A global view of regulatory networks in lung cancer: An approach to understand homogeneity and heterogeneity

Jiapei Lu, William Wang, Menglin Xu, Yuying Li, Chengshui Chen, Xiangdong Wang

Department of Pulmonary Medicine, The First Affiliated Hospital, Wenzhou Medical University, Wenzhou, China

**A B S T R A C T**

A number of new biotechnologies are used to identify potential biomarkers for the early detection of lung cancer, enabling a personalized therapy to be developed in response. The combinatorial cross-regulation of hundreds of biological function-specific transcription factors (TFs) is defined as the understanding of regulatory networks of molecules within the cell. Here we integrated global databases with 537 patients with lung adenocarcinoma (ADC), 140 with lung squamous carcinoma (SCC), 9 with lung large-cell carcinoma (LCC), 56 with small-cell lung cancer (SCLC), and 590 without cancer with the understanding of TF functions. The present review aims at the homogeneity or heterogeneity of gene expression profiles among subtypes of lung cancer. About 5, 136, 52, or 16 up-regulated or 19, 24, 122, or 97 down-regulated type-special TF genes were identified in ADC, SCC, LCC or SCLC, respectively. DNA-binding and transcription regulator activity associated genes play a dominant role in the differentiation of subtypes in lung cancer. Subtype-specific TF gene regulatory networks with elements should be an alternative for diagnostic and therapeutic targets for early identification of lung cancer and can provide insightful clues to etiology and pathogenesis.

© 2016 Elsevier Ltd. All rights reserved.

**Contents**

1. Introduction ................................................................. 31
   1.1. Challenges of precision data mining ........................................ 32
   1.2. Comparisons of TF homogeneity and heterogeneity ...................... 32
   1.3. Identification of homogeneity TF genes .................................... 32
   1.4. Subtype heterogeneity TF genes ......................................... 33
   1.5. Heterogeneity of TF biological processes and molecular functions .......... 34
   1.6. Understanding of TF-dominated mechanisms ............................... 34
2. Conclusion ................................................................. 36
   Conflict of interest .................................................................. 36
   Acknowledgements .................................................................. 36
Appendix A. Supplementary data ....................................................... 37
References ............................................................................ 37

1. Introduction

Since the end of the 20th century, the morbidity and mortality of lung cancer has gradually increased over the past decades, as the leading cause of malignant tumor death [1]. Patients with lung cancer were considered to have similar clinical manifestations by which the cancer should be a homogeneity. While, such disease similarities were inevitably biased with the understanding of lung cancer gene-phenotype and sequencing. The large number of heterogeneities at different levels suggest a sophisticated gene–gene interactions and regulatory coordination [2,3]. The dynamics of the human gene network were suggested to reflect cellular processes and disease status. Interaction networks and dynamic networks of transcriptional factors (TF) have been considered as disease-
specific biomarkers to monitor disease development, progression or treatment [4–6].

Homogeneity and heterogeneity of lung cancer can be closely associated with transcriptomic behaviors, evidenced by our finding that human TFs regulate a large number of downstream genes [7]. More recent studies performed a genome-wide analysis by combining chromatin immunoprecipitation of individual TFs with high-throughput sequencing, to derive a small number of sub-networks of TF in pluripotency [8] or larger-scale networks in differentiation. TFs mainly regulate the transcription of target genes by binding to specific DNA sequences in promoter regions in genes. Further understanding of TF function will provide more opportunities to find the novel mechanism of diseases, while currently available studies, which mined different expression profiles, are considered to hardly unravel TF dysfunction, interactions, or pathogenesis of diseases [8]. The present review aims to investigate the framework to construct the lung cancer-related synergistic regulatory network and to analyze biological functions of TF network. The present review also aims to overview homogeneity or heterogeneity of TF gene profiles to understand the specificity and potential mechanisms of lung cancer subtypes. Subtype-specific TFs may drive the development of lung cancer differentiation and may provide useful information for validation.

1.1. Challenges of precision data mining

As the amount of global databases for genome-wide expression increases, there is an increasing challenge to ensure the selection of optimal tools and methods to analyze the quality of currently existing databases [9–11]. Gene Expression Omnibus ( GEO) developed at NCBI and ArrayExpress developed at EBI are the two main international repositories and are the commonly used practical databases where roughly 45% of microarray data published studies are deposited. The combination of data from different sources in meta-analysis studies reveals new aspects of biological processes even if data heterogeneity is still a great challenge. One of the challenges is to select a precise tool of increasing methodologies to mine and analyze increased number of global data. For example, GeneMania is expected to show relationships between genes by integrating information from various databases and to validate the gene interactions [11]. DisGeNet and FunDo are used to identify potential roles of genes in the disease [12–14]. The combination of literature text mining with microarray data were suggested as a systemic biological model to evaluate selected targets and potential molecular mechanisms [15]. GeneWizard can be used to generate and validate biological hypotheses [16] and can also be used to extract hidden information by hard, soft, or hierarchical clustering, and frequent pattern mining [17].

The present review utilizes two keywords for searching, ‘lung cancer’ and ‘human’. These keyword searches resulted in 504 GSE datasets; among which there were 27 GSE datasets containing comparative gene expression profiling of lung cancer and non-cancer lung tissues. The data was driven in community-derived reporting standards and with several critical study elements; e.g. raw data; processed data; and descriptive metadata. Selected data were analyzed from a large number of studies by Web-based tools and strategies [18]. We found 21 GSE datasets with pathological definitions; including ADC; SCC; LCC and SCLC; and 16 with data normalized. We mined TF profiles in 537 patients with ADC; 140 with SCC; 9 with LCC; 56 with SCLC; or 590 adjacent non-tumor tissue; as explained in Data 1.

1.2. Comparisons of TF homogeneity and heterogeneity

We selected TF gene expression profiling data of each lung cancer subtype from the databases and compared it with non-cancer tissues, respectively, to evaluate the TF genes with more than 2-folds up or down changes. Differential TF genes identified from each lung cancer subtype were further compared with other lung cancer subtypes to identify the subtype heterogeneity-specific genes. The TF genes homogeneity of lung cancer among subtypes were defined by criteria which requires that TF genes in all subtypes differed between lung cancer and normal tissues by statistical analyses with p-values less than 0.05 (t-test). Different expressions of TF genes were visually detected by the most popular visual platform for studying biological networks Cytoscape [19–22]. Other alternatives like BioLayoutExpress3D [23] and ArrayMining [24] were considered as references. CompNet, a GUI-based tool was suggested to compare multiple interaction networks in the form of edge-lists, node-lists which were overlaid on a background network, or path-lists [25]. Such a tool can be used for overlaying and subsequent comparative visualization and analysis of multiple networks. A customized Association Rule Mining algorithm was developed to identify sets of genes showing expression profiles correlated with a gene of interest [26]. The significance of the correlation was calculated together with the standard Association Rule Mining index and the χ2 p value. The curated selection of comparisons and the developed algorithm were applied for the category of TF gene expression profiling. Hierarchical clusters were analyzed by MultiExperiment Viewer v4.9.0. The Database for Annotation, Visualization and Integrated Discovery (DAVID v6.7) (http://david.abcc.ncifcrf.gov/home.jsp) were used to generate Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway [27] and Gene Ontology (GO) gene function annotation.

1.3. Identification of homogeneity TF genes

The number of up- and down-regulated TF genes in ADC, SCC, LCC, or SCLC were 36 and 145, 191 and 58, 56and 155, or 53and 184, as compared to non-cancer tissues (P < 0.05), respectively, as shown in Supplemental Table S1. The 42 TF genes that were identified as homogeneity genes were significantly expressed in all four subtypes of lung cancer. Of those, 4 homogeneity genes were consistent to be up-regulated and 38 were down-regulated. We mapped the reported substrate interactions of co-expression TF genes and revealed some of the pathways controlled by multiple TF genes (Fig. 1). The hierarchical cluster analysis demonstrated that those homogeneity TF genes showed independent expression profiles in ADC, SCC, LCC, or SCLC, respectively (Fig. 2A).

The most significant expression of HTLF as the homogeneity gene in the lungs suggests that it plays a critical role in tumor suppression via potential mechanisms by which the gene can be silenced through promoter hypermethylation or alternative mRNA splicing, leading to the expression of truncated proteins that lack DNA repair domains, similar to other cancers [28]. The methylation of HTLF was considered as a predicted factor for patient prognosis, as lung cancer patients with methylated HTLF gene showed an overall significantly shorter survival than those without it [29]. It was suggested that HTLF as a member of the candidate gene panel could be prognosticators for the clinical outcome of patients with lung cancer, although the exact mechanism of how HTLF methylation is involved in patient overall survival remains unclear.

FOXM1 as a member of the Forkhead box (Fox) family of transcription factors is expressed in actively differentiated cells, critical for cell cycle progression. A recent study showed FOXM1 was associated with poor prognosis for patients with non-small cell lung cancer (NSCLC) through the promotion of tumor metastasis or a direct effect on the prognosis [30]. There was clinical and molecular evidence to show that FOXM1 expression was directly correlated with lung adenocarcinoma progression and metastasis [31]. Lung adenocarcinoma cells with FOXM1-positive staining had a greater capacity for metastasis in patients after surgery and
were considered as a valuable prognostic marker. FOXM1 could transcriptionally activate the SNAIL gene as an important signaling pathway to directly affect the processes of epithelial-mesenchymal transition, invasion, and metastasis of lung adenocarcinoma cells [32].

1.4. Subtype heterogeneity TF genes

The SCLC and NSCLC are the two main pathological categories, where SCLC occurs mainly in the central airways, while NSCLC occurs in the central or peripheral locations. The variation of occurrences will influence different decisions in clinical management [33]. Clinical trials demonstrated that histological subtypes of NSCLC responded differently to targeted drugs or newly developed chemotherapies, possibly related to cell derivation and pathogenetic origins [34–37]. For example, lung adenocarcinomas with up-expression of ARNTL2 could predict poor patient survival by metastasis [38]. It was suggested that the up-expression of HMGA1, E2F6, IRF1, and TFDP1 or downregulation or no expression of SUV39H1, RBL1, and HNRPD in blood could be used to...
diagnose NSCLC subtypes, e.g. ADC or SCC [39]. The TF E2F6 was initially found to be upregulated in NSCLC blood samples. The down-regulation of HNF4α target genes were found to be the most common pathway specific to ADC, while the disruption of numerous histone modifying enzymes and TF E2F1 to SCC [33]. Recent study demonstrated that ectopic expression of a small cell lung cancer transcription factor, INSM1, could impair alveolar regeneration in lung development by increasing the air sac and cause alveolar hypoplasia [40].

The heterogeneity of ADC was represented by 5 up-regulation of ADC-specific genes, including PLAU, ZNF187, HNRPK, C1orf107, or GRFL1, and 19 down-regulation of ADC-specific genes. We also found 136 up-expression and 24 down-expression of SCC-specific genes, 52 up-expression and 122 down-expression of LCC-specific genes, or 16 up-expression and 97 down-expression of SCLC-specific genes (DATE 2). Of those subtype-specific genes, there was one gene fivefold down-regulated in ADC. The number of up- and down-regulated subtype-special TF genes were more than fivefold in SCC, LCC, or SCLC and was 51 and 4, 38 and 33, or 5 and 23, respectively. We found that the heterogeneity of subtype TF exists in lung cancer and suggest to consider those subtype-specific TF as the part of potential biomarkers to define subtypes of lung cancer. Genes contributing to subtype heterogeneity of lung cancer were presented in Fig. 2B.

1.5. Heterogeneity of TF biological processes and molecular functions

Biological processes and molecular functions of selected subtype-specific TF genes in lung cancer were analyzed and shown in Fig. 3, respectively. Results of KEGG pathway analyses were detailed in Table 1. Of those, we noticed that miRNAs as a group of gene regulators play important roles in biologic processes such as cell proliferation, differentiation, and apoptosis. A number of miRNA target genes were recently analyzed in patients with ADC by miRNA transcriptome sequencing and TaqMan quantitative PCR validation to identify deregulated microRNA-transcription factor networks [41]. MiRNA-TF synergistic regulation network was firstly identified and activated in patients with lung cancer, including MYC, NFKB1, miR-590, or miR-570 [42]. A particular study integrated miRNA expression and sequencing in the circulation and tissue from ADC patients with global miRNA and mRNA expression and identified a regulatory network including miR-15b and miR-155, and TFs with prognostic value in lung cancer. Moreover, miR-149 was found to directly target FOXM1 in lung cancer cells and to be involved in the epithelial-mesenchymal transition induced by TGF-beta1 [43]. Decreased expression of activating transcription factor 2 was observed in NSCLC tissues and proposed as one of the direct functional target genes of miR-204 which could suppress NSCLC by targeting activating transcription factor 2 and might serve as a diagnostic biomarker and therapeutic target of NSCLC [44]. Together with our finding, we call the special attention to TF roles of miRNA in the understanding of the homogeneity and heterogeneity among lung cancer subtypes.

TFs play critical roles in the regulation and involvement of DNA activity. An increase of nucleotide substitutions around transcription factor binding sites were attributed to DNA-binding proteins, which act as partial barriers to the polymerase δ mediated displacement of polymerase α synthesized DNA [45]. The mutation of CTCF/cohesin-binding sites were often noted in colorectal tumors and probably caused by challenged DNA replication under aberrant conditions [46]. In addition to changes in chromatin organization, DNA accessibility and replication timing, variations of DNA-binding proteins with nucleotide excision repair resulted in the development of DNA mutation at the protein binding sites [47]. It is possible that the subtype-TF heterogeneity may be responsible for the regulation of subtype initiation and formation by mutational and DNA repair processes. It indicates that the identification of lung cancer driver mutations in subtype heterogeneity is equally important to driver TFs for decisive strategies of precision, personalized, and targeting therapies, and also for the design of chemotherapeutic combinations with target therapies.

We found TF-dominated signal pathways, e.g. maturity onset, antigen processing and presentation, and Jak-STAT signaling pathway in ADC; melanogenesis in LCC; or cell cycle or TGF-beta signaling pathway in SCLC (Table 1). We also noticed three TF-down-regulated pathways related to cancer, immune reaction, or Toll-like receptor signaling pathway in ADC; MAPK signaling pathway or B cell receptor signaling pathway in SCC; Jak-STAT or Toll-like receptor signaling pathways in LCC, or Jak-STAT or cytosolic DNA-sensing signaling pathways in SCLC, as detailed in Table 1. It is still a challenge to identify the homogeneity and heterogeneity of signal pathways among lung cancer subtypes, even though the variation of pathways exist between subtypes.

1.6. Understanding of TF-dominated mechanisms

One of the important objectives is to understand potential roles of TF per se in the development and formation of homogeneity or heterogeneity of lung cancer subtypes. It seems easier to understand the existence of TF homogeneity and heterogeneity, while it is difficult to understand the degree and mechanism of TF heterogeneity in the total heterogeneity of lung cancer. A similar issue that should be considered is how the heterogeneity of genes can influence and regulate the heterogeneity of cancer cell functions. We found that most of the transcription factors in KEGG pathways for lung cancer were negatively correlated with the effects. We observed that FOS (Fig. 4) is one of the most involved pathways in NSCLC, while STAT5A in SCLC.

The Fos family of TFs includes c-Fos, FosB, Fra-1 and Fra-2, as well as smaller FosB splice variants Fos B2 and delta FosB [48], and their encoded leucine zipper proteins can dimerize with Jun family proteins to form the TF complex activating protein 1 [49]. C-Fos is a constituent of the early studied activating protein 1 protein complexes, often up-expressed in tumor cells [50]. c-Fos can increase the expression of COX-2 and mediate the TPA-induced COX-2 reactivation by recruiting KDM2A to the COX-2 promoter in human lung cancer cells [51]. The c-Fos protein was mainly distributed in the cytoplasm of airway epithelial cells as target cells, involving biological function of epithelial cells [52]. Fos was also involved in the suppression of the immune response in these tumors and coordinated with histone lysine-specific demethylase 2A to activate the expression of cyclooxygenase-2.

STAT5A was detected in a number of cancers as a new independent marker for poor prognosis in breast cancer or a predictor of responses to antiestrogen therapy [53], HER-TKIs exert a HER2 expression-dependent anti-cancer effect by blockage of STAT5A signaling [54]. STAT5A as a candidate therapeutic target protein in prostate cancer, coordinated with androgen receptor (AR) to activate the signaling of both proteins. Active STAT5A increased nuclear levels of both unliganded and antiandrogen-liganded AR, and protect against prostatic degradation in prostate cancer [55]. STAT5A was involved in the regulation of DNA damage in a p53 dependent manner [56], although the exact mechanisms by which STAT5A contributes to the heterogeneity in the lung cancer remains unclear. STAT5A was found to regulate cell proliferation and apoptosis in various cancers, as a direct and functional target of miR-1469 [57]. STAT5A partly inhibits the lung cancer cells apoptosis induced by miR-1469. However the understanding to effectively target the STAT5A pathway for lung cancer as an alternative therapy should be furthermore explored [57].
TF-dominated mechanisms play critical roles in advanced discovery and development of lung cancer and metastasis and provide potential opportunities to deeply understand the pathogenesis of lung cancer [58]. TF can be a group of biomarkers to monitor clinical applications of immunotherapies, gene therapies, radio-therapies, or target-oriented therapies for lung. TF-dominated mechanisms can be used to identify differential interactions of driver genes, cancer-predictive genes, subtype-specific genes, or disease-exclusive genes or gene pairs. TF-dominated mechanisms are consist of a number of molecule networks and contribute to...
biological functions of multi-signal pathways and development of clinical precision medicine [59,60]. Drug efficacy and toxicity are also associated with or dependent upon TF-dominated mechanisms by which the cell can be more sensitive or resistant to drugs [61–65].

2. Conclusion

The present review emphasizes the importance of TF homogeneity or heterogeneity in lung cancer subtypes and calls special attention to investigate molecular mechanisms of TF contributions to subtype heterogeneity. We integrated published findings with results we mined from global databases and overviewed the existence of TF homogeneity and heterogeneity of lung cancer subtypes on basis of TF-involved signal pathways, molecular functions, or biological processes. The TF-dominated transcription regulations and down-regulation of DNA-binding pathways were considered as the part of homogeneity among lung cancer subtypes. A number of subtype-specific TFs and TF-regulated networks were noticed and proposed as the alternatives for functional biomarkers or therapeutic targets for subtype lung cancer. We hope that the homogeneity and heterogeneity of TFs can be considered as one of the evidence to determine and categorize new molecular subtypes of lung cancer, design new therapeutic strategy of precision medicine, and to discover and develop TF-based new targeting drugs.

Conflict of interest

The authors declare that there are no conflicts of interest.

Acknowledgements

The work was supported by Shanghai Leading Academic Discipline Project (B115), Zhongshan Distinguished Professor Grant (XDW), The National Nature Science Foundation of China (91230204, 81270099, 81320108001, 81270131, 81400035,
Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.semcancer.2016.11.004.

