System biology of Activity-regulated cytoskeleton-associated protein (Arc/Arg 3.1)

Nguyen Xuan Tung
(Dated: June 28, 2022)

The activity-regulated cytoskeletal (Arc) is a complex protein resulting from interactions among genetic and other factors. The mechanism of Arc development remains unknown. This study proposes that protein–protein interactions (PPIs) and pathway interactions network, particularly those among proteins encoded by casual or susceptibility genes. The component of our constructed PPI network comprised 61 nodes with 505 edges, and pathway network contains 120 nodes and 782 edges. By using parameters of large degree (k), high betweenness centrality (BC), and eccentricity we find hub nodes for both networks and the most important protein contributed to the network.

INTRODUCTION

ARC protein is released from neurons in extracellular vesicles that mediate the transfer of ARC mRNA into new target cells, where ARC mRNA can undergo activity-dependent translation. ARC capsids are endocytosed and are able to transfer ARC mRNA into the cytoplasm of neurons. Acts as a key regulator of synaptic plasticity: required for protein synthesis-dependent forms of long-term potentiation (LTP) and depression (LTD) and for the formation of long-term memory[1].

However, the molecular basis of Arc protein function remains a puzzle. To archive its functions, proteins interact with each other. This interaction between proteins can be represented in the form of networks (FIG. 1). The interaction network between proteins is one of the significant way to reveal the function of proteins the relationship between them. To identify significant proteins associated with disease, a statistical approach is necessary and is an efficient solution against the experimental constraint. One of the solution is constructing and analyzing network topology of protein-protein interaction. This can be done because of the availability of protein-protein interaction data in large numbers and the advance tools. In this paper, we will discuss about the biophysical and structural properties of Arc and the role of protein involved in the Arc PPI network through hub nodes. In protein-protein interaction (PPI) networks hubs represent proteins with a large number of interactions, called hub proteins[5].

The biological pathway can tell us a lot about humans diseases, however, the biological pathways of Arc protein still remain unclear. In the second part of the project, we discuss about the interaction of pathways of Arc protein. The pathway interaction network (FIG. 3) is constructed by integrating protein-protein interactions and KEGG pathway information.

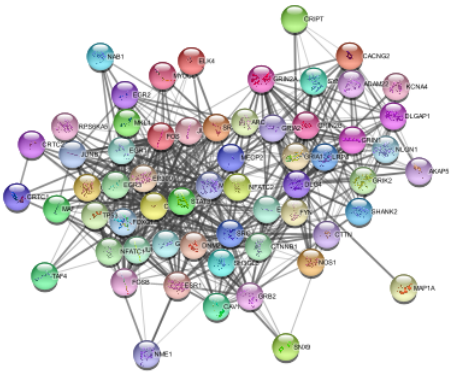

FIG. 1: The Arc PPI network. The network contains 61 nodes with 505 edges. Network was directly sent to Cytoscape.

METHODS

Collecting Data

This project used both public domain databases: STRING and UniProt. STRING is a database of known and predicted protein-protein interactions. String database includes manually curated protein interactions and uses confidence scoring to give an estimate of how likely an association is to occur[2]. In our study, the global PPI dataset containing 505 interactions among 61 unique human proteins was obtained with medium confidence score 0.4.

The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data. The public paper in Uniprot are used for reference in this project.

Preprocessing Data

We use the Search Tool for the Retrieval of Interacting Genes/Proteins database (STRING) to construct the PPI network associated with Arc. Given a name of the proteins as input (Arc), STRING can search for their neighbor interactors, the proteins that have direct interactions with the inputted proteins; then STRING can generate the PPI network consisting of all these proteins and all the interactions between them. Based on the seed proteins as input, we first constructed the PPI network (FIG. 1) with Cytoscape containing the seed proteins and their neighbors[3].

The topological analysis of the network

To evaluate the nodes in the PPI networks, we adopted several topological measures including degree (k), between centrality (BC), eccentricity, closeness centrality (CC), and clustering coefficient(ClusteringCoeff). The first two measures, degree (k) and BC, are often used for detecting the hub or bottleneck in a network. Degree (k) of a node is defined as the number of edges linked to it. A node with high degree (k) denotes a hub having many neighbors. BC of a node is the proportion of the number of shortest paths passing through it to the number of all the shortest paths in the network, quantifying how often a node acts as a bridge along the shortest paths between two other nodes. A node with high BC has great influence on what flows in the network and has more control over the network. It can represent the bottleneck in the network[3].

Eccentricity and CC of a node are the measures of centrality in the network, defined as the maximum distance from the node to all other nodes and the inverse of the average length of the shortest paths between the node and all other nodes, respectively. A node with lower eccentricity or higher CC is closer to the other nodes and more central in the network. Moreover, the maximum eccentricity is the diameter of a network; the minimum eccentricity is the radius of a network. The center of a network is the set of nodes of eccentricity equal to the radius. Clustering coefficient of a node is the proportion of the edges to all the possible edges within its neighbors, quantifying the closeness among its neighbors, and evaluating how small its neighbors' world is. A node with higher clustering coefficient has its neighbors closer to one another, and the world of its neighbors is smaller[3].

Pathway analysis under PPI data

A pathway interaction network is constructed base on protein protein interaction and cellular pathway. In order to find the pathway of the protein interaction network of Arc, we use Cytoscape plugin named StringApp to generate the enrichment map. StringApp can retrieve functional enrichment for Gene Ontology terms, KEGG, Reactome and Wiki Pathways, DISEASES, COMPARTMENTS, TISSUES, and protein domains at a user-specified significance threshold and show the results in a new table in the Table Panel. In this project, we only select the KEGG pathway to analysis[4].

FIG. 2: The Arc PPI network. The network contains 61 nodes with 505 edges. Data was downloaded from String database in TSV file and was generated by Cytoscape.

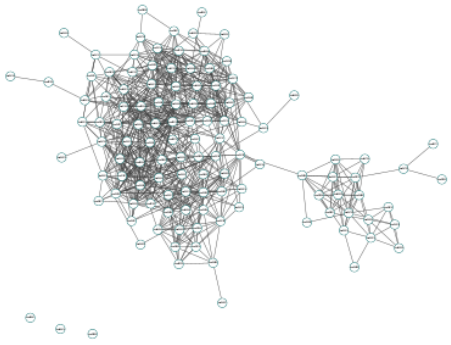

FIG. 3: Pathway interaction network. Nodes were on behalf of pathways and edges stood for the interaction between any two pathways.

RESULTS

Protein interaction networks

The component (FIG. 2) of the PPI network generated by STRING consisted of 61 nodes and 505 edges. The results of the topological analyses of each node were list in Table 1, including degree, BC, eccentricity, CC, clustering coefficient, etc. The number of edges is larger than the expected for random network of the same size significantly; the nodes were more connected than randomly.

The results of the topological analyses of each node in Table 1 showed that CREB1 was a hub (with the largest degree k=40) and bottleneck (with the second highest BC=0.103257) in the PPI network; DLG4 was a bottleneck (with the highest BC=0.1111) and also a hub (with higher degree than average). Another node can be considered is FOS with the second largest degree k and third largest BC (38 and 0.0722 respectively).

By using 3 sample KEGG pathways: amphetamine, cocain, and nicotine addiction, we find the proteins involved in the pathways. TABLE III shows us name of the protein associated with the pathways together with its description.

name	BC	CC	ClusteringCoeff	Degree	Eccentricity
DLG4	0.111	0.681	0.395	32	2
CREB1	0.103	0.75	0.397	40	2
FOS	0.072	0.731	0.425	38	2
CTNNB1	0.069	0.681	0.441	32	2
JUN	0.039	0.674	0.483	31	2
GRIN2B	0.037	0.645	0.481	27	2
EP300	0.037	0.625	0.477	29	3
ARC	0.036	0.606	0.533	21	2
SRC	0.032	0.645	0.524	27	2

TABLE I: Table contain protein with highest degree and BC in the protein intercation network.

Pathway interaction network

To identify the relevant pathways changed in Arc interaction network, we use a statistical approach on pathway level. The significance analysis of pathway is based on the protein-protein interaction database. The component (FIG. 3) of the pathway network generated consisted of 120 nodes and 782 edges. The results of the topological analyses of each node were list in Table 2, including degree, BC, eccentricity, CC, clustering coefficient, etc (only degree is shown in this paper, full table can be found in attached excel file). The impact analysis method yielded many significant pathways containing Choline metabolism in cancer, Colorectal cancer, Endocrine resistance, Apoptosis, Prostate cancer and so on (TABLE II). TABLE II shows few numbers of significant pathways with largest degree (k).

From the result of the analysis (TABLE II), we can find that the hub pathway and the center of the network is Choline metabolism in cancer (with highest degree k = 31) together with second and third largest degree respectively, Colorectal cancer, Endocrine resistance (k = 29), and Apoptosis, Prostate cancer (k = 28).

PathwayID	name	degree
hsa05231	Choline metabolism in cancer	31
hsa05210	Colorectal cancer	29
hsa1522	Endocrine resistance	29
hsa04210	Apoptosis	28
hsa05215	Prostate cancer	28
hsa05165	Human papillomavirus infection	28
hsa05224	Breast cancer	28
hsa05214	Glioma	27
hsa05220	Chronic myeloid leukemia	27
hsa05133	Pertussis	24

TABLE II: Significant pathways

DISCUSSION

Many studies has been conducted on Arc protein, however, the mechanism of it remains unclear. The proteins encoded by susceptibility genes may determine an individual's susceptibility to Arc through their encoded PPIs. Here, we study the potential key proteins through topological analysis. We use degree (k) and BC as the main parameters for evaluating the nodes in the PPI network.

The first part of this paper is to investigate the PPI network of Arc protein and explore the contributions of the proteins encoded by the susceptibility genes associated with Arc. Initially, a total of 61 proteins were included in our network. In the PPI network, there are 10 proteins (CREB1, FOS, JUN, ARC, FOSB, GRIA2, GRIA1, GRIN2A, GRIN2B, GRIN1) involved in the Amphetamine addiction pathway, 8 proteins (JUN, CREB1, FOSB, GRIA2, GRIN2A, GRIN1, GRIN2B, DLG4) involved in the Cocain addiction and 5 proteins (GRIA2, GRIA1, GRIN2A, GRIN1, GRIN2B) involved in the in Nicotine addiction in KEGG pathway. In addition, KEGG pathway was shown in FIG. 4, FIG. 5 and FIG. 6.

The most important node among them all is FOS with the second largest degree and third largest BC. It is one of the center node in the PPI network and has been reported in amphetamine addiction. Another important nodes can be considered are GRIA2, GRIN2A, and GRIN2B which have been reported in all three analysis amphetamine, cocain, and nicotine addiction KEGG pathways. These proteins are also can be considered as backbone protein

since they are in top 10% protein which has highest degree (k).

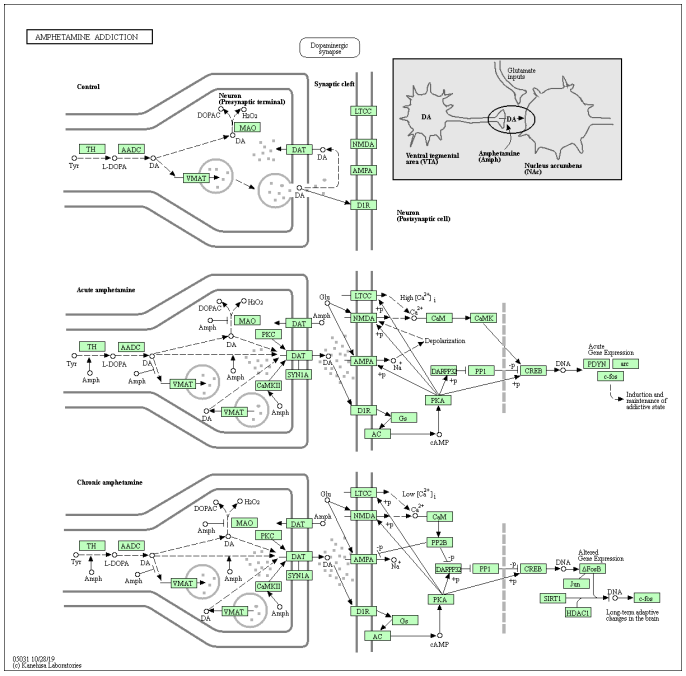

FIG. 4: Amphetamine addiction KEGG pathway[7]

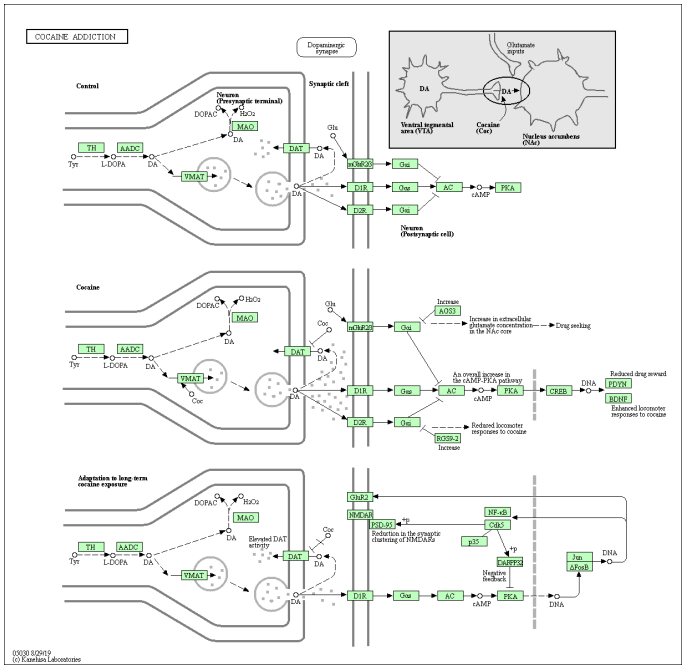

FIG. 5: Cocain addiction KEGG pathway[8]

In this second part of the project, we construct the pathway interaction network which described the crosstalks among pathways. In our network, several significant pathways were identified crosstalk with each other. Espcially, hsa05231 (Choline metabolism in cancer), hsa05210 (Colorectal cancer), hsa1522 (Endocrine resistance), hsa04210 (Apoptosis), and hsa05215 (Prostate cancer) are hub nodes, which suggesting that these pathways play an important role in the development of Arc protein.

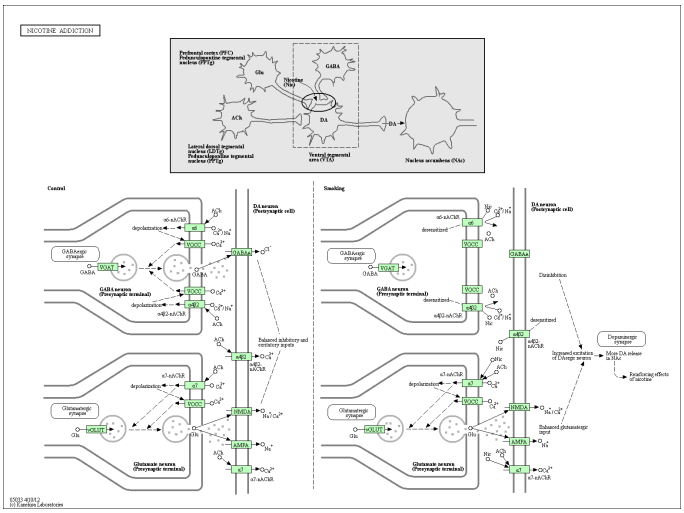

FIG. 6: Nicotine addiction KEGG pathway[9]

CONCLUSION

Our project find that Arc protein development through an integrated PPI network with CREB1 as one of its center which contribute in amphetamine and cocain addiction pathway. FOS is also involved in the development of amphetamine, cocain and nicotine addiction. It is shown that CREB1 and FOS are the keys protein for studying the formulation of protein interaction network of Arc.

In another hand, by constructing the pathway interaction network, we find that hsa05231 (Choline metabolism in cancer), hsa05210 (Colorectal cancer), hsa1522 (Endocrine resistance), hsa04210 (Apoptosis), and hsa05215 (Prostate cancer) are the hub nodes of the network and play an important roles in developing the Arc protein network. The pathway Choline metabolism in cancer was regarded as the seed pathway since it is the center of the network (with lowest eccentricity and sixth largest CC) which can be concluded that it contributes a crucial role in the network. However, the role of them in Arc still remains unclear and requires more studies.

[1] Uniprot, Activity-regulated cytoskeleton-associated protein · Homo sapiens (Human) · Gene: ARC (KIAA0278) , accessed 10 June 2022, https://www.uniprot.org/uniprotkb/Q7LC44/entry

[2] Wufeng Fan, Yuhan Zhou, and Hao Li (2017), Pathway In-

teraction Network Analysis Identifies Dysregulated Pathways in Human Monocytes Infected by Listeria monocytogenes
[3] Shaw-Ji Chen, Ding-Lieh Liao, Chia-Hsiang Chen, Tse-Yi Wang Kuang-Chi Chen (2019), Construction and Analysis of Protein-Protein Interaction Network of Heroin Use Disorder
[4] Cytoscape App Store, stringApp, accessed 11 June 2022, https://apps.cytoscape.org/apps/stringapp
[5] Zhang, J. (2013). Hub. In: Dubitzky, W., Wolkenhauer, O., Cho, KH., Yokota, H. (eds) Encyclopedia of Systems Biology. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9863-7-474
[6] String DB, accessed 10 June 2022, https://string-db.org/cgi/network?taskId=bUQCPpMqD4JmsessionId=buQPyb5mGgNG
[7] KEGG, accessed 11 June 2022, https://www.genome.jp/pathway/hsa05031
[8] KEGG, accessed 11 June 2022, https://www.genome.jp/pathway/hsa05030
[9] KEGG, accessed 11 June 2022, https://www.genome.jp/pathway/hsa05033

name	KEGG	Description
CREB1	Amphetamine	cAMP responsive element protein1
FOS	Amphetamine	FBJ murine osteosarcoma homolog
JUN	Amphetamine	Jun proto-oncogene
ARC	Amphetamine	Activity-regulated c-a protein
GRIN1	Amphetamine	Glutamate receptor ionotropic
FOSB	Amphetamine	Fosb proto-oncogene, ap-1
GRIA1	Amphetamine	GI receptor ampa subunit1
GRIA2	Amphetamine	GI receptor ampa subunit2
GRIN2A	Amphetamine	GR ionotropic, NMDA 2A
GRIN2B	Amphetamine	GR ionotropic, NMDA 2B
JUN	Cocain	Jun proto-oncogene
FOSB	Cocain	Fosb proto-oncogene, ap-1
CREB1	Cocain	cAMP responsive element binding protein1
GRIA2	Cocain	GI receptor ampa type subunit2
GRIN2A	Cocain	GR ionotropic, NMDA 2A
GRIN1	Cocain	GR ionotropic, NMDA 1
GRIN2B	Cocain	GR ionotropic, NMDA 2B
DLG4	Cocain	Disks large homolog 4
GRIA2	Nicotine	GI receptor ampa type subunit 2
GRIA1	Nicotine	GI receptor ampa type subunit 1
GRIN2A	Nicotine	GR ionotropic, NMDA 2A
GRIN1	Nicotine	GR ionotropic, NMDA 1
GRIN2B	Nicotine	GR ionotropic, NMDA 2B

TABLE III: Protein in PPI network involved in KEGG pathway of amphetamine, cocain and nicotine addiction.