Identification of biomarker in brain-specific gene regulatory network using structural controllability analysis

Zhihua Chen (✉ czhgd@gzhu.edu.cn)

Guangzhou University <https://orcid.org/0000-0002-8039-2141>

Siyuan Chen

Huazhong University of Science and Technology - Main Campus: Huazhong University of Science and Technology

Xiaoli Qiang

Guangzhou University

Research article

Keywords: brain-specific gene regulatory network, controllability analysis, biomarker, brain tumor research

Posted Date: July 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-684666/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Frontiers in Bioinformatics on January 26th, 2022. See the published version at <https://doi.org/10.3389/fbinf.2022.812314>.

RESEARCH

Identification of biomarker in brain-specific gene regulatory network using structural controllability analysis

Zhihua Chen¹, Siyuan Chen² and Xiaoli Qiang^{1*}

Abstract

Background: Brain tumor research has been stapled for human health while brain network research is crucial for us to understand brain activity.

Methods: Here the structural controllability theory is applied to study three human brain-specific gene regulatory networks, including forebrain gene regulatory network, hindbrain gene regulatory network and neuron associated cells cancer related gene regulatory network, whose nodes are neural genes and the edges represent the gene expression regulation among the genes. The nodes are classified into two classes: critical nodes and ordinary nodes, based on the change of the number of driver nodes upon its removal. Eight topological properties(out-degree DO , in-degree DI , degree D , betweenness B , closeness CA , in-closeness CI , out-closeness CO and clustering coefficient CC) are calculated in this paper and the results prove that the critical genes have higher score of topological properties than the ordinary genes. Then two bioinformatic analysis are used to explore the biologic significance of the critical genes.

Results: On the one hand, the enrichment scores in several kinds of gene databases are calculated and reveal that the critical nodes are richer in essential genes, cancer genes and the neuron related disease genes than the ordinary nodes, which indicates that the critical nodes may be the biomarker in brain-specific gene regulatory network. On the other hand, GO analysis and KEGG pathway analysis are applied on them and the results show that the critical genes mainly take part in 14 KEGG pathways that are transcriptional misregulation in cancer, pathways in cancer and so on, which indicates that the critical genes are related to the brain tumor. Finally, by deleting the edges or routines in the network, the robustness analysis of node classification is realized, and the robustness of node classification is proved.

Conclusion: The comparison of neuron associated cells cancer related GRN and normal brain-specific GRNs (including forebrain and hindbrain GRN) shows that the neuron-related cell cancer-related gene regulatory network is more robust than other types.

Keywords: brain-specific gene regulatory network; controllability analysis; biomarker; brain tumor research

1 Introduction

The world has opened its eyes to the threat posed by cancer [1]. Brain tumor is a mass or growth of abnormal cells in the human brain. It can begin in the human brain (primary brain tumors), or begin in other parts of body and spread to brain (secondary, or metastatic, brain tumors) [2]. Brain tumor accounts for 85% to 90% of all primary central nervous system (CNS) tumors [3]. There are new cases and deaths from brain tumor and other nervous system tumors estimated around the world every year. Approximately 256,213 new cases of brain and other CNS tumors were diagnosed in the year 2012, with an estimated 189,382 deaths [4], and there are 296,851 new cases and 241,037 deaths in 2018 [5]. Furthermore, the cause of most adult brain and spinal cord tumors is not known. It is urgent to study the pathogenic mechanism and treatment for brain tumors.

Many studies have focused on the role of single molecule or single pathway in regulating tissue-specific nuclear structure and gene expression. For example, SATB1, a cell-

type specific nuclear protein, can recruit chromatin remodeling factors and regulate many genes during thymocyte differentiation. And it is proposed by Cai-s et al. as a novel gene regulator, which can provide sites for tissue-specific and region specific histone modification of DNA sequences [6]. Fass D M et al. [7] studied the role of GPCR-kinase interacting protein 1 (GIT1), and found that GIT1 deletion interferes with the specific network of GIT1 interacting synapses. Although the understanding of individual molecules is crucial, the focus is on understanding the entire gene regulatory network at the system level. Because the properties of gene regulatory networks cannot be fully understood by studying single molecules [8].

Marbach D et al. Developed a comprehensive resource of 394 cell types and tissue-specific gene regulatory networks with 37 genome wide association studies (GWAS), which clarifies the genome-wide connectivity among transcription factors, enhancers, promoters and genes [9]. McKenzie A T et al. identified a novel set of brain cell consistent signatures and robust networks from the integration of multiple data sets, so it goes beyond the limitations associated with each individual research specific technical problem [10]. Therefore, it is feasible to construct brain specific gene reg-

*Correspondence: qiangxl@gzhu.edu.cn

¹Institute of Computing Science and Technology, Guangzhou University, 510006 Guangzhou, China

Full list of author information is available at the end of the article

ulatory network based on relevant data and study it based on network analysis.

Network Science has become an emerging and highly interdisciplinary research area that aims to increase our understanding of complex networks [11–15]. At the same time, with the increasing of massive genomic, proteomic, and metabolomics data, the formation of multi-layer biological molecular network is promoted, which lays a foundation for the analysis of biological problems by network science [16–19]. Detailed maps of mammalian brains could lead to a revolution in brain science, which allows us to understand and find the cure of numerous neurological and brain diseases. With that, network science could be applied in brain research widely [20–22]. It has been applied in many kinds of biological networks, such as mouse inter-region brain networks and human transcription factor regulatory networks [23]. Liu et al. used the control theory of structural controllability to analyse numerous models of real networks, for instance, the directed human protein interaction network, which helped us to identify disease genes and drug targets [24], the human signaling network to identify driver nodes [25] and so on. It has been proved that a system's behaviour can be guided towards a desired state with a suitable choice of control signals to some significant nodes [26–29]. Therefore, it's feasible that we can apply controllability theory of network science on the analysis of brain-specific gene regulatory network.

Network structural controllability analysis [27, 29] has been a general framework in identifying critical nodes that have crucial roles in controlling the state of the whole system. By applying this framework to human liver metabolic networks [30], the critical driver metabolites tend to be essential metabolites. Moreover, this framework can be used to predict potential drug-targets [25]. It is fair to expect that there are some possible connections between the structural controllability theory and the human brain-specific gene regulatory networks, which could provide valuable information on the brain-specific gene regulatory networks, such as identifying essential genes, brain-related disease genes and drug targets.

In this work, we apply structural controllability method to analysis large-scale directed human brain-specific gene regulatory networks, where nodes are neurons and edges represent the gene expression regulation among the genes. The weight of each edge ranges from 0 to 1, measures the normalized activity levels of the enhancer. We classify the nodes into two classes: critical genes and ordinary genes. Then we do the topological and biological analysis of these two kinds of genes, and find that critical genes tend to be essential genes. By calculating eight topological properties (out-degree, in-degree, degree, betweenness, closeness, in-closeness, out-closeness and clustering coefficient), we can see that critical genes have higher score of topological properties than ordinary genes. Moreover, the enrichment in several kinds of gene databases are explored,

which shows that critical nodes are richer in essential gene, cancer gene and the neuron related disease gene than ordinary nodes. Besides, GO analysis and KEGG pathway analysis also help to infer that critical nodes are useful for us to explore more significant biological information and to identify the biomarker for brain tumor research. Finally, since the gene regulatory network is not complete or there may be some false links, we do sensitivity analysis of the results by perturbing the network. We find that the result of node classification is quite robust, and the neuron associated cells cancer related gene regulatory network is more robust than the health networks. The findings in our paper could help identify potential essential, cancer and neuron related disease genes.

2 results

2.1 Classification by driver nodes

2.1.1 Description of brain-specific gene regulatory network

We construct three brain-specific gene regulatory networks, which consist of forebrain gene regulatory network (forebrain GRN), hindbrain gene regulatory network (hindbrain GRN) and neuron associated cells cancer related gene regulatory network (neuron associated cells cancer GRN). They are directed networks and their nodes are neuronal genes. Their edges represent the regulation among neuronal genes, specifying the genome-wide connectivity among transcription factors, enhancers, promoters and genes [9]. Forebrain GRN and hindbrain GRN are normal and health human's brain-specific gene regulatory networks, while neuron associated cells cancer related GRN is cancer patient's brain-specific gene regulatory network. In these networks, the edge direction corresponds to the hierarchy of signal flow between the interacting genes and the edge weight corresponds to the confidence of the predicted direction. Generally, we delete the edges whose edge weight is smaller than 0.05 for our study. By this way, the human forebrain GRN consists of 14435 genes (nodes) and 222867 directed edges, the human hindbrain GRN consists of 14601 genes and 228708 directed edges, and the neuron associated cells cancer related GRN consists of 15320 genes and 256434 edges.

2.1.2 Classification of brain-specific gene regulatory networks' nodes

By structural controllability theory, the minimum set of driver nodes is identified and the size of it is calculated as N_D . Then the nodes are classified as critical or ordinary, based on the change of N_D upon their removal. The node is critical if N_D has no change or increases because of its removal, or the node is ordinary if N_D decreases. The results are shown in Table I. The number of forebrain GRN's critical nodes is 642, while the number of hindbrain GRN's and neuron associated cells cancer related GRN's critical nodes are both 643.

2.2 Topological analysis

Different centralities of each gene in the three brain-specific gene regulatory networks are calculated, which including out-degree *DO*, in-degree *DI*, degree *D*, betweenness *B*, closeness *CA*, in-closeness *CI*, out-closeness *CO*, betweenness *B* and clustering coefficient *CC*. The average values of each topological property's value are shown in Table II. It is clearly that there are some similar topological properties among the three brain-specific gene regulatory networks, so we can make topological analysis from two aspects.

On one aspect of single network, taking the hindbrain GRN into account, the ordinary nodes' out-degree is 0, as shown in fig.1 (b) and (d), which means that the ordinary genes in the networks are all just signal receivers. And the degree distribution suggests that critical nodes have higher degree than ordinary nodes (degree distribution of forebrain GRN and neuron associated cells cancer related GRN are shown in fig.s1, it shows the same result as hindbrain GRN). The average of the ordinary nodes' betweenness is also 0. What's more, the average out-closeness of critical nodes is much bigger than ordinary nodes, and it is proved convincing by significance testing (Mann Whitney U test, p-value is smaller than 0.05, significance level = 0.05, see Table III). These results imply that critical nodes are more important or useful for further study.

On the other aspect of comparison of the three brain-specific gene regulatory networks, the average in-degree, average out-degree, average betweenness and average clustering coefficient of neuron associated cells cancer related GRN are bigger than forebrain GRN and hindbrain GRN, no matter critical or ordinary nodes. It can be explained that the connection of neuron associated cells cancer related GRN is tighter than the normal brain network.

2.3 Biological enrichment analysis

The enrichment scores of the critical genes and ordinary genes in different biofunctional gene databases are calculated, and GO analysis and KEGG pathway analysis are adopted on critical genes to explore the biological significance.

2.3.1 Enrichment score calculating analysis

the nodes in the databases which consist of essential genes, cancer genes, and the neuron related disease genes respectively are characterized as critical and ordinary nodes. Essential genes are necessary for cellular survivor. The gene essentiality analysis indicates that critical nodes are enriched in essential genes, whereas essential genes are underrepresented among ordinary nodes (fig.2). Furthermore, the critical nodes are enriched in cancer genes and the related disease genes in the fig.2, and it indicates that the disease genes are most likely among the critical nodes. Fig.s2 and s3 show the biological enrichment analysis of forebrain

GRN and neuron associated cells cancer related GRN, respectively. These results indicate that the proposed classification method is a reliable and useful tool for the prediction of brain tumor biomarkers. The critical genes mined in this way have bioinformatics significance and have a strong correlation with brain tumors.

2.3.2 GO analysis and KEGG pathway analysis on critical genes

GO (Gene Ontology) analysis [31] has the largest resource for cataloguing gene function, which is subdivided in three non-overlapping ontologies, Molecular Function (MF), Biological Process (BP) and Cellular Component (CC). KEGG (Kyoto Encyclopedia of Genes and Genomes) [32] is a knowledge base for systematic analysis of gene functions, linking genomic information with higher order functional information. More detailed biological function of the genes can be obtained from GO analysis and KEGG pathway analysis.

David, the online biological enrichment analysis software, is used to realize the GO analysis and KEGG pathway analysis on critical genes. The results are shown in fig.3 and table IV. GO analysis points out that critical genes are rich in nucleus (93.4%), nucleoplasm (42.7%), cytoplasm (33.3%) and so on. Their molecular functions mostly consist of transcription factor activity, sequence-specific DNA binding and protein binding. What's more, the critical genes mainly play the important roles in 7 biological processes, which are positive regulation of transcription from RNA polymerase II promoter (56.1%), DNA-templated (56.0%), etc. KEGG pathway analysis shows that the critical genes are rich in 14 KEGG pathways, which are confirmed with brain tumor from the Comparative Toxicogenomics Database (CTD, <http://ctdbase.org/>).

From the above analysis, we can find that the critical genes we find are important and have an association with brain tumor. And GO analysis and KEGG pathway analysis provide the reference for further study on brain tumor.

2.4 Robustness analysis of node classification

Since the critical nodes are vital and significant, it's necessary to know whether the classification is robustness if the network is attacked. Therefore, the robustness of node classification is systematically tested by deleting edges or nodes.

2.4.1 Deleting edges

Over 96% edges' weights are between 0 and 0.10 in the three brain-specific gene regulatory networks. The detail information can be seen in table V. Therefore, the edges, whose edge weight are smaller than 0.01, 0.02, ..., 0.10, are deleted respectively to get new networks. Then the same method is used to identify the new networks' critical nodes and ordinary nodes. Finally, the proportions of the critical nodes of the new networks to the original network are

compared in fig.4(a) and fig.s4. It can be seen that the node classification is robust with respect to deleting edges in the three brain-specific gene regulatory networks. What's more, through comparing the three curves in fig.4(b), we can know that the neuron associated cells cancer related GRN has higher value than the others, which means the neuron associated cells cancer related GRN is more robust than the health brain-specific gene regulatory networks.

2.4.2 Deleting nodes by critical nodes

In this part, the change of giant strongly connected component (GSCC) and giant weakly connected component (GWCC) are compared under three deleting strategies. It is applied on the three brain-specific gene regulatory networks in the same way, so we take hindbrain GRN for example. Firstly, deleting nodes randomly in the whole hindbrain GRN (noted as *randomly - all*). Secondly, deleting critical nodes randomly (noted as *critical*). Finally, deleting the same number of nodes as the second strategy in the whole hindbrain GRN, and whose nodes are deleted is randomly (noted as *randomly - critical - all*). Compared the curve of *critical* with *randomly - all* and *randomly - critical - all*, the results shown in fig.5(a) and (b) indicate that the maximal connected subgraph becomes smaller and smaller with the decrease of the critical nodes. It means that the critical nodes are crucial for the connection of the whole network. It is consistent with the results of forebrain GRN and neuron associated cells cancer related GRN from fig.s5. Moreover, the maximal connected subgraph is bigger than the others as described in fig.5(c)-(h), which are the comparisons among three brain networks in the same strategy. It means that neuron associated cells cancer related GRN is more robust than normal and health brain-specific gene regulatory network.

3 discussion

In this paper, the gens in human brain-specific gene regulatory networks (forebrain GRN, hindbrain GRN and neuron associated cells cancer related GRN) are divided into the critical genes and ordinary genes. By calculating eight topological properties (out-degree DO , in-degree DI , degree D , betweenness B , closeness CA , in-closeness CI , out-closeness CO and clustering coefficient CC), we find that the critical genes play important roles in the human brain-specific GRN networks. For example, the critical genes have larger score of CO than the ordinary genes. Biological enrichment analysis in essential genes database shows that critical genes are richer than the ordinary genes, so we predict that critical genes are more significant for us to explore biological information. Furthermore, the enrichments in cancer genes database and neuron related disease genes database are explored, and it is consistent with our prediction. Because the critical nodes are richer than the ordinary nodes in these gene databases. It indicates that the critical

genes can contribute to identifying the disease genes related to brain. GO analysis and KEGG pathway analysis indicate that critical genes are associated with brain tumor and hint that they are rich in transcriptional misregulation in cancer, pathways in cancer and so on, which provide references for further study on brain tumor. Finally, tests show that the nodes classification method is robust when the network is attacked. And the tests indicate that the neuron associated cells cancer related GRN is more robust than normal and health brain-specific gene regulatory networks (forebrain GRN and hindbrain GRN), which is straightaway that a person gets sick easily but regains health difficultly.

In conclusion, controllability theory is also an useful tool to analyse human brain-specific gene regulatory network. It can provide a feasible direction for biologists to study whether the biomarker mined by the proposed method is related to brain tumor or not. In addition, the research work also raises a number of questions. For instance, how can we quantify the influence of each critical genes for the network? Can the work expand to the structural of function brain network? Answers to these questions can further provide theoretical foundation for designing experiments.

4 method

4.1 Brain-specific gene regulatory networks

There are three kinds of brain-specific gene regulatory networks. For convenience, we called them hindbrain GRN, forebrain GRN and neuron associated cells cancer related GRN respectively. Hindbrain GRN and forebrain GRN are normal and health adult's brain-specific gene regulatory networks, while the neuron associated cells cancer related GRN is the patient's brain-specific gene regulatory network, who suffers from brain tumor. They all are dealt with by deleting edges whose edge weights are smaller than 0.05 before we use to analyse in this paper.

4.2 Structural Controllability and its applications to biological networks

Biological networks are complex nonlinear systems. The controllability of nonlinear systems is structurally similar to that of linear systems [27, 33]. We study a system with canonical linear, time-invariant dynamics formulated by [34]

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t), \quad (1)$$

where the vector $x(t) = (x_1(t), x_2(t), \dots, x_N(t))^T$ describes the states of the N nodes of the networked system at time t . The $N \times N$ matrix A is the transposition of the adjacency matrix and captures the wiring diagram of the system and the interaction strengths between nodes. The $N \times M$ matrix B is the input matrix ($N \geq M$) that identifies the nodes into which the input signals are

injected, M is the number of input signals, and $u(t) = (u_1(t), u_2(t), \dots, u_m(t))^T$ is the input vector.

In control theory, a system is controllable if it can be driven from any initial state to any desired final state during a finite time period [35]. According to Kalman's controllability rank condition [35], the system represented by Eq. (1) is controllable if and only if the $N \times NM$ controllability matrix C has full rank, i.e.,

$$\text{rank}(C) = \text{rank}[B, AB, A^2B, \dots, A^{N-1}B] = N. \quad (2)$$

This controllability rank condition indicates that to control the full network we must identify the number of signals and the nodes into which the signals are injected, called driver nodes. Liu et al. [27] recently showed that a full system can be structurally controlled by inputting signals into a minimum set of driver nodes. A system is structurally controllable if it is possible to choose non-zero weights in A and B such that Eq. (2) holds [27]. The minimum number of driver nodes for controlling a full network is denoted N_D and the minimum driver node density is $n_D = N_D/N$. The minimum driver node density required to control the full complex network quantifies its structural controllability [27, 36].

Structural controllability analysis has been applied to some biological networks, where interesting properties on the biological system and drug-targets have been discovered [24, 25, 30]. According to frequency of appearing in the minimum driver node sets, or the impact of removal a node on the minimum number of driver nodes, the nodes can be classified in to different classes: critical, redundant and ordinary (as explained in the following subsection). In biological molecular networks, biological molecules can be classified with different roles. By doing biological enrichment analysis of these different biological roles, candidate essential genes or drug-targets can be identified.

4.3 Node classification

According to the control theory, a dynamical system is controllable if, with a suitable choice of inputs, it can be driven from any initial state to any desired final state within finite time [35] [37]. By using the analytical tool developed by Liu et al. [27], we can identify the set of driver nodes in an arbitrary complex directed network, with time-dependent control that can guide the systems entire dynamics to study its controllability. Moreover, the minimum number of driver nodes is determined for a determined network. Hence, the mathematical framework and analytical tools that has been developed by Vinayagam et al. [24] can be used to compute the minimum number, and denote it as N_D . After removing a node, we denote the minimum number of driver nodes of the damaged network as N'_D . Then we classify the node by comparing N_D and N'_D . A node is critical if $N'_D > N_D$ or $N'_D = N_D$, and ordinary if $N'_D < N_D$. For example, hindbrain GRN's $N_D = 13959$, the deleted node is critical if $N'_D = 13959$, ordinary if $N'_D = 13958$.

4.4 Biological enrichment score calculating analysis

Biological enrichment score calculating analysis is a method for enrichment analysis of gene sets, which is used to identify gene classes that are over-expressed in a large group of genes and may be related to disease phenotypes. This method uses statistical methods to identify significantly enriched or missing genomes. Microarray and proteomic results usually identify thousands of genes for analysis [38].

As described in fig.6, the oval part represents the gene set of the network under study, which is represented by S . here, it refers to the genes that need to be analyzed in the brain gene regulatory network (the network can be the forebrain GRN, hindbrain GRN and neuron associated cells cancer related GRN respectively; the gene can be the critical gene set or the ordinary gene set). The rectangular part represents some known functional gene databases, such as essential genes, cancer genes, conserved genes or other disease-related genes respectively, which is represented by DB . Overlap means that there will be some intersection of genetic data between them. Biological enrichment score calculating analysis is used to quantify the ratio of the critical genes after classification to the known functional gene databases.

Then the z score is calculated to estimate the enrichment by the equation:

$$z \text{ score} = \frac{(S_{DB} - \text{mean of } R_{DB})}{SD \text{ of } R_{DB}}$$

where S_{DB} is the number of genes from DB overlapping with S , R_{DB} is the number of genes from R_{DB} overlapping with random set of genes of same size as the number of critical genes (noted as N_F , N_H , N_C), and the mean of R_{DB} and SD of R_{DB} are computed from 1000 simulations of random sets. It's obvious that critical genes are rich in database DB if z score is bigger than 0.

Essential genes are obtained from DEG database, whose number is 8254 for human [39]. Cancer genes are in COSMIC database [40], and we collect 616 cancer genes for our work. Finally, the related disease genes are found in the Gene database (NCBI, <https://www.ncbi.nlm.nih.gov/>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Essential genes are obtained from DEG database, whose number is 8254 for human [39]. Cancer genes are in COSMIC database [40], and we collect 616 cancer genes for our work. The related disease genes are found in the Gene database (NCBI, <https://www.ncbi.nlm.nih.gov/>).

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work was supported by Grants from the National Natural Science Foundation of China (Grant No.61876047).

Author's contributions

Zhihua Chen and Siyuan Chen did genetic analysis, performed the statistical analysis and drafted the manuscript. Xiaoli Qiang designed this study and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to Xueming Liu for revision and editing English-writing....

Author details

¹Institute of Computing Science and Technology, Guangzhou University, 510006 Guangzhou, China. ²School of Automation, Huazhong University of Science and Technology, 430074 Wuhan, China.

References

- McGuire, S.: World cancer report 2014. geneva, switzerland: World health organization, international agency for research on cancer, who press, 2015 (2016)
- Cheng, Y., Morshed, R.A., Auffinger, B., Tobias, A.L., Lesniak, M.S.: Multifunctional nanoparticles for brain tumor imaging and therapy. *Advanced drug delivery reviews* **66**, 42–57 (2014)
- Mehta, M., Vogelbaum, M.A., Chang, S., Patel, N.: Neoplasms of the central nervous system. *Cancer: principles and practice of oncology* **9**, 1700–49 (2011)
- Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D., Forman, D., Bray, F.: Cancer incidence and mortality worldwide: larc cancerbase. *GLOBOCAN 2012 v10* **11** (2012)
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **68**(6), 394–424 (2018)
- Cai, S., Han, H.-J., Kohwi-Shigematsu, T.: Tissue-specific nuclear architecture and gene expression regulated by satb1. *Nature genetics* **34**(1), 42 (2003)
- Fass, D.M., Lewis, M.C., Ahmad, R., Szucs, M.J., Zhang, Q., Fleishman, M., Wang, D., Kim, M.J., Biag, J., Carr, S.A., et al.: Brain-specific deletion of *git1* impairs cognition and alters phosphorylation of synaptic protein networks implicated in schizophrenia susceptibility. *bioRxiv*, 290312 (2018)
- Kitano, H.: Systems biology: a brief overview. *science* **295**(5560), 1662–1664 (2002)
- Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z., Bergmann, S.: Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nature methods* **13**(4), 366 (2016)
- McKenzie, A.T., Wang, M., Hauberg, M.E., Fullard, J.F., Kozlenkov, A., Keenan, A., Hurd, Y.L., Dracheva, S., Casaccia, P., Roussos, P., et al.: Brain cell type specific gene expression and co-expression network architectures. *Scientific reports* **8** (2018)
- Barabási, A.-L., et al.: Scale-free networks: a decade and beyond. *Science* **325**(5939), 412 (2009)
- Börner, K., Sanyal, S., Vespignani, A.: Network science. *Annual review of information science and technology* **41**(1), 537–607 (2007)
- Liu, X., Stanley, H.E., Gao, J.: Breakdown of interdependent directed networks. *Proceedings of the National Academy of Sciences* **113**(5), 1138–1143 (2016)
- Gao, J., Liu, X., Li, D., Havlin, S.: Recent progress on the resilience of complex networks. *Energies* **8**(10), 12187–12210 (2015)
- Peng, H., Zhao, D., Liu, X., Gao, J.: Collective motion in a network of self-propelled agent systems. *PloS one* **10**(12) (2015)
- Liu, X., Pan, L., Stanley, H.E., Gao, J.: Multiple phase transitions in networks of directed networks. *Physical Review E* **99**(1), 012312 (2019)
- Ortmayr, K., Dubuis, S., Zampieri, M.: Metabolic profiling of cancer cells reveals genome-wide crosstalk between transcriptional regulators and metabolism. *Nature Communications* **10**(1) (2019)
- Malod-Dognin, N., Petschnigg, J., Windels, S.F.L., Povh, J., Hemmingway, H., Ketteler, R., Przulj, N.: Towards a data-integrated cell. *Nature Communications* **10**(1) (2019)
- Liu, X., Maiorino, E., Halu, A., Loscalzo, J., Gao, J., Sharma, A.: Robustness and Lethality in Multilayer Biological Molecular Networks. doi:[10.1101/818963](https://doi.org/10.1101/818963)
- Sporns, O., Tononi, G., Kötter, R.: The human connectome: a structural description of the human brain. *PLoS computational biology* **1**(4), 42 (2005)
- Barabási, A.-L., et al.: *Network Science*. Cambridge university press, ??? (2016)
- Liu, X., Pan, L.: Predicating candidate cancer-associated genes in the human signaling network using centrality. *Current Bioinformatics* **11**(1), 87–92 (2016)
- CHANG, L.: Local controllability of biological networks. PhD thesis (2015)
- Vinayagam, A., Gibson, T.E., Lee, H.-J., Yilmazel, B., Roesel, C., Hu, Y., Kwon, Y., Sharma, A., Liu, Y.-Y., Perrimon, N., et al.: Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets. *Proceedings of the National Academy of Sciences* **113**(18), 4976–4981 (2016)
- Liu, X., Pan, L.: Identifying driver nodes in the human signaling network using structural controllability analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* **12**(2), 467–472 (2015)
- Liu, X., Pan, L.: Controllability of the better chosen partial networks. *Physica A: Statistical Mechanics and its Applications* **456**, 120–127 (2016)
- Liu, Y.-Y., Slotine, J.-J., Barabási, A.-L.: Controllability of complex networks. *nature* **473**(7346), 167 (2011)
- Yan, G., Vértés, P.E., Towilson, E.K., Chew, Y.L., Walker, D.S., Schafer, W.R., Barabási, A.-L.: Network control principles predict neuron function in the *caenorhabditis elegans* connectome. *Nature* **550**(7677), 519 (2017)
- Liu, X., Pan, L., Stanley, H.E., Gao, J.: Controllability of giant connected components in a directed network. *Physical Review E* **95**(4), 042318 (2017)
- Liu, X., Pan, L.: Detection of driver metabolites in the human liver metabolic network using structural controllability analysis. *BMC systems biology* **8**(1), 51 (2014)
- du Plessis, L., Škunca, N., Dessimoz, C.: The what, where, how and why of gene ontology¹a primer for bioinformaticians. *Briefings in bioinformatics* **12**(6), 723–735 (2011)
- Kanehisa, M., Goto, S.: *Kegg: kyoto encyclopedia of genes and genomes*. *Nucleic acids research* **28**(1), 27–30 (2000)
- Slotine, J.-J.E., Li, W., et al.: *Applied Nonlinear Control* vol. 199. Prentice-hall Englewood Cliffs, NJ, ??? (1991)
- Lombardi, A., Hörnquist, M.: Controllability analysis of networks. *Physical Review E* **75**(5), 056110 (2007)
- Kalman, R.E.: Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics, Series A: Control* **1**(2), 152–192 (1963)
- Liu, X., Peng, H., Gao, J.: Vulnerability and controllability of networks of networks. *Chaos, Solitons & Fractals* **80**, 125–138 (2015)
- Luenberger, D.G.: *Introduction to Dynamic Systems: Theory, Models, and Applications* vol. 1. Wiley New York, ??? (1979)
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**(43), 15545–15550 (2005)
- Zhang, R., Ou, H.-Y., Zhang, C.-T.: Deg: a database of essential genes. *Nucleic acids research* **32**(suppl.1), 271–272 (2004)
- Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., Stratton, M.R.: A census of human cancer genes. *Nature Reviews Cancer* **4**(3), 177 (2004)

Figures

Tables

appendix

Supplementary Information

Table 1: Classification By driver nodes

Network	Number of nodes		
	critical	ordinary	total
Forebrain GRN	642(4.45%)	13793	14435
Hindbrain GRN	643(4.40%)	13958	14601
Neuron associated cells cancer GRN	643(4.20%)	14677	15320

Table 2: Topological analysis of each gene in human brain network

	Forebrain GRN		Hindbrain GRN		Neuron associated cells cancer GRN	
	critical	ordinary	critical	ordinary	critical	ordinary
D	364.402	15.354	375.426	15.476	425.121	16.319
DI	17.257	15.355	19.737	15.476	26.313	16.318
DO	347.145	0	355.689	0	398.809	0
B	25212.564	0	25862.054	0	26124.840	0
CA	2.912e-05	2.779e-05	2.924e-05	2.758e-05	2.872e-05	2.661e-05
CI	4.943e-09	5.008e-09	4.848e-09	4.898e-09	4.407e-09	4.439e-09
CO	1.987e-07	4.800e-09	2.449e-07	4.690e-09	3.012e-07	4.260e-09
CC	0.029	0.279	0.034	0.275	0.040	0.297

Table 3: The significance test of *CO*

Network	Mann Whitney U test	
	P-value	W
Forebrain GRN	2.2e-16	8696500
Hindbrain GRN	2.2e-16	8870300
Neuron associated cells cancer GRN	2.2e-16	9356600

Table 4: KEGG Pathway analysis on critical genes

KEGG Pathway	Gene Number	P value	Genes
hsa05202:Transcriptional misregulation in cancer	45(7.56%)	1.58e-32	LMO2,TFE3,PPARG,...
hsa05166:HTLV-I infection	33(5.55%)	1.32e-13	STAT5A,SPI1,ELK1,...
hsa05200:Pathways in cancer	40(6.72%)	5.29e-13	PPARG,SPI1,FOXO1,...
hsa04550:Signaling pathways regulating pluripotency of stem cells	24(4.03%)	1.89e-12	NANOG,HNF1A,OTX1,...
hsa05203:Viral carcinogenesis	24(4.03%)	4.99e-09	EGR3,EGR2,SP100,...
hsa05030:Cocaine addiction	9(1.51%)	4.29e-05	ATF4,CREB3,RELA,...
hsa04390:Hippo signaling pathway	15(2.52%)	5.27e-05	TCF7,SOX2,SMAD4,...
hsa04919:Thyroid hormone signaling pathway	13(2.18%)	5.80e-05	THRA,THRB,RXRB,...
hsa04022:cGMP-PKG signaling pathway	14(2.35%)	3.25e-04	MEF2D,MEF2B,ATF4,...
hsa04668:TNF signaling pathway	10(1.68%)	0.00233	ATF4,CEBPB,CREB3,...
hsa04152:AMPK signaling pathway	10(1.68%)	0.00592	SREBF1,HNF4A,CREB3,...
hsa04068:FoxO signaling pathway	10(1.68%)	0.01020	EP300,FOXG1,SMAD4,...
hsa04310:Wnt signaling pathway	10(1.68%)	0.01222	TCF7,EP300,TP53,...
hsa04110:Cell cycle	9(1.51%)	0.01882	E2F4,EP300,TP53,...

Table 5: The Edge weight ratios of three GRN networks

Network	Number of edges	
	all	edge weight (0~0.10)
Forebrain GRN	2662324	2591577(97.3%)
Hindbrain GRN	2606819	2533623(97.2%)
Neuron associated cells cancer GRN	2654287	2565486(96.7%)

Figure 1: Characterizing the controllability of human brain-specific gene regulatory network and topological analysis. In the figure, hindbrain GRN is hindbrain gene regulatory network, forebrain GRN is forebrain gene regulatory network, neuron associated cells cancer related GRN is neural gene regulatory network associated cells cancer. (Table II)The values in table II is the average of those topological characteristics that consist of out-degree DO , in-degree DI , degree D , betweenness B , closeness CA , in-closeness CI , out-closeness CO , betweenness B and clustering coefficient CC . (a)In-degree distribution of hindbrain GRN. (b)Out-degree distribution of hindbrain GRN. (c)Average in-degree for ordinary and critical nodes in hindbrain GRN. (d)Average out-degree for ordinary and critical nodes in hindbrain GRN.

Figure 2: Biological enrichment score calculating analysis of hindbrain GRN. (a) Enrichment analysis of essential genes. Numbers of essential genes overlapping with critical and ordinary nodes are shown in red arrows. (b) Enrichment analysis of cancer genes. Numbers of cancer genes overlapping with critical and ordinary nodes are shown in red arrows. (c) Biological enrichment analysis of neuron associated cells cancer genes. Numbers of these genes overlapping with critical and ordinary nodes are shown in red arrows.

Figure 3: Go analysis on critical genes. (a)CC. (b)MF. (c)BP.

Figure 4: Robustness analysis of node classification. (a)Plot showing the fraction of critical nodes in new networks that overlaps with the original hindbrain GRN, the new networks are obtained by deleting edges from original hindbrain gene regulatory networks. (b)The comparison of three networks in case of deleting edges.

Figure 5: Robustness analysis of node classification.

Here we noted the three strategies by the following ways: Firstly, deleting nodes randomly in the whole hindbrain GRN (noted as *randomly - all*). Secondly, deleting critical nodes randomly (noted as *critical*). Lastly deleting the same number of nodes as the second strategy in the whole hind-brain GRN, and which nodes are deleted is randomly (noted as *randomly - critical - all*). (a) The change of giant strongly connected component (GSCC) in hindbrain GRN with three methods that are different in the way of deleting critical nodes. (b) The change of giant weakly connected component (GWCC) in hindbrain GRN with three methods that are different in the way of deleting critical nodes. (c and d) Comparison in the way of deleting nodes randomly in the whole brain-specific GRN respectively. (e and f) Comparison in the way of *critical*. (g and h) Comparison in the way of *randomly - critical - all*.

Figure 6: The method of Biological enrichment analysis.

The circle represents genes of the network which represents forebrain GRN, hindbrain GRN and neuron associated cells cancer related GRN respectively. And the rectangle represents the functional gene database which is on behalf of essential genes, conserved genes, cancer genes or the related genes respectively.

Figure S1: Characterizing the controllability of human brain-specific gene regulatory network and topological analysis. (a) In-degree distribution of forebrain GRN. (b) Out-degree distribution of forebrain GRN. (c) Average in-degree for ordinary and critical nodes in forebrain GRN. (d) Average out-degree for ordinary and critical nodes in forebrain GRN. (e) In-degree distribution of neuron associated cells cancer related GRN. (f) Out-degree distribution of neuron associated cells cancer related GRN. (g) Average in-degree for ordinary and critical nodes in neuron associated cells cancer related GRN. (h) Average out-degree for ordinary and critical nodes in neuron associated cells cancer related GRN.

Figure S2: Biological enrichment of forebrain GRN. (a) Biological enrichment analysis of essential genes. Numbers of essential genes overlapping with critical and ordinary nodes are shown in red arrows. (b) Biological enrichment of cancer genes. Numbers of cancer genes overlapping with critical and ordinary nodes are shown in red arrows. (c) Biological enrichment analysis of neuron associated cell-s cancer genes. Numbers of these genes overlapping with critical and ordinary nodes are shown in red arrows.

Figure S3: Biological enrichment analysis of neuron associated cells cancer related GRN. (a) Enrichment analysis of essential genes. Numbers of essential genes overlapping with critical and ordinary nodes are shown in red arrows. (b) Enrichment analysis of cancer genes. Numbers of cancer genes overlapping with critical and ordinary nodes are shown in red arrows. (c) Enrichment analysis of neuron associated cells cancer genes. Numbers of these genes overlapping with critical and ordinary nodes are shown in red arrows.

Figure S4: Robustness analysis of node classification. (a) Plot showing the fraction of critical nodes in new networks that overlaps with the original forebrain gene regulatory network, the new networks are obtained by deleting edges from original hindbrain gene regulatory networks. (b) Plot showing the fraction of critical nodes in new networks that overlaps with the original neuron associated cells cancer related gene regulatory network, the new networks are obtained by deleting edges from original hindbrain gene regulatory networks.

Figure S5: Robustness analysis of node classification. (a) The change of giant strongly connected component(GSCC) in forebrain GRN with three methods that are different in the way of deleting critical nodes. (b) The change of giant weakly connected component(GWCC) in forebrain GRN with three methods that are different in the way of deleting critical nodes. (c) The change of giant strongly connected component(GSCC) in neuron associated cells cancer related GRN with three methods that are different in the way of deleting critical nodes. (d) The change of giant weakly connected component(GWCC) in neuron associated cells cancer related GRN with three methods that are different in the way of deleting critical nodes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [graphicabstract.pdf](#)