

# Complex impact of DNA methylation on transcriptional dysregulation across 22 human cancer types

Zishan Wang<sup>1,†</sup>, Jiaqi Yin<sup>1,†</sup>, Weiwei Zhou<sup>1,†</sup>, Jing Bai<sup>1,†</sup>, Yunjin Xie<sup>1</sup>, Kang Xu<sup>1</sup>, Xiangyi Zheng<sup>1</sup>, Jun Xiao<sup>1</sup>, Li Zhou<sup>4</sup>, Xiaolin Qi<sup>2,\*</sup>, Yongsheng Li<sup>1,2,3,\*</sup>, Xia Li<sup>1,2,3,\*</sup> and Juan Xu<sup>1,2,3,\*</sup>

<sup>1</sup>College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China, <sup>2</sup>Key Laboratory of Tropical Translational Medicine of Ministry of Education, Hainan Medical University, Haikou, Hainan 571199, China, <sup>3</sup>College of Biomedical Information and Engineering, Hainan Medical University, Haikou, Hainan 570100, China and <sup>4</sup>Department of Nephrology, Affiliated Hospital of Chengde Medical College, Chengde, Hebei Province, China

Received December 13, 2019; Editorial Decision January 12, 2020; Accepted January 14, 2020

## ABSTRACT

Accumulating evidence has demonstrated that transcriptional regulation is affected by DNA methylation. Understanding the perturbation of DNA methylation-mediated regulation between transcriptional factors (TFs) and targets is crucial for human diseases. However, the global landscape of DNA methylation-mediated transcriptional dysregulation (DMTD) across cancers has not been portrayed. Here, we systematically identified DMTD by integrative analysis of transcriptome, methylome and regulome across 22 human cancer types. Our results revealed that transcriptional regulation was affected by DNA methylation, involving hundreds of methylation-sensitive TFs (MethTFs). In addition, pan-cancer MethTFs, the regulatory activity of which is generally affected by DNA methylation across cancers, exhibit dominant functional characteristics and regulate several cancer hallmarks. Moreover, pan-cancer MethTFs were found to be affected by DNA methylation in a complex pattern. Finally, we investigated the cooperation among MethTFs and identified a network module that consisted of 43 MethTFs with prognostic potential. In summary, we systematically dissected the transcriptional dysregulation mediated by DNA methylation across cancer types, and our results provide a valuable resource for both epigenetic and transcriptional regulation communities.

## INTRODUCTION

Sophisticated transcriptional regulation plays critical roles in the temporal and spatial expression of genes to achieve tissue specificity. Gene expression is mainly governed by transcriptional factors (TFs) in a sequence-specific manner (1,2). With the development of high-throughput sequencing technologies, particularly chromatin immunoprecipitation followed by sequencing (ChIP-Seq), it has become possible to determine TF binding sites (TFBSs) on a genome-wide scale (3–6). ChIP-Seq has provided a powerful way to reliably identify TFBSs, and it has been broadly used in current studies, thus accumulating large-scale datasets in different contexts (7–10). TFs, as ‘master regulators’, control many processes that determine the fate of cells (11,12) and specific pathways, such as immune response (13). Perturbation in TFs or TFBSs is likely to induce many diseases, including cancer (14–17). Although transcriptional dysregulation is closely associated with cancer, the underlying molecular mechanisms remain to be discovered.

Accumulating evidence has demonstrated that regulation of a TF to its target genes was modulated by various epigenetic mechanisms (18–22), such as DNA methylation. Methylation of DNA at cytosine within *cis*-regulatory elements can block or promote the recruitment of TFs. Increasing studies have reported that numerous TFs have DNA methylation-dependent regulatory activity, whereas some other TFs are not affected by methylation status (20,21,23,24). MeDReaders is the first resource for collecting information on the interactions between methylated DNA and TFs (24). Yin *et al.* analyzed 542 human TFs and found that there are numbers of TFs that prefer methylated sequences (20). Moreover, Hu *et al.* surveyed the human TF family and found that methylated CpG site (mCpG)-

\*To whom correspondence should be addressed. Email: xujuanbiocc@ems.hrbmu.edu.cn  
Correspondence may also be addressed to Xia Li. Tel: +86 451 86615922; Fax: +86 451 86615922; Email: lixia@hrbmu.edu.cn  
Correspondence may also be addressed to Yongsheng Li. Email: liyongsheng@ems.hrbmu.edu.cn  
Correspondence may also be addressed to Xiaolin Qi. Email: 13503213635@163.com

†The authors wish it to be known that, in their opinion, the first four authors should be regarded as Joint First Authors.

dependent binding activities of TFs were a widespread phenomenon (21). Zuo *et al.* described Methyl-Spec-seq to quantitatively assess the effects of methylation on TF binding affinity (25), which facilitate the study of the consequence for gene regulation of DNA methylation. However, the experimental method is time and cost consuming for use on a genome-wide scale. Considering the variety of DNA methylation and transcription regulation across tissues or cell types, it is necessary to identify DNA methylation-mediated transcriptional dysregulation (DMTD) across different contexts.

In addition, it is well known that aberration of DNA methylation contributes to carcinogenesis (26), and it frequently occurs in the promoter region of genes (27). Several studies have demonstrated that aberrant methylation in the promoter region could be used as a potential biomarker for cancer detection, diagnosis and prognosis (28–30). Moreover, aberration of DNA methylation perturbs the regulation of many TFs and consequently affects the downstream genes in cancer (31). DNA methylome has been incorporated in the transcriptional regulation circuitry and gene expression perturbation in cancer (32). However, DMTD has not been systematically explored across different cancer types. Comprehensive analysis of TF regulation dependency on target genes among cancer types will facilitate revealing the common and cancer-specific regulatory mechanisms. Furthermore, TFs have been found to regulate gene expression in a combinatorial manner across cell types. Besides cooperative binding (aiding each other in binding DNA) (33), TF cooperation is closely related to the chromatin state, such as DNA methylation. Domcke *et al.* have proposed a TF indirect collaboration model, in which TFs with demethylation function removed methylation of the local DNA sequence followed by binding of other TFs (34). This model required TFs with demethylation function, which have been scarcely reported in other biological environments (35–37). However, DNA methylation-mediated TF cooperativity in cancer has not been systematically investigated.

Fortunately, The Cancer Genome Atlas (TCGA) project covers large cancer patient cohorts and provides multi-dimensional omics data, including DNA methylation and expression profiles of both TF and target genes. It provides us an unprecedented opportunity to identify DMTD across cancer types in a systematic way. Thus, we systematically identified and analyzed DMTD (defined as transcriptional dysregulation in which the regulatory activity of the TF showed a significant difference between the high-methylation and low-methylation groups of the target gene) across 22 human cancer types. We first proposed a computational framework that integrated sample-paired expression and DNA methylation profiles, as well as the genome-wide transcriptional regulatory network. Comparison analysis among cancer types revealed pan-cancer methylation-sensitive TFs (MethTFs), which showed dominant characteristics, including presence in more cancer types, complex regulation, large proportion of known cancer genes and frequent differential expression. Analysis of the DNA methylation-mediated pattern in MethTFs revealed the complex impact of DNA methylation on the MethTF's regulatory activity. Finally, we explored MethTF cooperativity

mediated by DNA methylation, as well as their cooperative modes, thus revealing a survival-related network module. In summary, our genome-wide analysis revealed a comprehensive perturbation of DNA methylation in transcriptional regulation and provided novel insights into the impact of epigenetic signal on genetic regulation across cancer types.

## MATERIALS AND METHODS

### Sample-paired expression and DNA methylation profiles across 22 human cancer types

Genome-wide mRNA expression profiles across 22 human cancer types were downloaded from TCGA project. In total, expression levels of 19 810 mRNAs were quantified as fragments per kilobase per million reads mapped (FPKM). First, we removed the mRNAs whose FPKM was 0 (not detected) in >30% of the samples. Next, the expression value was  $\log_2(\text{FPKM} + 0.05)$  transformed. We also downloaded the genome-wide mRNA expression profiles (19 810 mRNAs) of normal samples across 14 cancer types available in TCGA. These expression values were also  $\log_2(\text{FPKM} + 0.05)$  transformed.

Genome-wide DNA methylation profiles of 21 human cancer types, except for ovarian cancer (OV), generated by the Illumina Infinium HumanMethylation450 BeadChip array were downloaded from TCGA. Genome-wide DNA methylation profile of OV was generated by the Illumina Infinium HumanMethylation27 BeadChip array. The methylation level of each probe was measured by the  $\beta$ -value, which ranges from 0 to 1 (representing unmethylated to fully methylated, respectively). Similarly, we removed the probes whose  $\beta$ -values were missing in >30% of the samples. The remaining probes with missing  $\beta$ -values were imputed by using the k-nearest neighbors (KNN) method. We next mapped the probes to the promoter regions of genes, which were defined as  $\pm 2$  kb regions around the transcription start sites (TSS). Next, the DNA methylation level of a gene was defined as the average  $\beta$ -values of probes that mapped to its promoter region. Only cancer types with >150 samples were kept in our analysis. Finally, sample-paired mRNA expression and DNA methylation profiles were analyzed which involved 18 549 genes and 8119 samples across 22 human cancer types.

All 22 human cancer types were further classified into nine classes, including lung (lung adenocarcinoma, LUAD; lung squamous cell carcinoma, LUSC), digestive system (colon adenocarcinoma, COAD; esophageal carcinoma, ESCA; liver hepatocellular carcinoma, LIHC; pancreatic adenocarcinoma, PAAD; stomach adenocarcinoma, STAD), endocrine system (pheochromocytoma and paraganglioma, PCPG; thyroid carcinoma, THCA), skin (skin cutaneous melanoma, SKCM), brain (brain lower grade glioma, LGG), reproductive system (breast invasive carcinoma, BRCA; cervical squamous cell carcinoma and endocervical adenocarcinoma, CESC; ovarian serous cystadenocarcinoma, OV; prostate adenocarcinoma, PRAD; testicular germ cell tumors, TGCT; uterine corpus endometrial carcinoma, UCEC), urinary system (bladder urothelial carcinoma, BLCA; kidney renal clear cell carcinoma, KIRC; kidney renal papillary cell carcinoma, KIRP), head

and neck (head and neck squamous carcinoma, HNSC) and soft tissue (sarcoma, SARC).

### Identification of DMTD in cancer

We proposed a computational framework to identify cancer-context transcriptional dysregulation mediated by DNA methylation. Our hypothesis was that DNA methylation of the target gene could change the regulatory activity of TF on it. This pipeline mainly consisted of the following five steps: construction of genome-wide transcriptional regulation (I), identification of cancer-context transcriptional regulation based on TF-gene correlation in expression (II-III) and identification of cancer-context DMTDs (IV-V).

First, we directly downloaded ChIP-Seq peak regions close to the TSS (30 kb upstream to 10 kb downstream of the TSS) of 468 TFs from ChIPBase v2.0 (10). These peaks were intersected with the promoter regions of genes ( $\pm 2$  kb regions around the TSS). The genes whose promoter regions overlapped with the TF binding peaks were identified as candidate targets.

Second, expression correlation was widely used to evaluate the potential regulation between regulators and genes (38,39). In this univariate regression model, changes in mRNA  $i$  expression,  $y_i$ , was used as a linear function of TF  $\mu$  expression,  $x_{\mu}$ , among  $n$  cancer samples in a given cancer type:

$$y_i = \beta_0 + \beta_{\mu}x_{\mu,i} + \varepsilon_i, i = 1, \dots, n.$$

where  $\beta_0$  was the intercept, and  $\beta_{\mu}$  was the regression coefficient for the TF expression variable. We estimated the TF coefficient by using the ordinary least squares method to test whether the expression level of a gene was associated with changes in the expression level of a TF. TF-gene regulation with Benjamini and Hochberg (BH) adjusted  $P$ -value  $< 0.01$  was identified as TF-gene regulation. As TFs can activate or repress the expression of target genes, both positively or negatively correlated TF-gene pairs were analyzed.

Third, we identified DMTDs from the cancer-context TF-gene regulation in each cancer type based on the changes in the TF's regulatory activity mediated by DNA methylation. TFs and target genes were filtered based on their expression variation across samples. Individual TFs and target genes with high variation ( $\log_2$  IQR  $> 0.58$ ) across samples were first selected. For each TF-gene regulation, the cancer samples were ranked based on the methylation level of the target gene. Here, we defined the methylation level of the gene as the average beta values of the CpG sites located in the promoter region (40,41). The top and bottom 25% of the samples were defined as the high-methylation group (H-group) and the low-methylation group (L-group), respectively. Next, we calculated the difference in the average methylation level of target genes between the H-group and the L-group. We only analyzed the top 25% TF-gene regulation in terms of the difference in the average methylation level of target genes between the H-group and the L-group. In addition, the TF-gene regulation fulfilled the following two criteria: (i) the TF was not differentially expressed between the H-group and the L-group; (ii) the target gene was differentially expressed between the

H-group and the L-group. Next, we tested each TF-gene regulation to determine whether it was affected by DNA methylation. We used the Spearman correlation coefficient (SCC) between the expression of TF and gene to represent the regulatory activity of the TF in the H-group ( $R_{\text{high}}$ ) and the L-group ( $R_{\text{low}}$ ), separately. This was performed by the `cor.test` function in R. To ensure that TF regulated the gene in at least one condition, we required that an absolute value of either  $R_{\text{high}}$  or  $R_{\text{low}}$  was  $> 0.3$ . We selected TF-gene regulation with the absolute value of difference between  $R_{\text{high}}$  and  $R_{\text{low}}$   $> 0.3$  for further analysis. Next, Fisher's exact test was used to evaluate the difference between the two SCCs as describe in a previous study (42). Briefly, we transformed these SCCs between TFs and genes as follows:

$$F(R) = \frac{1}{2} \ln \frac{1+R}{1-R}$$

Then, the rewiring score ( $\text{rewire}_{\text{TF-gene}}$ ) was calculated, which ranges from 0 to 1. A larger value indicates a more perturbed effect between TF and gene.

$$\text{rewire}_{\text{TF-gene}} = P \left( |X| \leq \left| \frac{F(R_{\text{high}}) - F(R_{\text{low}})}{\sqrt{\frac{1.06}{n_{\text{high}}-3} + \frac{1.06}{n_{\text{low}}-3}}} \right| \right), X \sim N(0, 1),$$

where  $n_{\text{high}}$  and  $n_{\text{low}}$  were the number of samples in the H-group and the L-group, separately. Finally, we randomly shuffled the samples' labels and recalculated the rewiring score 1000 times. We defined the  $P$ -value as the fraction of the rewiring score under random conditions that was greater than the real ones. TF-gene regulation with BH adjusted  $P$ -value  $< 0.05$  was regarded as DMTDs.

### Definition of the pattern for DMTDs

We defined the pattern of DMTDs based on the changes in the MethTF's regulatory activity between the low-methylation group and the high-methylation group. There were three major types and they were further classified into six subtypes. The detailed definition and their interpretations are shown in Supplementary Table S1.

The pattern of MethTF was defined on the basis of the frequency of patterns that involved DMTDs as regulators. It was defined as the dominant pattern whose frequency was not  $< 1/3$  for three major pattern types and  $1/6$  for six pattern subtypes. Similarly, we also defined the pattern mediated by DNA methylation for a MethTF in a pan-cancer background. Notably, a MethTF can be affected by DNA methylation in multiple patterns.

### Identification of MethTFs with demethylation function

To identify MethTFs with demethylation function in a cancer type, we calculated the SCC between the expression level of a MethTF and the average methylation level of its candidate target genes. If they were negatively correlated with each other ( $P < 0.05$ ), the MethTF might induce demethylation of target genes. The SCC and  $P$ -value were calculated by the `cor.test` function in R.

### Identification of cooperative MethTFs mediated by DNA methylation

We defined two MethTFs as a cooperative pair in a cancer type, only if they significantly shared target genes and were enriched in at least one common biological process in Gene Ontology (GO). The hypergeometric test was used for this procedure. For each candidate MethTF cooperative pair, cooperative mode was defined based on the patterns of two corresponding MethTFs. If two MethTFs have common pattern(s), the two MethTFs cooperate in the shared pattern(s), and if two MethTFs have no common pattern, the two MethTFs might cooperate via demethylation, in which one has an enhanced pattern and can induce target genes demethylation, while the other MethTF has an attenuated pattern. The remaining MethTF cooperative pairs not belonging to any of the above conditions were defined as having an unknown cooperative mode. Notably, two MethTFs can also cooperate in multiple modes, which was named as multiple mode.

### Identification of differentially expressed TFs in cancer

We used two methods to identify differentially expressed TFs based on the genome-wide mRNA expression profiles across 14 cancer types available in TCGA. For TFs with wide-range expression (the proportion of zero value < 30% of samples), we used the *t*-test followed by BH-corrected. TFs with fold change >2 and false discovery rate (FDR) <0.01 were regarded as differentially expressed. Otherwise, Fisher's exact test was used. First, we transformed the expression of a TF into a binary state: ON (expressed), OFF (not detected). The frequency of the two states was calculated in both cancer and normal samples. TFs with the ON/OFF state twice more frequent in cancer samples than in normal samples and FDR <0.01 were also regarded as differentially expressed.

### Functional enrichment analysis

The hypergeometric test was used to perform functional enrichment analysis for testing whether a gene set of interest were overrepresented in a GO term gene set. We particularly focused on cancer hallmarks-related GO terms, which were commonly used in previous studies (39,43–46). The human genes annotated to GO terms were downloaded from National Center for Biotechnology Information (NCBI) file 'gene2go'. It was a requirement that there were at least three target genes enriched in the specific GO terms. GO terms with *P*-values <0.05 were considered significant.

### Prognostic potential assessment of the MethTF cooperativity module

To explore the prognostic power of a MethTF cooperative module, univariate Cox regression analysis was first used to evaluate the association between survival time and the expression level of each MethTF in the module. The MethTFs with a *P*-value <0.05 were regarded as survival-related MethTFs. For each patient, we integrated the Cox

regression coefficient and the expression of the survival-related MethTF in the MethTF cooperativity module to calculate a risk score as follows:

$$\text{Risk score } (i) = \sum_{k=1}^n \beta_k * e_{ki},$$

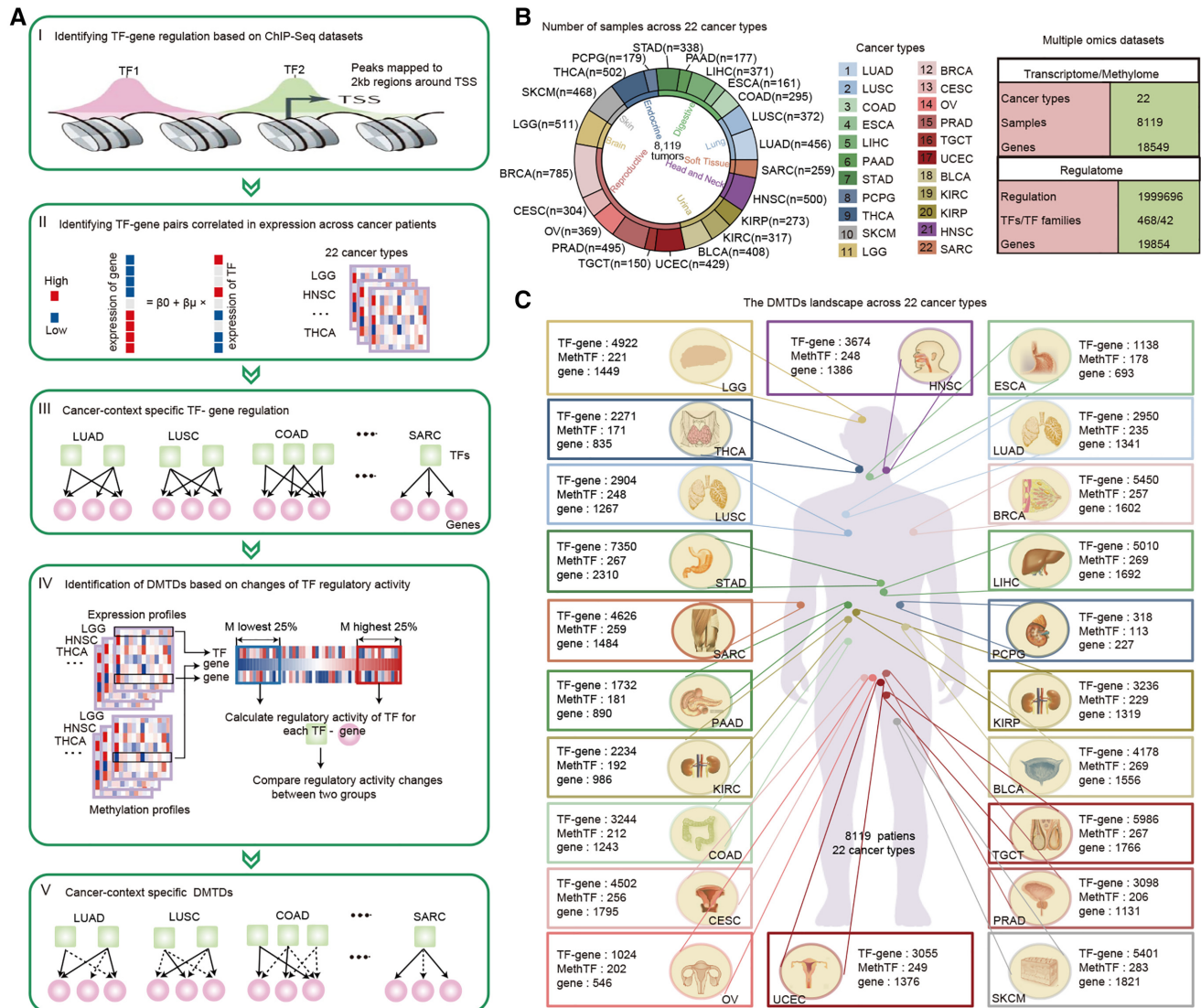
where  $\beta$  was the Cox regression coefficient,  $n$  was the number of survival-related MethTFs in the module, and  $e_{ki}$  was the expression level of MethTF  $k$  in patient  $i$ . Patients were randomly divided into the training and testing patient sets without age and sex differences (47). We used the training patient sets to train the parameters of the model, which were directly applied to both the training and testing sets. Patients were dichotomized into the low-risk group and the high-risk group. Log rank test was used to evaluate the survival difference between the two groups.

## RESULTS

### Transcriptional regulation is prevalently affected by DNA methylation across cancers

We proposed a computational framework to systematically identify transcriptional regulation mediated by DNA methylation across 22 human cancer types. This framework mainly consisted of five steps (Figure 1A). First, we obtained 1 999 696 TF–gene regulations among 468 TFs (443 TFs belonged to 42 TF families) and 19 854 target genes from high-throughput ChIP-Seq datasets (Figure 1B). Although an increasing number of ChIP-Seq data are available for cell lines, it is still limited for a specific cancer type. We found that only three cancer types (BRCA, LUAD and PRAD) had more number of ChIP-Seq datasets (Supplementary Table S2). To obtain genome-wide transcriptional regulation for more TFs, we collected all available ChIP-Seq data in our analysis. Previous studies have established that transcriptional regulation is context dependent (38), and therefore, we used a linear regression model to identify cancer-context TF–gene regulation (Figure 1A). Furthermore, we identified the TF–gene regulation only based on the ChIP-Seq data in corresponding cancer. We found that >75% regulation in three cancer types can be recalled (Supplementary Table S3), suggesting that our analyses are reliable for identifying the TF–gene regulatory network in cancer. Next, we used an estimator to assess the statistical significance of changes in the TF's regulatory activity between the low-methylation and high-methylation groups. In total, we identified 318–7350 DMTDs across 22 human cancer types (Figure 1C), suggesting that the prevalence of DNA methylation has an impact on transcriptional regulation. These cancer-context DMTDs formed a complex transcriptional perturbation network in each cancer type (Supplementary Figure S1).

However, compared with the large number of cancer-context transcriptional regulation in each cancer type (Supplementary Figure S2, top panel), we observed that only a very small number of transcriptional regulations were perturbed by DNA methylation across 22 human cancer types (Supplementary Figure S2, bottom panel). These observations suggested that the backbone of the transcriptional regulation network was stable. Particularly, LGG and TGCT

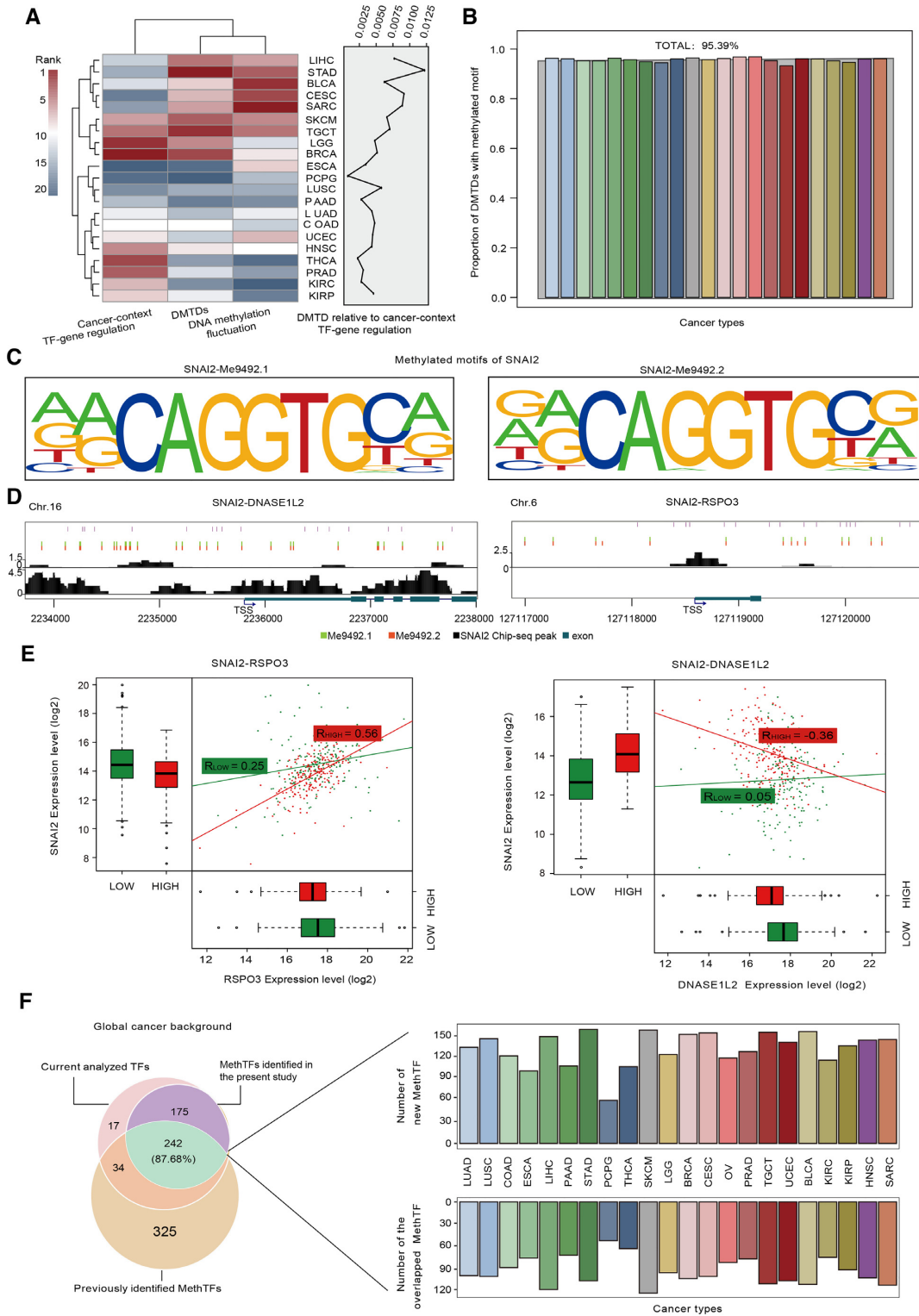


**Figure 1.** A systematically integrative framework for identifying DMTDs in cancer. (A) Schematic overview for identifying DMTDs in cancer. First, we integrated the expression profiles and TF-gene regulation for identifying cancer-context TF-gene regulation (steps I-III). Next, for each TF-gene regulation, two groups of patients were selected based on variation in the methylation level of the target gene. The regulatory activity of TF was determined to be altered or not in these two sample groups (step IV). DMTDs in each cancer were integrated into the regulatory network (step V). (B) The number of samples with paired expression and DNA methylation profiles across 22 human cancer types and the number of TF-gene regulation from ChIP-Seq datasets. Different cancer types correspond to different colors. All cancer types were sorted by their tissue of origin. (C) The number of DMTDs, MethTFs and genes in each cancer type.

together occupied ~50% cancer-specific transcriptional regulation and 15.88% cancer-specific DMTDs (Supplementary Figure S3A and B, right panel). These results might be explained by the higher tissue specificity of corresponding tissues (brain and testis), which were observed in previous studies (48). Although a TF can regulate different targets, dozens in some cases while hundreds in other cases, but depends on the cancer type involved. There were a significantly different number of DMTDs across cancer types (Supplementary Figure S2 and Supplementary Table S4). Identification of perturbations of the DNA methylation-mediated transcriptional regulation will provide novel insights in cancer. Thus, we have set up a web-based interface named DMTDB (<http://bio-bigdata.hrbmu.edu.cn/>

DMTDB) for users to search and download the detailed information of DMTDs or MethTFs of interest across cancer types.

To evaluate the dynamic changes in transcriptional regulation across cancer types, we ranked all of the cancer types based on the proportion of cancer-context TF-gene regulation or DMTDs, separately. We observed that the ranks by DMTDs of several cancer types increased significantly, while those of other cancer types decreased significantly after being compared by the cancer-context TF-gene regulation (Figure 2A). For example, the rank of STAD increased, which was consistent with the DNA methylation fluctuation range of STAD. On the contrary, THCA exhibited an opposite trend. These observations suggested



**Figure 2.** Validation of DMTDs across cancer types. (A) Rank of 21 cancer types (except for OV) based on the proportion of cancer-context TF-gene regulation, proportion of DMTDs and DNA methylation level fluctuation range (defined as IQR of genes' methylation), respectively. (B) Validating DMTDs based on the methylated motif derived from the methylated binding regions of MethTFs. (C) Two methylated motifs of MethTF SNAI2. (D) The methylation sites and SNAI2 binding peaks are shown in the genome browser. (E) The scatter plots showing the expression correlation between SNAI2 and DNASE1L2 (right panel) or RSP03 (left panel) in BRCA. The box plots showing the distribution of expression across cancer patients. The red and green lines were fitted based on the patients with high (red) and low (green) methylation levels of target genes. (F) Venn plot showing the overlap of MethTFs identified in the present study with previously identified MethTFs. The number of overlapped TFs among three TF sets: TFs in the current analysis, predicted MethTFs, and previously identified MethTFs (left panel). Number of newly identified MethTFs and validated MethTFs across 22 human cancer types (right panel).

that the DNA methylation level exhibited a determined impact on the transcriptional regulation to a certain extent in cancer. In addition, we calculated the number of cancers in which transcriptional regulation and DMTDs were observed. We found that only 0.28% (189/66 772) of all of the DMTDs occurred in >4 cancer types and 86.75% (57 925/66 772) of DMTDs occurred in just one cancer type (Supplementary Figure S3B). These results revealed the specificity of DMTDs across cancer types. The DMTDs occurring in more than five cancer types formed a dense TF–gene regulatory network (Supplementary Figure S3C). However, the proportion of MethTFs was relatively high, and it was 8.15% in one cancer type and 24.7% in >19 cancer types (Supplementary Figure S3D). When considering the MethTF families, the proportion was much higher (Supplementary Figure S3E). We also found that TF families were found in different number of cancer types. For example, TF members in the C2H2 ZF family were affected by DNA methylation in more cancer types compared with those in the Homeodomain family (Supplementary Figure S3F), which was consistent with their expression patterns and biological functions (49).

#### DMTDs are enriched in methylated motifs

Evidence has proved that a TF has different binding regions when the sequence is methylated or not (21). To validate the transcriptional regulation mediated by DNA methylation, we assessed whether DMTDs were enriched in the known methylated motifs (24). Methylated motifs of 229 MethTFs were obtained, which included 30 303 DMTDs (methylated DMTDs) across 22 cancer types. As a result, we found that ~95% of DMTDs had at least one methylated motif (Figure 2B). In addition, approximately two-thirds of methylated DMTDs showed an overlap between the methylated motif and the binding peaks of corresponding MethTFs at the promoter regions of target genes (Supplementary Figure S4). For example, SNAI2 had two methylated motifs and one or both of them were presented at all the target gene promoters (Figure 2C). In BRCA, the methylated motifs were overlapped with the binding peaks for two DMTDs of SNAI2 (Figure 2D), accompanied by a changeable regulatory activity between the low-methylation and high-methylation groups (Figure 2E). We next assessed our results based on the previously identified MethTFs (24). The majority of overlapped MethTFs (87.68%) were validated on a pan-cancer background. Moreover, the regulatory activity of 175 TFs was newly found to be affected by DNA methylation (Figure 2F).

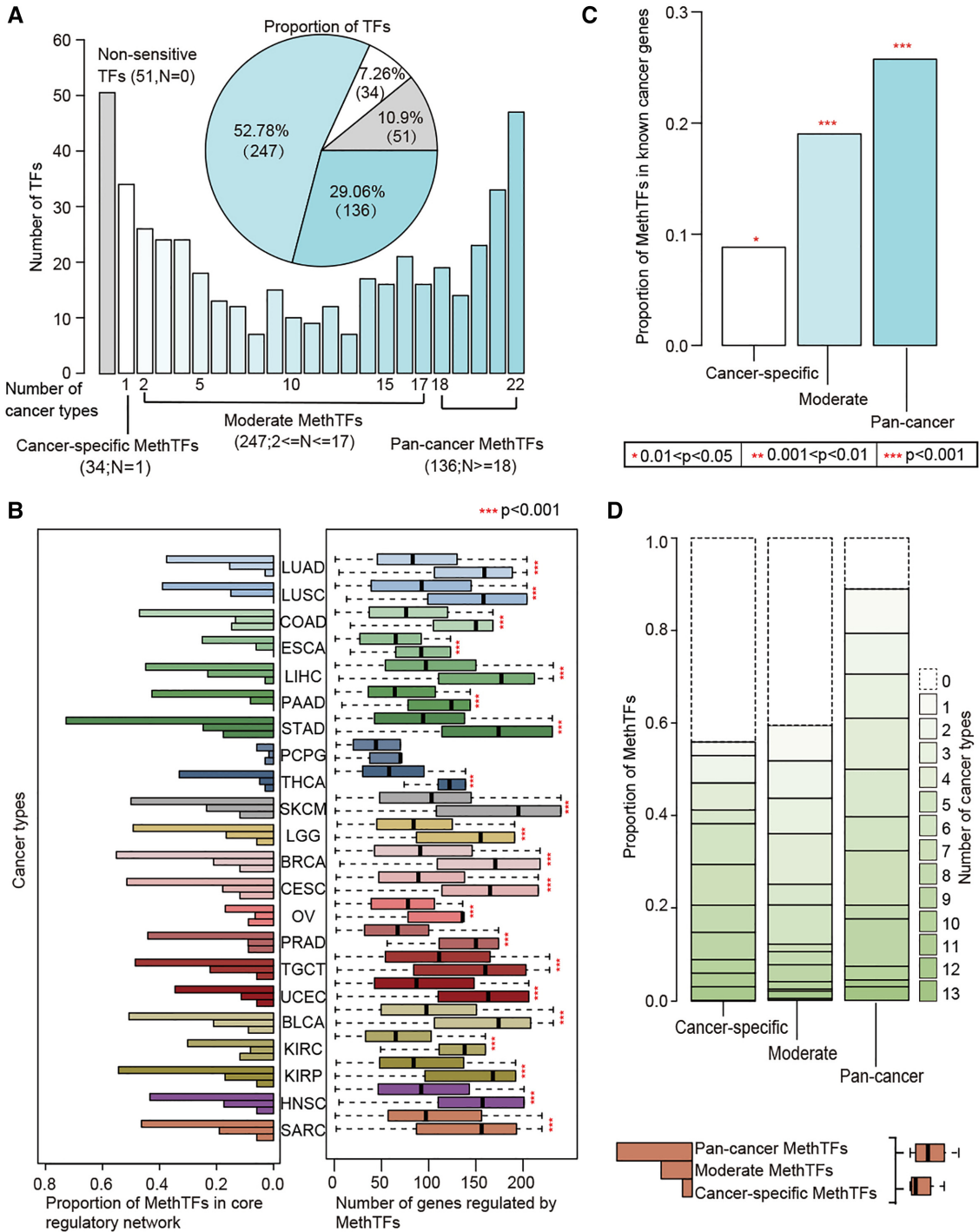
#### Pan-cancer MethTFs play central roles in cancer

Systematic identification of genome-wide DMTDs enabled us to explore the roles of TF across different cancer types. First, we explored the contribution of TFs by calculating the number of cancer types that each TF was observed. MethTF was defined as a TF whose regulatory activity was affected by DNA methylation of one or more target genes. We found that the distribution showed a bimodal distribution (Figure 3A), suggesting both cancer specificity and high cancer conservation of TFs. The regulatory activity of

89.1% (417/468) of TFs was influenced by DNA methylation in cancer, and only 10.9% (51/468) of TFs were not affected in any cancer type (Figure 3A). Interestingly, we observed that a substantial number of TFs (29.06%, 136/468) were influenced by DNA methylation across multiple cancer types (>17 cancer types) and a small proportion of TFs (7.26%, 34/468) was observed in only one cancer type (Figure 3A). The distribution was robust when we defined MethTF based on different thresholds (Supplementary Figure S5A). Thus, TFs were divided into the following four categories according to the number of cancer types: non-sensitive TFs, cancer-specific MethTFs, moderate MethTFs and pan-cancer MethTFs (Figure 3A).

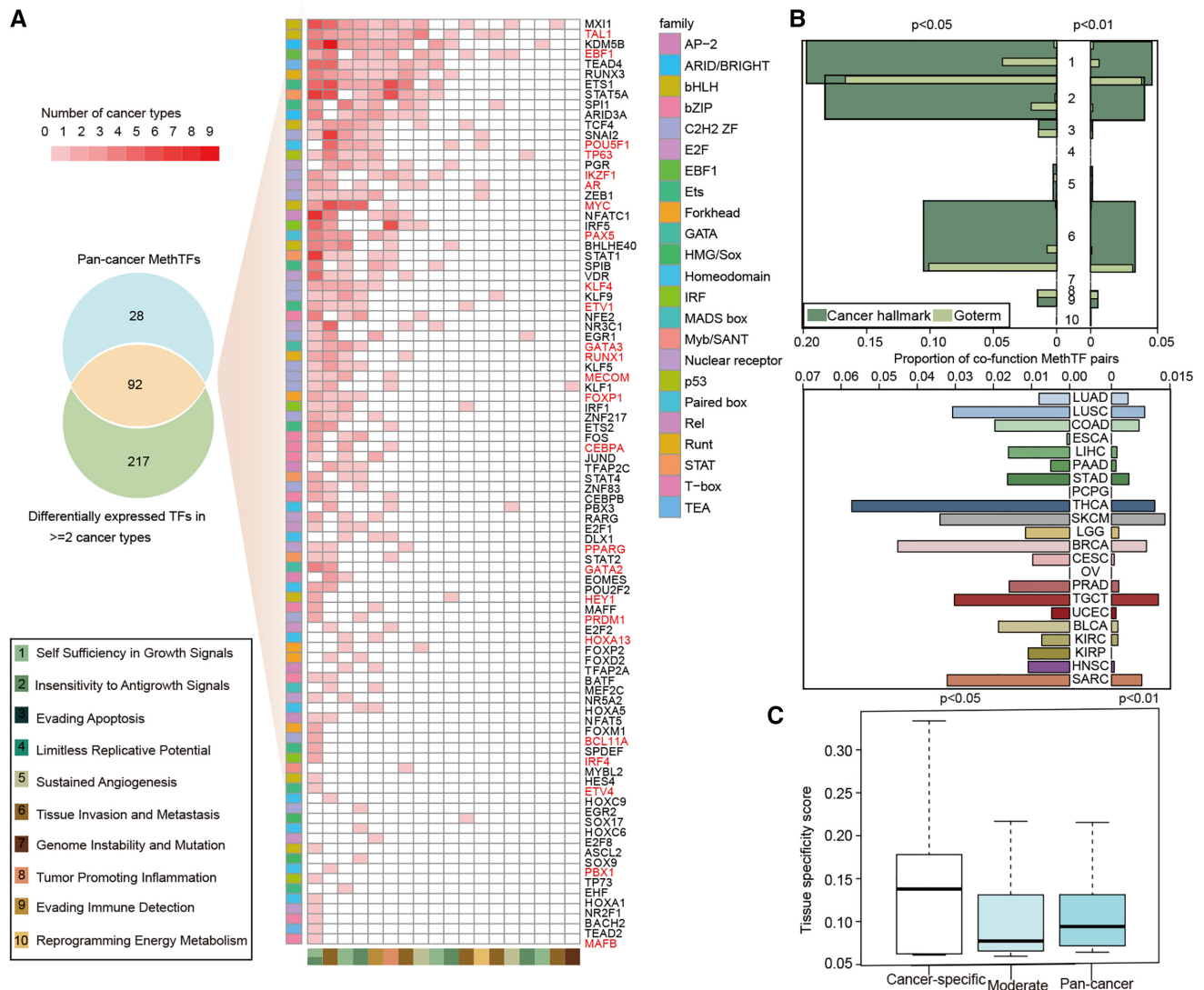
Previous network analyses have revealed that a regulator with more regulated targets tends to locate in the central position of a network and plays important roles (47,50). Thus, we calculated the degree of MethTFs, which was defined as the number of target genes they regulated in each cancer type. We found that pan-cancer MethTFs regulated more target genes than the other two categories (Figure 3B, right panel). Moreover, pan-cancer MethTFs had a reasonably large number of target genes in many cancer types (Supplementary Figure S5B). Interestingly, there was a higher proportion of pan-cancer MethTFs participating in the MethTF core regulatory network, in which the target genes were also MethTFs (Figure 3B, left panel and Supplementary Figure S5C). These observations suggested that pan-cancer MethTFs tend to be affected by other MethTFs or DNA methylation of more target genes, implying their centrality in the DNA methylation-mediated transcriptional network. Moreover, we found that all of the MethTFs in the three categories were significantly enriched in known cancer genes. Particularly, pan-cancer MethTFs showed the highest proportion (Figure 3C). We obtained similar results based on the genome-wide association study (GWAS)-associated genes (Supplementary Figure S5D).

Moreover, abnormal biological processes in cancer were closely associated with perturbation of expression of related genes. We thus assessed whether MethTFs tend to differentially express in cancer. Notably, 88.97% of pan-cancer MethTFs exhibited differential expression across different cancer types (Figure 3D and Supplementary Figure S6A). In addition, the direction of differential expression of many pan-cancer MethTFs was consistent (Supplementary Figure S6B). To further investigate the roles of pan-cancer MethTFs, we selected those exhibiting differential expression in at least two cancer types. Their dysregulated target genes were used to perform functional enrichment analysis in cancer hallmark-associated GO terms (Figure 4A). These MethTFs were found to be enriched in two particular GO terms that were related to three cancer hallmarks: signal transduction and cell adhesion (Figure 4A and B). Signal transduction was related to two cancer hallmarks (self-sufficiency in growth signals and insensitivity to the antigrowth signal), while cell adhesion was related to tissue invasion and metastasis. These results demonstrated that pan-cancer MethTFs tend to disturb cancer hallmark-associated functions. One example was KLF4 in the C2H2 ZF family, which has been confirmed to suppress cell proliferation and cell cycle in many cancers, such as cancer of the colon (51), stomach (52), and lung (53). Previous



**Figure 3.** Characterization of MethTFs in cancer. (A) Bar plots showing the number of TFs observed in different number of cancer types. Based on the number of cancer types, TFs were classified into the following four categories: non-sensitive TFs, cancer-specific MethTFs, moderate MethTFs and pan-cancer MethTFs. (B) The barplots in the left panel showing the proportion of MethTFs in different categories that were observed in the core regulatory network across cancer types. The box plots in the right panel showing the number of target genes for MethTFs in different categories. (C) Proportion of MethTFs in known cancer genes across the three MethTF categories. Known cancer genes were obtained from the Cancer Gene Census. (D) Proportion of differently expressed MethTFs in different number of cancer types. Different colors represent different number of cancer types in which MethTFs were differentially expressed.





**Figure 4.** Function of MethTFs across cancer types. (A) Venn plot showing the overlap between pan-cancer MethTFs and those differentially expressed in  $\geq 2$  cancer types. The heatmap in the right panel showing the function enrichment of MethTFs in cancer hallmarks. Each row represents a MethTF and each column represents a GO term. The color of the value represents the number of combinations between cancer types and GO terms. MethTFs were significantly enriched in the GO term. MethTFs were ranked based on the number of combinations between cancer types and GO terms. MethTFs in known cancer genes were labelled. (B) Proportion of co-function MethTF pairs in different GO terms related to 10 cancer hallmarks and cancer types. A  $P$ -value less than 0.05 (left panel) or 0.01 (right panel) are shown. (C) The distribution of tissue specificity scores for the three MethTF categories.

studies have proved that KLF4 could bind to methylated or unmethylated CpG sites in a different sequence context based on independent domains (21). Here, we found that KLF4, as a known cancer gene, dysregulates number of genes mediated by DNA methylation, which were associated with many cancer hallmarks (Supplementary Figure S7A). We also observed decreased expression of KLF4 in many cancer types (Supplementary Figure S7B). Another example was ETV4 in the Ets family, which regulates genes associated with the tissue invasion and metastasis hallmark (Supplementary Figure S7C). In addition, ETV4 showed increased expression in many cancer types (Supplementary Figure S7D), including colonic cancer, gastric cancer, and endometrial cancer. ETV4 has been found to be involved in the epithelial mesenchymal transition (EMT) process that

induces the invasive ability of tumor cells (54,55) and its somatic mutations were found to be involved in the progression of colonic, gastric, endometrial, and squamous cell carcinomas (56). Therefore, all these results indicated the important roles of MethTFs, particularly pan-cancer MethTFs, in carcinogenesis, and they were mediated by DNA methylation.

Tissue- or cell-type-specific expression of genes, including TFs, is often indicative of the corresponding specific roles in the physiological process or diseases (57). Thus, an entropy-based method (58) was applied to evaluate the tissue specificity score of each MethTF across 30 tissues from the Genotype-Tissue Expression (GTEx) project. Compared with cancer-specific MethTFs, MethTFs in the other two categories exhibited a relatively low score of tissue speci-

ficity, although it was not significant (Figure 4C). Finally, we assessed the enrichment of TF families among the three MethTF categories. We found that both Homeodomain and THAP finger TF families tend to be cancer specific (Supplementary Figure S7E), which was consistent with the roles of TF members in these two TF families (11,12,59). On the other hand, the Ets family was enriched in the pan-cancer category (Supplementary Figure S7E), in which only one TF member ELF4 was not present in any cancer, and the remaining TF members were present in multiple cancer types (Supplementary Figure S7F). Previous studies have reported that TF members in the Ets family were associated with different cancer types (16). Our results suggested that TF members in the Ets family tend to promote carcinogenesis in a methylation-dependent manner. Together, all these observations suggested that MethTFs have widespread effects on transcription regulation, particularly pan-cancer MethTFs, which play central roles in promoting carcinogenesis.

### **Pan-cancer MethTFs exhibit complex DNA methylation-mediated patterns in cancer**

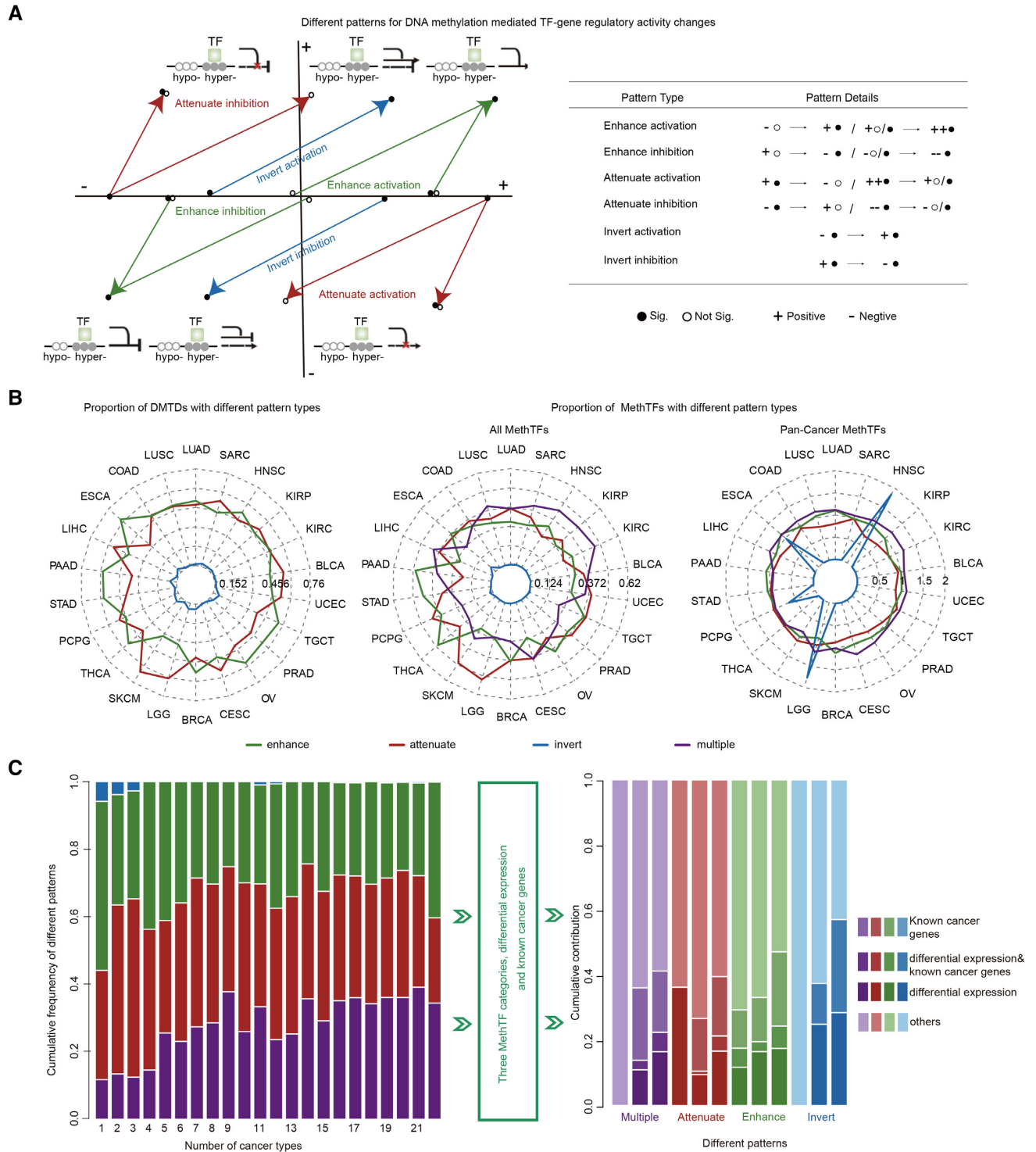
It is well known that TFs activate or repress the expression of target genes mainly depending on the local sequence and availability of cofactors (60,61). DNA methylation can affect the binding of many TFs in a sophisticated manner to perturb the transcriptional regulation. To explore the pattern mediated by DNA methylation on transcriptional regulation in cancer, we classified all DMTDs into three major patterns, including attenuate, enhance and invert (Figure 5A and Supplementary Table S1). The enhance group represented increased MethTF's regulatory activity with the DNA methylation level of the target gene promoter going up, while the attenuate group represented decreased MethTF's regulatory activity and the invert group represented reversed MethTF's regulatory activity. For example, in BRCA, the regulatory activity of SNAI2 was enhanced by the DNA methylation level of the both RSPO3 and DNASE1L2 (Figure 2D). Moreover, DMTDs could be further classified into six patterns by considering the direction of regulatory activity and significance of SCC (Figure 5A and Supplementary Table S1). We observed that most cancer types (19/22) have higher proportions in a fine manner, but not the inverted regulatory activity (Figure 5B, left panel). In addition, DNA methylation-mediated patterns of a MethTF were dependent on its target genes in different cancer types, such as SNAI2 in BRCA (Figure 2E). We further determined the DNA methylation-mediated pattern of a MethTF by summarizing the impact of all its related DMTDs in an individual cancer type or in a pan-cancer background (see details in Materials and Methods). Similarly, each MethTF was also classified into one of the following four patterns: attenuate, enhance, invert and multiple. Multiple pattern represents that the regulatory activity of a MethTF was affected by DNA methylation in two or three major patterns. We found that MethTFs with a multiple or enhance pattern exhibited a relatively high stability across 22 human cancer types (Supplementary Figure S8A). Moreover, the majority of MethTFs were found to be affected by DNA methylation in multiple pattern either

in several cancer types (Figure 5B, middle panel) or in the pan-cancer background (Supplementary Figure S8B, left panel). The proportion of multiple pattern for pan-cancer MethTFs was even higher (Figure 5B, right panel and Supplementary Figure S8B, right panel). These results demonstrated the complexity of the DNA methylation-mediated pattern of MethTF in cancer, which was a slightly more remarkable when focusing on pan-cancer MethTFs. Moreover, we found that the more the cancer types in which a MethTF was involved, the more complex was the pattern of the MethTF (Figure 5C, left panel). We thus defined pan-cancer MethTFs with multiple pattern in a pan-cancer background as 'complex pan-cancer MethTFs'. Furthermore, we investigated the contribution of differential expression and known cancer genes to different patterns. We found that the contributions increased with the number of occurred cancer types (Figure 5C, right panel and Supplementary Figure S8C and D).

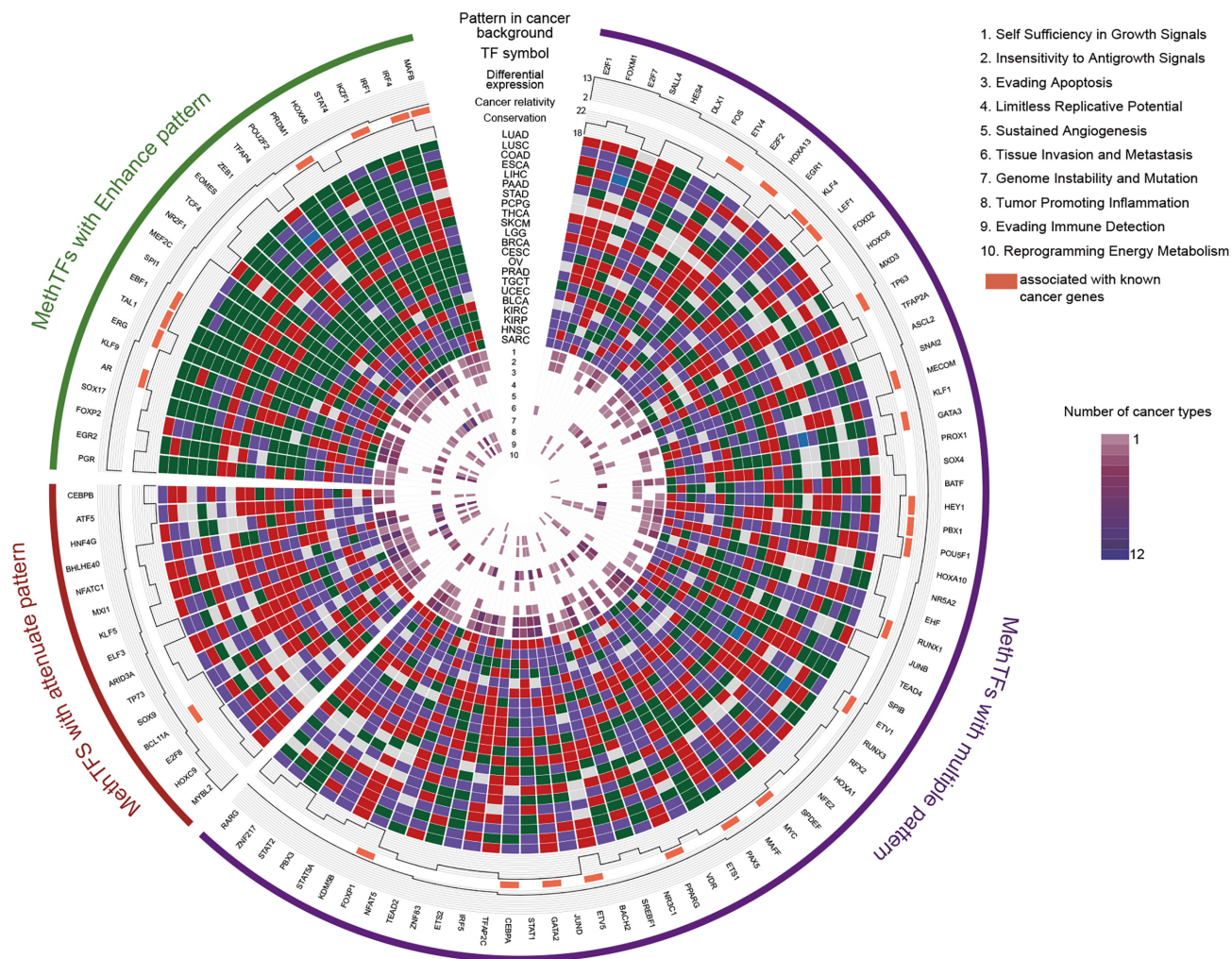
In addition, we explored the roles of MethTFs with different patterns by performing functional enrichment analysis via their target genes. We found that the three previous enriched cancer hallmarks were significantly enriched by all the patterns although with different enrichment level (Figure 6). We also observed that KLF4 affected different cancer hallmarks through target genes in different patterns (Supplementary Figure S9). The regulatory activity of KLF4 was attenuated or inverted by DNA methylation in pan-cancer background. Many target genes with an attenuate or invert pattern controlled particularly two cancer hallmarks: insensitivity to antigrowth signals and tissue invasion and metastasis. Together, these results illustrated the complex impact of DNA methylation on the regulatory activity of MethTFs, particularly pan-cancer MethTFs, and they showed how the impact was delivered to contribute to carcinogenesis.

### **Pan-cancer MethTFs function in a combinational manner with prognostic potential**

Researchers have found there are ~1600 TFs in human genome (49) to control transcription of the whole genome, which contains tens of thousands of genes. The limited TFs are thought to control the expression of the larger gene sets in a combinational manner, in which multiple TFs collaborate to regulate individual genes. The above function analysis found that a biological process was regulated by multiple MethTFs (Figure 4A and B), suggesting that MethTFs might work cooperatively to promote carcinogenesis. Thus, we explored MethTF cooperativity mediated by DNA methylation based on the following two criteria: significantly sharing targets and enriched in at least one common function (Figure 7A). Totally, we identified 6156 MethTF cooperative pairs across 22 human cancer types, forming a complex cooperative network. Moreover, many cancer hallmarks were generally co-regulated by multiple MethTFs, particularly the three cancer hallmarks described above (Supplementary Figures S10A and 4A, B). We further classified the cooperative mode of MethTF cooperative pairs into six modes based on the patterns of the corresponding two MethTFs. As shown in Figure 7B, the modes of cooperative MethTFs were complex, but



**Figure 5.** Complex transcriptional regulatory pattern mediated by DNA methylation in cancer. (A) The model in the left panel showing the TF-gene perturbation patterns mediated by DNA methylation. It was classified into three major types (attenuate, enhance, invert), in which each pattern type was further divided into two subtypes (attenuated inhibition, attenuated activation; enhanced inhibition, enhanced activation; inverted inhibition, inverted activation). The table in the right panel showing the details for the definition of each pattern. (B) The proportion of DMTDs (left panel), all MethTFs (middle panel) and pan-cancer MethTFs (right panel) with different patterns across 22 human cancer types. Lines in different color represent distinct patterns. (C) Contribution of different patterns for MethTFs in different number of cancer types (left panel). Contribution of aberrant expression and known cancer genes, known cancer genes and other genes (right panel). From the bottom to top, each cumulative bar plot represents the contribution of differentially expressed genes, differentially expressed known cancer genes, known cancer genes and other genes.

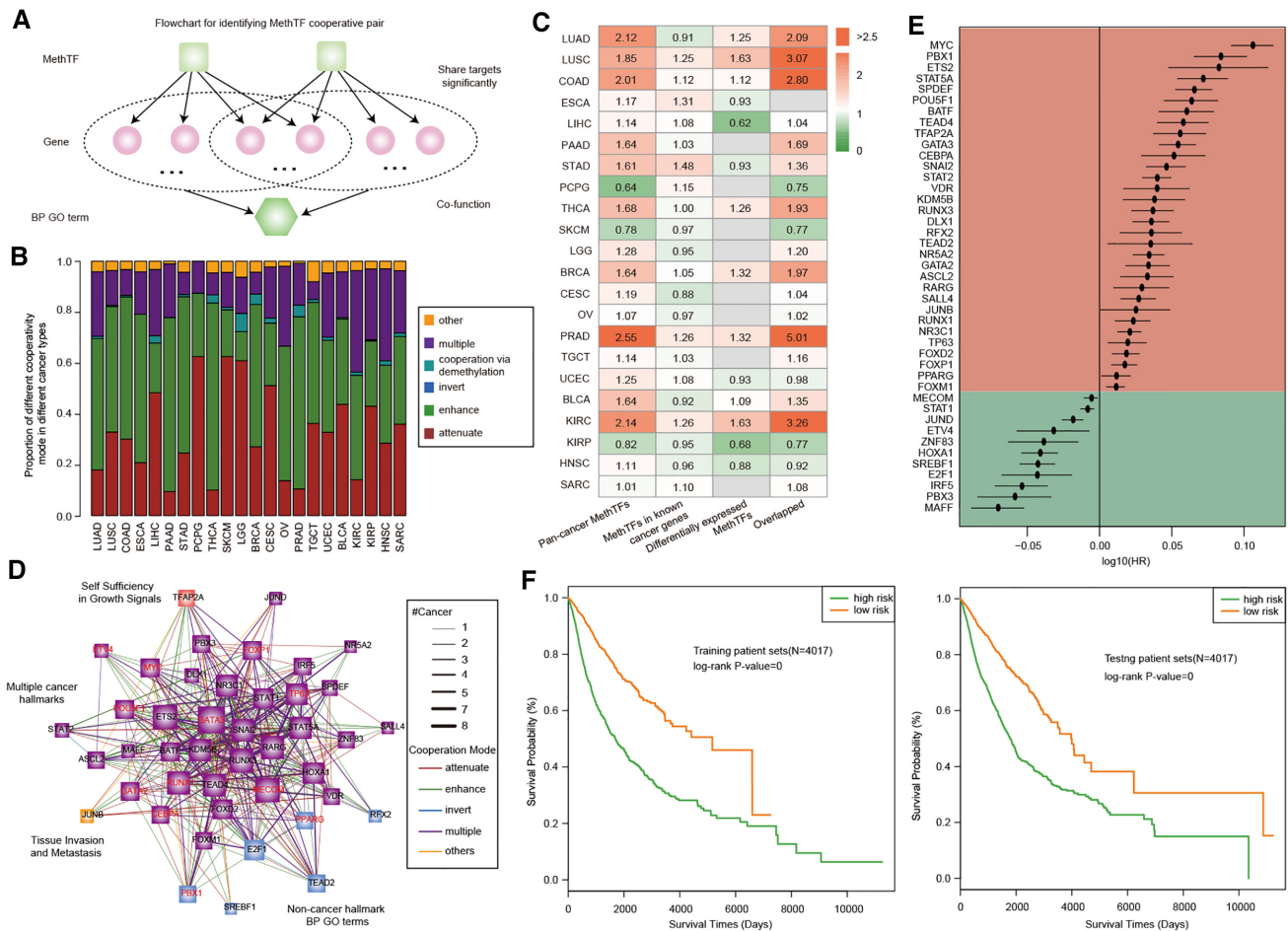


**Figure 6.** Global overview of the pan-cancer MethTFs. The circo plot showing detailed information of pan-cancer MethTFs that were differentially expressed in at least two cancer types. From the outside to inside layers, the following information was included: patterns in the pan-cancer background, MethTF symbols, number of cancer types with differential expression, known cancer gene or not (red for known cancer gene), number of cancer types that observed the MethTF, patterns in individual cancer types, and enrichment of 10 cancer hallmarks. The orders of the cancer hallmarks are shown in the top right corner.

there was a dominant mode in several cancer types. For example, the regulatory activity of cooperative MethTFs was mainly enhanced by DNA methylation of their co-targets in PAAD, whereas the attenuate mode was dominant in PCPG (Figure 7B). When focusing on cancer-related genes, such as pan-cancer MethTFs, known cancer genes, and differentially expressed MethTFs, we found that MethTFs tend to cooperate with other MethTFs in different modes (Figure 7C and Supplementary Figure S10B, C). These results suggested that MethTFs, particularly pan-cancer MethTFs, tend to cooperatively regulate cancer hallmarks in complex patterns mediated by DNA methylation.

In addition, many studies have demonstrated the importance of methylation and transcriptional regulation in cancer prognosis. Thus, we further investigated the prognostic potential of the MethTF cooperative module comprised by ‘complex pan-cancer MethTFs’, which were differentially expressed in at least two cancer types. Simi-

lar to the previous method (47), we identified a survival-related MethTF cooperative module containing 43 complex pan-cancer MethTFs (Figure 7D and E). We found that 81.40% (35/43) of MethTFs cooperate in multiple cooperative modes to regulate multiple cancer hallmarks. They tended to locate at the center of the cooperative network while the remaining MethTFs controlling only one hallmark tended to locate at the margin of the cooperative work (Figure 7D). Moreover, besides 12 MethTFs annotated in Cancer Gene Census, many other MethTFs have been reported to be associated with cancers. For example, E2F1 has been confirmed to function in tumor invasion and metastasis (62–64). We found that this module could stratify patients into distinct groups with significantly different survival times (Figure 7F). Moreover, we found that there was higher proportion of survival-associated TFs in 43 complex pan-cancer MethTFs, when compared with either other pan-cancer MethTFs (Supplementary Figure S11A) or moderate and specific MethTFs using different



**Figure 7.** MethTFs cooperativity mediated by DNA methylation in cancer. (A) A schematic graph showing the identification of a cooperative MethTF pair mediated by DNA methylation. Two criteria were used, sharing targets significantly ( $P < 0.05$ ) and both enriched at least one common BP GO term ( $P < 0.05$ ). (B) Proportion of cooperative MethTF pairs with different cooperativity modes across 22 human cancer types. (C) Fold enrichment between MethTFs belonging to the pan-cancer class, known cancer genes, or differentially expressed genes versus other MethTFs across 22 human cancer types. The last column was folds between the overlapped MethTFs among the three MethTF sets vs other MethTFs. (D) A survival-related MethTF cooperative module comprised 43 complex pan-cancer MethTFs. Known cancer genes were labelled in red. (E) The survival portrait of 43 complex pan-cancer MethTFs in the survival-related MethTF cooperative module. (F) Kaplan–Meier plot for TCGA patients with different risk scores in both the training and testing patient sets.

cutoffs (Supplementary Figure S11B–D). These results suggest that these ‘complex pan-cancer MethTFs were likely to be associated with patient survival. Collectively, our observations indicated the importance of the methylation-dependent combinatorial association among MethTFs in carcinogenesis and led to a promising clinical use of the MethTFs cooperative biomarkers.

## DISCUSSION

In this study, we have proposed a computational framework to identify DMTDs in cancer. Genome-wide DMTDs and MethTFs were identified across 22 human cancer types. The diversity of DMTDs and MethTFs demonstrated the existence of heterogeneity across different cancer types. The majority of DMTDs were cancer specific, while a larger number of MethTFs were common across cancer types. All TFs were divided into four categories and comparison analysis was performed between three MethTF categories. We re-

vealed that pan-cancer MethTFs had vital functional characteristics. In addition, we dissected the patterns mediated by DNA methylation on the regulatory activity of MethTFs and we found that MethTFs were influenced in different patterns depending on the target genes. A considerable number of MethTFs were affected in multiple patterns, highlighting the complexity of DNA methylation mediated patterns. Moreover, we investigated MethTF cooperativity mediated by DNA methylation and we found a methylation-dependent combinatorial pattern for MethTFs to promote carcinogenesis with prognostic potential.

It is important to determine the TF–gene regulatory network before identifying the perturbations of DNA methylation-mediated transcriptional regulatory network. However, there is still no gold standard regulatory network across different cancer types. Evidence has shown that integration of regulator occupancies data and mRNA expression correlation can improve the methods for inferring gene regulatory network (50,65–67). Thus, the first two steps

of our method were used to identify the cancer context-specific TF–gene regulatory network. Next, we identified the DNA methylation-mediated TF–gene perturbations in each cancer. This process may overlook several cases of TF–gene pairs that showed no correlation in expression. Thus, we also provided the DNA methylation perturbed gene-regulatory network in each cancer by not considering the second step (downloaded from the website). These data provided a valuable resource for users who do not care much about the cancer specificity of gene regulatory networks. In the computational pipeline, we used the expression correlation changes to evaluate the TF–gene regulation perturbation. It is better to identify the differential binding TFs based on ChIP-Seq data. However, it is technically challenging to obtain these data from primary tumors, and they are currently unavailable for many cancer types. In contrast, gene expression data are more widely available for various types of cancer. With the development of biotechnology and increase in ChIP-Seq data in various cancers, we can extend this pipeline by using standard differential binding methods. Moreover, evidence has shown that cancer types with similar tissue origin or within the same system share multiple molecular characteristics (38,68). To test this concept in the transcriptional regulation mediated by DNA methylation, a cancer-type-paired similarity score (measured by the Jaccard index) was calculated based on both DMTDs and MethTFs across 22 cancer types. As a result, we observed that cancer types with similar tissue origins (KIRC vs KIRP) or within the same system (COAD vs STAD) were more consistent (Supplementary Figure S12A). Notably, Jaccard indices of DMTD from the same systems were significantly higher compared with DMTD from different systems ( $P = 0.013$ ; Supplementary Figure S12B–E). These results suggested that cancer types with similar tissue origin or within the same system exhibited a similar transcriptional regulation mechanism mediated by DNA methylation.

Moreover, we identified a type of MethTFs, pan-cancer MethTFs, whose regulatory activity was affected by DNA methylation in multiple cancer types. These MethTFs were involved in more DMTDs and were commonly identified in various cancer types. In individual cancer types, the regulatory activity of these MethTFs was affected by DNA methylation of more target genes and these MethTFs were likely to affect or be affected by other MethTFs, reflecting the centrality of these MethTFs in the DNA methylation-mediated transcriptional network. In addition, we observed that three MethTF categories were enriched in known cancer genes and pan-cancer MethTFs had the highest proportion. Pan-cancer MethTFs tended to be differentially expressed in many cancer types. All these observations revealed the important roles of pan-cancer MethTFs in carcinogenesis. Furthermore, functional analysis indicated that these MethTFs were mainly involved in two functional pathways, signal transduction and cell adhesion, involved in the following three cancer hallmarks: self-sufficiency in the growth signal, insensitivity to the antigrowth signal, and tissue invasion and metastasis. We found that the patterns mediated by DNA methylation of MethTFs became more complex if the MethTFs occurred in more cancer types. We observed that MethTFs took part in different hallmarks depending

on different target genes with distinct patterns mediated by DNA methylation. For example, the regulatory activity of KLF4 was attenuated by methylation of FBXO2 to regulate insensitivity to the antigrowth signal hallmark; while it was enhanced by methylation of CDH11 to regulate the tissue invasion and metastasis hallmark (Supplementary Figure S9). These observations provided a novel mechanism behind TF regulation mediated by DNA methylation in cancer, and deepens our understanding of how DNA methylation influenced transcriptional dysregulations to contribute to carcinogenesis.

The classical view was that methylation at CpG prohibited the recruitment of most TFs, but many recent studies have identified hundreds of TFs that preferred CpG-methylated sequences. Yin *et al.* used HT-SELEX to explore the impact of cytosine methylation on DNA binding specificities of TF in humans (20), confirming the pattern mediated by DNA methylation in many TFs in our work (Supplementary Table S5). However, a considerable degree of difference exists, which may be explained by the following reasons: On one hand, Yin *et al.* carried on an in vitro study while our work was performed in a cancer background. On the other hand, our work was based on changes in regulatory activity of TF at varied methylation levels, identifying both the direct impact of DNA methylation on the MethTF's regulatory activity through interference with base recognition and the indirect impact through other known or unknown mechanisms. Furthermore, Yin *et al.* found that DNA bindings of the TF member in some TF families was likely to be inhibited or enhanced by mCpG, while in our work, the regulatory activity of TF members in different families was influenced by more complex patterns in a different cancer type (Supplementary Figure S13). To test whether a specific TF structural family has common DNA methylation-mediated transcriptional characteristics across different cancer types, we performed a comparison analysis. We observed that MethTF cooperative pairs in a same family, of which two MethTFs belonged to the same family, preferred to have similar pattern types compared with MethTF cooperativity pairs in different families and this was a non-random phenomenon across cancer types (Supplementary Figure S14). Thus, we delineated a map on which TF members of 27 TF families preferred to have different methylation-mediated patterns across 22 cancer types by performing a hypergeometric analysis between TFs of a family and a pattern type (Supplementary Figure S15). These observations demonstrated that a specific TF structural family has common DNA methylation-mediated transcriptional characteristics across different cancer types.

In addition, we observed that several top-ranked biological processes were significantly enriched by MethTF cooperative pairs including many immune associated GO terms, which suggests the widespread perturbations of immune function in cancer (Supplementary Figure S16). Recently, immunotherapy has received the most attention in the field of cancer treatment, which has made pleasing progress. There are many immune-related genes (69) in our identified MethTFs, which could deepen our understanding of immunity in cancer from the perspective of the DNA methylation layer, and the immune-related MethTFs are provided in Supplementary Table S6 for researchers to perform further

analysis. Finally, we stratified the cancer patients into distinct subgroups with different clinical characteristics based on our identified MethTF cooperative pairs across 22 human cancer types.

In summary, we systematically identified the perturbations of the transcription regulation mediated by DNA methylation across 22 human cancer types based on a computational framework. This study explored the impact of DNA methylation on the regulatory activity of TF and further investigated their contribution to carcinogenesis, thus providing a highly valuable resource for further investigation of the impact of DNA methylation on the transcription regulation and for deepening our understanding of carcinogenesis in epigenetic layer.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Key R&D Program of China [2018YFC2000100]; National Natural Science Foundation of China [61873075, 31871338, 31970646]; Heilongjiang Touyan Innovation Team Program; Natural Science Foundation for Distinguished Young Scholars of Heilongjiang Province [JQ2019C004]. Funding for open access charge: National Key R&D Program of China [2018YFC2000100]; National Natural Science Foundation of China [61873075, 31871338, 31970646]; Heilongjiang Touyan Innovation Team Program; Natural Science Foundation for Distinguished Young Scholars of Heilongjiang Province [JQ2019C004].  
*Conflict of interest statement.* None declared.

## REFERENCES

- Won, K.J., Ren, B. and Wang, W. (2010) Genome-wide prediction of transcription factor binding sites using an integrated model. *Genome Biol.*, **11**, R7.
- Whitfield, T.W., Wang, J., Collins, P.J., Partridge, E.C., Aldred, S.F., Trinklein, N.D., Myers, R.M. and Weng, Z. (2012) Functional analysis of transcription factor binding sites in human promoters. *Genome Biol.*, **13**, R50.
- Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D. and Ren, B. (2005) A high-resolution map of active promoters in the human genome. *Nature*, **436**, 876–880.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E. *et al.* (2000) Genome-wide location and function of DNA binding proteins. *Science (New York, N.Y.)*, **290**, 2306–2309.
- Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M. and Brown, P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, **409**, 533–538.
- Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J. *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, 99–104.
- Farnham, P.J. (2009) Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.*, **10**, 605–616.
- Park, P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.*, **10**, 669–680.
- Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F. *et al.* (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, **457**, 854–858.
- Zhou, K.R., Liu, S., Sun, W.J., Zheng, L.L., Zhou, H., Yang, J.H. and Qu, L.H. (2017) ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res.*, **45**, D43–D50.
- Dunwell, T.L. and Holland, P.W. (2016) Diversity of human and mouse homeobox gene expression in development and adult tissues. *BMC Dev. Biol.*, **16**, 40.
- Burglin, T.R. (2011) Homeodomain subtypes and functional diversity. *Subcell. Biochem.*, **52**, 95–122.
- Singh, H., Khan, A.A. and Dinner, A.R. (2014) Gene regulatory networks in the immune system. *Trends Immunol.*, **35**, 211–218.
- Lee, T.I. and Young, R.A. (2013) Transcriptional regulation and its misregulation in disease. *Cell*, **152**, 1237–1251.
- Sur, I. and Taipale, J. (2016) The role of enhancers in cancer. *Nat. Rev. Cancer*, **16**, 483–493.
- Sizemore, G.M., Pitarresi, J.R., Balakrishnan, S. and Ostrowski, M.C. (2017) The ETS family of oncogenic transcription factors in solid tumours. *Nat. Rev. Cancer*, **17**, 337–351.
- Li, Y., Zhang, Y., Li, X., Yi, S. and Xu, J. (2019) Gain-of-function mutations: an emerging advantage for cancer biology. *Trends Biochem. Sci.*, **44**, 659–674.
- Chen, C.C., Xiao, S., Xie, D., Cao, X., Song, C.X., Wang, T., He, C. and Zhong, S. (2013) Understanding variation in transcription factor binding by modeling transcription factor genome-epigenome interactions. *PLoS Comput. Biol.*, **9**, e1003367.
- Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y. and Pritchard, J.K. (2011) Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.*, **21**, 447–455.
- Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F. *et al.* (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science (New York, N.Y.)*, **356**, pii: eaaj2239.
- Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C. *et al.* (2013) DNA methylation presents distinct binding sites for human transcription factors. *Elife*, **2**, e00726.
- Li, Y., McGrail, D.J., Xu, J., Mills, G.B., Sahni, N. and Yi, S. (2018) Gene regulatory network perturbation by genetic and epigenetic variation. *Trends Biochem. Sci.*, **43**, 576–592.
- Xuan Lin, Q.X., Sian, S., An, O., Thieffry, D., Jha, S. and Benoukraf, T. (2019) MethMotif: an integrative cell specific database of transcription factor binding motifs coupled with DNA methylation profiles. *Nucleic Acids Res.*, **47**, D145–D154.
- Wang, G., Luo, X., Wang, J., Wan, J., Xia, S., Zhu, H., Qian, J. and Wang, Y. (2018) MedReaders: a database for transcription factors that bind to methylated DNA. *Nucleic Acids Res.*, **46**, D146–D151.
- Zuo, Z., Roy, B., Chang, Y.K., Granas, D. and Stormo, G.D. (2017) Measuring quantitative effects of methylation on transcription factor-DNA binding affinity. *Sci. Adv.*, **3**, eaao1799.
- Cheung, H.H., Lee, T.L., Rennert, O.M. and Chan, W.Y. (2009) DNA methylation of cancer genome. *Birth Defects Res. C Embryo Today*, **87**, 335–350.
- Robertson, K.D. (2005) DNA methylation and human disease. *Nat. Rev. Genet.*, **6**, 597.
- Dong, Y., Zhao, H., Li, H., Li, X. and Yang, S. (2014) DNA methylation as an early diagnostic marker of cancer (Review). *Biomed. Rep.*, **2**, 326–330.
- Li, M., Chen, W.D., Papadopoulos, N., Goodman, S.N., Bjerregaard, N.C., Laurberg, S., Levin, B., Juhl, H., Arber, N., Moinova, H. *et al.* (2009) Sensitive digital quantification of DNA methylation in clinical samples. *Nat. Biotechnol.*, **27**, 858–863.
- Laird, P.W. (2003) The power and the promise of DNA methylation markers. *Nat. Rev. Cancer*, **3**, 253.
- Dhingra, P., Martinez-Fundichely, A., Berger, A., Huang, F.W., Forbes, A.N., Liu, E.M., Liu, D., Sboner, A., Tamayo, P., Rickman, D.S. *et al.* (2017) Identification of novel prostate cancer drivers using RegNetDriver: a framework for integration of genetic and epigenetic alterations with tissue-specific regulatory network. *Genome Biol.*, **18**, 141.
- Liu, Y., Liu, Y., Huang, R., Song, W., Wang, J., Xiao, Z., Dong, S., Yang, Y. and Yang, X. (2019) Dependency of the cancer-specific

- transcriptional regulation circuitry on the promoter DNA methylome. *Cell Rep.*, **26**, 3461–3474.
33. Morgunova, E. and Taipale, J. (2017) Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.*, **47**, 1–8.
  34. Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L. and Schubeler, D. (2015) Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature*, **528**, 575–579.
  35. Suzuki, T., Shimizu, Y., Furuhashi, E., Maeda, S., Kishima, M., Nishimura, H., Enomoto, S., Hayashizaki, Y. and Suzuki, H. (2017) RUNX1 regulates site specificity of DNA demethylation by recruitment of DNA demethylation machineries in hematopoietic cells. *Blood Adv.*, **1**, 1699–1711.
  36. de la Rica, L., Rodriguez-Ubrea, J., Garcia, M., Islam, A.B., Urquiza, J.M., Hernando, H., Christensen, J., Helin, K., Gomez-Vaquero, C. and Ballestar, E. (2013) PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome Biol.*, **14**, R99.
  37. Suzuki, T., Maeda, S., Furuhashi, E., Shimizu, Y., Nishimura, H., Kishima, M. and Suzuki, H. (2017) A screening system to identify transcription factors that induce binding site-directed DNA demethylation. *Epigenet. Chromatin*, **10**, 60.
  38. Li, Y., Li, L., Wang, Z., Pan, T., Sahni, N., Jin, X., Wang, G., Li, J., Zheng, X., Zhang, Y., Ulrich, A., Schirmacher, P., Herpel, E. et al. (2018) LncMAP: Pan-cancer atlas of long noncoding RNA-mediated transcriptional network perturbations. *Nucleic Acids Res.*, **46**, 1113–1123.
  39. Li, Y., McGrail, D.J., Xu, J., Li, J., Liu, N.N., Sun, M., Lin, R., Pancsa, R., Zhang, J., Lee, J.S. et al. (2019) MERIT: Systematic analysis and characterization of mutational effect on RNA Interactome Topology. *Hepatology*, **70**, 532–546.
  40. Klett, H., Balavarca, Y., Toth, R., Gigic, B., Habermann, N., Scherer, D., Schrotz-King, P., Ulrich, A., Schirmacher, P., Herpel, E. et al. (2018) Robust prediction of gene regulation in colorectal cancer tissues from DNA methylation profiles. *Epigenetics*, **13**, 386–397.
  41. Yagound, B., Smith, N.M.A., Buchmann, G., Oldroyd, B.P. and Remnant, E.J. (2019) Unique DNA methylation profiles are associated with cis-Variation in honey bees. *Genome Biol. Evol.*, **11**, 2517–2530.
  42. Jiang, W., Mitra, R., Lin, C.C., Wang, Q., Cheng, F. and Zhao, Z. (2016) Systematic dissection of dysregulated transcription factor-miRNA feed-forward loops across tumor types. *Brief. Bioinform.*, **17**, 996–1008.
  43. Plaisier, C.L., Pan, M. and Baliga, N.S. (2012) A miRNA-regulatory network explains how dysregulated miRNAs perturb oncogenic processes across diverse cancers. *Genome Res.*, **22**, 2302–2314.
  44. Li, Y., Li, S., Chen, J., Shao, T., Jiang, C., Wang, Y., Chen, H., Xu, J. and Li, X. (2014) Comparative epigenetic analyses reveal distinct patterns of oncogenic pathways activation in breast cancer subtypes. *Hum. Mol. Genet.*, **23**, 5378–5393.
  45. Uhlen, M., Zhang, C., Lee, S., Sjostedt, E., Fagerberg, L., Bidkhorji, G., Benfeytas, R., Arif, M., Liu, Z., Edfors, F. et al. (2017) A pathology atlas of the human cancer transcriptome. *Science*, **357**, eaan2507.
  46. Lu, J., Xu, J., Li, J., Pan, T., Bai, J., Wang, L., Jin, X., Lin, X., Zhang, Y., Li, Y. et al. (2018) FACER: comprehensive molecular and functional characterization of epigenetic chromatin regulators. *Nucleic Acids Res.*, **46**, 10019–10033.
  47. Li, Y., Xu, J., Chen, H., Bai, J., Li, S., Zhao, Z., Shao, T., Jiang, T., Ren, H., Kang, C. et al. (2013) Comprehensive analysis of the functional microRNA-mRNA regulatory network identifies miRNA signatures associated with glioma malignant progression. *Nucleic Acids Res.*, **41**, e203.
  48. Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A. et al. (2015) Proteomics. Tissue-based map of the human proteome. *Science*, **347**, 1260419.
  49. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R. and Weirauch, M.T. (2018) The human transcription factors. *Cell*, **172**, 650–665.
  50. Xu, J., Feng, L., Han, Z., Li, Y., Wu, A., Shao, T., Ding, N., Li, L., Deng, W., Di, X. et al. (2016) Extensive ceRNA-ceRNA interaction networks mediated by miRNAs regulate development in multiple rhesus tissues. *Nucleic Acids Res.*, **44**, 9438–9451.
  51. Zhao, W., Hisamuddin, I.M., Nandan, M.O., Babbitt, B.A., Lamb, N.E. and Yang, V.W. (2004) Identification of Kruppel-like factor 4 as a potential tumor suppressor gene in colorectal cancer. *Oncogene*, **23**, 395–402.
  52. Wei, D., Kanai, M., Huang, S. and Xie, K. (2006) Emerging role of KLF4 in human gastrointestinal cancer. *Carcinogenesis*, **27**, 23–31.
  53. Hu, W., Hofstetter, W.L., Li, H., Zhou, Y., He, Y., Pataer, A., Wang, L., Xie, K., Swisher, S.G. and Fang, B. (2009) Putative tumor-suppressive function of Kruppel-like factor 4 in primary lung carcinoma. *Clin. Cancer Res.*, **15**, 5688–5695.
  54. Bieche, I., Tozlu, S., Girault, I., Onody, P., Driouch, K., Vidaud, M. and Lidereau, R. (2004) Expression of PEA3/E1AF/ETV4, an Ets-related transcription factor, in breast tumors: positive links to MMP2, NRG1 and CGB expression. *Carcinogenesis*, **25**, 405–411.
  55. Hiroumi, H., Dosaka-Akita, H., Yoshida, K., Shindoh, M., Ohbuchi, T., Fujinaga, K. and Nishimura, M. (2001) Expression of E1AF/PEA3, an Ets-related transcription factor in human non-small-cell lung cancers: its relevance in cell motility and invasion. *Int. J. Cancer*, **93**, 786–791.
  56. Oh, S., Shin, S. and Janknecht, R. (2012) ETV1, 4 and 5: an oncogenic subfamily of ETS transcription factors. *Biochim. Biophys. Acta*, **1826**, 1–12.
  57. Yu, X., Lin, J., Zack, D.J. and Qian, J. (2006) Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues. *Nucleic Acids Res.*, **34**, 4925–4936.
  58. Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. and Rinn, J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.*, **25**, 1915–1927.
  59. Aguilo, F., Zakirova, Z., Nolan, K., Wagner, R., Sharma, R., Hogan, M., Wei, C., Sun, Y., Walsh, M.J., Kelley, K. et al. (2017) THAP1: role in mouse embryonic stem cell survival and differentiation. *Stem Cell Rep.*, **9**, 92–107.
  60. Meijnsing, S.H., Pufall, M.A., So, A.Y., Bates, D.L., Chen, L. and Yamamoto, K.R. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science (New York, N. Y.)*, **324**, 407–410.
  61. Wong, K.H. and Struhl, K. (2011) The Cyc8-Tup1 complex inhibits transcription primarily by masking the activation domain of the recruiting protein. *Genes Dev.*, **25**, 2525–2539.
  62. Pillai, S., Kovacs, M. and Chellappan, S. (2010) Regulation of vascular endothelial growth factor receptors by Rb and E2F1: role of acetylation. *Cancer Res.*, **70**, 4931–4940.
  63. Minato, Y., Tashiro, E., Kanai, M., Nihei, Y., Kodama, Y. and Imoto, M. (2007) Transcriptional regulation of a new variant of human platelet-derived growth factor receptor alpha transcript by E2F-1. *Gene*, **403**, 89–97.
  64. Li, Z., Guo, Y., Jiang, H., Zhang, T., Jin, C., Young, C.Y. and Yuan, H. (2014) Differential regulation of MMPs by E2F1, Sp1 and NF-kappa B controls the small cell lung cancer invasive phenotype. *BMC Cancer*, **14**, 276.
  65. Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R. et al. (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature*, **489**, 91–100.
  66. Angelini, C. and Costa, V. (2014) Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems. *Front. Cell Dev. Biol.*, **2**, 51.
  67. Gao, Y., Wang, P., Wang, Y., Ma, X., Zhi, H., Zhou, D., Li, X., Fang, Y., Shen, W., Xu, Y. et al. (2019) Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res.*, **47**, D1028–D1033.
  68. Li, Y., Sahni, N., Pancsa, R., McGrail, D.J., Xu, J., Hua, X., Coulombe-Huntington, J., Ryan, M., Tychon, B., Sudhakar, D. et al. (2017) Revealing the determinants of widespread alternative splicing perturbation in cancer. *Cell Rep.*, **21**, 798–812.
  69. Bhattacharya, S., Andorf, S., Gomes, L., Dunn, P., Schaefer, H., Pontius, J., Berger, P., Desborough, V., Smith, T., Campbell, J. et al. (2014) ImmPort: disseminating data to the public for the future of immunology. *Immunol. Res.*, **58**, 234–239.