European Review for Medical and Pharmacological Sciences 2015; 19: 4784-4790

The PPI network and clusters analysis in glioblastoma

W.-Q. HE, J.-W. GU, C.-Y. LI[1], Y.-Q. KUANG, B. KONG, L. CHENG, J.-H. ZHANG, J.-M. CHENG, Y. MA

Department of Neurosurgery, Chengdu Military General Hospital, Chengdu, People's Republic of China
[1]Department of Transfusion, Chengdu Military General Hospital, Chengdu, People's Republic of China

Abstract. – OBJECTIVE: Glioblastoma is the most aggressive tumor of the brain. To further understand its molecular mechanism, we carried out a systemic bioinformatics study of gene chips downloaded from Gene Expression Omnibus database.

MATERIALS AND METHODS: LIMMA package in R language was used to identify the differentially expressed genes (DEGs) between glioblastoma samples and normal controls.

RESULTS: Further, we constructed protein-protein interaction networks by mapping the DEGs into PPI data and identified network clusters in these networks. The results revealed that expression of 516 genes, which are mainly involved in phosphate metabolic process and signal transduction, were altered in glioblastoma samples. LYN, CD22 and LCP2 form a densely protein complex in the PPI network.

CONCLUSIONS: Our results suggest that LYN, CD22 and LCP2 play important roles in the occurrence and progression of glioblastoma.

Key Words:
Glioblastoma, Protein-Protein interaction network, ClusterONE analysis, MCODE analysis.

Introduction

Glioma, makes up 30%-40% of all brain tumors, is the most aggressive tumor of the brain[1]. Glioblastoma is the most aggressive form of glioma, accounting for approximately 50% of all glial tumor types[2]. A series of reconstruction features on pathology organization exist in the process of glioblastomas occurrence and progression, including malignant proliferation and aggressive growth of new blood vessels. Recent studies on histopathology, cytogenetic, and relevant molecular changes during tissue remodeling process of glioblastoma may offer valuable references for studying on molecular etiology, malignant progression, molecular subtyping, target screening for gene therapy, and prognostic evaluation against glioblastoma. In recent years, with the deep analysis of the gene mapping profile related to glioblastoma, the linked research on molecular and cell levels laps further.

With the knowledge of the cell and molecular changes during the pathogenesis, reoccurrences and malignant progress, several groups attempt to study the protein-protein interaction (PPI) network related to cancer[3,4], such as in human liver cancer[5], breast cancer[6,7] and colorectal cancer[8]. And luckily, some tools for systematic analysis have been developed[9,10]. All of these studies on cancer PPI network make means for related cell apoptosis[11] and aberrant methylation research[12], novel Single Nucleotide Polymorphisms (SNPs) discovering[13], discrimination of cancer phenotype[14], diagnostic and prognostic markers finding[15] and promising drug devising[16].

In this study, we focus on the molecular and cytology research of glioblastoma by viewing tumor as a whole. We here apply systematic methods of PPI and cluster analysis on glioblastoma and unfold some interesting discoveries.

Materials and Methods

All human studies have been approved by China Ethics Committee and performed in accordance with the ethical standards.

Affymetrix Mcroarray Data

The transcription profile of GSE6014[17] was obtained from National Center of Biotechnology Information Gene Expression Omnibus (GEO)

Corresponding Author: Jian-wen Gu, MD; e-mail: jianwengu1122@hotmail.com

Database (http://www.ncbi.nlm.nih.gov/geo/) which is based on the Affymetrix Human Genome U133A Array. Total 4 gene chips were available for analysis.

Protein-Protein Interaction (PPI) Data

The Human Protein Reference Database (HPRD)[18] is a protein database accessible through the internet. The Biological General Repository for Interaction Datasets (BioGRID)[19] is a curated biological database of protein-protein and genetic interactions.

The PPI data from the HPRD and BIOGRID were collected for further analysis. A total of 326119 unique PPI pairs were collected, in which 39240 pairs are from HPRD and 379426 pairs are from BIOGRID.

Differentially Expressed Genes (DEGs) analysis

For the GSE6014 dataset, LIMMA package[20] in R language was used to identify DEGs between glioblastoma samples and the control samples. The original expression datasets from all conditions were extracted into expression estimates, and then a linear model was constructed. The DEGs only with the fold change value larger than 2 and *p*-value less than 0.05 were selected.

PPI Network Construction

For demonstrating the potential PPI relationship, the DEGs were mapped to the PPI data that have been collected from HPRD and BIOGRID. The Pearson Correlation Coefficient (PCC)[21] was calculated for all pair-wise comparisons of gene-expression values between normal genes and the DEGs. The PPI relationships whose absolute PCC are larger than 0.75 were considered as significant. Then, a PPI network was constructed using the Cytoscape[22] based on the PPI relationships.

ClusterONE Analysis

ClusterONE http://www.cs.rhul.ac.uk/home/tamas/assets/files/cl1/cl1-cytoscape-0.1.html) strives to discover densely connected and possibly overlapping regions within the Cytoscape network you are working with[23]. ClusterONE works by "growing" dense regions out of small seeds (typically one or two vertices), driven by a quality function called cohesiveness. The quality of a group is evaluated by the number of internal edges divided by the number of edges involving nodes of the group. Based on the PPI network, sub-graphs smaller than 5 or having a density (number of edges within the cluster divided by the number of theoretically possible edges) less than 0.3 were discarded.

MCODE Analysis

Molecular Complex Detection (ftp://ftp.mshri.on.ca/pub/BIND/Tools/MCODE; MCODE) also detects densely connected regions in large PPI networks that may represent molecular complexes[24]. In this study, we used MCODE to further mine the core protein complex in clusters identified by ClusterONE.

Gene Ontology (GO) and KEGG Pathway Analysis

DAVID (The database for annotation, visualization and integrated discovery), a high-throughput and integrated data-mining environment, analyzes gene lists derived from high-throughput genomic experiments[25]. In David, a cumulative hypergeometric distribution is used for calculating the probability of getting at least n successes in the hypergeometric experiment. To compute a cumulative hypergeometric probability, we may need to add one or more individual probabilities.

$$P = 1 - \sum_{i=0}^{k-1} \frac{\binom{f}{i}\binom{n-f}{m-i}}{\binom{n}{m}}$$

where n is the number of proteins in PPI network, f is the number of proteins which were enriched in GO terms, m is the number of proteins which involve in KEGG pathways and k stands for the frequency of GO-ID emergency. We used the threshold of FDR (false discovery rate, adjusted by Benjamini-Hochberg method[26]) < 0.05 to identify the over-represented GO categories in biological process and KEGG pathway analysis based on the cumulative hypergeometric distribution.

Results

Identification of DEGs between DCA treated samples and untreated controls.

Publicly available microarray data set GSE6014 were obtained from GEO database. A total of 561 DEGs with the fold change > 2 and *p*-value < 0.05 were selected using the LIMMA package. All of these DEGs are up-regulated in glioblastoma samples.

PPI network Construction

In order to systemically analyze the functions of DEGs in glioblastoma samples, we mapped these DEGs to PPI data and obtained some PPI networks (Figure 1). At a PCC > 0.75, a total of 171 relationships between 76 DEGs and 137 normal genes were identified. In Figure 1, the DEGs of PTPN11 (degree =22), PAK1 (degree = 12), LCK (degree = 10) and LCP2 (degree = 8) formed local networks with high degrees (the number of interactions).

Network Clustering

The PPI network included hundreds of relationships and it is hard for us to know which relationships are crucial. Therefore, we identified the densely connected or possibly overlapping regions in the PPI network by using ClusterONE. By setting minimum size as 5 and minimum density as 0.3, five network clusters was found in PPI network (Figure 2).

Core Protein Complex Clustering

To obtain higher overexpression clusters, we further analyzed the network clusters by MCODE. MCODE does not provide any statistical score on the resulting clusters but can be used as a discovery tool in network analysis. Interactions among three nodes were filtered, including two DEGs (LCP2 and CD22) and 1 normal genes (LYN) (Figure 3).

Function Analysis of the PPI Network

DAVID was used to describe the function of the PPI network. Several GO categories were enriched among these genes in the PPI network, such as protein amino acid phosphorylation (GO: 0006468), phosphorylation (GO: 0016310) and phosphate metabolic process (GO: 0006796) (Table I). Table I only lists the top10 enriched GO terms.

In addition, several KEGG pathways were enriched among these genes in the PPI network, including natural killer cell mediated cytotoxic-

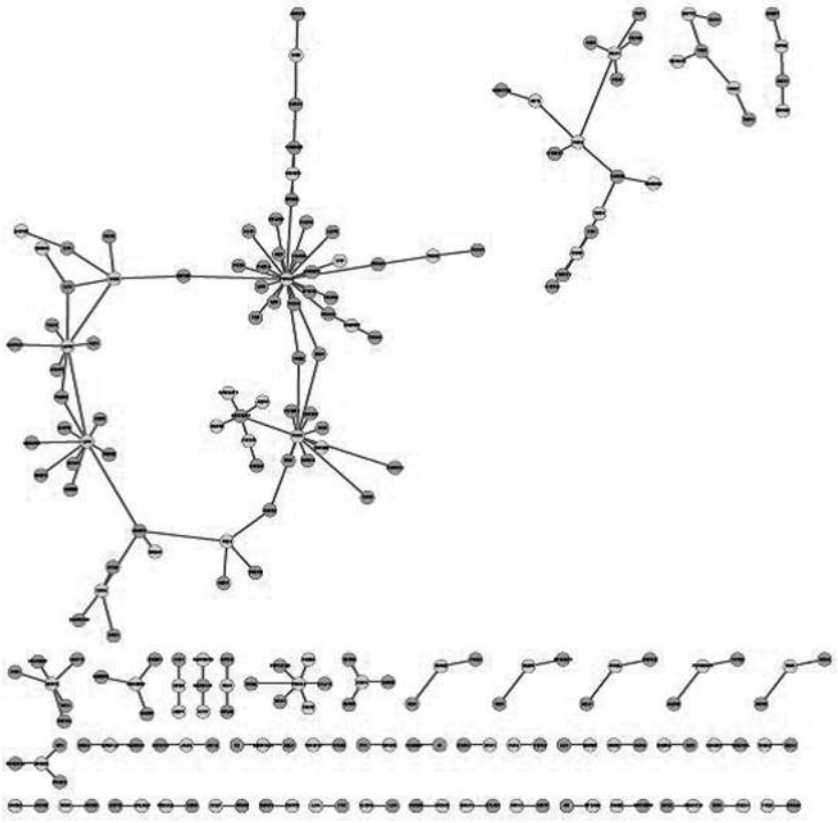

Figure 1. PPI Network constructed by Cytoscape. The yellow nodes stand for DEGs and the pink nodes stand for normal genes.

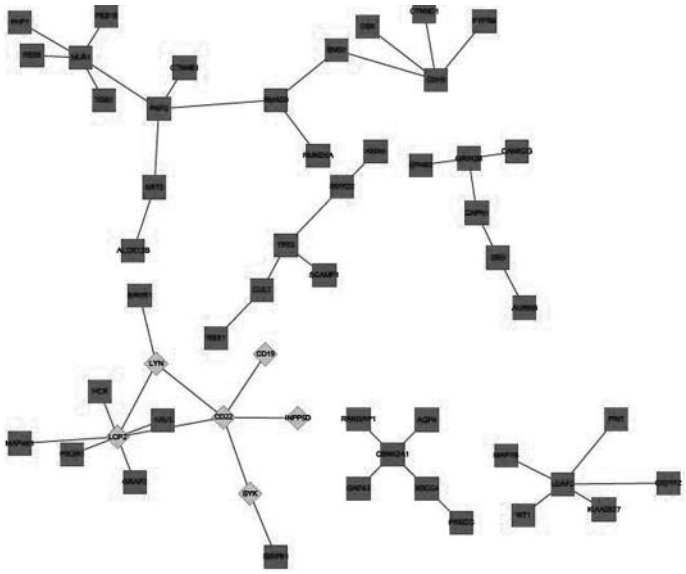

Figure 2. Network clusters identified from PPI network. The rhombic nodes stand for DEGs and the round nodes stands for normal genes.

ity (hsa04650), ErbB signaling pathway (hsa04012) and pathways in cancer (hsa05214) (Table II). Table II lists the KEGG pathways based on FDR < 0.01.

Discussion

Malignant glioma progression is a cumulative genetic change with multiple-steps. In this study, we investigated the molecular mechanism of glioblastoma using bioinformatics methods. We found that the expression of 516 genes, which are mainly involved in phosphate metabolic process and signal transduction, were altered in glioblastoma samples. Pathways of natural killer cell mediated cytotoxicity, ErbB signaling pathway and neurotrophin signaling pathway were dysregulated in glioblastoma. In addition, we constructed PPI networks, and studied the properties of this network by identifying network clusters: sets of genes that together involved in a biological process..

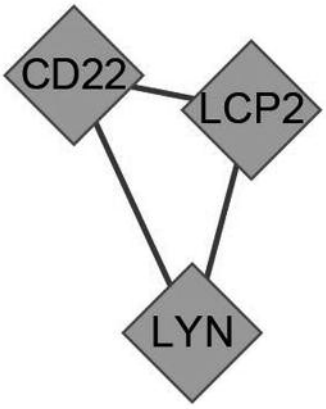

Figure 3. MCODE Cluster.

The PPI network is un-weighted, since each PPI was kept only once. As it is too large to yield any interesting information, it is necessary to divide it into connected sub-networks that might represent functional modules or protein sub-complexes. By clustering twice using ClusterONE and MCODE, a core protein complex was identified, including LYN, CD22 and LCP2.

Tyrosine-protein kinase LYN is a member of the Src family which is mainly expressed in hematopoietic cells, neural tissues, liver and adipose tissue[27-29]. LYN was reported to be a mediator of epithelial-mesenchymal transition and a target of dasatinib in breast cancer[30], and mediates cell motility and tumor growth in EGFR vIII-expressing head and neck cancer[31], involved in CD24-induced ERK1/2 activation in colorectal cancer[32]. In the clinical area, LYN is also reported to be target of treatment of prostate cancer[33] and Ewing's sarcoma[34]. In embryonic brain, LYN activation by integrin is required for oligodendrocyte progenitor proliferation[35]. Stettner et al[36] suggest that LYN kinase activity is significantly elevated in glioblastoma tumors and promotes the malignant phenotype in these tumors. In addition, a recent study by Feng et al[37] suggest

Table I. The top 10 significantly enriched GO terms in glioblastoma.

Term	Description	Count	FDR
GO:0006468	Protein amino acid phosphorylation	48	2.35E-17
GO:0016310	Phosphorylation	49	6.36E-15
GO:0006796	Phosphate metabolic process	52	1.89E-13
GO:0006793	Phosphorus metabolic process	52	1.89E-13
GO:0007166	Cell surface receptor linked signal transduction	70	9.07E-12
GO:0051174	Regulation of phosphorus metabolic process	34	8.42E-11
GO:0019220	Regulation of phosphate metabolic process	34	8.42E-11
GO:0007167	Enzyme linked receptor protein signaling pathway	29	9.17E-11
GO:0042325	Regulation of phosphorylation	33	1.66E-10
GO:0007242	Intracellular signaling cascade	54	3.02E-10

FDR indicates false discovery rate adjusted by Benjamini-Hochberg method.

that LYN largely mediate EGFRvIII stimulation of Rac1 activity and glioblastoma cell migration through inducing phosphorylation of the tyrosine residue 722 of Dock180, a guanine nucleotide exchange factor for Rac1.

CD22, a kind of cell membrane related antigen differentiation, is also a kind of well-studied tumor cell surface target. With special feature of B-cell-restricted phosphoglycoprotein of the immunoglobulin superfamily, CD22 has gained considerable interest as a therapeutic target for B-cell-directed therapies. It is reported to be used as a tumor marker for hairy cell leukemia[38]. Although CD22 is not a DEG in glioblastoma, it still play important roles in the occurrence and progression of glioblastoma by interacting with LYN. Our study is in accordance with a previous study which suggested that CD22 functions as a molecular 'scaffold' that specifically coordinates the docking of multiple effector molecules, in a context necessary for B cell antigen receptor-dependent Src homology-2 domain-containing inositol polyphosphate-5'-phosphatase activity and c-Jun N-terminal kinase stimulation[39].

LCP2 (Lymphocyte cytosolic protein 2), also known as SLP-76, was identified by association with the SH3 domain of the Grb2 adapter protein in T cells[40]. In our study, LCP2 was interacted with LYN and CD22. This result is consistent with a previous immunoprecipitation study which suggested that LCP2 is associated with FYN, LYN and Fc receptor γ-chain in collagen-related peptide- stimulated platelets[41]. Gross et al[41] suggest tyrosine phosphorylation of LCP2 is mediated by either FYN or LYN. As expected, the GO enrichment analysis in our study suggests DEGs are involved in phosphate metabolic process.

Conclusions

We constructed some PPI networks, and filtered a core protein complex in the network. Functional analyses suggested that interactions among LYN, CD22 and LCP2 play important roles in the progression of glioblastoma, mainly by involving in phosphate metabolic process and

Table II. The significantly enriched KEGG pathways in glioblastoma.

Term	Description	Count	FDR
hsa04650	Natural killer cell mediated cytotoxicity	18	1.03E-04
hsa04012	ErbB signaling pathway	14	5.95E-04
hsa05200	Pathways in cancer	27	7.14E-04
hsa05220	Chronic myeloid leukemia	13	7.79E-04
hsa04722	Neurotrophin signaling pathway	16	0.001276
hsa04666	Fc gamma R-mediated phagocytosis	14	0.00168
hsa05214	Glioma	11	0.007429

FDR indicates false discovery rate adjusted by Benjamini-Hochberg method.

signal transduction. Our analysis may aid in understanding the molecular mechanism of glioblastoma. However, further experimental studies are needed because our study is based on gene chips from a small sample size.

Acknowledgements

This work was supported by a grant from the National Natural Science Foundation of China (No.81071037).

Conflict of Interest

The Authors declare that there are no conflicts of interests.

References

1) Kleihues P, Burger PC, Scheithauer BW. The new WHO classification of brain tumours. Brain Pathol 1993; 3: 255-68.
2) Molina JR, Hayashi Y, Stephens C, Georgescu MM. Invasive glioblastoma cells acquire stemness and increased Akt activation. Neoplasia 2010; 12: 453-463.
3) Kar G, Gursoy A, Keskin O. Human cancer protein-protein interaction network: a structural perspective. PLoS Comput Biol 2009; 5: e1000601.
4) Sun J, Zhao Z. A comparative study of cancer proteins in the human protein-protein interaction network. BMC Genomics 2010; 11(Suppl 3): S5.
5) Zhou HJ, Liu YK, Li Z, Yun D, Shun QL, Guo K. Analysing protein-protein interaction networks of human liver cancer cell lines with diverse metastasis potential. J Cancer Res Clin Oncol 2007; 133: 663-672.
6) van den Akker EB, Verbruggen B, Heijmans BT, Beekman M, Kok JN, Slagboom PE, Reinders MJ. Integrating protein-protein interaction networks with gene-gene co-expression networks improves gene signatures for classifying breast cancer metastasis. J Integr Bioinform 2011; 8: 188.
7) Sanz-Pamplona R, Garcia-Garcia J, Franco S, Messeguer X, Driouch K, Oliva B, Sierra A. A taxonomy of organ-specific breast cancer metastases based on a protein-protein interaction network. Mol Biosyst 2012; 8: 2085-2096.
8) Li BQ, Huang T, Liu L, Cai YD, Chou KC. Identification of colorectal cancer related genes with mRMR and shortest path in protein-protein interaction network. PLoS One 2012; 7: e33393.
9) Sanz-Pamplona R, Berenguer A, Sole X, Cordero D, Crous-Bou M, Serra-Musach J, Guino E, Angel Pujana M, Moreno V. Tools for protein-protein interaction network analysis in cancer research. Clin Transl Oncol 2012; 14: 3-14.
10) Xu Y, Hu W, Chang Z, Duanmu H, Zhang S, Li Z, Yu L, Li X. Prediction of human protein-protein interaction by a mixed Bayesian model and its application to exploring underlying cancer-related pathway crosstalk. J R Soc Interface 2011; 8: 555-567.
11) Chu LH, Chen BS. Construction of a cancer-perturbed protein-protein interaction network for discovery of apoptosis drug targets. BMC Syst Biol 2008; 2: 56.
12) Liu H, Su J, Li J, Lv J, Li B, Qiao H, Zhang Y. Prioritizing cancer-related genes with aberrant methylation based on a weighted protein-protein interaction network. BMC Syst Biol 2011; 5: 158.
13) Savas S, Geraci J, Jurisica I, Liu G. A comprehensive catalogue of functional genetic variations in the EGFR pathway: protein-protein interaction analysis reveals novel genes and polymorphisms important for cancer research. Int J Cancer 2009; 125: 1257-1265.
14) Xiong J, Liu J, Rayner S, Li Y, Chen S. Protein-protein interaction reveals synergistic discrimination of cancer phenotype. Cancer Inform 2010; 9: 61-66.
15) Su J, Yoon BJ, Dougherty ER. Identification of diagnostic subnetwork markers for cancer in human protein-protein interaction network. BMC Bioinformatics 2010; 11(Suppl 6): S8.
16) Shangary S, Wang S. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. Annu Rev Pharmacol Toxicol 2009; 49: 223-241.
17) Bonnet S, Archer S. Ialunis-Turner J, Haromy A, Beaulieu C, Thompson R, Lee CT, Lopaschuk GD, Puttagunta L, Bonnet S, Harry G, Hashimoto K, Porter CJ, Andrade MA, Thebaud B, Michelakis ED. A mitochondria-K+ channel axis is suppressed in cancer and its normalization promotes apoptosis and inhibits cancer growth. Cancer Cell 2007; 11: 37-51.
18) Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R, Pandey A. Human Protein Reference Database—2009 update. Nucleic Acids Res 2009; 37: D767-772.
19) Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X, Reguly T, Rust JM, Winter A, Dolinski K, Tyers M. The BioGRID Interaction Database: 2011 update. Nucleic Acids Res 2011; 39: D698-704.
20) Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004; 3: Article3.
21) Derrick TR, Bates BT, Dufek JS. Evaluation of time-series data sets using the Pearson product-moment correlation coefficient. Med Sci Sports Exerc 1994; 26: 919-928.

22) Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, 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.

23) Bader GD HC. An automated method for finding molecular complexes in large protein interaction networks. 2003;

24) Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003; 4: 2.

25) Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44-57.

26) Benjamini YH, Y. Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Series B (Methodological) 1995; 57: 289-300.

27) Yamanashi Y, Mori S, Yoshida M, Kishimoto T, Inoue K, Yamamoto T, Toyoshima K. Selective expression of a protein-tyrosine kinase, p56lyn, in hematopoietic cells and association with production of human T-cell lymphotropic virus type I. Proc Natl Acad Sci U S A 1989; 86: 6538-6542.

28) Umemori H, Wanaka A, Kato H, Takeuchi M, Tohyama M, Yamamoto T. Specific expressions of Fyn and Lyn, lymphocyte antigen receptor-associated tyrosine kinases, in the central nervous system. Brain Res Mol Brain Res 1992; 16: 303-310.

29) Yamada E, Pessin JE, Kurland IJ, Schwartz GJ, Bastie CC. Fyn-dependent regulation of energy expenditure and body weight is mediated by tyrosine phosphorylation of LKB1. Cell Metab 2010; 11: 113-124.

30) Choi YL, Bocanegra M, Kwon MJ, Shin YK, Nam SJ, Yang JH, Kao J, Godwin AK, Pollack JR. LYN is a mediator of epithelial-mesenchymal transition and a target of dasatinib in breast cancer. Cancer Res 2010; 70: 2296-2306.

31) Wheeler SE, Morariu EM, Bednash JS, Otte CG, Seethala RR, Chiosea SI, Grandis JR. Lyn kinase mediates cell motility and tumor growth in EGFRvIII-expressing head and neck cancer. Clin Cancer Res 2012; 18: 2850-2860.

32) Su N, Peng L, Xia B, Zhao Y, Xu A, Wang J, Wang X, Jiang B. Lyn is involved in CD24-induced ERK1/2 activation in colorectal cancer. Mol Cancer 2012; 11: 43.

33) Goldenberg-Furmanov M, Stein I, Pikarsky E, Rubin H, Kasem S, Wygoda M, Weinstein I, Reuveni H, Ben-Sasson SA. Lyn is a target gene for prostate cancer: sequence-based inhibition induces regression of human tumor xenografts. Cancer Res 2004; 64: 1058-1066.

34) Guan H, Zhou Z, Gallick GE, Jia SF, Morales J, Sood AK, Corey SJ, Kleinerman ES. Targeting Lyn inhibits tumor growth and metastasis in Ewing's sarcoma. Mol Cancer Ther 2008; 7: 1807-1816.

35) Colognato H, Ramachandrappa S, Olsen IM, ffrench-Constant C. Integrins direct Src family kinases to regulate distinct phases of oligodendrocyte development. J Cell Biol 2004; 167: 365-375.

36) Stettner MR, Wang W, Nabors LB, Bharara S, Flynn DC, Grammer JR, Gillespie GY, Gladson CL. Lyn kinase activity is the predominant cellular SRC kinase activity in glioblastoma tumor cells. Cancer Res 2005; 65: 5535-5543.

37) Feng H, Hu B, Jarzynka MJ, Li Y, Keezer S, Johns TG, Tang CK, Hamilton RL, Vuori K, Nishikawa R, Sarkaria JN, Fenton T, Cheng T, Furnari FB, Cavenee WK, Cheng SY. Phosphorylation of dedicator of cytokinesis 1 (Dock180) at tyrosine residue Y722 by Src family kinases mediates EGFRvIII-driven glioblastoma tumorigenesis. Proc Natl Acad Sci U S A 2012; 109: 3018-3023.

38) Matsushita K, Margulies I, Onda M, Nagata S, Stetler-Stevenson M, Kreitman RJ. Soluble CD22 as a tumor marker for hairy cell leukemia. Blood 2008; 112: 2272-2277.

39) Poe JC, Fujimoto M, Jansen PJ, Miller AS, Tedder TF. CD22 forms a quaternary complex with SHIP, Grb2, and Shc. A pathway for regulation of B lymphocyte antigen receptor-induced calcium flux. J Biol Chem 2000; 275: 17420-17427.

40) Jackman JK, Motto DG, Sun Q, Tanemoto M, Turck CW, Peltz GA, Koretzky GA, Findell PR. Molecular cloning of SLP-76, a 76-kDa tyrosine phosphoprotein associated with Grb2 in T cells. J Biol Chem 1995; 270: 7029-7032.

41) Gross BS, Lee JR, Clements JL, Turner M, Tybulewicz VL, Findell PR, Koretzky GA, Watson SP. Tyrosine phosphorylation of SLP-76 is downstream of Syk following stimulation of the collagen receptor in platelets. J Biol Chem 1999; 274: 5963-5971.