Gene 681 (2019) 26–35

Contents lists available at ScienceDirect

Gene

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

Research paper

Identification of characteristic genes distinguishing neural stem cells from astrocytes☆

Shangyao Qin[a,1], Xiao Huang[a,1], Dan Wang[a,1], Xin Hu[b], Yimin Yuan[a], Xiu Sun[a], Zijian Tan[a], Yakun Gu[a], Xueyan Cheng[a], Cheng He[a,*], Zhida Su[a,*]

[a] Institute of Neuroscience, Key Laboratory of Molecular Neurobiology of Ministry of Education and the Collaborative Innovation Center for Brain Science, Second Military Medical University, Shanghai, China
[b] Department of Neurological Surgery, Xixi Hospital of Hangzhou, Hangzhou, China

ARTICLE INFO

Keywords:
Neural stem cells
Bioinformatics analysis
Characteristic genes
Biological property

ABSTRACT

Background: Neural stem cells (NSCs) have unique biological characteristics such as continuous proliferation and multipotential differentiation, providing a possible method for restoration of central nervous system (CNS) function after injury or disease. NSCs and astrocytes share many similar biological properties including cell morphology and molecular expression and can trans-differentiate into each other under certain conditions. However, characteristic genes specifically expressed by NSCs have not been well described.

Methods: To provide insights into the characteristic expression of NSCs, bioinformatics analysis of two microarrays of mouse NSCs and astrocytes was performed. Compared to astrocytes, the differentially expressed genes (DEGs) in NSCs were identified and annotated by GO, KEGG and GSEA analysis, respectively. Then key genes were screened by protein-protein interaction (PPI) network and modules analysis, and were verified using multiple high-throughput sequencing resources. Finally, the expression difference between the two cell types was confirmed by Real-time Quantitative PCR (qPCR), western blotting and immunochemical analysis.

Results: In the present study, 282 and 250 NSC-enriched genes from two microarrays were identified and annotated respectively, and the 77 overlapping DEGs were then selected. From the PPI network 24 key genes in three modules were screened out. Importantly, sequencing data of tissues showed that these 24 key genes tended to be highly expressed in NSCs compared with astrocytes. Furthermore, qPCR and western blot analysis of cultured NSCs and astrocytes showed two genes (*KIF2C* and *TOP2A*) were not only differentially expressed in RNA level but also at the protein level. Importantly, the NSC-specific genes *KIF2C* and *TOP2A* were validated by immunohistochemistry *in vivo*.

Conclusion: In present study, we identified 2 hub genes (*KIF2C* and *TOP2A*) that might serve as potential biomarkers for distinguishing NSCs from astrocytes, contributing to our comprehensive understanding of the biological properties and functions of NSCs.

1. Introduction

Neural stem cells (NSCs) are undifferentiated neural cells that have the capacity for self-renewal and can differentiate into the major classes of CNS cells, such as neurons and glial cells. The discovery of NSCs and neurogenesis in adult mammals provides a promising therapeutics for

Abbreviations: NSCs, neural stem cells; RGCs, embryonic radial glia cells; VZ, ventricular zone; GFAP, glial fibrillary acidic protein; GS, glutamine synthase; iPCs, intermediate progenitor cells; SVZ, subventricular zone; SGZ, subgranular zone; CP, cortical plate; GEO, Gene Expression Omnibus; RMA, Robust Multichip Average; DEGs, differentially expressed genes; PPI, protein-protein interaction; GO, Gene Ontology; KEGG, Kyoto encyclopedia of genes and genomes; GSEA, Gene Set Enrichment Analysis; qPCR, Real-time Quantitative PCR

* Corresponding authors.
E-mail addresses: chenghe@smmu.edu.cn (C. He), suzhida@smmu.edu.cn (Z. Su).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.gene.2018.09.044
Received 3 May 2018; Received in revised form 7 September 2018; Accepted 24 September 2018
Available online 25 September 2018
0378-1119/

CNS injury and disease (Goritz and Frisen, 2012). Recently, although our understanding of NSCs has dramatically increased, the characteristic genes expressed by NSCs remain incompletely defined.

NSCs, including embryonic radial glia cells (RGCs) and adult NSCs, are self-renewing and multi-potential cells that can generate neurons and glia cells (Kriegstein and Alvarez-Buylla, 2009; Flax et al., 1998; Laywell et al., 2000). During neurodevelopment, RGCs are the main subtype of NSCs that populate the ventricular zone (VZ). With the onset of neurogenesis, RGCs tend to acquire some hallmarks of glial cells, such as *vimentin*, the glial fibrillary acidic protein (*GFAP*) and glutamine synthase (*GS*) (Dimou and Gotz, 2014). RGCs can not only differentiate into neurons directly (direct neurogenesis), but also generate intermediate progenitor cells (iPCs) which consecutively produce neurons (indirect neurogenesis) (Song et al., 2002). In adults, NSCs are present mainly in two brain niches, the subventricular zone (SVZ) and the subgranular zone (SGZ) in the dentate gyrus (Bond et al., 2015). In earlier studies, *GFAP*-positive astrocytes in the SVZ region were considered to be the major type of adult NSCs (Song et al., 2002; Doetsch et al., 1999). As a special type of astrocytes, they were also called niche astrocytes (Doetsch et al., 1999; Gotz et al., 2016; Alvarez-Buylla and Lim, 2004). This evidence indicates that NSCs in the developing and adult brain have some astroglial properties.

Astrocytes are the most numerous cell type in the brain, playing multiple roles in maintaining physiological and pathological homeostasis of the CNS (Barres, 2008; Zuchero and Barres, 2015; Laywell and Steindler, 2002). Astrocytes generally remain in a non-active state in the CNS, but they can be activated by a variety of pathological stimuli, characterized by cell body hypertrophy, cytoplasmic processes extension, reentering into the cell cycle and proliferating (Colombo and Farina, 2016). Interestingly, activated astrocytes express some NSCs-related molecular bio-markers such as *GFAP*, *nestin*, *vimentin*, indicating that astrocytes in a particular state can acquire some "NSCs-like" properties (Song et al., 2002; Laywell and Steindler, 2002; Colombo and Farina, 2016). When cultured *in vitro*, activated astrocytes derived from an injured brain can form neurospheres and differentiate into neurons, astrocytes, and oligodendrocytes (Sirko et al., 2013; Shimada et al., 2012). These data suggest that astrocytes are different from other types of neural cells, such as microglia and oligodendrocytes, and can be trans-differentiated under certain conditions to obtain certain characteristics of NSCs.

NSCs, including RGCs and niche astrocytes, share many biological similarities with astrocytes in terms of cell morphology and molecular expression. For example, NSCs and astrocytes are both *GFAP*-positive cells (Kriegstein and Alvarez-Buylla, 2009; Colombo and Farina, 2016). *In vitro*, when FGF2 and EGF are added into the culture system, both NSCs and activated astrocytes can form neurospheres (Sofroniew, 2009; Gotz et al., 2015). Interestingly, they can be induced to transdifferentiate into each other bi-directionally under certain conditions (Hunter and Hatten, 1995; Faiz et al., 2015). In spite of their similar characteristics, NSCs and astrocytes are two different cell types. Except for the observation that NSCs and astrocytes are in different proliferation states (Kriegstein and Alvarez-Buylla, 2009; Laywell and Steindler, 2002), few distinct molecular signatures have been addressed in the literature. Therefore, it is important to provide further insights into the characteristic expression of NSCs.

In present study, multi-form bioinformatics as well as experimental methods were integrated in order to analyze the microarray data of NSCs and astrocytes. We here identified 2 key genes (*KIF2C* and *TOP2A*) that distinguish NSCs from astrocytes. These genes may serve as potential biomarkers for NSCs and contribute to our comprehensive understanding of their biological properties and functions.

2. Materials and methods

2.1. Microarray and RNA-seq data

Two mouse microarray datasets (GSE18765 and GSE69237) were obtained from the public database Gene Expression Omnibus (GEO) (Edgar and Barrett, 2006; Beckervordersandforth et al., 2010; Nakajima-Koyama et al., 2015). GSE18765 and GSE69237 were identified as DataSet1and DataSet2 respectively. We selected eight samples (four for NSCs and four for astrocytes) from DataSet1, and five samples (two for NSCs and three for astrocytes) from DataSet2. RNA-seq gene expression profiles were download from ENCODE project (http://encodeproject.org/ENCODE/) and GSE38805. All the data of these samples were analyzed *via* R (version 3.4.2). All the detailed information about these data was available in Supplementary Tables S1 and S2.

2.2. Identification and functional annotation of DEGs

The raw data of the two microarrays were analyzed by affy package (Gautier et al., 2004). The final expression matrix was obtained by preprocessing the DataSet with Robust Multichip Average (RMA), an R package, including missing values estimation, background correction, normalization and data summarization (Gautier et al., 2004). The DEGs were obtained based on their expression matrix and limma package (Ritchie et al., 2015). Statistically significant cut-off criteria were set for Fold Change (FC, logFC > 2) and p-value ($p < 0.05$). All computations of p-values were subjected to correction according to the Benjamini-Hochberg method. Gene Ontology (GO) (http://geneontology.org/) and Kyoto encyclopedia of genes and genomes (KEGG) (http://www.kegg.jp/) pathway annotation analysis were performed by clusterProfiler package (Yu et al., 2012). The up-regulated and the down-regulated DEGs in NSCs were subjected to enrichment analysis. A p-value < 0.05 was identified as cut-off criteria for this analysis.

2.3. Gene Set Enrichment Analysis (GSEA)

As a different analysis method, GSEA was used to analyze a specified set of genes within two groups and identify their biological significance (Subramanian et al., 2007). We used clusterProfiler package to perform GSEA analysis based on the default parameters.

2.4. PPI network construction

The DEGs obtained from DataSet1 and DataSet2 were compared, and overlapping genes which were both up-regulated or down-regulated in the NSCs group were screened out. The overlapping genes were used for protein-protein interaction (PPI) network construction.

The PPI network was constructed using the search tool for the retrieval of interacting genes (STRING) (Version 10.0; http://string.embl.de) (Szklarczyk et al., 2017). The DEGs enriched in NSCs were uploaded to STRING to construct PPI network based on two active interaction sources, textmining and experiments. A combined score > 0.4 was set as inclusion criteria. The visualization of PPI network was obtained from Cytoscape software (version 3.4.0) (Shannon et al., 2003). Furthermore, modularization analysis of associated genes was performed by MCODE in Cytoscape for finding modules in PPI networks (Saito et al., 2012). The key genes were identified with a degree > 10 in the modules.