Analyzing Integrated Network of Methylation and Gene Expression Profiles in Lung Squamous Cell Carcinoma

Yusri Dwi Heryanto (✉ yusri_dh@ims.u-tokyo.ac.jp)

The University of Tokyo

Kotoe Katayama

The University of Tokyo

Seiya Imoto

The University of Tokyo

Research Article

Keywords:

Posted Date: February 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1367606/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on September 22nd, 2022. See the published version at <https://doi.org/10.1038/s41598-022-20232-5>.

Analyzing Integrated Network of Methylation and Gene Expression Profiles in Lung Squamous Cell Carcinoma

Yusri Dwi Heryanto^{1,*}, Kotoe Katayama², and Seiya Imoto^{1, 2}

¹Division of Health Medical Intelligence, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

²Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

*yusri_dh@ims.u-tokyo.ac.jp

ABSTRACT

Gene expression, DNA methylation, and their organizational relationships are commonly altered in lung squamous cell carcinoma (LUSC). To elucidate these complex interactions, we reconstructed a differentially expressed genes (DEGs) network and differentially methylated cytosine (DMCs) network by partial information decomposition and inverse correlation algorithm, respectively. Then, we performed graph union to integrate the networks. The community detection and enrichment analysis of the integrated network revealed the tight interaction between the gene sets of the cell cycle, keratinization, immune system, and xenobiotic metabolism in LUSC. The DMCs analysis showed that the hypomethylation targeted the gene sets that were responsible for cell cycle, keratinization, and NRF2 pathway. In other hand, hypermethylated genes affected the circulatory system development, immune system, extracellular matrix organization, and cilium organization. By centrality measurement, we detected NCAPG2, PSMG3, and FADD as hub genes which were highly connected to other nodes and might play important roles in LUSC gene dysregulation. We also found that the high betweenness centrality genes affect the patients' survival. These results showed the integrated network analysis enabled us to obtain the global view of the interactions and regulations in LUSC.

Introduction

Lung squamous cell carcinoma (LUSC) is the second most common subtype of lung cancer after lung adenocarcinoma, accounting for the 20 percent of all lung cancer diagnoses¹. It is characterized by the keratinization and/or intercellular bridges of the lung epithelial cells². The progressive accumulation of mutations and epigenetic abnormalities are common and drive the LUSC progression³. The progress in LUSC research found the roles of genetic abnormalities of TP53, PI3KCA, FGFR1 and others in LUSC pathogenesis and treatments⁴. Epigenetic studies on LUSC also found the important drivers of cancer, such as the methylations of NFE2L2, SOX2, and TP63^{5,6}. However, only few studies explore and analyze the organizational and hierarchical interaction between these drivers in LUSC. The studies of the interaction between the genes and its regulators are vital to understand the pathogenesis and to aid the management of LUSC.

Network-based modelings are powerful approaches for analyzing the interactions between variables. A network or graph is a mathematical structure made up of vertices (or nodes) connected by edges (or links). The vertices and edges might hold some properties which describe its characteristics. Network-based modelings have been used to study the genes and cytosine methylations relationship. For example, the network study in leukemia found that both of the gene expression and methylation consistently affected the Ras, PI3K-Akt, and Rap1 signaling pathway⁷. Another study has identified novel cancer related pathway by integrating the methylation data and protein-protein interaction networks⁸. These studies used the networks that were obtained from open source databases.

In our study, we reconstructed the DEGs network and DMCs network on LUSC computationally using the Partial Information Decomposition and Context (PIDC)⁹ and Enhancer Linking by Methylation/Expression Relationships (ELMER) algorithm¹⁰, respectively. The advantage of this approach is it enables us to find novel interactions that have not been included in the databases. Then, we combined DEGs-DMCs network and extracted the giant component of the graph. This graph consists of nodes that represent genes or methylation probes, and undirected edges that represent the regulatory connections. Overall, the graph provides a blueprint of the molecular interactions in cancer that are significantly different to its normal tissue counterpart. We can get many informations by analyzing the graph topology, for example, the identification of the important regulatory genes by centrality measurements^{7,11} and the clustering of the similar nodes using the community detection analysis¹². By

integrating the DMCs network and DEGs network, we could study the coordination of cellular systems at gene and methylation level simultaneously. Our analysis may provide the basis for identification of novel interactions and core regulatory genes in LUSC.

Results

Network characteristics

Figure 1. The flowchart of the analyzes steps in our study.

Using the data derived from Genomic Data Commons-The Cancer Genome Atlas Lung Squamous Cell Carcinoma (GDC-TCGA-LUSC) datasets, we performed differential expression analyzes of gene expression and cytosine methylation. Then, we used PIDC and ELMER algorithm to reconstruct DEGs and DMCs network, respectively. Briefly, the Partial Information Decomposition decomposes the mutual information between variables into unique, redundant, and synergistic components. PIDC calculated the relationship between variables as the mean proportion of unique components. The PIDC will return all possible edges between variables and its ranks. The edges in the DEGs network represent the highest 1% of PIDC rank. For the methylation network, The ELMER algorithm selects the closest 10 upstream genes and the closest 10 downstream genes for each DMC. Then, it tested for the inverse correlation between the the DMC methylation and the genes expression. The edges in the DMCs network represent the significant (*adjusted P* < 0.01) inverse correlations. Next, we took the union of both graphs and extracted the giant component of the graphs. The flowchart in Fig. 1 summarizes the analyzes steps of our study.

The final result of the integrated network had 9748 nodes and 228246 edges. Out of 9748 nodes, 7903 were marked as DEGs and the remaining 1845 were DMCs probes. The edges consisted of 224149 gene-gene and 4097 probe-gene interaction edges (3369 hypomethylation and 728 hypermethylation edges) (Fig. 2). We listed all the nodes and edges in Supplementary Table S1 and S2.

Figure 2. The integrated network of cytosine methylation and gene expression. *Green nodes* are the genes and *red nodes* are the methylation probes. *Blue edges*, *red edges*, and *gray edges* are the hypomethylation, hypermethylation, and gene-gene relationships, respectively.

Community identification analysis

Using Leiden algorithm¹², we identified the 10 biggest communities that had at least 200 nodes and made of about 50% of total nodes. We named and ranked each of the community based on the number of the nodes in the community (e.g. the biggest community is Community 1, the second biggest is Community 2, and so on) (Fig. 3). The Gene set enrichment analysis revealed the functional classes of each community. For example, the biggest community, Community 1, was mainly genes for DNA replication and cell cycle. The next biggest communities, Community 2, 3, and 4, were responsible for keratinization, immune system, and complement-coagulation cascade pathways, respectively. We listed the 10 biggest communities and some of its functional class in the Table 1. The complete list is shown in Supplementary Table S3.

The Fig. 3 helps visualize the interaction between community. We used ForceAtlas2 as network layout algorithm to display the network in 2-dimensional image¹³. The ForceAtlas2 is a force-directed layout algorithm where nodes repulse each other like charged particles, while edges attract their nodes, like springs. In this algorithm, the stronger the interaction between the communities, the closer they are. For example, Community 3 have closer relationship to the Community 6 and 4 than to the Community 8. To quantitatively measure the strength of the interaction, we calculated $C_x(y)$ which are the ratio of the links connected between Community x and y to the total number of intercommunity links on Community x . The intercommunity links are the links that connect one community to another different community. The Fig. 4 shows the heatmap of the ratio $C_x(y)$ from Source community x to the Target Community y . We used the same as previous Community 3 as example. As we can see, the Community 3 in row 3 has most of its intercommunity links connects to the Community 4 (26%) and Community 6 (14%). We test the significance of $C_x(y)$ using network randomization test. We found that every $C_x(y)$ value on Fig. 4 were not random (Supplementary Table S4).

Figure 3. The interaction of the 10 biggest communities in the network. We included only nodes that have degree at least 15 to simplify the illustration. The size of the nodes are correlated with the betweenness centrality. The gray nodes are the nodes which are not the member of any 10 major communities. We also listed some of the functional classes of each community.

Centrality measure analysis

Centrality measure analysis was used to investigate the roles of some nodes and their impact on the networks. There are some centrality measurement such as degree centrality, closeness centrality, and betweenness centrality. Here, we focused on the betweenness centrality. Betweenness centrality of a node measures the number of the shortest paths that pass through that node. In gene regulation network, sometimes the most important nodes in the system are not the one with the highest number of edges but the middleman that connect groups or the ones who have the most control over the flow of the information. Betweenness centrality measures the amount of influence a node has over the flow of information and is mathematically formulated as follows:

$$B(u) = \sum_{s \neq u \neq t} \sigma_{st}(u) / \sigma_{st}$$

where u is a node, σ_{st} is the total number of shortest paths between nodes s and t , and $\sigma_{st}(u)$ is the number of the shortest paths between nodes s and t that pass node u . Betweenness centrality has been used to identify genes that have high impacts in leukemia patients' survival⁷ and core regulator in breast cancer cells¹¹. We calculated the betweenness centrality of every node (Supplementary Table S1). The 3 highest betweenness centrality genes are NCAPG2, PSMG3, and FADD.

In our study, we investigated the survival of two groups of genes set: high rank and low rank group, which represent genes in the top 10% and bottom 10% of the betweenness centrality ranking, respectively. By Kaplan Meier Survival analysis, we found that 192 from 875 ($\pm 21.9\%$) high rank genes were significantly ($P < 0.05$) affect survival. In contrast, only 92 from 764 ($\pm 12\%$) low rank genes were significantly ($P < 0.05$) affect survival (Fig. 5, Supplementary Table S5).

Figure 4. The heatmap of the ratio $C_x(y)$. Each row of the heatmap shows the relative proportion $C_x(y)$ of the intercommunity links from the source community x (vertical axis), to the target community y (horizontal axis). Every $C_x(y)$ value in the figure is not random by network randomization test.

Methylation analysis

There are 1842 hypomethylated genes and 313 hypermethylated genes in our networks. The hypomethylated genes and probes were dominant in Community 1, 2, 7, and 8, which regulated cell cycle, keratinization, NRF2 pathway, and tRNA modification in the mitochondrion, respectively. By contrast, the hypermethylated genes and probes were dominant in Community 3 (immune system), 4 (tube development, blood vessel morphogenesis), 6 (circulatory system development, extracellular matrix (ECM) organization), and 10 (cilium movement and organization) (Fig. 6).

When a group of DMCs acts as enhancers or silencers in a specific sample subset, this is often the result of an altered upstream master regulator transcription factor (TF)⁵. By using function `get.enriched.motif` and `get.TFs` in ELMER package, we identified the enriched motifs and master regulator of methylation change in LUSC. In hypomethylated probes, the top enriched motifs were FOSL2, FOSB, FOSL1, and FOS. We identified ZNF74, TP63, KLF5, TFAP4, and ZFP64 as master regulator TFs of hypomethylated genes. In hypermethylated probes, the ZBT14, E2F2, SP1, and SP2 were top enriched motif and CREB3L1, CXXC5, ETS1 were the master regulators. We also investigated the TFs for each community. Only communities that have at least 10 DMCs probe connected to it were considered (Table 2). All of top enriched motifs and master regulator TFs on global and community level were listed in Supplementary Table S6 and S7, respectively.

Discussion

In this study, we performed network-based modeling to study the interaction between genomic and methylomic profiles in LUSC. Because the nature of the methylation and gene expression data are different, we used 2 different methods to reconstruct the DEGs and DMCs network, which are PIDC and ELMER algorithm, respectively. The PIDC and ELMER are well-tested and have been used in many studies as network inference methods^{5,14,15}.

Many network inference methods such as GENIE3¹⁶, Partial correlation¹⁷, and SINCERITIES¹⁸ can also reconstruct DEGs network. However, we chose PIDC because it has a strong mathematical background, can detect noisy or non-linear relationship, reasonably fast, and does not need time series data to reconstruct gene network^{9,14}. The PIDC algorithm uses multivariate information measures to identify the relationship between genes. The information measurements can quantify dependence

Community	Source	Term name	Term size	Intersection size	P value
Community_1	REAC	Cell Cycle	679	181	2.42e-86
Community_1	REAC	Cell Cycle, Mitotic	549	154	3.22e-75
Community_1	REAC	Cell Cycle Checkpoints	290	96	8.81e-52
Community_2	REAC	Formation of the cornified envelope	128	65	1.52e-55
Community_2	REAC	Keratinization	215	66	2.75e-39
Community_2	WP	Hair Follicle Development: Cytodifferentiation - Part 3 of 3	89	24	2.16e-13
Community_3	REAC	Immune System	2038	217	6.93e-60
Community_3	REAC	Innate Immune System	1090	127	2.69e-33
Community_3	REAC	Neutrophil degranulation	476	76	6.74e-27
Community_4	KEGG	Complement and coagulation cascades	85	19	2.86e-10
Community_4	REAC	Surfactant metabolism	29	11	6.02e-8
Community_4	REAC	Diseases associated with surfactant metabolism	9	7	1.82e-7
Community_5	REAC	Glucuronidation	25	7	3.02e-6
Community_5	KEGG	Neuroactive ligand-receptor interaction	340	20	3.66e-6
Community_5	WP	miRNAs involved in DNA damage response	50	9	1.29e-5
Community_6	REAC	Extracellular matrix organization	298	47	5.96e-26
Community_6	REAC	Collagen formation	89	20	4.83e-13
Community_6	REAC	Collagen degradation	64	16	6.40e-11
Community_7	WP	NRF2 pathway	145	28	3.35e-16
Community_7	KEGG	Glutathione metabolism	56	13	8.42e-9
Community_7	WP	Nuclear Receptors Meta-Pathway	321	30	9.68e-9
Community_8	REAC	tRNA modification in the mitochondrion	8	3	1.01e-2
Community_8	KEGG	Serotonergic synapse	112	6	1.99e-2
Community_9	WP	Dopaminergic Neurogenesis	30	7	1.69e-5
Community_9	REAC	Transport of inorganic cations/anions and amino acids/oligopeptides	105	8	4.63e-3
Community_9	REAC	SLC-mediated transmembrane transport	248	11	1.80e-2
Community_10	WP	Ciliopathies	184	27	1.24e-30
Community_10	KEGG	Huntington disease	306	10	1.29e-5
Community_10	KEGG	Amyotrophic lateral sclerosis	363	10	6.30e-5

Table 1. The gene set enrichment analysis revealed the enriched pathways on each community. In this table, we only showed the Reactome (REAC), Kyoto Encyclopedia of Genes and Genomes (KEGG), and WikiPathways (WP) as the gene set data sources.

between variables without making assumptions about the nature of the dependencies¹⁹. It is ideal for the noisy and non-linear relationships which are usually seen in cancer transcriptomic datasets. The PIDC divides the information between variables into categories: redundant, unique, and synergistic. By doing so, PIDC can distinguish between unique information from a pair of variables within the group and redundant information shared by multiple variables⁹. The PIDC algorithm has shown to have more accuracy, stability, and efficiency compared to other network inference methods¹⁴. The limitations of PIDC are it does not provide the edges direction information and the effect of the relationship (e.g. inhibitory or excitatory). We advised the researchers that needs the edge direction and node sign information to use other network inference methods (e.g. GENIE3 for directed graph or partial correlation for signed networks).

ELMER uses methylation changes at cis-regulatory modules in tumors as the central hub of the DMCs network. Then, it uses correlation analysis to associate them with both upstream regulator transcription factors and downstream target genes¹⁰. The ELMER algorithm has been used in some studies to investigate the methylation landscape of many cancers^{5,15}. Thus, ELMER not only can reconstruct methylation network but also infer the master regulator transcription factors that bind to the methylation motif binding site.

The integration was performed at the network level by performing graph union of DEGs and DMCs network. We performed the community identification, centrality measurement, and gene set enrichment analysis to discover the relationship patterns in the integrated network.

The community detection analysis revealed the subnetworks communities, which has strong interactions between nodes in

Figure 5. The P value survival analysis distribution of the genes in the high rank (top 10%) and low rank (bottom 10%) betweenness centrality groups. The red line is the P value significant cut-off ($P = 0.05$). Some of the lowest P value genes from each group are shown in the figure.

the same community than nodes in different groups. Using functional enrichment analysis, we found that each community targeted the specific biological process or pathways. The pathways of DNA replication, cell cycle, ECM organization, and immune system are the common pathways altered in many cancers²⁰. The keratinization, cilium organization, and surfactant metabolism are the LUSC and lung cell characteristics. The complement-coagulation cascade pathway reflects the importance of complement in regulating tumor microenvironment^{21,22} and the risk of coagulation disorder in LUSC²³. We hypothesized that the genes in the major communities were heavily dysregulated in LUSC.

Another important finding is 2 from 10 major communities are related to detoxification-related pathways: Community 5 of glucuronidation pathway and Community 7 of NRF2 and glutathione metabolism pathway. The glucuronidation and glutathione metabolism pathways are related to phase II enzymes for metabolizing of xenobiotic²⁴. The NRF2 pathways are master regulator of antioxidant response²⁵. It might relate to the therapy resistance by abusing antioxidant response and the fact that smoking is the main risk factor for LUSC²⁶.

The network visualization in Fig. 3 revealed how the different communities interact in LUSC. From here after, we used the term Connection $x - y$ to describe the interaction between Community x and y . For example, we explored the relationship between Community 2 and 7 or Connection 2-7. The Community 2 affects keratinization and Community 7 has roles in detoxification, such as NRF2 and glutathione metabolism pathway. In Fig. 3, we can see the position of the Community 2 is very close to Community 7. The majority of intercommunity links ($\pm 23\%$) on Community 2 connect to the Community 7. The converse is also true. Most of the Community 7 intercommunity links ($\pm 34\%$) connect to the Community 2. It indicates the tight interaction between both of Community. In a well-written review by Ishitsuka *et al.*, they discussed the extensive importance of NRF2 in keratinization²⁵. They stated that the KEAP1/NRF2 pathway plays important roles on regulator of keratinization, squamous epithelial tissue external responses, and detoxification. Based on our finding, we hypothesized the high number of edges in Connection 2-7 showed the dysregulation of the capability of lung epithelial cell to response to external or toxic stimuli. This is supported by the studies that showed the NRF2 deficiency led to the chemical carcinogens susceptibility²⁷ and smoking alter the NRF2 and glutathione pathway^{26,28}.

Another example is the interaction between Community 3 of immune system, Community 4 of complement-coagulation

Figure 6. The number of differentially hypomethylated/hypermethylated genes and probes in each community.

cascade, and Community 6 of ECM organization. The ratio $C_3(4)$ and $C_3(6)$ are 0.26 and 0.14, respectively. The complement is one of a key actor in innate immunity and the coagulation system. In other hand, immune system cells such as tumor-associated-macrophages (TAM) and tumor-associated-neutrophils (TAN) have major roles in reshaping the tumor ECM^{29,30}. Many studies conducted to investigate the ECM-immune cell-complement interactions in lung cancer. Complement C1q proteins could activate and recruit TAM³¹. C9 were down-regulated in alveolar TAM that leads to lung cancer progression³². Macrophages could also regulate C3-independent C5a generation that promotes squamous carcinogenesis³³. Other studies about TANs showed C5a could recruit TAN by stimulating the release of leukotriene B4 and the production of IL-1^{34,35}. It was also shown that C3aR-dependent neutrophil extracellular traps (NETs) could accumulate TANs³⁶. Some studies have demonstrated lung cancer cells have higher concentrations and expression of C3a and of C5a than nonmalignant lung cells^{37,38}.

The next step is to find the highly connected nodes/genes or hub genes using centrality measurement. In the graph theory, the removal of the hub nodes in a network increases the proportion of unreachable groups of nodes than non-hubs genes. Hence, hub nodes are important for maintaining the global network structure. In many organisms, the removal of hub genes is more likely to be lethal than non-hub genes. This is known as the centrality-lethality rule³⁹. The examples of centrality measurements are degree centrality, closeness centrality, and betweenness centrality. Studies have shown that betweenness and degree centrality were better than closeness centrality to identify hub genes^{11,40}. Our result supports the centrality-lethality rule. The expressions of the genes that have high betweenness centrality in our network are more likely to affect the survival of the LUSC patient than the low-betweenness-genes. The NCAPG2, PSMG3, and FADD were the three highest betweenness centrality genes in our integrated network. The NCAPG2 protein is a subunit of condensin II complex, which has roles in mitotic chromosome assembly and segregation. The upregulation of NCAPG2 promotes proliferation of lung cancer cells⁴¹. The PSMG3 is the Chaperone protein, which promotes assembly of the 20S proteasome. As far as the author's knowledge, no papers have investigated the effect of the PSMG3 dysregulation on LUSC. However, the antisense long non-coding RNAs of PSMG3, PSMG3-AS1, is highly expressed in LUSC and its inhibition reduced the invasiveness⁴². In our study, we found that the expression PSMG3 or ENSG00000157778 affected the patient's survival significantly (Supplementary Table S5). The further study of PSMG3 may become a potential prognostic marker and therapeutic target to treat LUSC in the future. The Fas-associated death domain protein (FADD) transmits the apoptotic signal delivered by death receptors. The release of FADD by non-small cell lung cancer cell correlated with the aggressiveness and metastasis⁴³.

The methylation analysis of the network showed that the hypomethylated probes were targeted the cell cycle (Community

Community	Methylation type	Top motif	Top common TFs
Global	hypomethylation	FOSL2, FOSB, FOSL1, FOS	ZNF74, TP63, KLF5, TFAP4, ZFP64
1	hypomethylation	FOSB, P63, P53, FOS	ZNF74, TFAP4, KLF5, TP63, ZFP64
2	hypomethylation	FOSL2, FOSL1, FOS, P53	TP63, KLF5, SOX15
5	hypomethylation	P53	TP63, SOX2, KLF5
7	hypomethylation	P73, P53, P63	TP63, SOX2
8	hypomethylation	P53, ANDR, P63, P73	TP63, TFAP4
Global	hypermethylation	ZBT14, E2F2, SP1, SP2	CREB3L1, CXXC5, FOXP1, ETS1
1	hypermethylation	BHE41	VENTX, CXXC5, TBX5
3	hypermethylation	GMEB2, ZBT14, SP2	FLI1, SPI1, IRF4, MEF2C
4	hypermethylation	SP2, SP1, HME2	FOXP1, ETS1, FLI1, IRF4
5	hypermethylation	CENPB, SP1, MBD2	HNF1B, RORC, NFE2
6	hypermethylation	VEZF1, E2F5, ZF64A	ETS1, FLI1

Table 2. The community top motif and master regulator. Only communities that have at least 10 DMCs probe connected to it were considered. Global refers to the master regulator TFs of all hypomethylated or hypermethylated probes

1), NRF2 and glutathione metabolism pathway (Community 7), keratinization (Community 2), and tRNA modification in the mitochondrion pathway (Community 8). The up-regulation of these pathways has been linked to invasiveness, therapy resistance, smoking, and poor prognosis in many studies^{20,26,44–46}. In contrast, the hypermethylated DMCs probes down-regulated the pathways related to cancer inhibition and normal development/differentiation of tissue, such as immune system (Community 3), tube development and blood vessel morphogenesis (Community 4), circulatory system development and ECM organization (Community 6), and cilium organization (Community 10).

The master regulator TFs gain (for hypomethylated probes) or loss (for hypermethylated probes) can change the methylation status of DMCs. We used ELMER package to identify these upstream master regulator TFs. In our study, we found that TP63, KLF5, and SOX2 were over-expressed and became top TFs for hypomethylated probes at both of global or community level. This result is supported by previous chromatin immunoprecipitation sequencing study that found the TP63, SOX2, and KLF5 were core regulators which determined chromatin accessibility, epigenetic modifications, and gene expression patterns in esophageal squamous cell carcinoma⁴⁷. In contrast, we found that the suppressed expression of CXXC5 and FOXP1 acted as regulator of hypermethylation. CXXC5 is nuclear zinc-finger protein comprising DNA methyltransferases, DNA demethylases, histone methyltransferases, and histone demethylases, that contribute to transcriptional regulation by preferentially binding to unmethylated CpG islands⁴⁸. CXXC5 is a negative-feedback regulator of the Wnt/beta-catenin pathway⁴⁹, an inhibitor of liver cancer by promoting TGF-beta-induced cell cycle arrest⁵⁰, and required for DNA damage-induced p53 activation⁵¹. FOXP1 is a TF that belongs to the P subfamily of the Forkhead Box family. FOXP1 is a prostate cancer suppressor by regulating androgen receptor and FOXA1⁵². FOXP1 is also associated with improved survival in lung cancer⁵³.

Taken together, all the findings of our study suggest that integrating the DMCs network and DEGs network has potential to reveal the complex interactions between genes and its regulator (e.g. TFs and methylated cytosines). Our findings are supported by many experimental studies. Our analysis workflow can also work not only in LUSC, but also in other cancers and diseases. We believe that the deeper understanding of the global organizational structure of gene regulatory network will assist us in LUSC diagnostic and therapeutic management.

Methods

Lung cancer datasets preparation

The HTSeq-FPKM-UQ gene expressions data of 502 LUSC primary tumor sample and 49 normal tissue sample were downloaded from the NCI Genomic Data Commons (GDC) data portal using Bioconductor package TCGAAbiolinks⁵⁴. The gene expressions data were processed using the TCGAAbiolinks workflow from Silva *et al*⁵⁵. In short, we removed outliers, failed hybridizations, or mistracked samples by performing Array Array Intensity correlation using TCGAanalyze_Preprocessing function. Next, we normalized mRNA transcripts and filtered genes with low signal across samples using TCGAanalyze_Normalization and TCGAanalyze_Filtering functions, respectively. We selected the genes that differentially expressed two-fold by TCGAanalyze_DEA. Then, we performed $\log(1+x)$ transform and standardization to the expression of the selected genes.

TCGA level 3 DNA methylation data based on the Illumina Infinium HumanMethylation450 (HM450) BeadArray platform from the same sample were downloaded using getTCGA function from ELMER package¹⁰.

Networks reconstruction

We used the PIDC algorithm to reconstruct the DEGs network. In PIDC, we examined every gene triple, for example, *Source*₁, *Source*₂, and *Target*. The PIDC algorithm divides the information between *Source*₁ and *Target* into three categories: redundant, unique, and synergistic. Redundant information is the portion of information about *Target* that either *Source*₁ or *Source*₂ can provide. The unique information is uniquely contributed from *Source*₁ or *Source*₂ only. The synergistic information is the portion of information from both *Source*₁, *Source*₂. Then, the direct functional relationship between *Source*₁ and *Target* is calculated by using the mean proportion of the unique information between *Source*₁ and *Target* over all *Source*₂⁹.

We used the PIDC algorithm implemented in package `NetworkInference.jl`⁹. The Bayesian blocks algorithm and maximum likelihood were used as discretizer algorithm and estimator parameter, respectively. It returns all possible edges between genes and its ranks. Finally, a DEGs network was defined by keeping the highest 1% of ranked edges from the PIDC algorithm results.

DMCs network reconstruction was performed using ELMER package. The ELMER network reconstruction has 3 main steps:

1. identifying distal methylation probes on HM450 platform,
2. selecting distal probes with significantly different DNA methylation level between normal and tumor groups (i.e. both hyper and hypomethylation),
3. connect putative target genes and its differentially methylated cytosine (DMC) probes.

This process will return the probe-gene pairs who have significant (*adjusted P* < 0.01) inverse correlation between methylation of the probe and expression of the gene.

Motif and transcription factor identification

To identify transcription factor (TF) which act as upstream master regulator by binding to TF binding motif DNA methylation, we performed two additional steps:

1. identifying enriched motif in the differentially methylated probes
2. identify regulatory transcription factor (TF) whose expression associates with TF binding motif DNA methylation

We performed all the steps from DMCs network reconstruction until TF identification using `TCGA.pipe` function on unsupervised modes from the ELMER package.

Networks integration

To combine the DMCs and DEGs network, we performed graph union operation. Thus, the vertices and edges in the integrated network are the union of vertices and edges from DMCs and DEGs network. Then, we extracted the subgraph that has the largest number of connected component or the giant component of the graph.

Network analysis

We performed the community identification, genes set enrichment analysis, and betweenness centrality measurement on the integrated network. The community detection of the network was performed using the Leiden algorithm¹² through Python package `leidenalg` (<https://github.com/vtraag/leidenalg>). In this function, we used `CPMVertexPartition` as the objective function parameter and 0.0085 as the value of the resolution parameter. The genes set enrichment analysis from each of the communities were performed using `g:Profiler`⁵⁶. We used Kyoto Encyclopedia of Genes and Genomes, Reactome, WikiPathways, and Gene Ontology (molecular function, cellular component, biological process) as data sources in `g:Profiler`. The betweenness centrality was calculated through package `Graph.jl`⁵⁷.

To analyze the intercommunity interactions, we visualized the network in Gephi using ForceAtlas2 as network layout algorithm^{13,58}. Then, we calculated a ratio $C_x(y)$, which measures the proportion of the links that connected between Source Community x and Target Community y to the total of the links on the Source Community x that are not connected to the Source Community itself (e.g. intercommunity links). It is formulated as follows:

$$C_x(y) = \frac{L_x(y)}{\sum_{z \neq x} L_x(z)}$$

where $L_x(y)$ is the number of the links that connect the Community x and Community y . By definition, $L_x(y)$ is equal to $L_y(x)$. The denominator of $C_x(y)$ is the sum of the number of the links that connected the Community x to other community except the Community x itself.

We performed network randomization tests to determine whether the ratio $C_x(y)$ is happen by random chance or not. The steps of the network randomization tests is shown in Algorithm 1

```

Data: Original network, number of iteration  $N = 5000$ ;
Result:  $L$  = An array with length  $N$  of  $C_x(y)$  matrix ( $K \times K$ )
Count  $k \leftarrow$  the number of the unique community, ;
Count  $E_{inter} \leftarrow$  number of intercommunity links;
 $n \leftarrow 0$ ;
 $L \leftarrow$  empty list;
while  $n < N$  do
     $n \leftarrow n + 1$ ;
     $Network_{copy} \leftarrow$  copy of the original network;
    delete all intercommunity links in  $Network_{copy}$ ;
     $e \leftarrow 0$ ;
    while  $e < E_{inter}$  do
         $e \leftarrow e + 1$ ;
        Pick 2 random nodes from different community in  $Network_{copy}$  and create a link between them;
    end
    Calculate  $C_x(y)$  for every pair of community  $x$  and  $y$  in  $Network_{copy}$ ;
    Store the result on a  $K \times K$  matrix;
    Append the matrix on the  $L$ ;
end

```

Algorithm 1: Network randomization procedure

We then performed the one-sample-t-test on the observed $C_x(y)$ and the $C_x(y)$ from the network randomization procedure. We adjusted the P -values using Bonferroni correction.

Survival analysis

We selected the genes in the upper 90th-percentile and lower 10th-percentile of betweenness centrality ranking and grouped it into two. Then, we performed an univariate Kaplan-Meier survival analysis for each selected gene. We used the 33th-percentile and 67th-percentile as the quantile threshold to identify samples with low and high expression of a gene. We used TCGAanalyze_SurvivalKM in TCGAbiolinks package to perform survival analysis.

The source code to perform and replicate all analyses in our study are available at GitHub repository (https://github.com/yusri-dh/LUSC_integrated_network/).

References

1. Lewis, D. R., Check, D. P., Caporaso, N. E., Travis, W. D. & Devesa, S. S. Us lung cancer trends by histologic type. *Cancer* **120**, 2883–2892, DOI: <https://doi.org/10.1002/cncr.28749> (2014). <https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.1002/cncr.28749>.
2. Kumar, V., Abbas, A. K. & Aster, J. C. *Robbins & Cotran Pathologic Basis of Disease 10th Edition* (Elsevier, Philadelphia, PA, 2021).
3. Network, T. C. G. A. R. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525, DOI: [10.1038/nature11404](https://doi.org/10.1038/nature11404) (2012).
4. Perez-Moreno, P., Brambilla, E., Thomas, R. & Soria, J.-C. Squamous cell carcinoma of the lung: Molecular subtypes and therapeutic opportunities. *Clin. Cancer Res.* **18**, 2443–2451, DOI: [10.1158/1078-0432.ccr-11-2370](https://doi.org/10.1158/1078-0432.ccr-11-2370) (2012).
5. Yao, L., Shen, H., Laird, P. W., Farnham, P. J. & Berman, B. P. Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol.* **16**, DOI: [10.1186/s13059-015-0668-3](https://doi.org/10.1186/s13059-015-0668-3) (2015).
6. Yuan, G. *et al.* Elevated NSD3 histone methylation activity drives squamous cell lung cancer. *Nature* **590**, 504–508, DOI: [10.1038/s41586-020-03170-y](https://doi.org/10.1038/s41586-020-03170-y) (2021).
7. Sanchez, R. & Mackenzie, S. A. Integrative network analysis of differentially methylated and expressed genes for biomarker identification in leukemia. *Sci. Reports* **10**, DOI: [10.1038/s41598-020-58123-2](https://doi.org/10.1038/s41598-020-58123-2) (2020).
8. Li, J., Zhang, Q., Chen, Z., Xu, D. & Wang, Y. A network-based pathway-extending approach using DNA methylation and gene expression data to identify altered pathways. *Sci. Reports* **9**, DOI: [10.1038/s41598-019-48372-1](https://doi.org/10.1038/s41598-019-48372-1) (2019).

9. Chan, T. E., Stumpf, M. P. & Babbie, A. C. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* **5**, 251–267.e3, DOI: [10.1016/j.cels.2017.08.014](https://doi.org/10.1016/j.cels.2017.08.014) (2017).
10. Silva, T. C. *et al.* ELMER v.2: an r/bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics* **35**, 1974–1977, DOI: [10.1093/bioinformatics/bty902](https://doi.org/10.1093/bioinformatics/bty902) (2018).
11. Narang, V. *et al.* Automated identification of core regulatory genes in human gene regulatory networks. *PLOS Comput. Biol.* **11**, e1004504, DOI: [10.1371/journal.pcbi.1004504](https://doi.org/10.1371/journal.pcbi.1004504) (2015).
12. Traag, V. A., Waltman, L. & van Eck, N. J. From louvain to leiden: guaranteeing well-connected communities. *Sci. Reports* **9**, DOI: [10.1038/s41598-019-41695-z](https://doi.org/10.1038/s41598-019-41695-z) (2019).
13. Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLoS ONE* **9**, e98679, DOI: [10.1371/journal.pone.0098679](https://doi.org/10.1371/journal.pone.0098679) (2014).
14. Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A. & Murali, T. M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* **17**, 147–154, DOI: [10.1038/s41592-019-0690-6](https://doi.org/10.1038/s41592-019-0690-6) (2020).
15. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, DOI: [10.1126/science.aav1898](https://doi.org/10.1126/science.aav1898) (2018).
16. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5**, e12776, DOI: [10.1371/journal.pone.0012776](https://doi.org/10.1371/journal.pone.0012776) (2010).
17. Kim, S. Ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. Methods* **22**, 665–674 (2015).
18. Papili Gao, N., Ud-Dean, S. M. M., Gandrillon, O. & Gunawan, R. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. *Bioinformatics* **34**, 258–266 (2018).
19. Mahon, S. S. M. *et al.* Information theory and signal transduction systems: From molecular information processing to network inference. *Semin. Cell & Dev. Biol.* **35**, 98–108, DOI: [10.1016/j.semcdb.2014.06.011](https://doi.org/10.1016/j.semcdb.2014.06.011) (2014).
20. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674, DOI: [10.1016/j.cell.2011.02.013](https://doi.org/10.1016/j.cell.2011.02.013) (2011).
21. Kleczko, E. K., Kwak, J. W., Schenk, E. L. & Nemenoff, R. A. Targeting the complement pathway as a therapeutic strategy in lung cancer. *Front. Immunol.* **10**, DOI: [10.3389/fimmu.2019.00954](https://doi.org/10.3389/fimmu.2019.00954) (2019).
22. Zhang, R., Liu, Q., Li, T., Liao, Q. & Zhao, Y. Role of the complement system in the tumor microenvironment. *Cancer Cell Int.* **19**, DOI: [10.1186/s12935-019-1027-3](https://doi.org/10.1186/s12935-019-1027-3) (2019).
23. Tagalakakis, V. *et al.* High risk of deep vein thrombosis in patients with non-small cell lung cancer: A cohort study of 493 patients. *J. Thorac. Oncol.* **2**, 729–734, DOI: [10.1097/jto.0b013e31811ea275](https://doi.org/10.1097/jto.0b013e31811ea275) (2007).
24. Castell, J. V., Donato, M. T. & Gómez-Lechón, M. J. Metabolism and bioactivation of toxicants in the lung. the in vitro cellular approach. *Exp. Toxicol. Pathol.* **57**, 189–204, DOI: [10.1016/j.etp.2005.05.008](https://doi.org/10.1016/j.etp.2005.05.008) (2005).
25. Ishitsuka, Y., Ogawa, T. & Roop, D. The KEAP1/NRF2 signaling pathway in keratinization. *Antioxidants* **9**, 751, DOI: [10.3390/antiox9080751](https://doi.org/10.3390/antiox9080751) (2020).
26. Müller, T. & Hengstermann, A. Nrf2: Friend and foe in preventing cigarette smoking-dependent lung disease. *Chem. Res. Toxicol.* **25**, 1805–1824, DOI: [10.1021/tx300145n](https://doi.org/10.1021/tx300145n) (2012).
27. Xu, C. *et al.* Inhibition of 7, 12-dimethylbenz(a)anthracene-induced skin tumorigenesis in c57bl/6 mice by sulforaphane is mediated by nuclear factor κ B-related factor 2. *Cancer Res.* **66**, 8293–8296, DOI: [10.1158/0008-5472.can-06-0300](https://doi.org/10.1158/0008-5472.can-06-0300) (2006).
28. Gould, N. S., Min, E., Gauthier, S., Martin, R. J. & Day, B. J. Lung glutathione adaptive responses to cigarette smoke exposure. *Respir. Res.* **12**, DOI: [10.1186/1465-9921-12-133](https://doi.org/10.1186/1465-9921-12-133) (2011).
29. Madsen, D. H. *et al.* Tumor-associated macrophages derived from circulating inflammatory monocytes degrade collagen through cellular uptake. *Cell Reports* **21**, 3662–3671, DOI: [10.1016/j.celrep.2017.12.011](https://doi.org/10.1016/j.celrep.2017.12.011) (2017).
30. Deryugina, E. I. *et al.* Tissue-infiltrating neutrophils constitute the major in vivo source of angiogenesis-inducing MMP-9 in the tumor microenvironment. *Neoplasia* **16**, 771–788, DOI: [10.1016/j.neo.2014.08.013](https://doi.org/10.1016/j.neo.2014.08.013) (2014).
31. Benoit, M. E., Clarke, E. V., Morgado, P., Fraser, D. A. & Tenner, A. J. Complement protein C1q directs macrophage polarization and limits inflammasome activity during the uptake of apoptotic cells. *The J. Immunol.* **188**, 5682–5693, DOI: [10.4049/jimmunol.1103760](https://doi.org/10.4049/jimmunol.1103760) (2012).

32. Li, L. *et al.* Hypoxia restrains the expression of complement component 9 in tumor-associated macrophages promoting non-small cell lung cancer progression. *Cell Death Discov.* **4**, DOI: [10.1038/s41420-018-0064-3](https://doi.org/10.1038/s41420-018-0064-3) (2018).
33. Medler, T. R. *et al.* Complement c5a fosters squamous carcinogenesis and limits t cell response to chemotherapy. *Cancer Cell* **34**, 561–578.e6, DOI: [10.1016/j.ccell.2018.09.003](https://doi.org/10.1016/j.ccell.2018.09.003) (2018).
34. Allendorf, D. J. *et al.* C5a-mediated leukotriene b4-amplified neutrophil chemotaxis is essential in tumor immunotherapy facilitated by anti-tumor monoclonal antibody and β -glucan. *The J. Immunol.* **174**, 7050–7056, DOI: [10.4049/jimmunol.174.11.7050](https://doi.org/10.4049/jimmunol.174.11.7050) (2005).
35. Khameneh, H. J. *et al.* C5a regulates IL-1 β production and leukocyte recruitment in a murine model of monosodium urate crystal-induced peritonitis. *Front. Pharmacol.* **8**, DOI: [10.3389/fphar.2017.00010](https://doi.org/10.3389/fphar.2017.00010) (2017).
36. Guglietta, S. *et al.* Coagulation induced by c3ar-dependent NETosis drives protumorigenic neutrophils during small intestinal tumorigenesis. *Nat. Commun.* **7**, DOI: [10.1038/ncomms11037](https://doi.org/10.1038/ncomms11037) (2016).
37. Corrales, L. *et al.* Anaphylatoxin c5a creates a favorable microenvironment for lung cancer progression. *The J. Immunol.* **189**, 4674–4683, DOI: [10.4049/jimmunol.1201654](https://doi.org/10.4049/jimmunol.1201654) (2012).
38. Kwak, J. W. *et al.* Complement activation via a c3a receptor pathway alters CD4+ t lymphocytes and mediates lung cancer progression. *Cancer Res.* **78**, 143–156, DOI: [10.1158/0008-5472.can-17-0240](https://doi.org/10.1158/0008-5472.can-17-0240) (2017).
39. He, X. & Zhang, J. Why do hubs tend to be essential in protein networks? *PLoS Genet.* **2**, e88, DOI: [10.1371/journal.pgen.0020088](https://doi.org/10.1371/journal.pgen.0020088) (2006).
40. Raman, K., Damaraju, N. & Joshi, G. K. The organisational structure of protein networks: revisiting the centrality–lethality hypothesis. *Syst. Synth. Biol.* **8**, 73–81, DOI: [10.1007/s11693-013-9123-5](https://doi.org/10.1007/s11693-013-9123-5) (2013).
41. Zhan, P. *et al.* NCAPG2 promotes tumour proliferation by regulating g2/m phase and associates with poor prognosis in lung adenocarcinoma. *J. Cell. Mol. Medicine* **21**, 665–676, DOI: [10.1111/jcmm.13010](https://doi.org/10.1111/jcmm.13010) (2016).
42. Jin, E. *et al.* Expression of oncogenic long noncoding RNA PSMG3-antisense 1 in lung squamous cell carcinoma. *Oncol. Lett.* **22**, DOI: [10.3892/ol.2021.13012](https://doi.org/10.3892/ol.2021.13012) (2021).
43. Cimino, Y. *et al.* FADD protein release mirrors the development and aggressiveness of human non-small cell lung cancer. *Br. J. Cancer* **106**, 1989–1996, DOI: [10.1038/bjc.2012.196](https://doi.org/10.1038/bjc.2012.196) (2012).
44. Traverso, N. *et al.* Role of glutathione in cancer progression and chemoresistance. *Oxidative Medicine Cell. Longev.* **2013**, 1–10, DOI: [10.1155/2013/972913](https://doi.org/10.1155/2013/972913) (2013).
45. Park, H. J. *et al.* Keratinization of lung squamous cell carcinoma is associated with poor clinical outcome. *Tuberc. Respir. Dis.* **80**, 179, DOI: [10.4046/trd.2017.80.2.179](https://doi.org/10.4046/trd.2017.80.2.179) (2017).
46. qiong Huang, S. *et al.* The dysregulation of tRNAs and tRNA derivatives in cancer. *J. Exp. & Clin. Cancer Res.* **37**, DOI: [10.1186/s13046-018-0745-z](https://doi.org/10.1186/s13046-018-0745-z) (2018).
47. Jiang, Y.-Y. *et al.* TP63, SOX2, and KLF5 establish a core regulatory circuitry that controls epigenetic and transcription patterns in esophageal squamous cell carcinoma cell lines. *Gastroenterology* **159**, 1311–1327.e19, DOI: [10.1053/j.gastro.2020.06.050](https://doi.org/10.1053/j.gastro.2020.06.050) (2020).
48. Long, H. K., Blackledge, N. P. & Klose, R. J. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem. Soc. Transactions* **41**, 727–740, DOI: [10.1042/bst20130028](https://doi.org/10.1042/bst20130028) (2013).
49. Kim, H.-Y. *et al.* CXXC5 is a negative-feedback regulator of the wnt/ β -catenin pathway involved in osteoblast differentiation. *Cell Death & Differ.* **22**, 912–920, DOI: [10.1038/cdd.2014.238](https://doi.org/10.1038/cdd.2014.238) (2015).
50. Yan, X. *et al.* CXXC5 suppresses hepatocellular carcinoma by promoting TGF- β -induced cell cycle arrest and apoptosis. *J. Mol. Cell Biol.* **10**, 48–59, DOI: [10.1093/jmcb/mjx042](https://doi.org/10.1093/jmcb/mjx042) (2017).
51. Zhang, M. *et al.* The CXXC finger 5 protein is required for DNA damage-induced p53 activation. *Sci. China Ser. C: Life Sci.* **52**, 528–538, DOI: [10.1007/s11427-009-0083-7](https://doi.org/10.1007/s11427-009-0083-7) (2009).
52. ichi Takayama, K. *et al.* Integrative analysis of FOXP1 function reveals a tumor-suppressive effect in prostate cancer. *Mol. Endocrinol.* **28**, 2012–2024, DOI: [10.1210/me.2014-1171](https://doi.org/10.1210/me.2014-1171) (2014).
53. Feng, J. *et al.* High expression of FoxP1 is associated with improved survival in patients with non–small cell lung cancer. *Am. J. Clin. Pathol.* **138**, 230–235, DOI: [10.1309/ajcpdhqfnyjz01yg](https://doi.org/10.1309/ajcpdhqfnyjz01yg) (2012).
54. Colaprico, A. *et al.* TCGAbiolinks: an r/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71–e71, DOI: [10.1093/nar/gkv1507](https://doi.org/10.1093/nar/gkv1507) (2015).

55. Silva, T. C. *et al.* TCGA workflow: Analyze cancer genomics and epigenomics data using bioconductor packages. *F1000Research* **5**, 1542, DOI: [10.12688/f1000research.8923.2](https://doi.org/10.12688/f1000research.8923.2) (2016).
56. Raudvere, U. *et al.* g:profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198, DOI: [10.1093/nar/gkz369](https://doi.org/10.1093/nar/gkz369) (2019).
57. Fairbanks, J. *et al.* Juliagraphs//graphs.jl: an optimized graphs package for the julia programming language (2021).
58. Bastian, M., Heymann, S. & Jacomy, M. Gephi: An open source software for exploring and manipulating networks (2009).

Data Availability

We downloaded the publicly available LUSC dataset in: The National Cancer Institute (NCI) Genomic Data Commons (GDC) TCGA <https://gdc.cancer.gov/access-data/gdc-data-portal> by using TCGAbiolinks and ELMER package.

Author contributions statement

Y.D.H. was responsible for study conceptualization, data curation, analyses, visualization, and writing the original draft of the manuscript. K.K. was responsible for study conceptualization, supervision, and editing manuscript. S.I. was responsible for funding acquisition, project administration, supervision, and editing manuscript. All authors reviewed the manuscript.

Additional information

Competing interests

The authors declare no competing interests

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS1TCGALUSCnetworknodes.xlsx](#)
- [SupplementaryTableS2TCGALUSCnetworkedges.xlsx](#)
- [SupplementaryTableS3Communityanditssignificantgsea.xlsx](#)
- [SupplementaryTableS4pvaluecxymatrix.xlsx](#)
- [SupplementaryTableS5survivalanalysisbetweennesscentrality.xlsx](#)
- [SupplementaryTableS6globallevelenrichedmotifsandmethylationmasterregulator.xlsx](#)
- [SupplementaryTableS7communitylevelenrichedmotifsandmethylationmasterregulator.xlsx](#)