**Core proteins in exocytosis and endocytosis**

Abstract

Introduction

Within and around cells, materials are constantly being shipped one way or another across membranes. At any given moment, a particular substance may transport from one side of the membrane to the other side. This transport is very important to cellular function, and the disorder of transport system in the cell causes many diseases, such as Wilson Disease. Cellular transport contains many different kinds of process, such as diffusion, active transport and electron carriers However, exocytosis and endocytosis are the most essential part of the cellular transport system.

Exocytosis is the durable, energy-consuming process by which a cell directs the contents of secretory vesicles out of the cell membrane and into the extracellular space. The membrane-bound vesicles contain soluble proteins will be secreted to extracellular environment, membrane proteins and lipids will be sent to the cell membrane. There are basically five steps for the exocytosis, which are Vesicle trafficking, Vesicle tethering, Vesicle docking, Vesicle priming and Vesicle fusion.

Endocytosis is an energy-using process by which cells absorb molecules (such as proteins) by engulfing them. It is used by all cells of the body because most substances large polar molecules that cannot pass through the cell membrane. Endocytosis is more widely used than exocytosis in the cells. Endocytosis can be divided into four different subtypes: clathrin-mediated endocytosis, non-clathrin-mediated endocytosis, macropinocytosis and Phagocytosis. Clathrin-mediated and non-clathrin-mediated endocytosis are involved in the receptors on the cell membrane pathway. Macropinocytosis is the process of forming a pocket in the cellular membrane which then fills with extracellular fluid and molecules. Phagocytosis is the process by which cells bind and internalize particulate matter larger than around 0.75 µm in diameter via the usage of vesicles.

Exocytosis and endocytosis can be involved in the same cell in the same pathway. One example is in the synaptic signaling pathway. Synapses are very important for the human body, especially for the signal transmission in the neuron system. Exocytosis and endocytosis are the most important processes in the vesicles containing neuron-transmitters release in the pre-synapse, which are responsible for the signal transmission in the synaptic junction. Specifically, during the neuron-transmitters release process, the first step is the fusion of the vesicle to the membrane in the pre-synapse, while the second step is the re-sealing of the vesicle after releasing the signaling molecule (is this phrase right?), and the third and final step is the return of the vesicle to the cytoplasm. Thus, endocytosis occurs after exocytosis in the same pathway. Therefore, exocytosis and endocytosis are tightly related to each other. However, the full extent of this relationship is not understood. We therefore propose to utilize new methods to more fully investigate this relationship.

Exocytosis and endocytosis can take place in many different parts of cell, such as the endoplasmic reticulum, the Golgi apparatus, the endosome, the lysosome and the cell membrane. However, it is not known what the most important location for the connection between exocytosis and endocytosis is in the cell. Therefore, we also would like to investigate the most important part of the cell for the regulation between exocytosis and endocytosis.

The primary premise of this study is that there are a set of core driver proteins that are highly influential in the processes of exocytosis and endocytosis. These core driver proteins should be found in the pathways of both exocytosis and endocytosis, thus showing that their importance in both pathways. Similarly, it should be possible to identify these proteins given a protein-protein interaction (PPI) network by quantifying the strength of the protein’s connections.

Methods

Results

**Data Generation**

In order to generate our datasets, we utilized the human BIOGRID database.12 First, we searched the entire GeneOntology database to find proteins annotated to have function in exocytosis and endocytosis.3 Then, we identified the interactions of these proteins from BIOGRID and used those proteins and all proteins that interact with the original set of proteins to generate the full exocytosis and endocytosis network (full). Next, we filtered this network to include only the exocytosis PPIs (exocytosis). Finally, we filtered the full network to include only the endocytosis PPIs (endocytosis).

**Endocytosis network**

We first analyzed the endocytosis PPI network, as identified by BIOGRID and Gene Ontology. Both the FPC score and core/periphery analysis rankings were utilized in order to rank the “structurally dominant” nodes. The top five proteins, as determined by the FPC score, are as follows: TNK2 (activation of CDC42), CDC42 (GTPase of the Rho family), SRC (cell growth), GRB2 (EGF receptor), EGFR (EGF receptor). Figure 1 shows the full endocytosis network with the top FPC scoring proteins shown in red. Similar results to the FPC were found from the core/periphery analysis scoring.

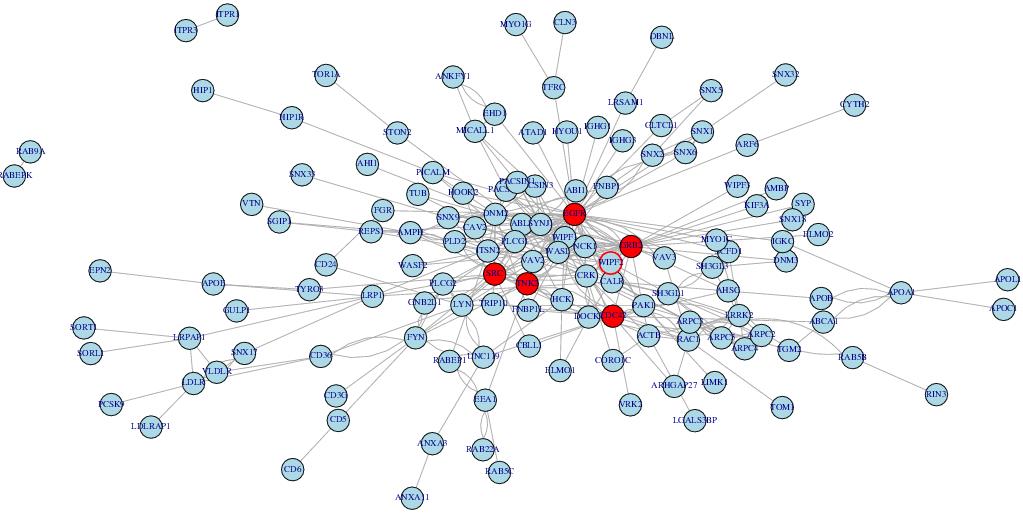


Figure 1: Endocytosis network with top 5 nodes as identified by FPC score marked in red

We notice the top 5 FPC score proteins in the endocytosis network are all located in the membrane. And all the proteins are associated with membrane signal receptor, such as EGF. These results are somewhat different from our expectation.

**Exocytosis network**

Next, we set up the exocytosis network from the Gene ontology and BioGRID database, then we analyze the database by both methods. Again, we rank the top 5 FPC scoring proteins in the exocytosis network. The top five proteins from this method are RPH3AL (Rab GTPase effector), STXBP1 (Syntaxin-binding protein 1), STX1A (vesicle fusion process), SYTL4 (Synaptotagmin-like protein 4), RAB27B (vesicle fusion process). These five nodes are shown in red in Figure 2, which shows the entire exocytosis protein network, as derived from Gene Ontology and BIOGRID. Again, similar results are found from the core/periphery analysis procedure.

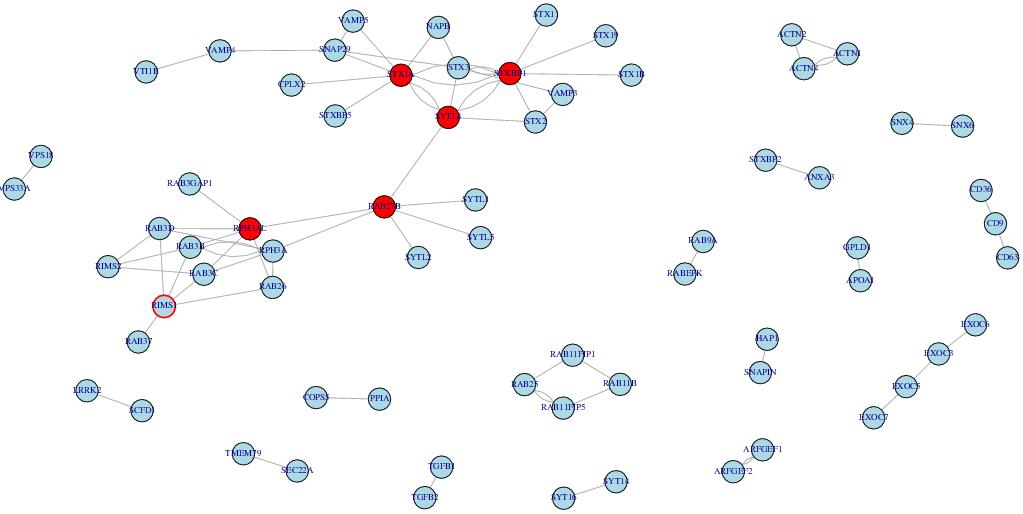


Figure 2: Exocytosis network

We notice the top 5 proteins are all belong to SNARE complex, which is responsible for the vesicle fusion to the cell membrane.

**Exocytosis or Endocytosis network**

We want to find the overlap of the endocytosis and endocytosis network. However, the identified overlap network is very small, and attempting to find a core for this network would be useless. Another issue is that the proteins derived from Gene Ontology is very limited, as there are many famous regulatory proteins, such as dynamin and sydapin, that are absent from either one or both endocytosis and exocytosis. Thus, we broadened our database to explore proteins implicated in either exocytosis or endocytosis. Surprisingly, the top scoring proteins are most similar to the endocytosis network, possibly showing the effects of large degree nodes on these methods.

We set up the exocytosis or endocytosis network from the Gene ontology and BioGRID database, and then we analyzed the network by our methods. Again, we rank the top scoring proteins in the full network. The top five proteins from the FPC scoring method are COPS5 (regulator of E3 ligase), LYN (association with receptor), SRC (cell growth), GRB2 (EGF receptor), EGFR (EGF receptor). Similar results are also seen from the core/periphery analysis, which is shown in Table (?).It is also worth noticing that COPS5 is a regulator of E3 ligase and located in ER, which is different from other proteins, as all the other proteins are located in the cellular membrane. From this point, we can say our method really work, because when we connect both exocytosis and endocytosis network, we can find some proteins can be easily omitted.

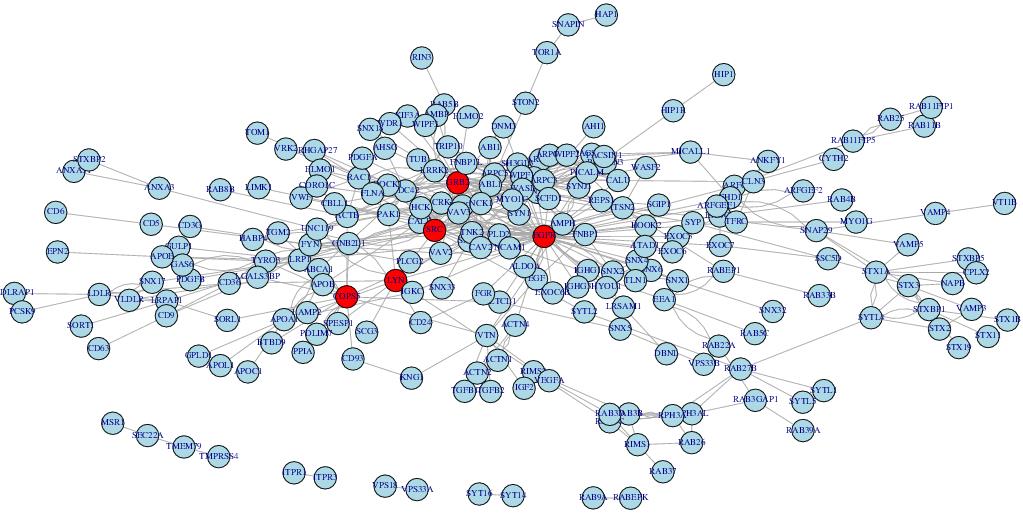


Figure 3: Exocytosis or endocytosis (full) network; top 5 FPC scored nodes are shown in red

**Vesicle network**

Since we find that Gene Ontology is incomplete for exocytosis and endocytosis, it is necessary to look for other data sources. Unfortunately, a full database consisting of exocytosis and endocytosis pathways and protein interactions is not available. However, it should be possible to utilize the vesicle network to show that Gene Ontology is unreliable. We set up out vesicle database from the Reactome Pathway Database and get all the protein-protein interactions from BioGRID . Then we analyzed the network by our methods. Again, we rank the top 5 FPC score proteins in the vesicle network. The top five proteins from the FPC scoring are ALB (binding to water, Ca(2+), Na(+), K(+)), CLTC (Clathrin heavy chain 1), ARRB1 (Arrestin beta 1), YWHAQ (mediate signal transduction), UBC (ubiquitin).

We can see the top 5 proteins in the vesicle network make more sense than the network derived from Gene ontology database, and that these five proteins are highly involved in the exocytosis and endocytosis. For example, UBC is ubiquitin being involved in all the proteins degradation, which is widely used in the cell activity.

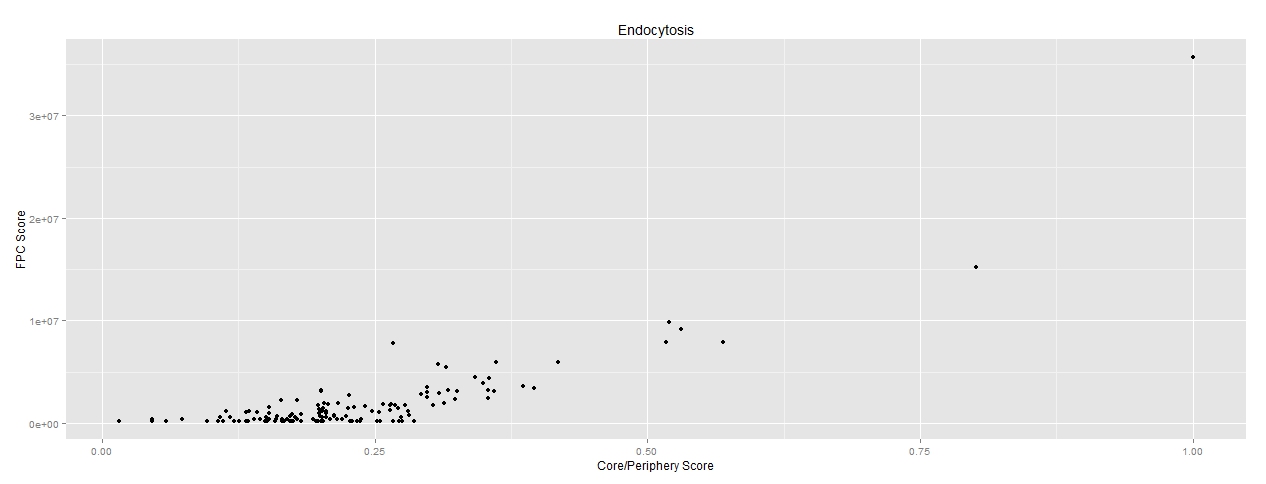
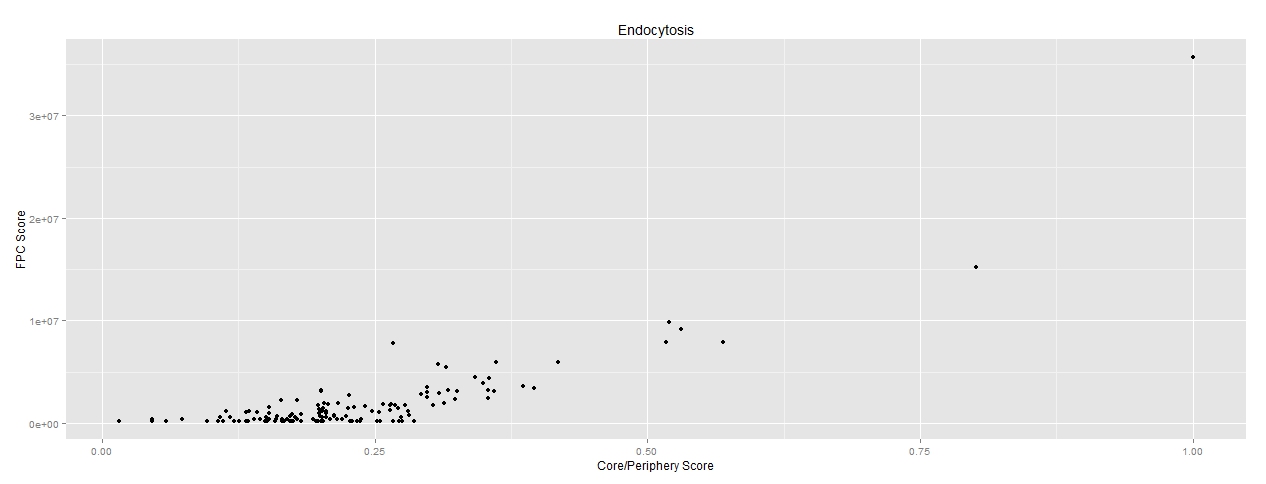
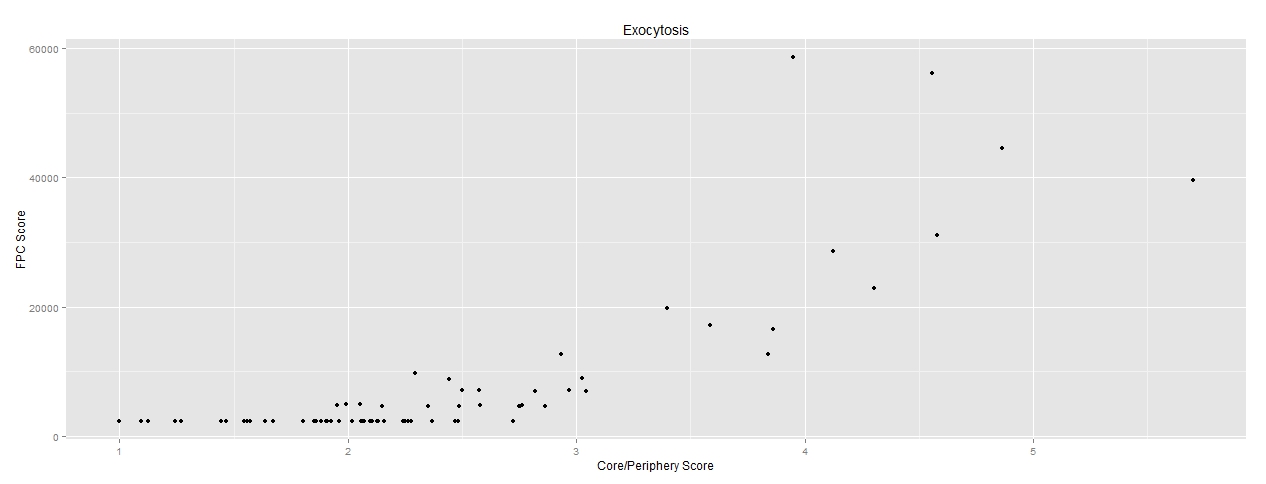
**Validation of Results**

We validated our results via two different methods. The first involved finding the correlation between the FPC score and the core/periphery analysis score. For each different network, we also show the relationship between the two scoring methods as well in Figures 4-6. The correlations between the two scoring methods are shown in Table 1. It is worth noting that there are significant correlations between the two measures

The second validation method involved comparing the resampled scores against the scores of the complete network. For each different network, the correlation was recorded for the resampled scores; these correlations are also shown in Table 1. Judging from the correlations, it seems like the methods were implemented correctly and the results are likely to be biologically valid as there are very high correlations for all of the completed correlations.[[1]](#endnote-1)

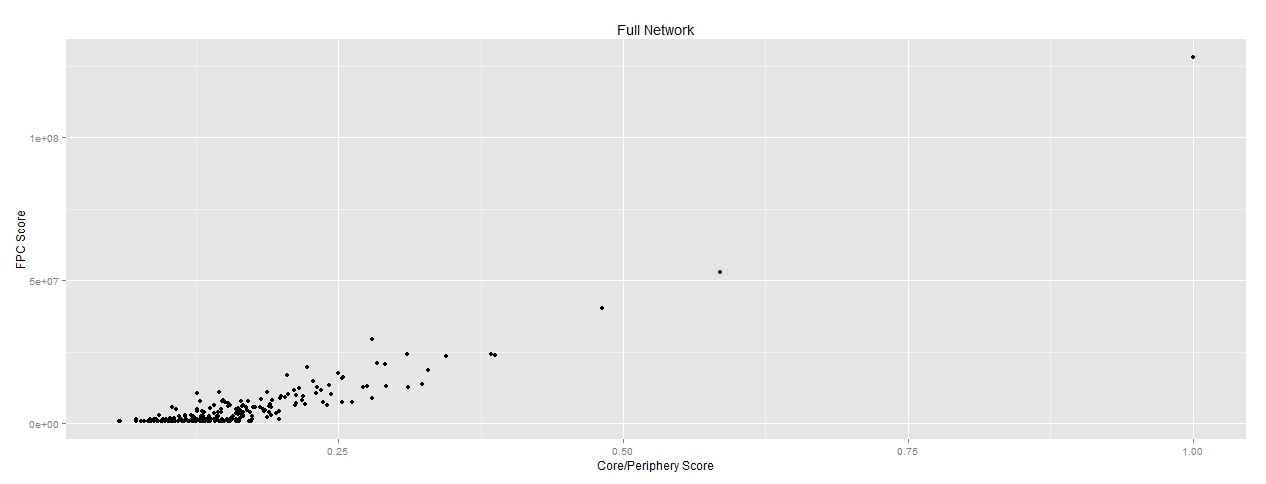
Table 3: Correlation coefficients showing the association for each different method and each network. Also see Footnote 2.

|  |  |  |
| --- | --- | --- |
| Correlation Coefficient | Comparison | Network |
| 0.924508103 | FPC, C/P | Full |
| 0.852576907 | FPC, C/P | Endocytosis |
| 0.820836959 | FPC, C/P | Exocytosis |
| 0.999999322 | FPC, FPC Validation | Full |
| 0.999997786 | FPC, FPC Validation | Endocytosis |
| 0.999994891 | FPC, FPC Valation | Exocytosis |
| 0.810518922 | C/P, C/P Validation | Exocytosis |
| 0.693445 | C/P, C/P Validation | Endocytosis |

*Figure 5: Exocytosis Network FPC Score vs Core/Periphery Score*

*Figure 4: Endocytosis network FPC score vs Core/Periphery score*



*Figure 6: Full network FPC Score vs Core/Periphery Score*

1. Note: there is no correlation recorded between the full network core/periphery score and its validated score due to computational constraints. After running this validation for several days, it did not finish, showing the large runtime of this type of validation procedure, even while utilizing large amounts of computational resources. [↑](#endnote-ref-1)