Quantitative Biology 2017, 5(4): 328–337
DOI 10.1007/s40484-017-0129-y

RESEARCH ARTICLE

Transcription regulation by DNA methylation under stressful conditions in human cancer

Sha Cao[1,†], Yi Zhou[1,†], Yue Wu[1], Tianci Song[1,2], Burair Alsaihati[1,3] and Ying Xu[1,2,*]

[1] Computational Systems Biology Lab, Department of Biochemistry and Molecular Biology, Department of Statistics and Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA
[2] College of Computer Science and Technology and School of Public Health, Jilin University, Changchun 130012, China
[3] National Center for Genomics Research (NCGR), King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia
* Correspondence: xyn@uga.edu

Received July 25, 2017; Revised September 26, 2017; Accepted October 16, 2017

***Background*: We aim to address one question: do cancer *vs.* normal tissue cells execute their transcription regulation essentially the same or differently, and why?**

***Methods*: We utilized an integrated computational study of cancer epigenomes and transcriptomes of 10 cancer types, by using penalized linear regression models to evaluate the regulatory effects of DNA methylations on gene expressions.**

***Results*: Our main discoveries are: (i) 56 genes have their expressions consistently regulated by DNA methylation specifically in cancer, which enrich pathways associated with micro-environmental stresses and responses, particularly oxidative stress; (ii) the level of involvement by DNA methylation in transcription regulation increases as a cancer advances for majority of the cancer types examined; (iii) transcription regulation in cancer *vs.* control tissue cells are substantially different, with the former being largely done through direct DNA methylation and the latter mainly done via transcriptional factors; (iv) the altered DNA methylation landscapes in cancer *vs.* control are predominantly accomplished by *DNMT1*, *TET3* and *CBX2*, which are predicted to be the result of persistent stresses present in the intracellular and micro-environments of cancer cells, which is consistent with the general understanding about epigenomic functions.**

***Conclusions*: Our integrative analyses discovered that a large class of genes is regulated via direct DNA methylation of the genes in cancer, comparing to TFs in normal cells. Such genes fall into a few stress and response pathways. As a cancer advances, the level of involvement by direct DNA methylation in transcription regulation increases for majority of the cancer types examined.**

Keywords: DNA methylation; transcriptional regulation; micro-environment stress

INTRODUCTION

Epigenetic regulation refers to chemical modifications to the genomic DNA or its binding histones that can influence gene expressions, which do not change their component nucleotides or amino acids [1]. Such modifications are reversible and some are short-term heritable across a few generations. The best studied epigenetic modifications are DNA methylation and histone modifications such as phosphorylation, acetylation and ubiquitination. It has been well established in a few organisms that the dynamics of an epigenome plays a vitally important role in the evolutionary adaptation of an organism to its environmental stresses, such as: (i) hypomethylation of the *NtGPDL* gene helps tobacco adapt to salt and cold stresses [2]; (ii) *Drosophila larvae* survives heat-shock and osmotic stresses through inheriting specific phosphorylation of the dATF-2 protein, which results in disruption of the heterochromatin [3]; and (iii) natural and social environments can alter human behavior

[†] These authors contributed equally to this work.

and psychology through epigenetic modification, which can pass on to the next generation [4,5].

Compared to these research fields, epigenomic research in cancer is somewhat behind as the published cancer epigenomic studies are largely at a stage of documenting which epigenome-related genes are mutated in cancer; which genes have their epigenetic levels altered in cancer *vs.* control tissues, such as DNA hypomethylation at a genome scale in cancer in general [6] and DNA hypermethylation in certain tumor suppresser genes [7]; and the possible functional effects of specific epigenetic changes. While some information has been accumulated about cancer epigenomes, we are yet to understand, in general, why the observed epigenomic changes take place in cancer. Hence, a result has been that the detected epigenomic changes in cancer remain largely functionally disconnected from each other without a common underflow that naturally links them. To the best of our knowledge, published studies have yet to link epigenomic modifications in cancer to intracellular or micro-environmental stresses like in the aforementioned-research areas. This is possibly due to the reality that our understanding about what stresses may contribute to cancer initiation, progression and metastases is quite limited, even though it has been widely observed that the general stress levels in cancer tend to be considerably higher than normal and other chronic disease cells. These general stresses are reflected by highly up-regulated general stress response pathways such as heat-shock pathway [8] and endoplasmic reticulum (ER) stress responses [9], as well as oxidative stress and hypoxia-related stress [10]. However, which stresses may play the defining roles in driving the disease remains largely unanswered.

Currently, the triggering signals, the responding mechanisms as well as the cellular functions of DNA methylation remain largely unknown. For example, the recent knowledge about the functional roles of DNA methylation in transcription regulation has challenged the traditional view of DNA methylation as simply a "silencing" mechanism of transcription of a gene. Other than just causing chromatin remodeling, hence resulting in the formation of heterochromatin [11,12], the functions of DNA methylation may vary across different genomic positions. For example, DNA methylation in promoters generally repress gene expression by preventing transcription factor binding [13], whereas methylation in the gene body has been found to positively correlate with the expression of the gene [14]. It was suggested that the knowledge of the detailed patterns of DNA methylation associated with genes of specific functions could be a key to understanding the relationship between DNA methylation and transcriptional regulation [15].

Here, we present an integrated computational study of cancer epigenomes and transcriptomes of 10 cancer types, to address four issues: (i) what genes, in terms of their functions, tend to have their transcriptions affected by their associated methylations in cancer? (ii) how do transcription factors and DNA methylation of a gene contribute to the transcription of the gene differently in cancer *vs*. control tissue cells and why? (iii) what are the possible causes of DNA methylation changes in cancer? and (iv) how do DNA methylations in different parts of a gene affect the gene's transcription? Through addressing these questions, we have discovered some general patterns of how stresses of multiple types contribute to the observed differences in transcription regulation in cancer *vs.* normal tissue cells.

RESULTS

The landscape of DNA methylation in cancer

We have conducted differential methylation analyses of ~400,000 CpG sites distributed across the entire human genome as well as differential expression analyses of ~14,000 genes, between cancerous and control tissue samples of 10 cancer types, namely BLCA (bladder urothelial carcinoma), BRCA (breast invasive carcinoma), COAD (colon adenocarcinoma), HNSC (head and neck squamous cell carcinoma), KIRP (kidney renal papillary cell carcinoma), LIHC (liver hepatocellular carcinoma), LUAD (lung adenocarcinoma), LUSC (lung squamous cell carcinoma), PRAD (prostate adenocarcinoma), and THCA (thyroid carcinoma). The sample size and the availability of the DNA methylation and gene-expression data for the 10 cancer types are detailed in the Section of Data and Methods.

A gene is defined as having cancer specific methylation-regulated expression (MRE) if (i) the gene is differentially expressed in cancer *vs.* control tissues; (ii) the gene is differentially methylated in at least one of its proximal CpG sites in cancer *vs.* control; and (iii) the differentially methylated (proximal) CpG site(s) have a significant contribution to the gene's expression in cancer. (i) and (ii) are examined by using the *t* test and Wilcoxon test, respectively; and the *p*-values are adjusted for false discovery rate by Holm corrections [16], and the significance cutoff is set to be 0.05. To address (iii), a penalized linear regression model is developed between a gene's expression and the methylation levels of all its proximal CpG sites. The regression model selects a subset of predictors, namely CpG sites, that achieves the highest cross-validation accuracy. A CpG site is said to have significant contribution to its gene's expression, if the site is selected by the regression model, with details given in in the Section of Data and Methods.

Table 1 summarizes the analysis results with the following note: among those genes that are significantly

Table 1. Differentially expressed genes and differentially methylated CpG sites.

Cancer type	#diff met CpG	#diff met genes	#diff expr genes	#diff expr & met genes	#MRE genes
BLCA	19,599	6,363	2,199	1,140	527
BRCA	81,841	12,337	10,237	8,877	5,355
COAD	33,375	8,006	8,585	4,984	1,627
HNSC	50,866	10,522	4,730	3,612	1,668
KIRP	49,094	10,759	7,458	5,848	1,781
LIHC	27,576	7,445	7,543	4,103	1,066
LUAD	38,544	9,295	9,347	6,117	4,848
LUSC	122,369	13,807	10,204	9,853	6,055
PRAD	50,377	11,333	5,936	4,846	2,585
THCA	25,920	8,313	7,044	4,254	1,577

Columns 1−6: (1) cancer type; (2) the total number of differentially methylated CpGs; (3) the number of genes with at least one of its CpG sites differentially methylated; (4) the number of differentially expressed genes; (5) the number of genes that are both differentially expressed and methylated in at least one of its CpG sites; and (6) the number of MRE genes.

differentially expressed, a majority has at least one of its CpG sites differentially methylated.

Three genes are common MREs shared by all ten cancer types: *ANK2*, *SYNPO2*, and *TGFBR3*, which are all related to cell morphology. We then conducted a pathway enrichment analysis over those MRE genes shared by at least 8 cancer types, totaling 56 genes, to understand the cellular functions of these genes, with detailed information of the 56 genes given in Supplementary Table S1. 29 pathways are statistically enriched (with p-value < 0.05) against the union of KEGG, REACTOME and Gene Ontology in the Msigdb database [17]. The details of the 29 enriched pathways can be found in Supplementary Table S2, and the MREs enriched pathways for each cancer type is listed in Supplementary Table S3. We noted that the two most significantly enriched pathways are N-linked glycosylation on proteins and *NFE2L2*-regulated genes, where N-linked protein glycosylation is reported to be induced by ER stress [18], possibly as a key exit for the substantially increased uptake of glucose in cancer in general; and *NFE2L2* is the master regulator of oxidative stress [19]. The rest of the pathways falls into the following categories: (i) cell morphogenesis and cell-cell adhesion; (ii) extracellular matrix; (iii) cell migration; (iv) post-translational modification; (v) nervous system development and signaling (axon guidance, axonogenesis, neurite development, neuron development, neuron differentiation, generation of neurons, and neurogenesis), and (vi) general signaling (G alpha I signaling events, response to biotic stimulus, and G protein coupled receptor activity). While these pathways can be activated by a wide range of conditions, one thing in common is: they can all be activated by intracellular stresses, particularly oxidative stress [20–22], and associated damages. Actually, all these pathways are strongly associated with responses to Fenton reactions as noted in our previous study [23]. Hence, we postulate that these pathways are transcriptionally regulated via DNA methylation in response to severe intracellular and micro-environmental oxidative stress.

To investigate if the contribution level by DNA methylation towards transcription regulation may change as a cancer advances, we have conducted a similar analysis to the above on cancer samples grouped based on their stages for each cancer type, and counted the average number of the MRE genes for each stage. The stages of the cancer samples are assigned based on the classification of the AJCC (American Joint Committee on Cancer) pathologic tumor stages. Figure 1 summarizes the result. Overall, as a cancer advances, the average number of MRE genes increases, except for the breast and thyroid cancers. This strongly suggests: as a cancer advances, it tends to utilize more epigenetic level regulation of the expressions of genes of certain functions, as way to adapt to severe oxidative stress and possibly others. We speculate that breast and thyroid cancers do not follow the same trend because other factors such as hormones may also contribute to the transcriptional regulations of the genes, aside from DNA methylation.

Genes whose transcription is partially regulated by DNA methylation in cancer *vs.* controls

We have then conducted an analysis to assess the level of contribution to the transcription of individual genes by transcription factors *vs.* by DNA methylation levels in the proximal CpG sites of each gene, respectively (see the Section of Data and Methods). The analysis was conducted on ~9,000 human genes with ~900 known and experimentally validated transcription factors (TFs), using the methylation levels of the gene's proximal CpG sites and its TFs' CpG sites, where these genes are

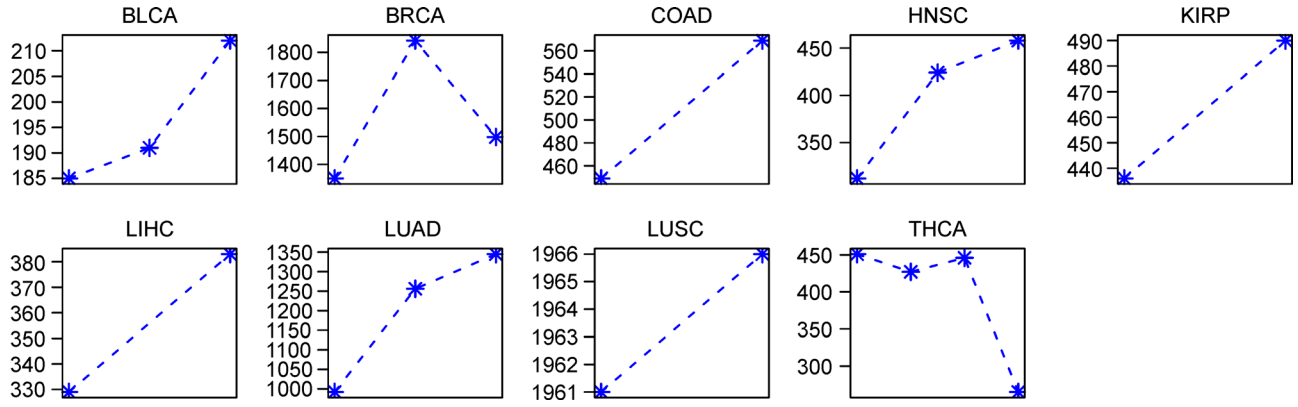

Figure 1. The number (y-axis) of MRE genes for each of the ten cancer types as a cancer advances from stage 1 through stage 4 along the x-axis. Note: prostate cancer (PRAD) is not included here since there is not enough information to derive the AJCC stages for the cancer type.

selected because they have at least one experimentally validated TF. For each gene, a linear regression model was built using its expression level as the response variable, and the methylation levels of the proximal CpGs of its TFs and of the gene itself as predictors. The procedure was applied to cancer and control samples, separately. Knowing that the numbers of the cancer and control samples differ considerably (a reality of the TCGA database where our tissue data are collected), we forced the regression model to select the same number of predictors, namely 2, 3, 4, 5, on both the cancer and the control samples to allow for fair comparisons. A goal of this analysis is to determine the possibly differential roles in genes transcription played by TFs and DNA methylation in cancer *vs.* control tissues. Figure 2 summarizes the numbers of predictors, which is fixed at 2, selected either in the proximity of each gene or its TFs, in control and cancer samples for each of the 10 cancer types. Similar plots for using 3, 4 or 5 predictors can be found in Supplementary Figure S1.

We see clearly from the figure that in controls, the predictors for a gene's expression are mostly its TFs' CpGs but in cancer tissues, the predictors are largely the methylation levels of CpGs of the gene itself, revealing that *for a large subset of human genes, the transcription regulation is done via different mechanisms in cancer vs. normal tissue cells*, the first such report to the best of our knowledge. Our interpretation of this observation is: cancer cells generally live in a more stressful environment, compared to normal tissue cells. Some of the stresses may be novel to the cells, making the stress-responsive system encoded at the genomic level ineffec-

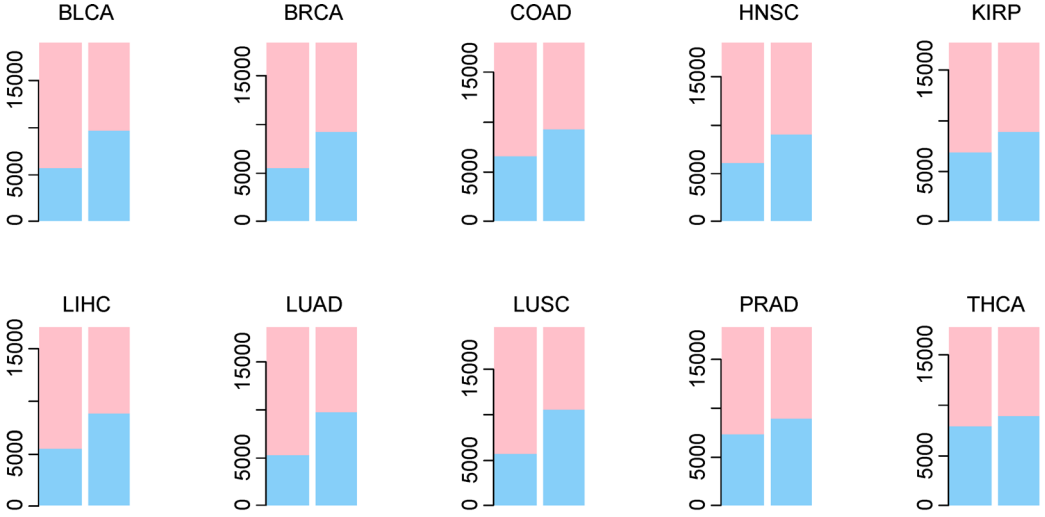

Figure 2. The numbers of selected predictors in control (left bar) and cancer (right bar) for each of the 10 cancer types, where the selected CpGs in the proximity of the genes are marked in blue and proximity of genes' TFs in pink. Note: the number of predictors selected for the linear models are fixed to be two.

tive or inefficient. It has been well established in the plant and *Drosophila* literature that novel stresses can be responded through epigenomic level activities [24], which can be passed on to the next few generations. We postulate: cancer tissue cells have to constantly deal with (novel) stresses, which may have led to increased epigenomic level stress-responses, such as DNA methylation as observed here. There is a clear advantage for doing so in cancer, as the rapidly proliferating cells in cancer will need to face similar stresses down the road, and such response mechanisms passed on from one generation to the next, would be more efficient for cancer cells to adapt and survive.

Identification of stresses that may have triggered DNA methylation/demethylation genes in cancer

To further investigate the functional roles of DNA methylation in transcription regulation of the above genes and pathways, we have conducted the following analysis. We compared the activity levels of DNA methylation (methyltransferases) as well as de-methylation genes (TETs and TDG) in cancer *vs.* the matching controls. We noted: majority of the DNA methylation and de-methylation genes is up-regulated significantly in cancer *vs.* controls (Supplementary Table S4), indicating that DNA methylation as well as de-methylation enzymes are more active in cancer than in controls, revealing that cancer cells utilize more epigenomic level regulations to cope with the (stressful) conditions in cancer. Interestingly, it has been well established that cancers tend to have reduced DNA methylation levels at the genome scale [6]; and our previous work has offered a mechanistic model to explain why this is the case [25]. Our discovery here clearly enriched the previous observation, namely, while the genome-scale DNA methylation is reduced, the activity levels of DNA methylation and de-methylation enzymes are more active in vicinity of the protein-encoding regions.

The DNA methylation enzyme genes used in the study are the DNA methyltransferase: *DNMT1*, *DNMT3A*, *DNMT3B* [26]; and the DNA de-methylation enzymes genes are: *TET1-3* and *TGD* [27,28]. In addition, the following genes are used to encode the Polycomb complex, a key regulator of epigenetic activities: *EZH2, EZH1, EED, SUZ12, RBBP4, RBBP7, RING1, RNF2, CBX2, CBX4, CBX6, CBX7, CBX8, PHC1, PHC2, PHC3, PCGF1, PCGF2, PCGF3, BMI1, PCGF5, PCGF6, ZNF134, SCMH1*. We started by searching for the determining factors of the global DNA methylation levels. We have conducted co-expression analyses among the genes encoding the DNA methylation machinery: DNA methylation enzyme, de-methylation enzymes and the Polycomb complex genes, for this search. We observed that *DNMT1*, *TET3* and *CBX2* tend to be highly co-expressed with other genes or highly connected in the co-expression networks across the 10 cancer types, hence suggesting their dominant roles among all these genes. We then built a linear model using the expressions of the three genes as the predictors and the total DNA methylation level in the relevant cancer samples as the response variable. We noted: the model can correctly predict the significantly positive and negative contributions to the total DNA methylation level by the expression levels of *DNMT1* and *TET3* in seven out of the ten cancer types; and that the expression levels of *DNMT1*, *TET3* and *CBX2* can well represent the overall methylation activities. Table 2 gives the coefficients and their statistical significances of the three genes in predicting the global DNA methylation level using the linear model across the 10 cancer types.

The intracellular and micro-environments of cancer cells may have specific stresses of the following types: persistent hypoxia, severe oxidative stress and stresses

Table 2. The coefficients in the linear models corresponding to the three enzymes for the 10 cancer types, along with the statistical significance of each predictor inside the parentheses.

Cancer type	Coefficient (Statistical significance)		
	TET3	*DNMT1*	*CBX2*
BLCA	−5.30e-03 (4.24e-02)	1.15e-02 (2.14e-04)	9.38e-04 (5.37e-01)
BRCA	−1.65e-03 (2.65e-01)	−8.06e-04 (6.44e-01)	1.29e-03 (2.87e-02)
COAD	−4.89e-03 (6.73e-02)	1.44e-02 (5.63e-06)	9.38e-05 (9.39e-01)
HNSC	−7.07e-03 (3.98e-05)	7.54e-03 (2.59e-04)	−1.27e-03 (2.94e-01)
KIRP	−5.24e-03 (2.37e-02)	6.05e-03 (3.85e-02)	1.79e-03 (7.20e-02)
LIHC	8.04e-03 (9.17e-02)	8.57e-04 (8.53e-01)	−1.24e-03 (6.29e-01)
LUAD	−7.22e-03 (2.49e-04)	4.53e-03 (1.96e-02)	−3.35e-03 (1.01e-05)
LUSC	−5.77e-03 (5.97e-02)	1.48e-04 (9.65e-01)	−8.88e-04 (5.69e-01)
PRAD	3.68e-03 (1.74e-02)	−4.22e-03 (8.51e-02)	5.70e-04 (4.92e-01)
THCA	−6.29e-03 (2.36e-10)	2.24e-03 (6.09e-02)	2.57e-04 (7.28e-01)

induced by rising intracellular pH [23,29–31]. We have previously demonstrated that persistent hypoxia will lead to a persistent gap in ATP demand and supply [32]. As a response, the affected cells tend to substantially increase their glucose uptake and metabolism, which serves as the basis of PET/CT based cancer diagnosis. Persistent oxidative stress, particularly that induced by innate immune cells, can give rise to persistent Fenton reactions and production of hydroxyl radicals [33], the most damaging molecules that human cells can generate, leading to persistent and extensive damages to the host cells, including proteins, nucleotides and lipids. Another and less studied consequence of Fenton reactions is the persistent production of OH^-, which will ultimately alter the intracellular pH [23]. All these represent novel stresses that tissue cells generally have not encountered before, at least not in a persistent manner, which could be key reasons for the increased utilization of epigenomic level activities. Hence, we hypothesize: *it is the combination of such novel stresses that leads to the increased epigenomic level activities, which determine how the transcription of genes of certain functions is executed, specifically having reduced needs for TFs in transcription regulation of these genes*.

We have conducted co-expression analyses between *TET3, DNMT1, CBX2* and the rest of all the ~14,000 genes except for the DNA (de)methylation and Polycomb complex, to predict processes that are strongly associated with DNA methylation activities. Our co-expression analyses revealed that: immune attack, DNA damage, and oxidative stress are the three processes that are most consistently and significantly associated with methylation activities (see the Section of Data and Methods for more details). Table 3 shows the significances of these three pathways. These data, coupled with the results in the previous section, provide strong evidence in support of our above hypothesis.

Elucidation of how DNA methylation in different locations contribute to transcription regulation

Here, we examine how the methylation level in different CpG sites around each gene, namely 5′ UTR, promoter, transcription start site (TSS), gene body, and 3′ UTR, may affect the expression level of the gene. Specifically, we have selected a subset of the CpG sites in each gene's vicinity to predict the gene expression level by using a penalized linear regression model (see the Section of Data and Methods). We observed that for those genes whose expression levels are significantly up-regulated, the methylation levels in their proximal CpG sites located in the gene body tend to postively contribute to its expression, but negatively by those CpGs in their promoters. For those genes whose expression levels are significanlty down-regulated, the methylation levels of their CpG islands in the genes bodies tend to negatively contribute to their expressions.

This observation indicates that the DNA methyaltion machineary can (significanlty) up-regulate a gene's expression by increasing the DNA methylations in its body and decreasing the DNA methylation in its promoter. Similarly, to (significanlty) down-regulate a gene's expression, the methylation machinery can accomplish this through increasing only the DNA methylation in the gene body. This finding is consistent with the current understanding that DNA methylaiton accumulated in gene body regions tend to activate or enhance the gene's expression, while those in the promoter regions inhibit or repress the gene's expression [13]. Figure 3 shows a detailed comparison across the 10

Table 3. Statistical significances of enriched pathways by genes having high statistical correlations with *TET3, DNMT1* and *CBX2*, where pathways with p-values < 0.05 are in bold.

Cancer type	Statistical significance		
	REACTOME_ADAPTIVE_IMMUNE_SYSTEM	RESPONSE_TO_DNA_DAMAGE_STIMULUS	MONOOXYGENASE_ACTIVITY
BLCA	**1.99E-04**	**2.12E-03**	**1.75E-03**
BRCA	**7.07E-04**	**2.23E-02**	9.10E-02
COAD	**4.67E-03**	2.77E-01	1.00E+00
HNSC	1.00E+00	1.00E+00	1.00E+00
KICH	**1.42E-07**	**4.60E-06**	**3.86E-02**
KIRC	**2.10E-05**	**6.11E-04**	1.16E-01
KIRP	**1.60E-04**	1.97E-02	**4.84E-03**
LIHC	**7.20E-05**	**1.47E-04**	**3.30E-02**
LUAD	9.85E-02	5.98E-01	1.00E+00
PRAD	**2.15E-07**	**1.71E-03**	**2.05E-02**
THCA	**2.19E-03**	**3.95E-02**	**3.73E-03**

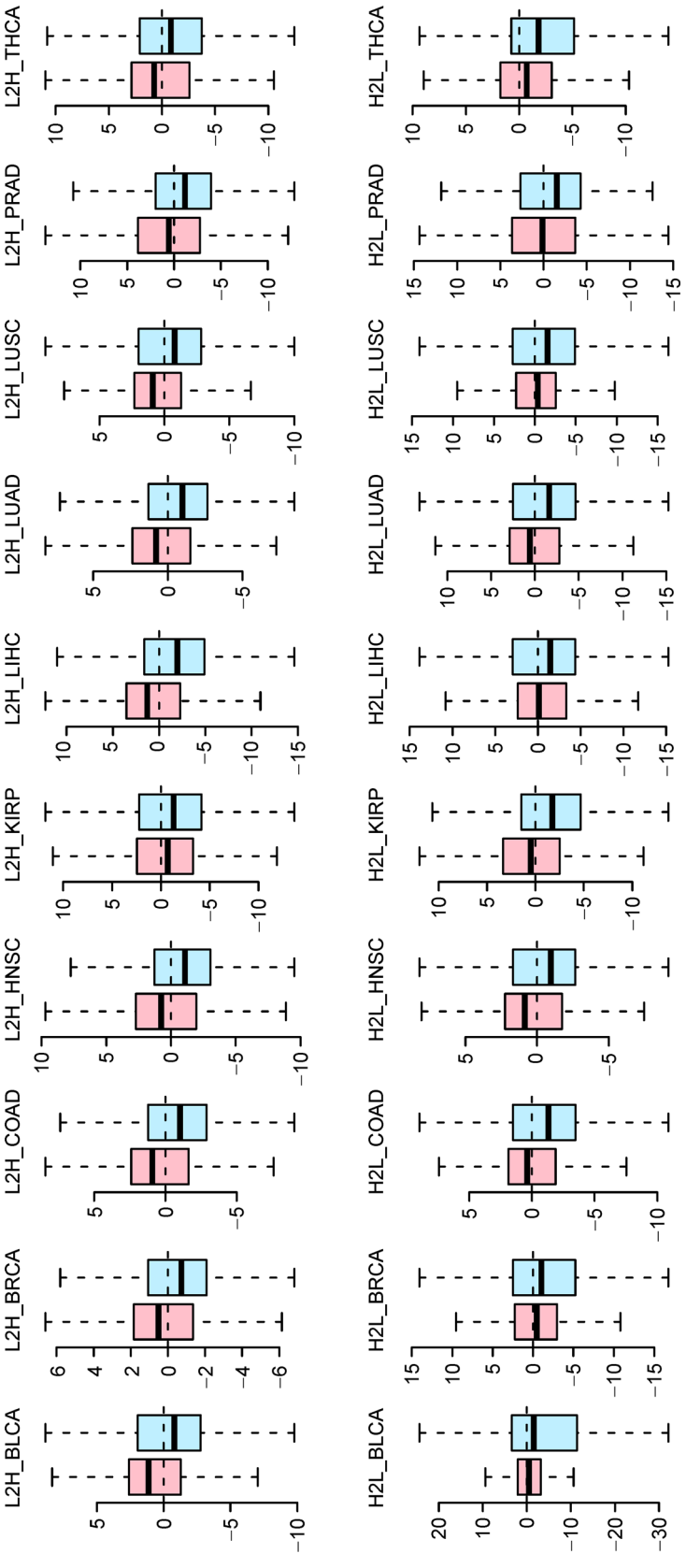

Figure 3. Coefficients of the predictors selected in gene body (pink boxes) and promoter (lightblue boxes) regions for 10 cancer types for those genes that are significantly up-regulated (upper panel) and down-regulated (lower panel).

cancer types.

DISCUSSION AND CONCLUSION

Our integrative analyses of gene expression and DNA-methylation data discovered that the regulation of expressions of a large class of genes is executed differently in cancer *vs.* normal tissue cells, with the former largely via direct DNA methylation of the genes and the latter by their TFs. Such genes fall into a few stress and response pathways. As a cancer advances, the level of involvement by direct DNA methylation in transcription regulation increases for majority of the cancer types examined. This is consistent with the general understanding of the roles played by epigenomics in responses to severe and novel stresses.

In the past two decades, substantial information has been accumulated regarding what stresses cancer cells may encounter at different stages of the disease, such as (i) oxidative stress associated with chronic inflammation that tends to take place in cancer sites; (ii) persistent hypoxia, which tends to take place with high oxidative stress due to the O_2 consumption by innate immune cells in support of their production of H_2O_2 and superoxide bursts [26]; (iii) the stress resulted by Fenton reactions, which threatens to increase the intracellular pH in a persistent manner [28,30,31]; and (iv) various general stresses such as heat-shock and unfolded protein responses. However, very little has been published regarding how these stresses affect the epigenomic level activities in cancer. Here, we have presented a new framework for studying the functional roles of epigenomic activities, specifically DNA methylation and demethylation activities in transcription regulation in cancer tissue cells. Our analyses revealed the possible roles of stresses in driving the increased methylation and demethylation activities, which give rise to the altered DNA methylation patterns proximal to genes of certain functional classes in cancer *vs*. control tissues. For the first time, we have clearly demonstrated that DNA methylation plays more significant roles in transcription regulation of genes of certain functional classes in cancer than in normal tissue cells, and offered an explanation to why this is the case. We postulate that as cancer tissue cells are constantly under severe stresses in their intracellular and micro-environment, DNA methylation represents an effective way in transcription regulation in cancer as a separate organism inside the human body.

DATA AND METHODS

Data

DNA methylation data measured using HumanMehtylation450 arrays for 10 cancer types are retrieved from the TCGA database [34]. We have also used RNA-Seq based gene expression data for 10 cancer types from the same database. Details are given in Table 4. Ten cancer types are selected because: (i) they have both normal and cancerous tissue samples; and (ii) they have a sufficiently large number of cancer samples whose gene expression and methylation data are both available. The gene expression data are log2-transformed. We have removed those genes when half of their log2 expression values are below 1. For methylation data, we have removed those CpG sites when half of their DNA methylation values are not observed, or the CpG sites were not in the proximity of any annotated genes. The procedures were done on expression and methylation data within cancer and control samples of each cancer type, respectively. And we have further discarded those genes either whose gene expressions or proximal CpG site DNA methylation data are not available. The remaining number of genes and CpG sites differ slightly from cancer to cancer, but on average, expression data of ~14,000 genes and methylation data of ~400,000 CpG sites are included in our analyses.

The methylation array data we used cover ~485,000

Table 4. Data used in our study.

Cancer type	Met (N)	Met (T)	Expr (N)	Expr (T)	Match (N)	Match (T)
BLCA	20	207	19	408	16	200
BRCA	96	685	113	1095	83	673
COAD	38	291	41	285	19	254
HNSC	50	426	44	520	20	418
KIRP	45	142	32	290	23	142
LIHC	50	125	50	371	41	122
LUAD	32	452	59	515	21	431
LUSC	42	359	51	501	8	359
PRAD	49	252	52	497	34	247
THCA	56	508	59	505	50	498

The columns are: (1) cancer types; sample sizes for (2) normal and (3) cancer methylation data; gene-expression data for (4) normal and (5) cancer tissues; and matching methylation data for (6) normal and (7) cancer tissue samples.

methylation sites that are distributed across the following genomic locations: TSS1500, TSS200, 5′ UTR, first exon, gene body, and 3′ UTR of the ~14,000 genes. Among these, ~300,000 sites are located in the near vicinity of each gene. Note: TSS1500 represents those CpG sites located between 200 nt to 1500 nt upstream the transcription start site (TSS); TSS200 are for those located within 200 nt upstream the TSS. This categorization is based on the Illumina Human Methylation 450 Platform [35].

Transcription factor and pathway databases

Experimentally validated human transcription factors (TFs) and their target genes are collected from multiple databases: TRED [36], Neph2012 [37], ENCODE [38], Marbach2016 [39] and TRRUST [40]. Overall, 919 transcription factors are retrieved. A TF-target relationship is considered as reliable only when the relationship appears in at least two of these databases. Overall, 72,407 pairs of such relationships are collected and used in our analyses.

The pathways and gene sets used in the enrichment analysis are C2, C5 and C6 collections in Msigdb [17], which represents curated gene sets, Gene Ontology gene sets and oncogenic gene sets.

Linear regression

For each response variable, we collected the corresponding predictors as the linear model predictors. We have built a linear regression model with L_1 penalty to select the most parsimonious subset of predictors that minimizes the following objective function:

$$\min_{\beta_0,\beta} \frac{1}{N}\sum_{i=1}^{N} (y_i - \beta_0 - \beta^T \boldsymbol{x}_i)^2 + \lambda \|\beta\|_1$$

where y_i, $\boldsymbol{x}_i$ are the i-th observed response and predictor, respectively; β_0, β are coefficients; and λ is the overall penalty parameter, which is selected based on cross-validation performance. In the case when the number of predictors are fixed, we select λ so that the selected predictors (fixed number) can explain the highest percentage of deviance.

Enrichment analysis

The enrichment analysis is performed using hypergeometric test, against selected pathways in Msigdb database. The p-values are FDR corrected.

Co-expression analysis

To identify the biological pathways that are most consistently and significantly associated with DNA methylation activities, we calculated correlations between genes *TET3, DNMT1, CBX2*, which is the direct and determining factor of DNA methylation activity, and the rest of all the ~14,000 genes, excluding the genes encoding the DNA (de)methylation enzymes and the Polycomb complex. We ranked all the genes based on their maximal (absolute) correlations with *TET3, DNMT1, CBX2*, and the top 1,000 genes with the highest (absolute) correlations are selected for enrichment analyses.

SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at DOI 10.1007/s40484-017-0129-y.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Victor Olman, formerly of the University of Georgia, for helpful discussion.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Sha Cao, Yi Zhou, Yue Wu, Tianci Song, Burair Alsaihati and Ying Xu declare they have no conflict of interests. All the data sets the authors used are from public repositories.

REFERENCES

1. Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet., 33, 245–254
2. Grativol, C., Hemerly, A. S. and Ferreira, P. C. G. (2012) Genetic and epigenetic regulation of stress responses in natural plant populations. Biochim. Biophys. Acta, 1819, 176–185
3. Seong, K.-H., Li, D., Shimizu, H., Nakamura, R. and Ishii, S. (2011) Inheritance of stress-induced, ATF-2-dependent epigenetic change. Cell, 145, 1049–1061
4. Ajonijebu, D. C., Abboussi, O., Russell, V. A., Mabandla, M. V. and Daniels, W. M. U. (2017) Epigenetics: a link between addiction and social environment. Cell. Mol. Life Sci., 74, 2735–2747
5. Wang, Y., Liu, H. and Sun, Z. (2017) Lamarck rises from his grave: parental environment-induced epigenetic inheritance in model organisms and humans. Biol. Rev., 92, 2084–2111
6. Feinberg, A. P. and Vogelstein, B. (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature, 301, 89–92
7. Esteller, M. (2002) CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. Oncogene, 21, 5427–5440
8. Calderwood, S. K., Khaleque, M. A., Sawyer, D. B. and Ciocca, D. R. (2006) Heat shock proteins in cancer: chaperones of tumorigenesis. Trends Biochem. Sci., 31, 164–172
9. Yadav, R. K., Chae, S. W., Kim, H. R. and Chae, H. J. (2014) Endoplasmic reticulum stress and cancer. J. Cancer Prev., 19, 75–88

10. Cairns, R. A., Harris, I. S. and Mak, T. W. (2011) Regulation of cancer cell metabolism. Nat. Rev. Cancer, 11, 85–95
11. Deaton, A. M. and Bird, A. (2011) CpG islands and the regulation of transcription. Genes Dev., 25, 1010–1022
12. Newell-Price, J., Clark, A. J. and King, P. (2000) DNA methylation and silencing of gene expression. Trends Endocrinol. Metab., 11, 142–148
13. Jones, P. A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet., 13, 484–492
14. Jjingo, D., Conley, A. B., Yi, S. V., Lunyak, V. V. and Jordan, I. K. (2012) On the presence and role of human gene-body DNA methylation. Oncotarget, 3, 462–474
15. Bird, A. (2002) DNA methylation patterns and epigenetic memory. Genes Dev., 16, 6–21
16. Holm, S. (1979) A simple sequentially rejective multiple test procedure. Scand. J. Stat., 6, 65–70
17. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA, 102, 15545–15550
18. Gerlach, J. Q., Sharma, S., Leister, K. J., Joshi, L. (2012) A Tight-Knit Group: Protein Glycosylation, Endoplasmic Reticulum Stress and the Unfolded Protein Response. In Endoplasmic Reticulum Stress in Health and Disease. Agostinis P., Afshin S. eds., pp. 23–39 Dordrecht: Springer
19. Nguyen, T., Nioi, P. and Pickett, C. B. (2009) The Nrf2-antioxidant response element signaling pathway and its activation by oxidative stress. J. Biol. Chem., 284, 13291–13295
20. Chiarugi, P., Pani, G., Giannoni, E., Taddei, L., Colavitti, R., Raugei, G., Symons, M., Borrello, S., Galeotti, T. and Ramponi, G. (2003) Reactive oxygen species as essential mediators of cell adhesion J. Cell Biol., 161, 933–944
21. Salim, S. (2017) Oxidative stress and the central nervous system. J. Pharmacol. Exp. Ther., 360, 201–205
22. Theccanat, T., Philip, J. L., Razzaque, A. M., Ludmer, N., Li, J., Xu, X. and Akhter, S. A. (2016) Regulation of cellular oxidative stress and apoptosis by G protein-coupled receptor kinase-2; The role of NADPH oxidase 4. Cell. Signal., 28, 190–203
23. Sun, H., Zhang, C., D, N., Sheng, T., and Xu, Y. (2017). Fenton Reactions Drive Nucleotide and ATP Syntheses in Cancer., (In review).
24. Stern, S., Fridmann-Sirkis, Y., Braun, E. and Soen, Y. (2012) Epigenetically heritable alteration of fly development in response to toxic challenge. Cell Reports, 1, 528–542
25. Cao, S., Zhu, X., Zhang, C., Qian, H., Schuttler, H. B., Gong, J. P., and Xu, Y. (2017) Competition between DNA methylation, nucleotide synthesis and anti-oxidation in cancer versus normal tissues. doi: 10.1158/0008-5472.CAN-17-0262
26. Valente, S., Liu, Y., Schnekenburger, M., Zwergel, C., Cosconati, S., Gros, C., Tardugno, M., Labella, D., Florean, C., Minden, S., *et al.* (2014) Selective non-nucleoside inhibitors of human DNA methyltransferases active in cancer including in cancer stem cells. J. Med. Chem., 57, 701–713
27. Rasmussen, K. D. and Helin, K. (2016) Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev., 30, 733–750
28. Wee, S., Dhanak, D., Li, H., Armstrong, S. A., Copeland, R. A., Sims, R., Baylin, S. B., Liu, X. S. and Schweizer, L. (2014) Targeting epigenetic regulators for cancer therapy. Ann. N. Y. Acad. Sci., 1309, 30–36
29. Khansari, N., Shakiba, Y. and Mahmoudi, M. (2009) Chronic inflammation and oxidative stress as a major cause of age-related diseases and cancer. Recent Pat. Inflamm. Allergy Drug Discov., 3, 73–80
30. Reuter, S., Gupta, S. C., Chaturvedi, M. M. and Aggarwal, B. B. (2010) Oxidative stress, inflammation, and cancer: how are they linked? Free Radic. Biol. Med., 49, 1603–1616
31. Fiaschi, T. and Chiarugi, P. (2012) Oxidative stress, tumor microenvironment, and metabolic reprogramming: a diabolic liaison. Int. J. Cell Biol., 2012, 762825
32. Zhang, C., Cao, S., Toole, B. P. and Xu, Y. (2015) Cancer may be a pathway to cell survival under persistent hypoxia and elevated ROS: a model for solid-cancer initiation and early development. Int. J. Cancer, 136, 2001–2011
33. Thomas, C., Mackey, M. M., Diaz, A. A. and Cox, D. P. (2009) Hydroxyl radical is produced via the Fenton reaction in submitochondrial particles under oxidative stress: implications for diseases associated with iron accumulation. Redox Rep., 14, 102–108
34. The Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J. M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet., 45, 1113–1120
35. Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., Delano, D., Zhang, L., Schroth, G. P., Gunderson, K. L., *et al.* (2011) High density DNA methylation array with single CpG site resolution. Genomics, 98, 288–295
36. Jiang, C., Xuan, Z., Zhao, F. and Zhang, M. Q. (2007) TRED: a transcriptional regulatory element database, new entries and other development. Nucleic Acids Res., 35, D137–D140
37. Neph, S., Stergachis, A. B., Reynolds, A., Sandstrom, R., Borenstein, E. and Stamatoyannopoulos, J. A. (2012) Circuitry and dynamics of human transcription factor regulatory networks. Cell, 150, 1274–1286
38. The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature, 489, 57–74
39. Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z. and Bergmann, S. (2016) Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat. Methods, 13, 366–370
40. Han, H., Shim, H., Shin, D., Shim, J. E., Ko, Y., Shin, J., Kim, H., Cho, A., Kim, E., Lee, T., *et al.* (2015) TRRUST: a reference database of human transcriptional regulatory interactions. Sci. Rep., 5, 11432

