Lung Cancer 85 (2014) 110–115

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

Review

Analysis of lncRNA expression profiles in non-small cell lung cancers (NSCLC) and their clinical subtypes☆

Jingcheng Yang[a,1], Jinyuan Lin[a,1], Tianxiao Liu[a,1], Ting Chen[b], Shangling Pan[c], Weiqiang Huang[a,*], Shikang Li[a,*]

[a] Department of Thoracic and Cardiovascular Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, China
[b] Department of Management Information System, College of Computer and Information Engineering, Guangxi Teachers Education University, Nanning, China
[c] Department of Pathophysiology, Guangxi Medical University, Nanning, China

ARTICLE INFO

Article history:
Received 30 January 2014
Received in revised form 4 April 2014
Accepted 11 May 2014

Keywords:
Non-small cell lung cancer
Biomarker
lncRNA
Expression profiling
Adenocarcinoma
Squamous cell carcinoma

ABSTRACT

Lung cancer is one of the most common human cancers worldwide. Among all lung cancer cases, non-small cell lung cancer (NSCLC) accounts for approximately 85%. Long non-coding RNAs (lncRNAs) are non-protein-coding transcripts that have been shown to play important roles in tumourigenesis and tumor progression. To reveal novel tumor-related lncRNAs in NSCLC and their associations with clinical subtypes, we herein identified 2935 probe sets mapped to lncRNAs on Affymetrix HG-U133 Plus 2.0 array with an lncRNA classification pipeline. We found 47 lncRNAs differentially expressed between normal lung tissues and tumor samples and 19 lncRNAs differed in expression between SCC and AC, two subtypes of NSCLC, after analyses of the gene expression profiles of five datasets downloaded from the gene expression omnibus (GEO) with a leave one dataset out validation process. The different lncRNA expression profiles between NSCLC and normal tissue and between the subtypes of NSCLC may have potential implications in the pathogenesis of this cancer. lncRNAs screening may be beneficial in the diagnosis, subclassification, and the personalized treatment of NSCLC.

1. Introduction

lncRNA (long non-coding RNA) is an RNA molecule that is longer than 200 nucleotides which is not translated into a protein [1]. Many identified lncRNAs are transcribed by RNA polymerase II (RNA pol II). They mainly locate within nucleus or cytosolic compartment [2]. lncRNAs can be further subcategorized into the following locus biotypes based on their location with respect to protein-coding genes: (1) sense, (2) antisense, (3) bidirectional, (4) intronic, and (5) intergenic [3]. They regulate gene expression through epigenetic regulation, splicing, imprinting, transcriptional regulation and subcellular transport [4]. Over the last few decades, researches have been focusing on the role of protein-coding genes in the pathogenesis of cancer [5], paying less attention to the possible effects of lncRNAs. However, in recent years emerging evidence indicate that lncRNAs are dysregulated and play important roles in tumourigenesis and tumor progression [6]. Some of the regulatory mechanisms of lncRNAs have been elucidated. For example, HOTAIR, an lncRNA locates in the HOXC locus on 12q13.13, was firstly described as having a fundamental role in human breast cancer [7]. It binds to PRC2, silencing a portion of the HOXD locus, inducing H3 lysine 27 trimethylation, then remodeling the gene expression pattern of breast epithelial cells as does in embryonic fibroblasts [4,8]. lncRNA expression profiles are also altered in other types of cancers, including human prostate cancer, renal carcinoma, ovarian cancer, and human lung adenocarcinoma, raising the possibility that lncRNAs may become a promising diagnostic biomarker [9].

Lung cancer is one of the most common human cancers worldwide with considerable high morbility and mortality [10,11]. Among all lung cancer cases, non-small-cell lung cancer (NSCLC) accounts for approximately 85% [12] whose most effective therapy is complete lung resection plus appropriate chemotherapeutic strategy [13]. Nevertheless, several studies have observed different therapeutic responses due to specific cell types of NSCLC [14]. This

☆ *Disclaimers*: The authors take full responsibility for the content of this publication, confirming that it reflects their viewpoint and medical expertise.

* Corresponding authors at: Department of Thoracic and Cardiovascular Surgery, The First Affiliated Hospital of Guangxi Medical University, 6 Shuangyong Road, Nanning 530021, Guangxi Zhuang Autonomous Region, China. Tel.: +86 13397705186; fax: +86 771 5352775.
E-mail address: shikangli@hotmail.com (S. Li).

[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.lungcan.2014.05.011
0169-5002/

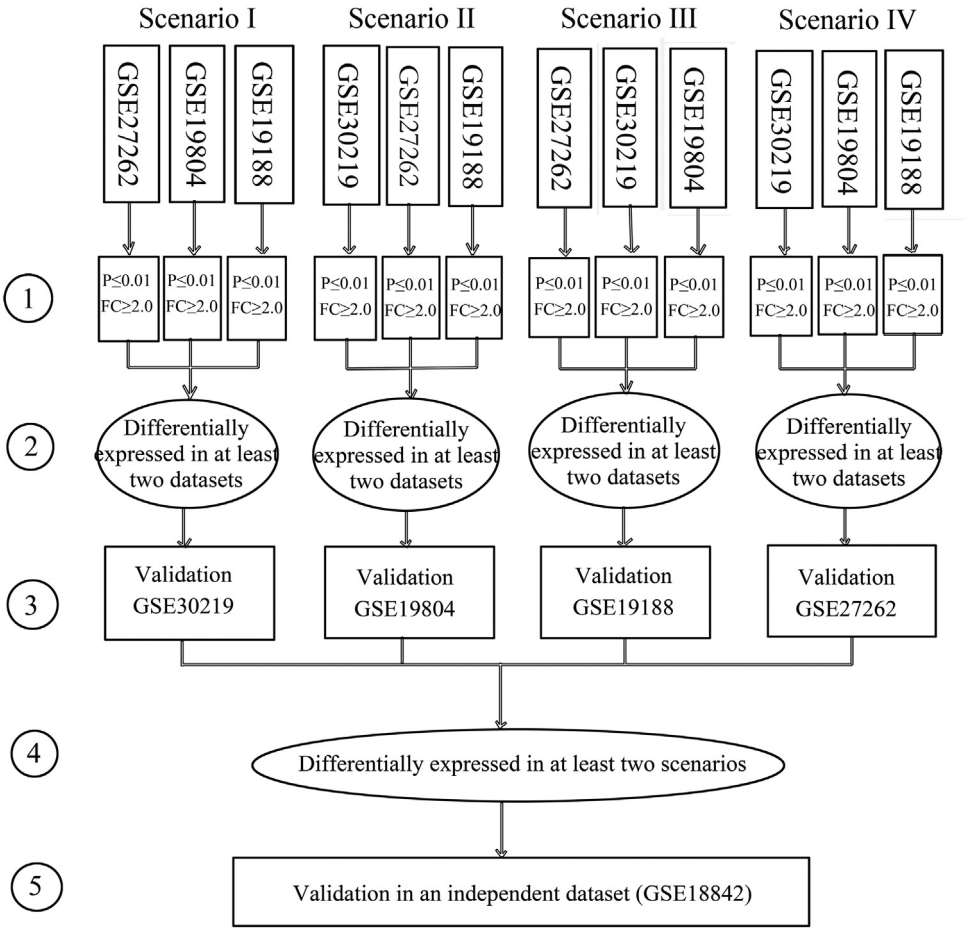

Fig. 1. Consistency-based meta-analysis process. The four datasets (GSE27262, GSE19804, GSE19188 and GSE30219) were switched around to create four scenarios. After the individual analysis of each dataset, three of the datasets were used to perform the signature and the result was then validated in an independent dataset in each scenario. The signatures from each scenario were then combined to be the final gene signatures, which were confirmed by the fifth dataset.

has placed a new emphasis on the importance of accurate subtyping [15]. The WHO classification of lung cancer is based upon tumor differentiation as seen in routinely produced sections examined by light microscopy [16]. While these methods provide satisfactory diagnostic accuracy for this malignancy histologically, they cannot subtype it precisely due to various factors such as small sample size, different sampling site, tumor heterogeneity, and the skill of the observer [15,17]. lncRNAs, as a novel class of functional molecules involved in cancer biology, may yield valuable information for precise subtyping of NSCLCs [18]. Previously published gene expression microarray studies have provided abundant profiling data for lung cancer [19–21].

In this study, we aimed to identify a set of lncRNA expression signatures in NSCLC, as well as in different histological subtypes by analyzing a cohort of previously published datasets from the gene expression omnibus (GEO), in an attempt to provide novel information on its lncRNA expression profiles which might be beneficial to the precise diagnosis, subcategorization (squamous cell carcinoma, SCC; adenocarcinoma, AC), and ultimately, the individualized treatment of NSCLC.

2. Materials and methods

2.1. GEO lung cancer gene expression data

To identify all relevant datasets, we searched GEO for NSCLC expression profiling studies. Studies were included in the systematic review if (i) they were gene profiling studies in patients with NSCLC; (ii) they used NSCLC tissue and normal lung tissue for comparison; (iii) they used the same platform; (iv) they contained more than three samples meeting the quality control standard in experimental and controlled group. As a result, 5 panels of NSCLC gene expression datasets were included: GSE27262, GSE19804, GSE19188, GSE30219 and GSE18842 [19–23]. While GSE27262, GSE19804, GSE19188 and GSE30219 were used in the *leave one dataset out validation process*, and GSE18842 served as an independent dataset to validate the gene signature derived from the meta-analysis. Of these datasets, GSE18842, GSE19188 and GSE30219 included information of SCC and AC. These datasets were selected to compare the lncRNA expression signatures among NSCLC subtypes.

2.2. lncRNA classification pipeline

To evaluate the lncRNA expressions in NSCLC gene expression data, we applied a pipeline as described by Zhang et al. [18] to identify the probe sets uniquely mapped to lncRNAs from the Affymetrix array by using the following steps. Firstly, we mapped Affymetrix HG-U133 Plus 2.0 probe set ID to the NetAffx Annotation Files (HG-U133 Plus 2.0 Annotations, CSV format, release 33, 10/30/12) (Table S1). The annotations contained the probe set ID, Ensembl gene ID, Refseq transcript ID gene symbol, gene title, gene symbol and other informative items for the specific probe set. Then the probe sets were filtered by the