

Article

# Transcription Factor and lncRNA Regulatory Networks Identify Key Elements in Lung Adenocarcinoma

Dan Li <sup>1</sup>, William Yang <sup>2</sup>, Jialing Zhang <sup>3</sup>, Jack Y. Yang <sup>1</sup>, Renchu Guan <sup>1</sup> and Mary Qu Yang <sup>1,\*</sup>

<sup>1</sup> Joint Bioinformatics Graduate Program, Department of Information Science, George W. Donaghey College of Engineering and Information Technology, University of Arkansas at Little Rock and University of Arkansas for Medical Sciences, 2801 S. University Ave, Little Rock, AR 72204, USA; dxli@ualr.edu (D.L.); wxyang1@ualr.edu (J.Y.Y.); rxguan@ualr.edu (R.G.)

<sup>2</sup> School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA; wyang1@andrew.cmu.edu

<sup>3</sup> Department of Genetics, Yale University, New Haven, CT 06520, USA; jialing.zhang@yale.edu

\* Correspondence: mqyang@ualr.edu; Tel.: +1-501-683-2035

Received: 19 September 2017; Accepted: 21 December 2017; Published: 5 January 2018

**Abstract:** Lung cancer is the second most commonly diagnosed carcinoma and is the leading cause of cancer death. Although significant progress has been made towards its understanding and treatment, unraveling the complexities of lung cancer is still hampered by a lack of comprehensive knowledge on the mechanisms underlying the disease. High-throughput and multidimensional genomic data have shed new light on cancer biology. In this study, we developed a network-based approach integrating somatic mutations, the transcriptome, DNA methylation, and protein-DNA interactions to reveal the key regulators in lung adenocarcinoma (LUAD). By combining Bayesian network analysis with tissue-specific transcription factor (TF) and targeted gene interactions, we inferred 15 disease-related core regulatory networks in co-expression gene modules associated with LUAD. Through target gene set enrichment analysis, we identified a set of key TFs, including known cancer genes that potentially regulate the disease networks. These TFs were significantly enriched in multiple cancer-related pathways. Specifically, our results suggest that hepatitis viruses may contribute to lung carcinogenesis, highlighting the need for further investigations into the roles that viruses play in treating lung cancer. Additionally, 13 putative regulatory long non-coding RNAs (lncRNAs), including three that are known to be associated with lung cancer, and nine novel lncRNAs were revealed by our study. These lncRNAs and their target genes exhibited high interaction potentials and demonstrated significant expression correlations between normal lung and LUAD tissues. We further extended our study to include 16 solid-tissue tumor types and determined that the majority of these lncRNAs have putative regulatory roles in multiple cancers, with a few showing lung-cancer specific regulations. Our study provides a comprehensive investigation of transcription factor and lncRNA regulation in the context of LUAD regulatory networks and yields new insights into the regulatory mechanisms underlying LUAD. The novel key regulatory elements discovered by our research offer new targets for rational drug design and accompanying therapeutic strategies.

**Keywords:** transcription factor; long non-coding RNA; gene regulation; cancer; network; systems biology

## 1. Introduction

Lung cancer is one of the leading causes of morbidity and mortality worldwide, especially in men and smokers. In fact, according to an American Cancer Society report in 2017, one in four cancer deaths

are from lung cancer [1]. It has been estimated that there will be approximately 222,500 new lung cancer cases and 155,870 lung cancer-related deaths in the United States in 2017. Lung adenocarcinoma is a major subtype of non-small cell lung cancer that accounts for 85% of lung cancer cases.

The initiation and progression of lung cancer is a complex process due to a number of factors, including environmental exposure [2], smoking [3,4], signaling pathways, as well as genomic variance. Previous studies have even discovered some genomic elements and suggest the involvement of genetic mutations [5,6], transcriptional dysregulation [7], and immunosuppression [8]. Additionally, long non-coding RNAs (lncRNAs) have also been discovered to play critical roles in the development of various types of cancers [9–13].

Transcription factors (TFs) play critical roles in regulating the expression of genes [14]. Recent studies have shown cancer stem cells (CSCs) to be a potential cause of lung cancer initiation and development [15,16], while TFs to be important markers of CSCs [15]. The study of TFs has also improved our understanding of the mechanisms involved in the dysregulation of gene expression in diseases. For example, researchers investigated the transcription factor NF- $\kappa$ B, a driver of small cell lung cancer progression in mice, by assessing its expression and regulation patterns. The results indicated its overexpression in metastatic high-grade neuroendocrine lung tumors [17].

Additionally, several computational studies have used interactions and gene expression to identify disease-associated TFs and their functions in lung cancer. For example, the interactions between TFs and microRNAs were useful for discovering the regulatory roles of TFs in lung cancer [18]. The expression patterns of the microRNAs, TFs, and the common genes regulated by them were used to construct lung cancer regulatory networks [19]. Currently, network-based methods are performed for studying cancer, including lung cancer. Research using co-expression analysis [11,20] identified gene modules with functional variations between different cancer conditions. Another approach combining multiple data sources to identify key genetic elements in breast cancer was also reported [21]. However, the regulatory patterns including the interactions between lncRNAs and TFs in lung cancer remain to be elucidated. More comprehensive studies are needed to address the divergent molecular mechanisms underlying lung cancer.

In this study, we focused on the identification of key regulatory elements, including TFs and lncRNAs, responsible for lung adenocarcinoma (LUAD) progression. An integrative systems biology approach combining somatic mutations, gene expression profiles, DNA-methylation, protein–lncRNA interactions, and TF–gene interactions was developed for identifying the regulatory networks in LUAD. Genes with similar expression patterns were clustered as gene modules by co-expression network analysis. We further used a Bayesian network approach to indicate the potential regulatory relationships (directed edges) between genes. The final regulatory networks were achieved by incorporating known TF and target interactions based on experimental evidence and curation. TFs and lncRNAs with driver somatic mutations and/or connected with regulatory networks enriched with TF target genes were further investigated as key regulatory candidates. Aside from the lung cancer-related key regulators that we identified and thoroughly investigated in our study, we also discovered several novel TFs and lncRNAs that we found to be key regulatory elements. Pan-cancer analysis of the transcriptomes from 16 solid-tissue cancers provided additional evidence of the functional roles of these key lncRNAs. In this study, we introduced a new comprehensive network-based analysis merging multi-layer genomic data for revealing regulatory relationships between genes in LUAD. The gene regulatory networks and key regulatory TFs and lncRNAs are now available for studies investigating lung cancer.

## 2. Materials and Methods

### 2.1. Multidimensional Genomics Datasets

We obtained RNA sequencing (RNA-seq) data, DNA methylation data, and somatic mutation profiles of LUAD from The Cancer Genome Atlas (TCGA, March 2017) data portal [22]. The data

comprehensively represented 464 LUAD patients. As some patients had more than one sequenced sample, the number of LUAD tissue samples was actually larger than the total number of patients. We used all the samples available for each data type. Overall, we collected 540 RNA-seq, 527 DNA methylation, and 540 somatic mutation datasets for our analysis.

The RNA-seq data were generated from 56 normal and 484 LUAD tissue samples. Both raw reads and normalized gene expression (Fragments Per Kilobase of transcript per Million) data were downloaded from the TCGA data portal for our study. We used the raw reads for the differential expression analysis (DEG) and the normalized data for co-expression module discovery. The somatic mutations were called by Mutect2 [23]. The Methyllumin Bioconductor package [24] was applied for normalization and calculation of the  $\beta$  values for DNA methylation data processing. The  $\beta$  values represent the ratio of the methylated probe intensity to the overall intensity at each CpG locus [25]. We filtered out the methylation samples that had no  $\beta$  values or no CpG island information or methylations that occurred beyond 1 kb upstream or downstream of the gene transcription start site (TSS). As a result, 446 DNA methylation samples consisting of 26 normal and 420 tumor samples were retained for the subsequent analysis.

## 2.2. Differential Analysis

We applied the edgeR Bioconductor package [26] to perform the differential expression analysis based on the RNA-seq data. The fold change (FC) and false discovery rate (FDR) were employed as the criteria for selecting the differentially expressed genes. After excluding the transcripts with median expression levels less than 0.5 FPKM (Fragments Per Kilobase of transcript per Million), the transcripts with  $|\log_2(\text{FC})| > 0.5$  and  $\text{FDR} < 0.05$  were chosen as differentially expressed genes. For DNA methylation, we used samr R package [27] to perform the differential methylation analysis.

## 2.3. Regulatory Network Building

The regulatory networks were obtained from multi-step analysis including co-expression module identification, Bayesian network analysis, TF target assessment, and further pruning to reduce false positives. The gene regulatory network identification procedure is illustrated in Figure 1.

First, we identified co-expression modules using weighted correlation network analysis (WGCNA) [28] based on RNA-seq expression profiles. The WGCNA algorithm yielded a Topological Overlap Measurement (TOM) and provided a generalized assessment of the edges between two gene nodes.

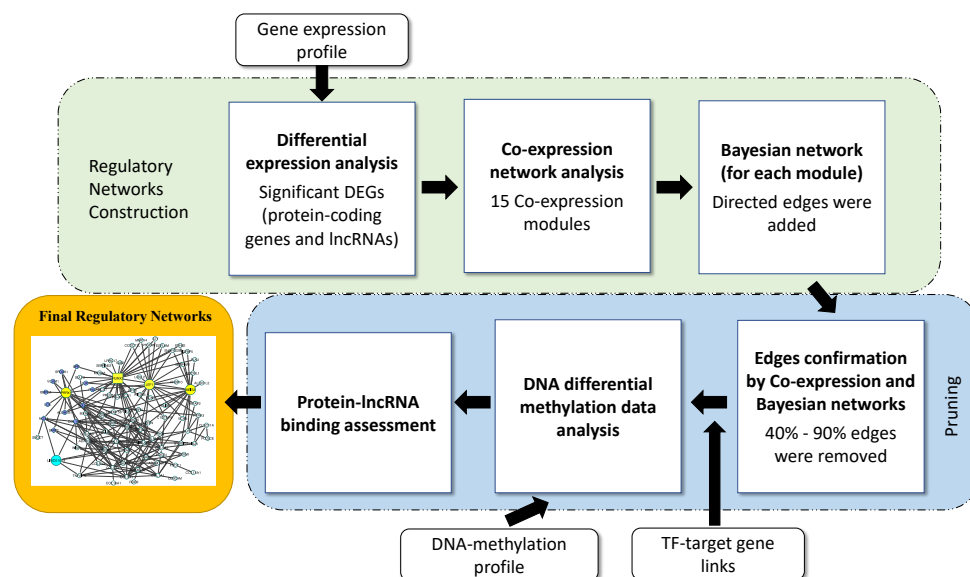
Then, Bayesian network analysis was conducted to infer potential regulatory relationships among genes in the co-expression modules. A Bayesian network is an NP-hard (non-deterministic polynomial-time hardness) problem [29] and often applies heuristic searching methods to reduce the search space size and computational complexity. Here, we used the hill climbing search approach to infer the direction of gene edges in the modules. This score-based optimization learning algorithm ranked the network structures according to a goodness-of-fit score and then identified the best network structures [30]. We used the bnlearn R package [31] to perform the analysis and measured the strength of the directed edges in the network. Then, all the edges were sorted in descending order of strength, and only those whose edges equaled 75% and above were retained. Additionally, edges whose connections were not confirmed by the co-expression were removed.

Additional regulatory relationships among genes were inferred based on TF and target gene interactions. We downloaded the TFs from four major databases, including JASPAR [32], AnimalTFDB 2.0 [33], Regulatorycircuits [34], and the Transcriptional Regulatory Element Database (TRED) [35]. The downstream target genes were selected based on a recent gene regulation study, and only lung cancer-specific TF and target interactions were used for network construction [34]. If a transcription factor and its target gene presented in the same module, a directed edge was added between the two gene nodes.

Each gene module contained protein-coding genes and lncRNAs. We calculated the binding scores to assess the potential binding between the lncRNAs and proteins [36]. The higher the score, the greater the potential that an lncRNA binds to a protein. Only interactions with binding scores equal to 25 and above (considered as real binding) were retained.

DNA methylation, perhaps because it blocks the promoters where transcription factors bind, was believed to play a crucial role in repressing gene expression. It has been observed that higher DNA-methylation in promoters corresponds to the lower expression of the corresponding genes. Therefore, we removed the edges between TF and target genes if target gene expression was correlated with its DNA methylation in the promoter.

We applied a bootstrapping method to assess the robustness of the networks to noise. In each bootstrapping iteration, we randomly deleted 2% of the nodes from the networks. We then evaluated the percentage of preserved edges (regulation relationships) after 100 iterations for each network.



**Figure 1.** Gene regulatory network identification procedure. The procedure includes co-expression identification, edge determination by transcription factor (TF) targets, Bayesian analysis, and long non-coding RNAs (lncRNA)–protein binding potential. The DNA methylation analysis was used to infer potential epigenetic regulations. DEGs: Differential Expressed Genes.

#### 2.4. Driver Somatic Mutation Identification

The somatic mutation profiles from LUAD patients were obtained from the TCGA data portal. All the identified somatic mutations were merged into a single VCF file. We then used the Cancer-Related Analysis of Variants Toolkit (CRAVAT 4.3) [37] to identify genes that harbored significant somatic mutations. CRAVAT combines two driver mutation predictors, CHASM [38] and VEST [39], to score the somatic mutations. Both the CHASM and VEST predictions were based on the Random Forest model and yielded  $p$ -values that were used to rank the significance of the somatic mutations in LUAD.

#### 2.5. Gene Enrichment Analysis

The odds ratios of the target genes of a specific TF inside the network and outside the network were calculated. Next, we applied the Fisher Exact Test to assess the statistical significance of the target gene enrichment in the individual regulatory networks. We applied DAVID [40] analysis to evaluate the pathway enrichment of the genes in the regulatory networks. The networks were plotted using Cytoscape v3.4.0 [41].

### 3. Results

#### 3.1. Differential Analysis of Multidimensional Genomic Profiles of Lung Adenocarcinoma

RNA-seq data generated from 56 normal and 484 lung adenocarcinoma tissue samples were obtained from the TCGA project. Differential expression analysis yielded 6220 differentially expressed genes, including 5934 protein-coding genes and 286 lncRNAs, ( $|\log_2(\text{FC})| > 0.5$  and  $\text{FDR} < 0.05$ ), where 78 lncRNAs and 2165 protein-coding genes were under-expressed in tumors. We also obtained the DNA methylation profile generated from 26 normal and 420 tumor samples from the same patient cohort. Differential methylation analysis revealed 1903 and 2992 genes had positive (hyper-) and negative (hypo-) methylation, respectively, in their promoter regions ( $q\text{-value} \leq 0.0075$  and  $\text{FDR} < 0.05$ ). We found two under-expressed lncRNAs and 281 protein-coding genes that had elevated methylation levels in their promoters. Furthermore, based on the somatic mutation profiles of the same LUAD patients, we found that 2835 genes harbored at least one significant somatic mutation ( $\text{FDR} < 0.05$ ).

#### 3.2. Disease Regulatory Network Identification

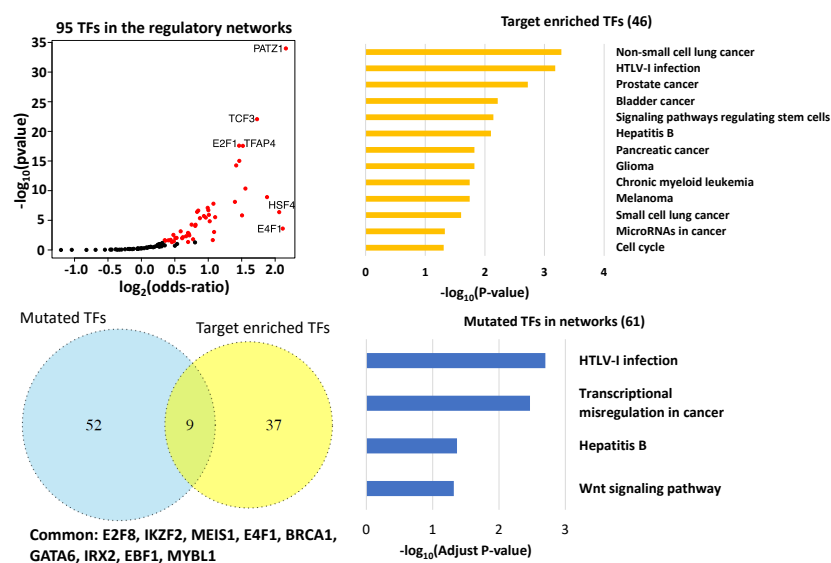
We conducted a co-expression analysis of with the differentially expressed genes in LUAD and revealed 15 co-expression gene modules. The modules contained 4012 differentially expressed protein-coding genes and 124 lncRNAs. Genes with similar or opposite expression patterns were clustered into the same module. TFs control target gene expression by interacting with *cis*-regulatory regions around these genes. To identify inference due to regulatory relationships among genes in the modules, we first searched for TFs and found that 297 of 1806 TFs from four major TF databases were present in the modules. The lung cancer-specific target genes of each TF were based on the annotation from a recent large-scale study [30]. Whenever a TF and corresponding target gene were present in the same module, we added a directed edge between them to account for the physical interaction and expression correlation. Additional edges were added by coupling Bayesian network and co-expression analyses. It has been widely accepted that promoter hypermethylation is associated with aberrant gene silencing in tumors [42,43]. To rule out epigenetic regulation of gene expression rather than regulation by transcription factors, we therefore disconnected all the edges from the genes in the tumor where there was an under-expression with a gain of promoter methylation.

As a result, we obtained 15 regulatory networks in which each gene member either had at least one incoming or outgoing edge. We found that the majority of the genes in the gene modules remained in the regulatory networks. Approximately 87% (13/15) of the networks contained more than 50% of the genes in the original modules, and 53.3% of the networks retained at least 80% of the genes in the modules. Eight of the regulatory networks consisted of more than 91% of the over-expressed genes, whereas five consisted of more than 91% of the under-expressed genes. The remaining two networks were mixed with up- and down-regulated genes and contained 41.85% and 25.22% of over-expressed genes, respectively. Pathway analysis showed that 14 out of the 15 networks were enriched by genes in at least one known cancer pathway as well as pathways that had not been recorded as cancer-related pathways in the literature, except network 10 (Table S1). Network 10 contained the highest percentage (28.9%) of lncRNAs compared to the other networks. For example, network 11 contained only one lncRNA (Figure S1), which may attribute to the absence of known cancer pathways in the network, as the function role of lncRNAs in cancer largely remains to be elucidated. We also conducted a bootstrapping experiment to examine the robustness of the regulatory networks in relation to noise. Out of 100 bootstrapping iterations, the median preserved rate of the edges in the 15 networks was 86.5% and 91.8% in at least 80 and 70 iterations, respectively. The majority of the edges in the networks demonstrated reasonable robustness to noise.

#### 3.3. Key Regulatory Elements in the Networks

Of the 95 TFs that have at least one target gene in the same network, we found 46 (46/95) in which targeted genes were significantly enriched in the networks ( $p < 0.05$  Fisher Exact Test, Figure 2a).

We also calculated the odds ratios to measure how strong the presence of a specific TF's target genes was in the network. *PATZ1*, *E4F1*, and *HSF4* had the largest odds ratios. Their odds ratios were larger than 4, suggesting that the presence of target genes inside the network are at least four times higher than those outside of it (Figure 2a). In a network, out-degree is defined as the number of outgoing edges emanating from a node. *PATZ1* (network 4, out-degree 192) is a DNA damage responsive TF that interacts with p53 [44]. The gene also regulates the expression of p53 target genes and is involved in cancer progression [45]. *E4F1* (network 8, out-degree 13) directly controls a transcriptional program involved in cell cycle checkpoints, metabolism, and mitochondrial homeostasis, as well as regulates the p53 response [46]. Inactivation of *HSF4* (network 8, out-degree 35), a heat shock factor, has been associated with tumorigenesis [44]. Additionally, the two TFs with the smallest *p*-values in the target gene enrichment test were also associated with tumors. *E2F1* (network 1, out-degree 147) can stimulate apoptosis and function as a tumor suppressor [47], while the *TCF3* fusion (network 4, out-degree 207) has been found in adenocarcinomas in situ [48].



**Figure 2.** Key transcription factors in the network. (a) A total of 95 TFs demonstrated significant expression level alterations in lung adenocarcinomas and had at least one target gene in the same network. (b) The pathways that were significantly enriched for the 46 key TFs. (c) Nine common TFs were revealed by overlapping 46 key TFs with 61 TFs carrying at least one somatic mutation. (d) The pathways that were significantly enriched for 61 TFs harboring significant somatic mutation(s) in lung adenocarcinoma (LUAD). HTLV-1: human T cell lymphotropic virus type 1.

The Database for Annotation, Visualization and Integrated Discovery (DAVID) pathway analysis suggested that the 46 TFs were significantly enriched in multiple cancer-related pathways (Benjamini  $p$ -adjust  $< 0.05$ , Figure 2b). These TFs were also prevalent in the measles and HTLV-I (human T cell lymphotropic virus type 1) infection pathways ( $p$ -adj =  $2.6 \times 10^{-2}$  and  $p$ -adj =  $1.6 \times 10^{-2}$ , respectively). Measles virus has been used for cancer therapy [49], while HTLV-1 plays a role in apoptosis [50]. Interestingly, the Hepatitis B pathway was also abundant with these TFs ( $p$ -adjust = 0.064,  $p$ -value =  $7.9 \times 10^{-3}$ , respectively), suggesting a putative relationship between this disease and LUAD.

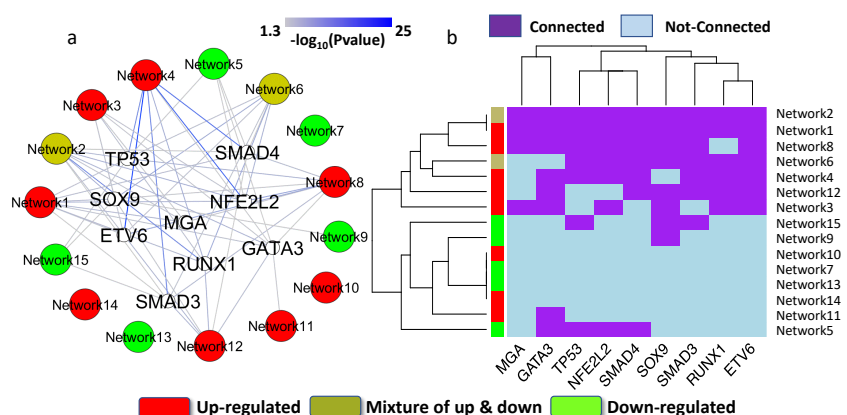
Sixty-one of 297 TFs in the network harbored at least one significant somatic mutation in LUAD. Nine of these mutated TFs, E2F8, IKZF2, MEIS1, E4F1, BRCA1, GATA6, IRX2, EBF1, and MYBL1, were also connected with a significant number of downstream targets in the network (Figure 2c,  $p < 0.05$  Fisher Exact Test), suggesting their key regulatory roles in LUAD. The following literature review confirmed that six of the nine TFs, including E2F8, MEIS, E4F1, BRCA1, GATA6, and IRX2, play essential roles in lung cancer progression [46,51–55]. The genetic deletion of *EBF1* is related to

LUAD pathogenesis [56]. Its role in LUAD remains to be elucidated. *IKZF2* and *MYBL1* represent novel LUAD genes. Moreover, we found that the 61 significantly mutated TFs were enriched in four pathways, including transcriptional misregulation in cancer, Wnt signal pathway, Hepatitis B, and HTLV-1 infection pathway (Figure 2d). The latter two pathways were also enriched for the TFs revealed by gene set enrichment analysis (Figure 2b), highlighting the roles that the two pathways play in LUAD.

### 3.4. Key Regulatory Elements Outside the Networks

We further investigated TFs whose expression levels did not change significantly in LUAD; however, their downstream targets were significantly abundant in the disease networks ( $p < 0.05$ ). We found that 63 TFs of this type also harbored one or more somatic mutation(s) (FDR  $< 0.05$ ) in LUAD patients. Functional analysis suggested that the 63 TFs were abundant in cancer related pathways, including transcriptional misregulation in cancer, pathways in cancer, Hepatitis B, colorectal cancer, and pancreatic cancer ( $p$ -adjust  $< 0.04$ ). The median number of networks potentially regulated by this type of TF was seven (Supplementary Figure S2).

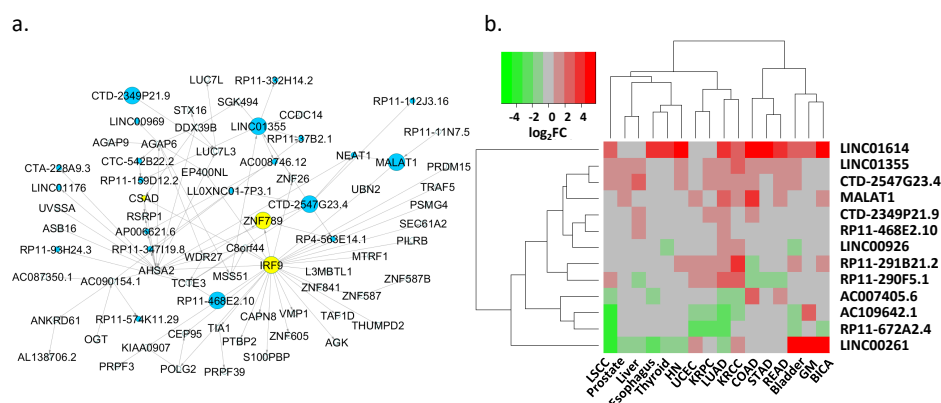
Previous studies have indicated that many identified mutations are related to cancer progression, but may only have an impact on tumor cells that have already emerged and on subsequent tumor growth [57]. This type of mutation is considered a passenger mutation. In contrast, mutations that cause cancer and promote tumor evolution are driver mutations. Hence, driver mutation information can further help us to prioritize key regulatory elements. OncoPrints is a function provided by cBioPortal, a widely used web tool in cancer research [58]. Coupling the driver mutations reported in OncoPrints with the results from target gene enrichment analysis, we identified nine critical transcription factors, encoded by *TP53*, *MGA*, *SOX9*, *ETV6*, *GATA3*, *NFE2L2*, *RUNX1*, *SMAD3*, and *SMAD4*. The nine genes that encode TFs appeared to be driver genes and acted as key network regulators. The collection of OncoPrints driver mutations consisted of a multiplicity of curated resources, including OncoKB, mutation hotspots, and recurrence in cBioPortal and COSMIC [59]. Each driver TF tends to mediate both under- and over-expressed networks (Figure 3a). We observed that the nine TFs clustered into several groups based on the networks that they potentially regulated (Figure 3b). For example, *TP53* and *NFE2L2* (also known as NRF2), were grouped together. It has been reported that p53 and NRF2 have similar functional roles, and both transcription factors enhance the capacity of cells to mitigate oxidative stress. It has also been found that NRF2 has an essential role in regulating p53 [60,61].



**Figure 3.** Key transcription factors outside the network. (a) The target genes of nine transcription factors that carried driver mutations were abundant in the regulatory networks. The color of the edge represents the  $p$ -value of the target gene enrichment analysis. The red-colored networks were over-expressed, whereas the green-colored networks were under-expressed in the disease. The yellow-colored networks were mixed with over-expressed and under-expressed genes. (b) The hierarchy clusters for the nine TFs and their regulated networks.

### 3.5. Regulation of Key lncRNAs

We also identified several lncRNAs that may play a key regulatory role in LUAD. Of the 44 lncRNAs that were present in the networks (Figure 4a), 13 lncRNAs demonstrated great potential for binding protein-coding genes and controlling their transcription as inferred by their binding scores and Bayesian analysis predictions (Table 1). These key lncRNAs included several known lung cancer lncRNAs, such as metastasis-associated lung adenocarcinoma transcript (MALAT1), LINC00261, and LINC01614 (Table 1, top three rows) [62]. For example, MALAT1, a critical regulator of the metastasis phenotype in lung cancer cells, potentially regulated the expression of *UBN2* and *NEAT1* (Figure 4b, Table 1, row two). *NEAT1* is an oncogenic lncRNA, and its elevated expression level has been associated with the progression of non-small-cell lung cancer. In contrast, *UBN2* is a protein-coding gene that serves a transcriptional regulatory function.



**Figure 4.** Key regulatory lncRNAs. (a) Network 10 consists of several lung cancer-related lncRNAs, such as MALAT1 and NEAT1. The blue-colored nodes represent lncRNA transcripts. (b) The expression alterations ( $\log_2(\text{FC})$ ) of 13 key lncRNAs in the 16 solid tissue cancer types. Red denotes over-expression while green represents under-expression of the lncRNAs in cancers. Grey refers to an insignificant gene expression change.

We further examined the expression correlations of these lncRNAs and their corresponding target genes in 484 LUAD and 56 matched normal tissue samples, as well as in 53 human tissues based on Genotype-Tissue Expression (GTEx) project RNA-seq data (Table S2). The median value of Pearson correlation coefficients for the lncRNA and target gene pairs in normal lung and LUAD tissue samples was 0.77, which was consistent with the results from the co-expression module analysis, and 0.7 in the 53 human tissues. AC007405.6, LINC00261, and RP11-672A2.4, which are located within genomic proximity of their target genes, exhibited high expression correlation; the corresponding correlation coefficients were 0.83, 0.74, and 0.93 in normal lung and LUAD tissues and 0.81, 0.70 and 0.92 in the 53 human tissue samples, respectively. The results indicated putative *cis*-regulatory roles of these lncRNAs in controlling the expression of their target genes. Furthermore, these target genes are all known cancer-related genes [63,64]. For example, RP11-672A2.4 and its target gene *LRR32* share the same promoter region. *LRR32* encodes GARP, and aberrant expression of GARP has been reported in human breast, lung and colon cancers [64].

We expanded our investigation on these lncRNAs to 16 solid-tissue tumor types including LUAD in the TCGA project. Eight out of 13 key lncRNAs were differentially expressed in at least eight tumor types ( $|\log_2(\text{FC})| > 1$  &  $\text{FDR} < 0.05$  Table 1, Figure 4b). LINC01614, CTD-2547G23.4, LINC01355, MALAT1, CTD-2349P21.9 and RP11-468E2.10 were over-expressed, whereas RP11-672A2.4 was under-expressed in cancers. The remaining six lncRNAs demonstrated mixed over-/under-expression patterns in different cancer types. Additionally, the lncRNA and target protein-coding gene correlation analysis suggested that some lncRNAs might promote tissue-specific regulation. For example,



AC109642.1 and FMO2 were only correlatively expressed in LUAD and LSCC (Lung Squamous Cell Carcinoma). In contrast, the other lncRNAs showed correlated expression patterns with their targets in most cancer types (Supplementary Figure S3). Consistent with the expression analysis in the 53 human tissue samples, the expression of AC007405.6, LINC00261, and RP11-672A2.4 and their targets were correlated in these 16 cancer types.

**Table 1.** Key regulatory long non-coding RNAs (lncRNAs), their corresponding downstream target protein-coding genes and lncRNAs, and the cancer types in which the key lncRNAs are differentially expressed.

Key lncRNA	Target Protein-Coding Gene(s) <sup>1</sup>	Target lncRNA(s) <sup>1</sup>	Differentially Expressed in Cancer Types <sup>2</sup>
LINC00261	<i>FOXA2</i> (0.74; 0.7)	NA	BICA, Bladder, Esophagus, GM, HN, KRCC, Liver, LSCC, LUAD, Prostate, Thyroid, UCEC
MALAT1	<i>UBN2</i> (0.49; 0.83)	NEAT1 (0.77; 0.84)	BICA, COAD, KRCC, LUAD, Prostate, READ, UCEC
LINC01614	<i>TNFAIP6</i> (0.64; 0.64), <i>NOX4</i> (0.75; 0.71)	NA	BICA, Bladder, COAD, Esophagus, GM, HN, KRCC, LSCC, LUAD, READ, STAD, Thyroid
AC007405.6	<i>ERICH2</i> (0.83; 0.81)	NA	COAD, Esophagus, KRCC, LSCC, LUAD, READ
AC109642.1	<i>FMO2</i> (0.82; 0.81), <i>ANGPT1</i> (0.79; 0.77)	NA	Bladder, GM, KRCC, KRPC, LSCC, LUAD, UCEC
RP11-672A2.4	<i>LRR32</i> (0.93; 0.92)	NA	BICA, Bladder, KRPC, LSCC, LUAD, UCEC
LINC01355	<i>WDR2</i> (0.72; 0.83), <i>SGK494</i> (0.67; 0.70), <i>CCDC14</i> (0.71; 0.84), <i>ZNF26</i> (0.66; 0.85), <i>PCGF3</i> (0.62; 0.72), <i>AC087350.1</i> (0.54; 0.72)	RP11-159D12.2 (0.68; 0.78), AC008746.12 (0.77; 0.85), RP11-332H14.2 (0.64; 0.65)	Bladder, COAD, HN, KRCC, KRPC, Liver, LSCC, LUAD, Prostate, READ, STAD
CTD-2547G23.4	<i>TCTE3</i> (0.78; 0.81), <i>HCG27</i> (0.71; 0.72)	NEAT1 (0.47; 0.70), LLOXNC01-7P3.1 (0.81; 0.82), LINC01355 (0.77; 0.85), RP4-563E14.1 (0.70; 0.82), RP11-112J3.16 (0.73; 0.72)	Bladder, COAD, HN, KRCC, KRPC, Liver, LSCC, LUAD, Prostate, READ
CTD-2349P21.9	<i>LUC7L3</i> (0.66; 0.65)	NA	COAD, KRPC, Liver, LUAD
RP11-468E2.10	<i>TCTE3</i> (0.65; 0.70)	NA	Liver, LUAD
LINC00926	<i>TRAF3IP3</i> (0.67; 0.70), <i>TNFRSF13C</i> (0.77; 0.80), <i>FDCSP</i> (0.44; 0.38)	NA	Bladder, KRCC, LUAD, Thyroid
RP11-290F5.1	<i>FCRL5</i> (0.84; 0.84), <i>PIM2</i> (0.84; 0.77), <i>DERL3</i> (0.76; 0.66)	NA	COAD, KRCC, KRPC, Liver, LSCC, LUAD, READ, STAD, UCEC
RP11-291B21.2	<i>ZNF683</i> (0.80; 0.75)	AC002331.1 (0.74; 0.71)	BICA, Bladder, COAD, HN, KRCC, KRPC, LUAD, UCEC

<sup>1</sup> The first number in parentheses represents the Pearson correlation coefficient of the key regulator and the corresponding target gene in the normal lung and lung adenocarcinoma (LUAD) tissues, while the second number represents their expression correlation in the 53 human tissues. <sup>2</sup> BICA: Breast Invasive Carcinoma; COAD: Colon Adenocarcinoma; HN: Head and Neck; GM: Glioblastoma Multiforme; KRCC: Kidney Renal Clear Cell Carcinoma; KRPC: Kidney Renal Papillary Cell Carcinoma; LSCC: Lung Squamous Cell Carcinoma; READ: Rectum Adenocarcinoma; STAD: Stomach Adenocarcinoma; UCEC: Uterine Corpus Endometrial Carcinoma.

Collectively, we confirmed three known cancer-associated lncRNAs and revealed nine other key lncRNAs that have yet to be reported in the literature and most likely have an essential regulatory role in LUAD (Table 1, Figure 4b).

#### 4. Discussion

In this study, we integrated various types of genomic data to identify key regulatory elements in lung adenocarcinoma at the systems biology level. The genomic datasets, including whole-exome sequencing, RNA-seq, and DNA-methylation data, were obtained from the same patient cohort derived from the TCGA project. As cancer is a heterogeneous and complex disease, integrating genomics data from the same patient group can reduce the false positives that might arise from variations in individual genomic makeup rather than disease-related genetic alterations. We found that CHASM, a tool for disease driver predictions based on a random forest algorithm, yielded a large driver gene set with over 2000 genes being predicted as drivers at  $FDR < 0.05$ . On the other hand, VEST reported that the somatic mutations in over 12,000 genes had significant pathogenicity at  $FDR < 0.05$ . A majority of predicted driver genes (approximately 98%) by CHASM also showed significant pathogenicity by VEST. To pinpoint drivers with high specificity, we used a list of cancer driver genes that were experimentally verified and carefully curated to remove false positives. Our further regulation analysis of these genes in the network context offers novel insights into the mechanisms of the disease.

We revealed nine key TFs by combining co-expression modules, target gene enrichment, and somatic mutation analysis. The functional roles of six of these TFs in LUAD are supported by published studies. *E2F8* is a therapeutic target for lung cancer [48]. *MEIS1* has been found to inhibit non-small-cell lung cancer [52]. *E4F1* has a critical role in cancer cell survival and could be a target for cancer therapy [46]. *BRCA1* is a breast-cancer-susceptibility gene, and a recent study indicated that this gene could be a potential molecular marker in non-small-cell lung cancer [53]. *GATA6* is an inhibitor of LUAD metastatic progression [54]. Hypermethylation of the *IRX2* promoter frequently occurs in LUAD [55]. *EBF1*, *IKZF2* and *MYBL1* are novel candidates that have regulatory roles in LUAD and could be used for further experimental validation. Our in-silico approach enables the integration of multi-dimensional experimental data to effectively infer key regulatory elements in the disease.

lncRNAs often have low expression levels, and the majority of lncRNAs lack sequence conservation [65,66]; consequently, most lncRNAs are not yet well characterized [10,67]. Although several lncRNAs have been studied in cancer research, much more work remains to be completed. Here, we used a very stringent threshold to define key lncRNAs in lung adenocarcinoma. Only 13 lncRNAs were reported as being key regulators of cancer by our study, and we might be underestimating the role of other lncRNAs. However, these lncRNAs could be well-defined targets for further experimental examination and help us gain new insights into lncRNA regulation in cancer.

Our study highlighted the association of Hepatitis B in lung adenocarcinoma development. Currently, only a few studies have focused on the hepatitis virus in cancer [68–70]. Our results along with other previous reports indicate that certain viral infections could serve as mechanisms for the initiation and progression of lung adenocarcinoma, necessitating further investigation regarding the contribution of viruses to lung carcinogenesis.

#### 5. Conclusions

TFs and lnc RNAs are critical regulatory elements involved in lung cancer progression. The integrated analysis of multidimensional genomic data, including somatic mutations, gene expression, DNA methylation, TF-DNA interactions, and protein-lncRNA interactions, has enabled a deeper investigation into cancer development. Our study developed an integrative computational framework that applies network approaches to identify key regulatory elements that promote the initiation and progression of lung cancer. The regulatory networks were generated and refined by various genetic features. Key regulators revealed by multi-layer genomic data provided confident targets for other researchers for further experimental verification that could potentially be new targets for therapeutics and drug development.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2073-4425/9/1/12/s1](http://www.mdpi.com/2073-4425/9/1/12/s1). Figure S1: The regulatory networks contained lncRNAs. (a) The percentage of lncRNAs in the networks. Network 10 appeared to contain the highest number of lncRNAs compared to the other networks. The lncRNAs were absent in five networks: 3, 4, 6, 12, and 15, respectively. (b) Network 11 contain only one lncRNA (blue) LINC01614, a known lung cancer lncRNA. The yellow box, RUNX2, is a key TF harboring lung cancer driver somatic mutation. TFs often regulate multiple genes in the network. For instance, the purple-colored nodes are potential target *PRRX1* genes. The purple circles represent the target *PRRX1* genes reported by Marbach et al.; whereas, the purple diamonds represent the *PRRX1* target genes identified by Bayesian network analysis. Figure S2: The number of significantly connected networks of mutated 63 TFs outside networks. Figure S3: The expression correlation of the key lncRNAs and the putative target genes in 16 solid-tissue cancer types. Table S1: The significantly enriched pathways of 15 gene regulatory network in lung adenocarcinomas. Table S2: The RNaseq of 53 human tissues used by GTEx project.

**Acknowledgments:** This study was partially supported by United States National Institutes of Health (NIH) Academic Research Enhancement Award 1R15GM114739 and National Institute of General Medical Sciences (NIH/NIGMS) 5P20GM103429, Arkansas Science and Technology Authority (ASTA) Basic Science Research 15-B-23 and 15-B-38 and the Food and Drug Administration (FDA), contract No. HHSF223201510172C and HHSF223201610111C. However, the information contained herein represents the position of the author(s) and not necessarily that of the NIH and FDA.

**Author Contributions:** M.Q.Y. conceived of the project. M.Q.Y., D.L., J.Y.Y., and J.Z. designed the experiments, D.L., W.Y., and M.Q.Y. performed the analysis, M.Q.Y., D.L., W.Y., J.Z., J.Y.Y., and R.G. participated in the discussion and writing of the manuscript. All of the authors read and approved the manuscript.

**Conflicts of Interest:** The authors declare that they have no competing interests.

## References

1. American Cancer Society. Key Statistics on Lung Cancer. Available online: [www.cancer.org/cancer/non-small-cell-lung-cancer/about/key-statistics.html](http://www.cancer.org/cancer/non-small-cell-lung-cancer/about/key-statistics.html) (accessed on 15 August 2017).
2. Brown, T. Silica exposure, smoking, silicosis and lung cancer-complex interactions. *Occup. Med.* **2009**, *59*, 89–95. [[CrossRef](#)] [[PubMed](#)]
3. Govindan, R.; Ding, L.; Griffith, M.; Subramanian, J.; Dees, N.D.; Kanchi, K.L.; Maher, C.A.; Fulton, R.; Fulton, L.; Wallis, J.; et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **2012**, *150*, 1121–1134. [[CrossRef](#)] [[PubMed](#)]
4. Sun, S.; Schiller, J.H.; Gazdar, A.F. Lung cancer in never smokers—A different disease. *Nat. Rev. Cancer* **2007**, *7*, 778–790. [[CrossRef](#)] [[PubMed](#)]
5. Lee, W.; Jiang, Z.; Liu, J.; Haverty, P.M.; Guan, Y.; Stinson, J.; Yue, P.; Zhang, Y.; Pant, K.P.; Bhatt, D.; et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **2010**, *465*, 473–477. [[CrossRef](#)] [[PubMed](#)]
6. Sharma, S.V.; Bell, D.W.; Settleman, J.; Haber, D.A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **2007**, *7*, 169–181. [[CrossRef](#)] [[PubMed](#)]
7. Chen, C.-Y.; Jan, Y.-H.; Juan, Y.-H.; Yang, C.-J.; Huang, M.-S.; Yu, C.-J.; Yang, P.-C.; Hsiao, M.; Hsu, T.-L.; Wong, C.-H. Fucosyltransferase 8 as a functional regulator of nonsmall cell lung cancer. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 630–635. [[CrossRef](#)] [[PubMed](#)]
8. Carbone, D.P.; Gandara, D.R.; Antonia, S.J.; Zielinski, C.; Paz-Ares, L. Non-small-cell lung cancer: Role of the immune system and potential for immunotherapy. *J. Thorac. Oncol.* **2015**, *10*, 974–984. [[CrossRef](#)] [[PubMed](#)]
9. Sun, C.; Li, S.; Zhang, F.; Xi, Y.; Wang, L.; Bi, Y.; Li, D. Long non-coding RNA NEAT1 promotes non-small cell lung cancer progression through regulation of MIR-377-3p-e2f3 pathway. *Oncotarget* **2016**, *7*, 51784. [[CrossRef](#)] [[PubMed](#)]
10. Lin, C.-C.; Jen, J.; Lai, W.-W.; Tang, Y.-A.; Wang, Y.-C.; Lu, Y.-H. Oct4 transcriptionally regulates the expression of long non-coding RNAs NEAT1 and MALAT1 to promote lung cancer progression. *Mol. Cancer* **2017**, *16*, 104.
11. Iyer, M.K.; Niknafs, Y.S.; Malik, R.; Singhal, U.; Sahu, A.; Hosono, Y.; Barrette, T.R.; Prensner, J.R.; Evans, J.R.; Zhao, S.; et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **2015**, *47*, 199–208. [[CrossRef](#)] [[PubMed](#)]
12. Li, T.; Xie, J.; Shen, C.; Cheng, D.; Shi, Y.; Wu, Z.; Deng, X.; Chen, H.; Shen, B.; Peng, C.; et al. Upregulation of long noncoding RNA ZEB1-AS1 promotes tumor metastasis and predicts poor prognosis in hepatocellular carcinoma. *Oncogene* **2016**, *35*, 1575–1584. [[CrossRef](#)] [[PubMed](#)]

13. Chen, X.; Guo, W.; Xu, X.-J.; Su, F.; Wang, Y.; Zhang, Y.; Wang, Q.; Zhu, L. Melanoma long non-coding RNA signature predicts prognostic survival and directs clinical risk-specific treatments. *J. Dermatol. Sci.* **2017**, *85*, 226–234. [[CrossRef](#)] [[PubMed](#)]
14. Lee, T.I.; Young, R.A. Transcriptional regulation and its misregulation in disease. *Cell* **2013**, *152*, 1237–1251. [[CrossRef](#)] [[PubMed](#)]
15. Sławek, S.; Szmyt, K.; Fularz, M.; Dziudzia, J.; Boruczkowski, M.; Sikora, J.; Kaczmarek, M. Pluripotency transcription factors in lung cancer—a review. *Tumor Biol.* **2016**, *37*, 4241–4249. [[CrossRef](#)] [[PubMed](#)]
16. Qu, H.; Li, R.; Liu, Z.; Zhang, J.; Luo, R. Prognostic value of cancer stem cell marker CD133 expression in non-small cell lung cancer: A systematic review. *Int. J. Clin. Exp. Pathol.* **2013**, *6*, 2644. [[PubMed](#)]
17. Semenova, E.A.; Kwon, M.-C.; Monkhorst, K.; Song, J.-Y.; Bhaskaran, R.; Krijgsman, O.; Kuilman, T.; Peters, D.; Buikhuisen, W.A.; Smit, E.F.; et al. Transcription factor NFIB is a driver of small cell lung cancer progression in mice and marks metastatic disease in patients. *Cell Rep.* **2016**, *16*, 631–643. [[CrossRef](#)] [[PubMed](#)]
18. Tagne, J.-B.; Mohtar, O.R.; Campbell, J.D.; Lakshminarayanan, M.; Huang, J.; Hinds, A.C.; Lu, J.; Ramirez, M.I. Transcription factor and microRNA interactions in lung cells: An inhibitory link between NK2 HOMEBOX 1, MIR-200C and the developmental and oncogenic factors NFIB and MYB. *Respir. Res.* **2015**, *16*, 22. [[CrossRef](#)] [[PubMed](#)]
19. Mitra, R.; Edmonds, M.D.; Sun, J.; Zhao, M.; Yu, H.; Eischen, C.M.; Zhao, Z. Reproducible combinatorial regulatory networks elucidate novel oncogenic microRNAs in non-small cell lung cancer. *RNA* **2014**, *20*, 1356–1368. [[CrossRef](#)] [[PubMed](#)]
20. Cogill, S.B.; Wang, L. Co-expression network analysis of human lncRNAs and cancer genes. *Cancer Inform.* **2014**, *13*, 49. [[CrossRef](#)] [[PubMed](#)]
21. Hamed, M.; Spaniol, C.; Zapp, A.; Helms, V. Integrative network-based approach identifies key genetic elements in breast invasive carcinoma. *BMC Genomics* **2015**, *16* (Suppl. 5), S2. [[CrossRef](#)] [[PubMed](#)]
22. National Cancer Institute. Genomic Data Commons Data Portal. Available online: <https://portal.gdc.cancer.gov/> (accessed on 6 June 2017).
23. Cibulskis, K.; Lawrence, M.S.; Carter, S.L.; Sivachenko, A.; Jaffe, D.; Sougnez, C.; Gabriel, S.; Meyerson, M.; Lander, E.S.; Getz, G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **2013**, *31*, 213–219. [[CrossRef](#)] [[PubMed](#)]
24. Davis, S.; Du, P.; Bilke, S.; Triche, T.; Bootwalla, M. Methyumi: Handle Illumina methylation data. R package version 2.4. 0. 2012. Available online: [www.bioconductor.org/packages/release/bioc/html/methyumi.html](http://www.bioconductor.org/packages/release/bioc/html/methyumi.html) (accessed on 10 July 2017).
25. Du, P.; Zhang, X.; Huang, C.-C.; Jafari, N.; Kibbe, W.A.; Hou, L.; Lin, S.M. Comparison of beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinform.* **2010**, *11*, 587. [[CrossRef](#)] [[PubMed](#)]
26. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. Edger: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)] [[PubMed](#)]
27. Li, J.; Tibshirani, R. Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-seq data. *Stat. Methods Med. Res.* **2013**, *22*, 519–536. [[CrossRef](#)] [[PubMed](#)]
28. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)] [[PubMed](#)]
29. Knuth, D.E. Postscript about NP-hard problems. *ACM SIGACT News* **1974**, *6*, 15–16. [[CrossRef](#)]
30. Gámez, J.A.; Mateo, J.L.; Puerta, J.M. Learning bayesian networks by hill climbing: Efficient methods based on progressive restriction of the neighborhood. *Data Min. Knowl. Discov.* **2011**, *22*, 106–148. [[CrossRef](#)]
31. Scutari, M. Learning bayesian networks with the bnlearn R package. *J. Stat. Softw.* **2010**, *35*, 1–22.
32. Mathelier, A.; Fornes, O.; Arenillas, D.J.; Chen, C.-Y.; Denay, G.; Lee, J.; Shi, W.; Shyr, C.; Tan, G.; Worsley-Hunt, R.; et al. JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **2016**, *44*, D110–D115. [[CrossRef](#)] [[PubMed](#)]
33. Zhang, H.-M.; Liu, T.; Liu, C.-J.; Song, S.; Zhang, X.; Liu, W.; Jia, H.; Xue, Y.; Guo, A.-Y. AnimalTFDB 2.0: A resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res.* **2014**, *43*, D76–D81. [[CrossRef](#)] [[PubMed](#)]

34. Marbach, D.; Lamparter, D.; Quon, G.; Kellis, M.; Kutalik, Z.; Bergmann, S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat. Methods* **2016**, *13*, 366–370. [[CrossRef](#)] [[PubMed](#)]
35. Jiang, C.; Xuan, Z.; Zhao, F.; Zhang, M.Q. TRED: A transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res.* **2007**, *35*, D137–D140. [[CrossRef](#)] [[PubMed](#)]
36. Lu, Q.; Ren, S.; Lu, M.; Zhang, Y.; Zhu, D.; Zhang, X.; Li, T. Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genom.* **2013**, *14*, 651. [[CrossRef](#)] [[PubMed](#)]
37. Douville, C.; Carter, H.; Kim, R.; Niknafs, N.; Diekhans, M.; Stenson, P.D.; Cooper, D.N.; Ryan, M.; Karchin, R. CRAVAT: Cancer-related analysis of variants toolkit. *Bioinformatics* **2013**, *29*, 647–648. [[CrossRef](#)] [[PubMed](#)]
38. Carter, H.; Chen, S.; Isik, L.; Tyekucheva, S.; Velculescu, V.E.; Kinzler, K.W.; Vogelstein, B.; Karchin, R. Cancer-specific high-throughput annotation of somatic mutations: Computational prediction of driver missense mutations. *Cancer Res.* **2009**, *69*, 6660–6667. [[CrossRef](#)] [[PubMed](#)]
39. Douville, C.; Masica, D.L.; Stenson, P.D.; Cooper, D.N.; Gygax, D.M.; Kim, R.; Ryan, M.; Karchin, R. Assessing the pathogenicity of insertion and deletion variants with the variant effect scoring tool (VEST-Indel). *Hum. Mutat.* **2016**, *37*, 28–35. [[CrossRef](#)] [[PubMed](#)]
40. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [[CrossRef](#)] [[PubMed](#)]
41. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)] [[PubMed](#)]
42. Selamat, S.A.; Galler, J.S.; Joshi, A.D.; Fyfe, M.N.; Campan, M.; Siegmund, K.D.; Kerr, K.M.; Laird-Offringa, I.A. DNA methylation changes in atypical adenomatous hyperplasia, adenocarcinoma in situ, and lung adenocarcinoma. *PLoS ONE* **2011**, *6*, e21443. [[CrossRef](#)] [[PubMed](#)]
43. Zöchbauer-Müller, S.; Minna, J.D.; Gazdar, A.F. Aberrant DNA methylation in lung cancer: Biological and clinical implications. *Oncologist* **2002**, *7*, 451–457. [[CrossRef](#)] [[PubMed](#)]
44. Jin, X.; Eroglu, B.; Cho, W.; Yamaguchi, Y.; Moskophidis, D.; Mivechi, N.F. Inactivation of heat shock factor HSF4 induces cellular senescence and suppresses tumorigenesis in vivo. *Mol. Cancer Res.* **2012**, *10*, 523–534. [[CrossRef](#)] [[PubMed](#)]
45. Zheng, J.; Xiong, D.; Sun, X.; Wang, J.; Hao, M.; Ding, T.; Xiao, G.; Wang, X.; Mao, Y.; Fu, Y.; et al. Signification of hypermethylated in cancer 1 (HIC1) as tumor suppressor gene in tumor progression. *Cancer Microenviron.* **2012**, *5*, 285–293. [[CrossRef](#)] [[PubMed](#)]
46. Rodier, G.; Kirsh, O.; Baraibar, M.; Houles, T.; Lacroix, M.; Delpech, H.; Hatchi, E.; Arnould, S.; Severac, D.; Dubois, E.; et al. The transcription factor E4F1 coordinates CHK1-dependent checkpoint and mitochondrial functions. *Cell Rep.* **2015**, *11*, 220–233. [[CrossRef](#)] [[PubMed](#)]
47. Hung, J.-J.; Hsueh, C.-T.; Chen, K.-H.; Hsu, W.-H.; Wu, Y.-C. Clinical significance of E2F1 protein expression in non-small cell lung cancer. *Exp. Hematol. Oncol.* **2012**, *1*, 18. [[CrossRef](#)] [[PubMed](#)]
48. El-Aarag, S.A.; Mahmoud, A.; Hashem, M.H.; Elkader, H.A.; Hemeida, A.E.; El Hefnawi, M. In silico identification of potential key regulatory factors in smoking-induced lung cancer. *BMC Med. Genom.* **2017**, *10*, 40. [[CrossRef](#)] [[PubMed](#)]
49. Russell, S.J.; Peng, K.W. Measles virus for cancer therapy. In *Measles*; Griffin, D.E., Oldstone, M.B.A., Eds.; Springer: Heidelberg, Germany, 2009; pp. 213–241.
50. Taylor, J.M.; Nicot, C. HTLV-1 and apoptosis: Role in cellular transformation and recent advances in therapeutic approaches. *Apoptosis* **2008**, *13*, 733. [[CrossRef](#)] [[PubMed](#)]
51. Park, S.-A.; Platt, J.; Lee, J.W.; López-Giráldez, F.; Herbst, R.S.; Koo, J.S. E2F8 as a novel therapeutic target for lung cancer. *J. Natl. Cancer Inst.* **2015**, *107*. [[CrossRef](#)] [[PubMed](#)]
52. Li, W.; Huang, K.; Guo, H.; Cui, G. MEIS1 regulates proliferation of non-small-cell lung cancer cells. *J. Thorac. Dis.* **2014**, *6*, 850. [[PubMed](#)]
53. Reguart, N.; Cardona, A.F.; Carrasco, E.; Gomez, P.; Taron, M.; Rosell, R. BRCA1: A new genomic marker for non-small-cell lung cancer. *Clin. Lung Cancer* **2008**, *9*, 331–339. [[CrossRef](#)] [[PubMed](#)]
54. Cheung, W.K.; Zhao, M.; Liu, Z.; Stevens, L.E.; Cao, P.D.; Fang, J.E.; Westbrook, T.F.; Nguyen, D.X. Control of alveolar differentiation by the lineage transcription factors GATA6 and HOPX inhibits lung adenocarcinoma metastasis. *Cancer Cell* **2013**, *23*, 725–738. [[CrossRef](#)] [[PubMed](#)]

55. Sato, T.; Arai, E.; Kohno, T.; Takahashi, Y.; Miyata, S.; Tsuta, K.; Watanabe, S.I.; Soejima, K.; Betsuyaku, T.; Kanai, Y. Epigenetic clustering of lung adenocarcinomas based on DNA methylation profiles in adjacent lung tissue: Its correlation with smoking history and chronic obstructive pulmonary disease. *Int. J. Cancer* **2014**, *135*, 319–334. [[CrossRef](#)] [[PubMed](#)]
56. Liao, D. Emerging roles of the EBF family of transcription factors in tumor suppression. *Mol. Cancer Res.* **2009**, *7*, 1893–1901. [[CrossRef](#)] [[PubMed](#)]
57. Greenman, C.; Stephens, P.; Smith, R.; Dalglish, G.L.; Hunter, C.; Bignell, G.; Davies, H.; Teague, J.; Butler, A.; Stevens, C.; et al. Patterns of somatic mutation in human cancer genomes. *Nature* **2007**, *446*, 153–158. [[CrossRef](#)] [[PubMed](#)]
58. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **2013**, *6*, pl1. [[CrossRef](#)] [[PubMed](#)]
59. Forbes, S.A.; Beare, D.; Boutselakis, H.; Bamford, S.; Bindal, N.; Tate, J.; Cole, C.G.; Ward, S.; Dawson, E.; Ponting, L.; et al. COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **2016**, *45*, D777–D783. [[CrossRef](#)] [[PubMed](#)]
60. Zhang, Y.; Fan, H.; Fang, S.; Wang, L.; Chen, L.; Jin, Y.; Jiang, W.; Lin, Z.; Shi, Y.; Zhan, C.; et al. Mutations and expression of the NFE2L2/KEAP1/CUL3 pathway in Chinese patients with lung squamous cell carcinoma. *J. Thorac. Dis.* **2016**, *8*, 1639. [[CrossRef](#)] [[PubMed](#)]
61. Rotblat, B.; Melino, G.; Knight, R.A. NRF2 and p53: Januses in cancer? *Oncotarget* **2012**, *3*, 1272. [[CrossRef](#)] [[PubMed](#)]
62. White, N.M.; Cabanski, C.R.; Silva-Fisher, J.M.; Dang, H.X.; Govindan, R.; Maher, C.A. Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. *Genome Biol.* **2014**, *15*, 429. [[CrossRef](#)] [[PubMed](#)]
63. Li, C.M.-C.; Gocheva, V.; Oudin, M.J.; Bhutkar, A.; Wang, S.Y.; Date, S.R.; Ng, S.R.; Whittaker, C.A.; Bronson, R.T.; Snyder, E.L.; et al. Foxa2 and Cdx2 cooperate with Nkx2-1 to inhibit lung adenocarcinoma metastasis. *Gene. Dev.* **2015**, *29*, 1850–1862. [[CrossRef](#)] [[PubMed](#)]
64. Metelli, A.; Wu, B.X.; Fugle, C.W.; Rachidi, S.; Sun, S.; Zhang, Y.; Wu, J.; Tomlinson, S.; Howe, P.H.; Yang, Y.; et al. Surface expression of TGF- $\beta$  docking receptor GARP promotes oncogenesis and immune tolerance in breast cancer. *Cancer Res.* **2016**, *76*, 7106–7117. [[CrossRef](#)] [[PubMed](#)]
65. Johnsson, P.; Lipovich, L.; Grandér, D.; Morris, K.V. Evolutionary conservation of long non-coding RNAs: sequence, structure, function. *Biochim. Biophys. Acta* **2014**, *1840*, 1063–1071. [[CrossRef](#)] [[PubMed](#)]
66. Quinn, J.J.; Zhang, Q.C.; Georgiev, P.; Ilik, I.A.; Akhtar, A.; Chang, H.Y. Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. *Gene. Dev.* **2016**, *30*, 191–207. [[CrossRef](#)] [[PubMed](#)]
67. Zhao, Y.; Li, H.; Fang, S.; Kang, Y.; Hao, Y.; Li, Z.; Bu, D.; Sun, N.; Zhang, M.Q.; Chen, R. NONCODE 2016: An informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.* **2016**, *44*, D203–D208. [[CrossRef](#)] [[PubMed](#)]
68. Hassan, M.M.; Li, D.; El-Deeb, A.S.; Wolff, R.A.; Bondy, M.L.; Davila, M.; Abbruzzese, J.L. Association between hepatitis B virus and pancreatic cancer. *J. Clin. Oncol.* **2008**, *26*, 4557–4562. [[CrossRef](#)] [[PubMed](#)]
69. Perz, J.F.; Armstrong, G.L.; Farrington, L.A.; Hutin, Y.J.; Bell, B.P. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J. Hepatol.* **2006**, *45*, 529–538. [[CrossRef](#)] [[PubMed](#)]
70. Lin, M.V.; King, L.Y.; Chung, R.T. Hepatitis C virus–associated cancer. *Annu. Rev. Pathol.* **2015**, *10*, 345–370. [[CrossRef](#)] [[PubMed](#)]

