Gene 575 (2016) 600–614

Contents lists available at ScienceDirect

Gene

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

Research paper

Systematically identify key genes in inflammatory and non-inflammatory breast cancer

Fan Chai [1], Yan Liang [1], Fan Zhang, Minghao Wang, Ling Zhong, Jun Jiang *

Breast Disease Center, Southwest Hospital, Third Military Medical University, Chongqing 400038, PR China

ARTICLE INFO

Article history:
Received 21 July 2015
Received in revised form 27 August 2015
Accepted 11 September 2015
Available online 25 September 2015

Keywords:
Tumor microenvironment
Inflammatory breast cancer
non-inflammatory breast cancer
PDGFRβ, SUMO1 and COL1A1

ABSTRACT

Although the gene expression in breast tumor stroma, playing a critical role in determining inflammatory breast cancer (IBC) phenotype, has been proved to be significantly different between IBC and non-inflammatory breast cancer (non-IBC), more effort needs to systematically investigate the gene expression profiles between tumor epithelium and stroma and to efficiently uncover the potential molecular networks and critical genes for IBC and non-IBC. Here, we comprehensively analyzed and compared the transcriptional profiles from IBC and non-IBC patients using hierarchical clustering, protein–protein interaction (PPI) network, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database analyses, and identified PDGFRβ, SUMO1, COL1A1, FYN, CAV1, COL5A1 and MMP2 to be the key genes for breast cancer. Interestingly, PDGFRβ was found to be the hub gene in both IBC and non-IBC; SUMO1 and COL1A1 were respectively the key genes for IBC and non-IBC. These analysis results indicated that those key genes might play important role in IBC and non-IBC and provided some clues for future studies.

1. Introduction

Breast cancer, categorized into two main subtypes: inflammatory breast cancer (IBC) and non-inflammatory breast cancer (non-IBC), is the second leading cause of cancer death among women (McGregor and Antoni, 2009). IBC, exhibiting epidemiological and histoclinical differences with non-IBC, is an aggressive form of breast cancer (Yamauchi et al., 2012). And IBC patients have worse survival outcomes compared with non-IBC patients even with multiple improved treatments, such as polychemotherapy and endocrine regimens (Dawood et al., 2011). Simultaneously, IBC patients have higher risk of cancer in the contralateral breast compared to non-IBC patients with comparable stage (Schairer et al., 2011). It seems that IBC represents different characteristics in different races (Hance et al., 2005). For example, black women were more vulnerable to IBC than the white women and showed higher incidence rate, poorer survival rate and shorter survival time once they're diagnosed as IBC (Hance et al., 2005).

A highly dynamic epithelium and the underlying stroma, the two main cellular compartments, comprise the mammary gland (Su et al., 2011). Once breast cancer initiates due to genetic mutations and epigenetic modifications on oncogenes, DNA repair genes and tumor suppressors, the normal growth regulatory pathways are altered in mammary epithelial cells (Su et al., 2011). The tumor stromal cells, physically close to epithelial cells, support the epithelial progression to malignancy by affecting neighboring microenvironment if molecular and phenotypic changes occur within the stroma (Su et al., 2011). For example, cancer-associated fibroblasts (CAFs), one of the main mediators in the tumor stroma, can promote breast cancer progression and metastasis by inducing epithelial–mesenchymal transition in breast epithelial cells, promoting angiogenesis in the primary site, increasing breast cancer stem cells by secreting chemokine CCL2 and activating other related processes (Mao et al., 2013).

Previous study showed that there were numerous different tumor stromal genes between IBC and non-IBC while there was no obvious difference in the tumor epithelial gene expression profiles between IBC and non-IBC, implying that the gene signature of the tumor stroma rather than that of the tumor epithelium may determine the IBC phenotype (Boersma et al., 2008). However, the gene expression profiles between tumor epithelium and stroma haven't been investigated in this study. And it's also critical to understand which genes may play the critical roles in the whole tumor microenvironment. Herein, we comprehensively compared the different gene expression profiles and related pathways between tumor epithelium and tumor stroma in IBC and non-IBC, respectively. Two comparisons were made, including IBC tumor epithelium versus IBC stroma (comparison 1) and non-IBC tumor epithelium versus non-IBC stroma (comparison 2). Differentially expressed genes (DEGs) were analyzed and GO biological process and KEGG enrichment analysis of DEGs were also conducted on these different genes. Some key genes, such as PDGFRβ, SUMO1 and COL1A1, and related key pathways, such as focal adhesion, response to organic substance and extracellular matrix organization, were identified in our systematical analysis.

* Corresponding author.
E-mail address: jiangjunchq@163.com (J. Jiang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.gene.2015.09.025
0378-1119/

Table 1
Statistical distribution of DEGs in comparison 1.

	Probe		Gene	
All	10,327		7521	
$\|Log_2(\text{fold change})\| > 1$ & p value < 0.05	440	129(Up) 311(Down)	368	114(Up) 254(Down)

The up or down level was obtained by comparing IBC tumor epithelium cells to surrounding stromal cells.

2. Methods

2.1. Data collection

We obtained the microarray expression dataset (GSE5847) by searching Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) and four groups of samples were available: 13 IBC tumor epithelium samples, 13 IBC tumor stroma samples, 35 non-IBC tumor epithelium samples and 34 non-IBC tumor stroma samples (Boersma et al., 2008). Affymetrix Human Genome U133A Array and the probe annotation files were performed to analyze the collected unprocessed data (.cel files).

2.2. Data processing and filtering

In order to quantify microarray signal, we applied GCRMA in this research (Wu et al., 2005). There are three main steps in the normalization process including model-based background correction, quantile normalization and summarizing.

The gene filter package in R language was applied to filter out uninformative data such as control probesets and other internal controls as well as removing genes which were expressed uniformly close to background detection levels (Gentleman et al., 2011). However, the filter does not remove probesets without Entrez Gene identifiers or with identical Entrez Gene identifiers.

2.3. Differentially expressed genes (DEGs) analysis

We performed two statistical comparisons: comparison 1 was between tumor epithelium and stroma for IBC samples, and comparison 2 was between tumor epithelium and stroma for non-IBC samples. Limmain R language was applied to identify the DEGs (Smyth, 2005). Only genes with $|\log_2(\text{fold change})| > 1$ and p value < 0.05 were recognized as statistically differentially expressed between two groups. The adjusted p value was obtained by applying Benjamini and Hochberg's (BH) false discovery rate correction on the original p value, and fold change threshold was selected based on our purpose of focusing on significantly differentially expressed genes.

2.4. Hierarchical clustering

We performed hierarchical clustering to classify analyzed samples based on gene expression profiles (Tavazoie et al., 1999). The DEGs, which were classified in GO biological processes and KEGG pathways, were further extracted, and the expression pattern of those DEGs was characterized. Furthermore, we obtained heat maps using R package for the DEGs classified in targeted GO biological processes or KEGG pathways.

Table 2
Statistical distribution of DEGs in comparison 2.

	Probe		Gene	
All	10,327		7521	
$\|Log_2(\text{fold change})\| > 1$& p value < 0.05	317	26 (Up) 291 (Down)	248	24 (Up) 224 (Down)

The up or down level was obtained by comparing non-IBC tumor epithelium cells to surrounding stromal cells.

Table 3
Nine commonly up-regulated genes in comparison 1&2 ($\log_2(\text{fold change}) > 1$ & p value <0.05).

Gene symbol	Comparison 1	P value	Comparsion 2	P Value
	Log_2(fold change)		Log_2(fold change)	
ERMP1	1.11	0.03	1.06	0.00
MREG	1.15	0.03	1.02	0.00
SPINT2	1.20	0.02	1.07	0.00
AP1M2	1.29	0.03	1.15	0.00
DHCR24	1.50	0.03	1.32	0.00
ERBB3	1.69	0.03	1.08	0.02
NEDD4L	1.69	0.01	1.41	0.00
NEBL	1.95	0.01	1.04	0.03
SQLE	2.05	0.00	1.03	0.00

2.5. GO and KEGG pathway analyses

We utilized R packages — GO.db, KEGG.db and KEGGREST to detect GO categories and KEGG pathways with significant overrepresentation in DEGs comparing with the whole genome (Carlson et al., 2007; Tenenbaum, 2013). The significantly enriched biological processes were identified when the p value was less than the threshold value 0.01. As to KEGG pathway, p value was set to less than 0.05.

2.6. Construction of biological network

Pair interactions, included in any of the three databases from HPRD (Human Protein Reference Database), BIOGRID (Biological General Repository for Interaction Datasets) and PIP (Human Protein–Protein Interaction Prediction Database) were chosen to be included in our PPI database (Keshava Prasad et al., 2009), (Chatr-Aryamontri et al., 2013), (McDowall et al., 2009). Consequently, 561,405 pairs of interactions are covered in our database. We utilized Cytoscapeto construct interaction network (Smoot et al., 2011). After functional enrichment analysis, the DEGs specified in dramatically altered GO biological processes and KEGG pathways were mapped to corresponding networks respectively to analyze their interactions.

3. Results

3.1. Differential expression analysis

Both tumor epithelium and tumor stroma play important roles in the progression of breast cancer (Su et al., 2011). Therefore, it is necessary to investigate the molecular characteristics between tumor epithelium and stroma in IBC and non-IBC for more comprehensive understanding of IBC and non-IBC. Herein, two comparisons were made, including IBC tumor epithelium versus IBC stroma (comparison 1) and non-IBC tumor epithelium versus non-IBC stroma (comparison 2), to analyze the gene profiles between tumor stroma and epithelium in IBC and non-IBC respectively. Accordingly we identified 368DEGs in comparison 1, among which 114 genes were up-regulated and 254 genes were down-regulated (Table 1). And 248 DEGs were obtained in comparison 2, including 24 up-regulated genes and 224 down-regulated genes (Table 2). To further investigate the commonly or specifically changed genes in IBC and non-IBC, we carefully conducted the analysis of these two comparisons, 156 commonly DEGs were identified, including 9 up-regulated genes and 147 down-regulated genes (Tables 3 and 4); 212 DEGS changed only in comparison 1 were identified, including 105 up-regulated genes and 107 down-regulated genes (Tables 5 and 6); while 92 DEGs changed only in comparison 2, including 15 up-regulated and 77 down-regulated (Tables 7 and 8); none was up-