32. Akhtar-Zaidi B, Cowper-Salhari R, Corradin O, Saiakhova A, Bartels CF, Balasubramanian D, Myeroff L, Lutterbaugh J, Jarrar A, Kalady MF, Willis J, Moore JH, Tesar PJ, Laframboise T, Markowitz S, Lupien M, Scacheri PC: **Epigenomic enhancer profiling defines a signature of colon cancer.** *Science* 2012, **336**(6082):736-739.
 33. Oehler VG, Guthrie KA, Cummings CL, Sabo K, Wood BL, Gooley T, Yang T, Epping MT, Shou Y, Pogossova-Agadjanyan E, Ladne P, Stirewalt DL, Abkowitz JL, Radich JP: **The preferentially expressed antigen in melanoma (PRAME) inhibits myeloid differentiation in normal hematopoietic and leukemic progenitor cells.** *Blood* 2009, **114**(15):3299-3308.
 34. Rodelsperger C, Guo G, Kolanczyk M, Pletschacher A, Kohler S, Bauer S, Schulz MH, Robinson PN: **Integrative analysis of genomic, functional and protein interaction data predicts long-range enhancer-target gene interactions.** *Nucleic Acids Res* 2011, **39**(7):2492-2502.
 35. Tekpli X, Zienoldiny S, Skaug V, Stangeland L, Haugen A, Mollerup S: **DNA methylation of the CYP1A1 enhancer is associated with smoking-induced genetic alterations in human lung.** *Int J Cancer* 2012, **131**(7):1509-1516.
 36. Steinmann K, Richter AM, Dammann RH: **Epigenetic silencing of erythropoietin in human cancers.** *Genes Cancer* 2011, **2**(1):65-73.
 37. Angrisano T, Sacchetti S, Natale F, Cerrato A, Pero R, Keller S, Peluso S, Perillo B, Avvedimento VE, Fusco A, Bruni CB, Lembo F, Santoro M, Chiariotti L: **Chromatin and DNA methylation dynamics during retinoic acid-induced RET gene transcriptional activation in neuroblastoma cells.** *Nucleic Acids Res* 2011, **39**(6):1993-2006.
 38. **A user's guide to the encyclopedia of DNA elements (ENCODE).** *PLoS Biol* 2011, **9**(4):e1001046.
 39. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR: **Human DNA methylomes at base resolution show widespread epigenomic differences.** *Nature* 2009, **462**(7271):315-322.
 40. Yue Y, Finley T, Radlinski F, Joachims T: **A support vector method for optimizing average precision.** *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval: July 23-27 2007; Amsterdam, The Netherlands* New York: ACM; 2007, 271-278.
 41. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z: **GORilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.** *BMC Bioinformatics* 2009, **10**:48.

doi:10.1186/gb-2013-14-3-r21

Cite this article as: Aran et al.: DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biology* 2013 **14**:R21.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

