CALCIUM SIGNALING SYSTEM IN MULTIPLE SPECIES: STRUCTURAL AND FUNCTIONAL ANALYSES OF PROTEIN INTERACTION NETWORKS

by

Senem Tiveci

B.S., Chemical Engineering, Boğaziçi University, 2002

M.S., Chemical Engineering, Boğaziçi University, 2004

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Chemical Engineering

Boğaziçi University

2011

to my precious, my mom…

ACKNOWLEDGEMENTS

First and foremost I owe my deepest gratitude to my thesis supervisor Prof. Kutlu Ülgen for her continuous patience, assistance, motivation and support. She believed in me more than I believed in myself. This thesis would not have been possible unless she did not give up on me. I also gratefully acknowledge my thesis co-supervisor Prof. Türkan Haliloğlu for advice, support and encouragement in finishing this dissertation. I would also like to thank to the other members of my thesis committee, Prof. Pemra Doruker, Prof. Dilek Kazan and Assist. Prof. A. Nevra Özer for their time devoted to read and comment on my thesis.

I would like to thank to all academic, administrative and technical staff of B.Ü. Department of Chemical Engineering for their help and support throughout my education. It is a pleasure for me to thank Prof. Betül Kırdar since she was the one who convinced me to become a chemical engineer in the first place. I especially thank to my wonderful friend, F. Betül Kavun (Özbayraktar) who is always there for me with her love, kindness, moral support and soothing conversations.

I am grateful to the Analytical Services Manager of Şişecam, Dr. Eşref Aydın for his endless understanding and support, giving me the time and space whenever I asked for. I will always be honoured for being an employee under his supervision.

My parents, relatives and friends deserve special mention for their inseparable support and prayers. My mother, Gülhan Bayrı, in the first place is the person whose dedication, love and persistent confidence in me have taken the load off my shoulder. Words fail me to express my love, my admiration and appreciation to her. I am nothing without you… And I am also thankful to my father, Aras Deveci, simply for being my beloved dad and for always being there with love and relieving support. Thanks and deepest prayers to my grandmother for raising me up, for loving me like her own daughter and for being able to make me grieve and smile at the same time. I miss you so much… Thanks to my uncle, Gökhan Bayrı, for his love, support and for being like an older brother

to me. Thanks to İpek, Tayfun, Göker, Emrah, Fidan, Tuğba, Pınar and all my lovely friends for the joy they brought to my life.

I would like to express my special thanks to M. Onur Delikanlı, whose magical love, understanding and continuous support turned my life into a dream which will go on ‘til death do us part…

Finally, I would like to thank everybody who was important to the successful realization of thesis, as well as expressing my apology that I could not mention personally one by one.

ABSTRACT

CALCIUM SIGNALING SYSTEM IN MULTIPLE SPECIES: STRUCTURAL AND FUNCTIONAL ANALYSES OF PROTEIN INTERACTION NETWORKS

Calcium is an important second messenger to have a key role in many processes occurring in the cell. Understanding the mechanisms of calcium signaling in multiple species i.e. yeast and worm may help to untangle calcium signaling pathways in human and suggest some drug targets to cure calcium related diseases. For these aims, the calcium signaling pathways in the aforementioned species using an algorithm that counts for physical interaction cascades based on gene ontology annotations. The resultant interaction networks are investigated in macroscopic and microscopic details to find out the important pivotal elements of these networks. By ontology enrichment study, it is found that all three networks are a part of transcription, and in more general, cell proliferation and growth. Using graph theoretical approaches, the networks are compared according to their topological characteristics, revealing that yeast and human calcium signaling networks are alike having almost the same degree and average path lengths. Calcium signaling in human brain shows more robust behaviour than the whole human brain interactome and the calcium signaling in human body in opposite of yeast and worm; implying that it has a distinct behaviour and duty to start and control brain and body functions. Worm, having a more imcomplete interactome compared to other two, shows unique topological properties. Modular segregation of the networks gives opportunity to functional annotation to those with unknown function. By a combination of neighborhood counting and module search methods, 17 yeast proteins are annotated with functional terms. Human and worm modules fail to qualify for such annotation. Finally, the search for diseases like Alzheimer's disease and Huntington disease caused by dysregulation of calcium signaling in human brain reveals the incompleteness of tissue specific protein interactome data, since the envisaged pathways for these diseases in literature and information gathered from various data sets fails to coincide.

ÖZET

CANLILARDA KALSİYUM SİNYAL MEKANİZMASI: PROTEİN ETKİLEŞİM AĞLARININ YAPISAL VE FONKSİYONEL ÖZELLİKLERİ

Kalsiyum, hücrede gerçekleşen birçok süreçte anahtar rol oynayan önemli bir ikincil taşıyıcıdır. Maya, solucan gibi canlılardaki kalsiyum sinyal mekanizmasını anlamak, insandaki kalsiyum sinyal yolizlerindeki bilinmeyen kısımları açıklığa kavuşturma ve kalsiyum ile ilgili hastalıkların tedavisine yönelik ilaç hedefleri bulma konusunda yardımcı olacaktır. Bu amaçla, bahsi geçen canlılardaki kalsiyum sinyal etkileşim ağlarını elde etmek için, fiziksel protein-protein etkileşimlerini Gene Ontology veri tabanındaki gen tanımlamalarına gore süzgeçten geçirerek protein etkileşim ağı oluşturulması amacıyla geliştirilmiş bir metot kullanılmıştır. Bu metot aracılığıyla oluşturulan etkileşim ağları makro ve mikro analiz yöntemleriyle detaylı olarak incelenmiş ve ağlardaki önemli proteinler tespit edilmiştir. Ağları oluşturan proteinler, hücrede görev aldıkları süreçler açısından incelendiğinde, üç ağın da hücre gelişimi, büyüme ve tranksripsiyon timeline dayalı işlemlere yöneldiği tespit edilmiştir. Grafik teorisi kullanılarak yapılan topolojik çalışmalar, insan ve maya ağlarının ortalama yol/yön uzunlukları, dereceleri ve çaplarının yakın olduğunu ve benzer özellikler sergilediğini göstermiştir. İnsan beynindeki kalsiyum sinyal ağı vücuttaki süreçleri yöneten organ olan beyinde yer alması sebebiyle, maya ve soluncanda görülen duruma ters olarak, vücudun geri kalanındaki kalsiyum sinyal ağına gore daha hızlı ve daha gürbüz olan ağ olarak bulunmuştur. Solucan etkileşim ağı ise, veri eksikliğinden olması muhtemel farklılıklar sergilemektedir. Oluşturulan kalsiyum sinyal ağlarını alt modüllerine bölerek inceleme yapılması ile görevi bilinmeyen proteinlere görev atamak mümkün olmuş ve komşu sayma metodu ile desteklenen moduler analiz sonucu mayadaki 17 proteine görev atanabilmiştir. Solucan ve insan ağlarında ise görev atanması için yeterli bulgu sağlanamamıştır. Son olarak, beyinde kalsiyum düzensizliğinden kaynaklı hastalıklar incelenmiş, elde edilen protein etkileşim ağı ile karşılaştırılmış ve literatürde öngörülen hastalık yolizleri ile dokuya özgü protein verileri arasında bilgi eksikliğine bağlı olduğu düşünülen uyuşmazlıklar saptanmıştır.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv
ABSTRACT vi
ÖZET vii
LIST OF FIGURES xii
LIST OF TABLES xix
1. INTRODUCTION 1
2. THEORY 4
2.1. Calcium Signaling in Yeast 4
2.1.1. Under Normal (Resting) Conditions 4
2.1.2. Under Stress 6
2.2. Calcium Signaling in Worm 7
2.2.1. Calcium IP3 Signaling in the C. Elegans Hermaphrodite Gonad 8
2.2.2. Oscillatory Calcium Signaling in the C. elegans Intestine 10
2.2.3. Calcium Signaling in C. Elegans Neurons and Muscle 11
2.3. Calcium Signaling in Human 13
2.3.1. Ca On Mechanisms 14
2.3.2. Ca Off Mechanisms 15
2.3.3. Signaling Functions of Calcium 16
2.3.4. Neuronal Calcium Signaling 17
2.3.5. Ca Signaling in KEGG 22
2.3.6. Brain Regions and Functions in General 23
2.3.7. Neurodegenerative Disorders and Calcium Signaling 25
2.3.7.1. Neuronal Ca^{2+} signaling and aging 25
2.3.7.2. Neuronal Ca^{2+} signaling and Alzheimer's Disease 26
2.3.7.3. Neuronal Ca^{2+} signaling and Parkinson Disease 30
2.3.7.1. Neuronal Ca^{2+} signaling and ALS 34
2.3.7.1. Neuronal Ca^{2+} signaling and HD 36
2.4. Calcium Related Domains in General 38
3. MATERIALS & METHODS 42
3.1. Databases 42

3.1.1. Interactome Sources 42
3.1.2. Sequence, Structure & Annotation Sources 43
3.2. Tools 49
3.2.1. Sequence & Structure Related Tools 49
3.2.2. Topology & Visualization Related Tools 51
3.3. Selective Permissibility Algorithm (SPA) 55
3.4. Graph Theory 56
3.5. Prediction of Protein Function 62
3.5.1. Neighborhood Counting 63
3.5.2. Module- Assisted Method 63
3.6. Utilization of Methods, Databases and Tools 65
4. RESULTS & DISCUSSION FOR CALCIUM SIGNALING IN YEAST 69
4.1. Core Proteins and Resultant Protein Interaction Network 69
4.1.1. Determination of Core Proteins for SPA 69
4.1.2. Resultant Protein Interaction Networks 73
4.2. Analysis of Reconstructed Network 77
4.2.1. Evolutionary History of Core Proteins 77
4.2.2. Reliability of GO Annotations 83
4.2.3. Annotation Table Enrichment in Yeast PIN 85
4.3. Topology Based Analysis of Yeast PIN 90
4.3.1. Topological Analysis of Yeast Network 90
4.3.2. Hub Proteins of Yeast Network 97
4.3.3. Module Detection and Proteins with Unknown Function 104
4.3.4. Calcium Related Pathways in Yeast Under Stress 131
5. RESULTS & DISCUSSION FOR CALCIUM SIGNALING IN WORM 139
5.1. Core Proteins & Resultants Protein Interaction Network 139
5.1.1. Determination of Core Proteins in SPA 139
5.1.2. Resultant Protein Interaction Network 140
5.2. Analysis of Reconstructed Network 142
5.2.1. Evolutionary History of Core Proteins 142
5.2.2. Annotation Table Enrichment in Worm PIN 143
5.3. Topology Based Analysis of Worm PIN 147
5.3.1. Topological Analysis of Worm PIN 147

5.3.2. Hub Proteins of the Worm Network 154
5.3.3. Module Detection and Proteins with Unknown Function 156
6. RESULTS & DISCUSSION FOR CALCIUM SIGNALING IN HUMAN 158
6.1. Core Proteins & Resultant Protein Interaction Network 158
6.1.1. Determination of Core Proteins for SPA 158
6.1.2. Detection of Brain Specific Proteome/Interactome Data 166
6.1.3. Resultant Protein Interaction Network 167
6.2. Analysis of Reconstructed Network 168
6.2.1. Evolutionary History of Core Proteins 168
6.2.2. Annotation Table Enrichment in Human PIN 172
6.3. Topology Based Analysis of Human PIN 176
6.3.1. Topological Analysis of Human Network 176
6.3.2. Hub Proteins of Human Network 185
6.3.3. Module Detection and Proteins with Unknown Function 190
6.4. Neurodegenerative Disorders and Human PIN 190
6.5. Comparison of Calcium Signaling in Human Brain and in Human in General 197
6.5.1. Comparison of Core Proteins & Resultant Protein Interaction Network 197
6.5.2. Topology Based Comparison of General PIN and Brain PIN 199
7. NETWORK COMPARISON 203
7.1. Topological Aspects 203
7.2. GO Enrichments of Yeast, Worm and Human 206
7.2.1. Comparison of GO Enrichments in Core Proteins 206
7.2.2. Comparison of GO Enrichments in PINs 207
7.2.3. Comparison of GO Enrichments in hubs 209
7.3. Orthology Search for Core and Hub Proteins 210
8. CONCLUSION 222
8.1. Methods & Databases 222
8.2. Calcium Signaling Networks 223
8.2.1. Core Proteins and Reconstructed Networks 223
8.2.2. Enriched Terms for Reconstructed Networks 223
8.2.3. Topological Properties 224

8.3. Future Prospects and Recommendations 225
8.4. Dynamic Networks 225
8.5. Domain Based Interaction Prediction 226
8.6. Gene Ontology 227
APPENDIX A 229
A.1 Interaction Classification with respect to Experiment Type 229
A.2 Gene Ontology Evidence Code Sources 230
A.3. Annotation Table Construction Code Algorithm 232
A.4. SPA Code Algorithm 233
APPENDIX B 240
B.1. Calcium Related Gene Ontology Terms 240
B.2. Annotation Tables for Yeast 245
B.3. Homology & Conservation Search of 12 Core Proteins for Yeast 258
APPENDIX C 263
C.1. Hub Proteins of Yeast PIN with 12 core 263
APPENDIX D 278
D.1 Module List Obtained from 12 yeast core PIN 278
D.2. Module Figures for Yeast 282
D.3. Annotations of Functionally Unknown Yeast Proteins with Evidence Scores 292
D.4 Domains of Function Assigned Yeast Proteins 302
APPENDIX E 308
E.1.1. Details of Worm Core Proteins 308
E.1.2. Annotations Tables for Worm 322
E.2. Homology Search of 17 Worm core Proteins 326
E.3. Modules for Worm 329
E.3.1. Module List Obtained from 17 worm core PIN 329
E.3.2. Module Figures for Worm 330
APPENDIX F 333
F.1. Core and KEGG Proteins for Human 333
F.2. Annotation Tables for Human 339
F.3. Modules for Human 348
F.4. Neurodegenerative Diseases 351

F.4.1. Alzheimer's Disease Pathway..... 351
F.4.2. Parkinson's Disease Pathway..... 355
F.4.3. ALS Pathway..... 358
F.4.4. Huntington's Disease Pathway..... 360

LIST OF FIGURES

Figure 2.1. Calcium signaling in yeast in general..... 5

Figure 2.2. Schematic diagram showing anatomical features of an adult C. elegans hermaphrodite. 8

Figure 2.3. Schematic diagram of IP3 signaling pathways identified in the adult hermaphrodite gonad. 9

Figure 2.4. C. elegans defecation cycle 10

Figure 2.5. Model illustrating possible role of intracellular Ca in regulating defecation cycle. 11

Figure 2.6. Neuronal Calcium Signaling Mechanism..... 19

Figure 2.7. Lobes of Cerebrum..... 24

Figure 2.8. The model of Ca^{2+} dysregulation in AD..... 29

Figure 2.9. The model of Ca^{2+} dysregulation in PD. 32

Figure 2.10. The model of excitotoxicity and Ca^{2+} dysregulation in ALS. 35

Figure 2.11. The model of Ca^{2+} dysregulation in HD. 37

Figure 3.1. The geometrical representation of the selection step of SPA..... 56

Figure 3.2. Representation of network structures: (a) random, (b) scale-free and, (c) hierarchical..... 58

Figure 3.3. Direct versus module-assisted approaches for functional annotation. 65

Figure 4.1. Core Protein Sources 70

Figure 4.2. Yeast Core protein network consisting of 63 & 12 proteins. 74

Figure 4.3. PIN with 63 cores vs. Whole Genome for Yeast 75

Figure 4.4a. SPA resultant Yeast PIN obtained from 63 core proteins with 3226 nodes and 20167 edges. 76

Figure 4.5. Network reconstruction steps in SPA for 63 & 12 core Proteins 77

Figure 4.6. The Conserved Residue Analysis of 12 yeast core proteins by Conseq 80

Figure 4.7. Reliability of Annotation Table for 12 core proteins of yeast 84

Figure 4.8. Annotation Table Comparison of core proteins and PIN for yeast 85

Figure 4.9. Predominant “Biological Process” terms in yeast Core Proteins 87

Figure 4.10. Predominant “Biological Process” terms in yeast PIN 87

Figure 4.11. Predominant “Molecular Function” terms in yeast Core Proteins 88

Figure 4.12. Predominant “Molecular Function” terms in yeast PIN 88

Figure 4.13. Predominant “Cellular Component” terms in yeast Core Proteins 89

Figure 4.14. Predominant “Cellular Component” terms in yeast PIN 89

Figure 4.15. Degree distribution in the yeast signaling network 91

Figure 4.16. Log transform of degrees versus log transform of P(k) (probability of having connectivity, k) in yeast signaling network 91

Figure 4.17. Degree centrality distribution of yeast PIN 94

Figure 4.18. Betweenness centrality distribution of yeast PIN 94

Figure 4.19. Closeness centrality distribution of yeast PIN 95

Figure 4.20. Shortest path distribution of yeast PIN 95

Figure 4.21. Predominant GO terms in yeast PIN hubs 100

Figure 4.22. The conserved residue analysis of yeast hub proteins 100

Figure 4.23. The hub proteins and their interactions in the yeast PIN 103

Figure 4.24. Module 1 used in function assignment to YDL115C 105

Figure 4.25. Module 2 used in function assignments to YIR025W and YDR383C 107

Figure 4.26. Module 4 used in function assignments to YDR202C 110

Figure 4.27. Module 5 used in function assignments to YDL076C (A), YNL166C (B) and YNL078W (C) 112

Figure 4.28. Module 6 used in function assignments to YNL253W (A), YML067C (B), YLR033W and YGR275W (C) 116

Figure 4.29. Module 8 used in function assignment to YKR035W-A 120

Figure 4.30. Module 9 used in function assignment to YGR134-W 122

Figure 4.31. Module 10 used in function assignment to YBL097W 124

Figure 4.32. Module 12 used in function assignment to YNL263C 125

Figure 4.33. Module 11 used in function assignment to YGL124C 127

Figure 4.34. Module 13 used in function assignment to YPL233W 127

Figure 4.35. Module 14 used in function assignment to YBR103W 127

Figure 4.36. HOG (High Osmolarity Glycerol) Pathway of Yeast 132

Figure 4.37. The shortest paths of Sln1 and Cdc42 found by Netsearch. 134

Figure 4.38. Mating Pheromone Pathway of Yeast 135

Figure 4.39. The shortest paths of Gpa1, Far1 and Cdc42 found by Netsearch. 136

Figure 4.40. PKC Pathway of Yeast 137

Figure 4.41. The shortest paths of Tor2 and Rho1 found by Netsearch. 138

Figure 5.1. SPA resultant Worm PIN obtained from 17 core proteins with 1714 nodes and 3174edges.. 140

Figure 5.2. Details of network reconstruction steps in SPA for 17 core Proteins for Worm 141

Figure 5.3. Annotation Table Comparison of core proteins and PIN for worm 144

Figure 5.4. Predominant “Biological Process” terms in worm PIN 145

Figure 5.5. Predominant “Molecular Function” terms in worm PIN 146

Figure 5.6. Predominant “Cellular Component” terms in worm PIN 146

Figure 5.7. Degree distribution in the worm signaling network 148

Figure 5.8. Log transform of degrees versus log transform of P(k) (probability of having connectivity, k) in worm signaling network and whole interactome 148

Figure 5.9. Degree distribution of worm PIN 150

Figure 5.10. Betweenness centrality of worm PIN 151

Figure 5.11. Closeness centrality of worm PIN 151

Figure 5.12. Shortest path distribution of worm PIN 152

Figure 5.13. Hub proteins and their interaction in worm PIN 155

Figure 5.14. Predominant GO terms in worm hub proteins 156

Figure 6.1. Calcium signaling in human 165

Figure 6.2. Human brain calcium signaling core protein network consisting of 8 nodes and 6 edges. 24 lonely proteins are not shown.. 167

Figure 6.3. SPA resultant Human brain calcium signaling protein network consisting of 2593 nodes and 9890 edges.. 167

Figure 6.4. ConSeq results for Human core proteins 169

Figure 6.5. Predominant “Biological Process” terms in Human PIN 173

Figure 6.6. Predominant “Biological Process” terms in Human core 173

Figure 6.7. Predominant “Cellular Component” terms in Human PIN 174

Figure 6.8. Predominant “Cellular Component” terms in Human core..... 174

Figure 6.9. Predominant “Molecular Function” terms in Human PIN 175

Figure 6.10. Predominant “Molecular Function” terms in Human core..... 175

Figure 6.11. Degree distribution in the human signaling network 177

Figure 6.12. Log transform of degrees versus log transform of P(k) (probability of having connectivity, k) in human brain calcium signaling network..... 177

Figure 6.13. Log transform of degrees versus log transform of P(k) (probability of having connectivity, k) in human brain whole network 178

Figure 6.14. Degree Distribution of Human Brain Calcium Signaling PIN..... 180

Figure 6.15. Betweenness Centrality Distribution of Human Brain Calcium Signaling PIN..... 180

Figure 6.16. Closeness Centrality Distribution of Human Brain Calcium Signaling PIN 181

Figure 6.17. Shortest Path Distribution of Human Brain Calcium Signaling PIN 181

Figure 6.18. Top 10 Hub proteins of Human PIN 185

Figure 6.19. Predominant GO terms of hub proteins of Human PIN 186

Figure 6.20. ConSeq Results of hub proteins of human PIN..... 186

Figure 6.21. Alzheimer’s Disease Pathway (taken from KEGG)..... 191

Figure 6.22. Amyotrophic Lateral Sclerosis Pathway (taken from KEGG) 192

Figure 6.23. Huntington's Disease Pathway (taken from KEGG) 193

Figure 6.24. Parkinson's Disease Pathway (taken from KEGG)..... 194

Figure 6.25. Comparison of Predominant "Cellular Component" terms in Human PIN in general and in brain 198

Figure 6.26. Comparison of Predominant "Molecular Function" terms in Human PIN in general and in brain 198

Figure 6.27. Comparison of Predominant "Biological Process" terms in Human PIN in general and in brain 199

Figure 6.28. Log transform of degrees versus log transform of P(k) (probability of having connectivity, k) in general calcium signaling network in human 200

Figure 6.29. Comparison of Shortest path distributions 202

Figure 7.1. Predominant biological process terms of yeast and human core proteins... 206

Figure 7.2. Predominant molecular function terms of yeast and human core proteins . 207

Figure 7.3. Predominant biological process terms of PINs 208

Figure 7.4. Predominant molecular function terms of PINs 208

Figure 7.5. Predominant biological process terms of hubs..... 209

Figure 7.6. Predominant molecular function terms of hubs 210

Figure A. 2 Go Evidence Code Decision Tree (Taken from Gene Ontology website) 231

Figure B.3.1. Conseq scores vs Residue position for cmd1. The more negative score a residue has, the more it is conserved..... 262

Figure D.2.1. Module 1 for yeast..... 282

Figure D.2.2. Module 2 for yeast..... 282

Figure D.2.3. Module 3 for yeast..... 283

Figure D.2.4. Module 4 for yeast..... 283

Figure D.2.5. Module 5 for yeast..... 284

Figure D.2.6. Module 6 for yeast..... 284

Figure D.2.7. Module 7 for yeast..... 285

Figure D.2.8. Module 8 for yeast..... 285

Figure D.2.9. Module 9 for yeast..... 286

Figure D.2.10. Module 10 for yeast..... 286

Figure D.2.11. Module 11 for yeast..... 287

Figure D.2.12. Module 12 for yeast..... 287

Figure D.2.13. Module 13 for yeast..... 288

Figure D.2.14. Module 14 for yeast..... 288

Figure D.2.15. Module 15 for yeast 288

Figure D.2.16. Module 16 for yeast 289

Figure D.2.17. Module 17 for yeast 289

Figure D.2.18. Module 18 for yeast 290

Figure D.2.19. Module 19 for yeast 290

Figure D.2.20. Module 20 for yeast 290

Figure D.2.21. Module 21 for yeast 290

Figure D.2.22. Module 22 for yeast 290

Figure D.2.23. Module 23 for yeast 290

Figure D.2.24. Module 24 for yeast 291

Figure D.2.25. Module 25 for yeast 291

Figure D.2.26. Module 26 for yeast 291

Figure D.2.27. Module 27 for yeast 291

Figure D.2.28. Module 28 for yeast 291

Figure D.2.29. Module 29 for yeast 291

Figure D.2.30. Module 30 for yeast 291

Figure E.3.2.1 Module 1 for worm 330

Figure E.3.2.2 Module 2 for worm 330

Figure E.3.2.3 Module 3 for worm 331

Figure E.3.2.4 Module 4 for worm 331

Figure E.3.2.5 Module 5 for worm 332

Figure E.3.2.6 Module 6 for worm 332

Figure F.3.1 Module 1 for human 348

Figure F.3.2 Module 2 for human 349

Figure F.3.3 Module 3 for human 349

Figure F.3.4 Module 4 for human 350

Figure F.3.5 Module 5 for human 350

Figure F.3.6 Module 6 for human 350

LIST OF TABLES

Table 3.1. The percentage of top n ranked proteins, identified from the yeast protein interaction dataset (yeast-2007/01/07) by three scoring methods and by the double screening scheme (DMNC || MNC, Hubba algorithm), to be yeast essential proteins... 53

Table 3.2. Databases and tools used in network reconstruction and analyses..... 68

Table 4.1. Yeast Core Protein List..... 71

Table 4.2. Comparison of 63 & 12 Core Proteins for Yeast and Their Resultant PINs . 75

Table 4.3. Homologues of yeast core proteins in human..... 79

Table 4.4. ConSeq results of yeast core proteins and their domains 82

Table 4.5. Topological properties of the yeast signaling networks 90

Table 4.6. Topological properties of various networks found from literature..... 96

Table 4.7. Hub Proteins of the Yeast signaling network 101

Table 4.8. ConSeq results of yeast hub proteins and their domains 102

Table 4.9. Functional Enrichment in Module 1 106

Table 4.10. Functional Enrichment in Module 2 108

Table 4.11. Functional Enrichment in Module 4 111

Table 4.12. Functional Enrichment in Module 5 113

Table 4.13. Functional Enrichment in Module 6 117

Table 4.14. Functional Enrichment in Module 8 120

Table 4.15. Functional Enrichment in Module 9 122

Table 4.16. Functional Enrichment in Module 10 124

Table 4.17. Functional Enrichment in Module 12 125

Table 4.18. Functional Enrichment in Modules 11, 13 and 14 127

Table 4.19. Functions Assigned to the Proteins with Unknown Function 129

Table 4.20. The Number of Shortest Paths for HOG, Mating Pheromone and PKC Pathways 138

Table 5.1. Homologues of worm core proteins in human 142

Table 5.2. Topological properties of the worm signaling networks and whole genome 147

Table 5.3. Topological properties of various networks found from literature 153

Table 5.4. Hub Proteins of the Worm signaling network 154

Table 6.1. Human Core Protein List 160

Table 6.2. Interactome Data Decomposition 166

Table 6.3. Conservation Scores of Human core proteins 170

Table 6.4. Topological properties of the human signaling networks and whole brain interactome 176

Table 6.5. Topological properties of various networks found from literature..... 183

Table 6.6. Topological properties of various networks found from literature..... 184

Table 6.7. Hub proteins of the Human PIN 188

Table 6.8. Conservation Scores Hub proteins and their domains..... 189

Table 6.9. Disease related proteins in KEGG and PIN..... 190

Table 6.10. Topological properties of calcium signaling in brain, in general and the whole brain interactome..... 201

Table 7.1. Topological parameters of Yeast, Worm and Human Networks..... 204

Table 7.2. Orthologs of core proteins of yeast, worm and human..... 212

Table 7.3. Orthologs of hub proteins 220

1. INTRODUCTION

One of the main goals of functional genomics is to determine the function of genes predicted from the completely sequenced genomes. In the past decade, massive amounts of biological data have been accumulated from genome sequencing as well as from transcriptomes, proteomes and interactomes. It is a challenging task to integrate such relevant data sources to represent the comprehensive knowledge of genes within and between genomes, which provide specialized information to describe the biological roles of the proteins. The Gene Ontology (GO) is one such resource that is becoming the de facto standard for facilitating information search tasks across databases and for annotating products, based on the following two observations:

- interacting proteins often function in the same biological process, which assumes that two proteins acting in the same biological process are more likely to interact than two proteins involved in different processes, and moreover, proteins functioning in specific biological processes should be more likely to interact than proteins functioning in general processes;

- to interact physically, proteins must exist in close proximity, at least transiently, which suggests that co-localization may serves as an useful predictor for protein interactions.

Since proteins perform their functions by interacting with one another and with other biomolecules, reconstructing a map of the protein–protein interactions of a cell provides valuable insight into the cellular function and machinery of a proteome. However, the full potential of these data cannot be utilized for discovering signal transduction networks for several reasons: the incompleteness of the interactome data, i.e. presence of false-negatives (interactions that are not detected but do occur in the cell); low reliability of the protein-protein interaction data due to prevalence of false-positives (the interactions that are seen in an experiment but never occur in the cell); lack of suitable mathematical models for representation and analysis of signaling networks.

The main objective of the proposed Ph.D. work is to understand the Calcium Signaling pathway of multiple species as yeast, worm and human. In fact, the ultimate goal is to unravel the human calcium signaling pathway in detail to suggest some drug target proteins for the treatment of diseases like Alzheimer's disease, Huntington Disease, Parkinson Disease and many others caused by dysregulation of Calcium Signaling in human brain. But since, the yeast calcium signaling pathway is known in details and the two pathways of yeast and human resemble each other, it will be a chance to use the yeast pathway as well as worm pathway to shed light on the human pathway. The pathways are compared and their common and different points will be determined.

In order to broaden the understanding of the calcium signaling pathway, two perspectives are used. These are the investigation of the calcium signaling both with macroscopic and microscopic tools. The macroscopic study consists of the construction of a complete protein-protein interaction network of the pathway and analyses through graph theoretical approaches to realize the topology of the calcium signaling network. The aim of applying such methods to the calcium signaling pathway is to find the critical proteins of the pathway.

For drug target identification and drug design, the molecular structure of the proteins should also be studied. The structure information of the proteins under investigation provides an insight to the prediction of new protein-protein interactions and drug targeting mechanism of specific important proteins. This part of the work constitutes the microscopic point of view; it includes conservation analyses of sequences and domains, the use of some elastic models like Gaussian network model, and protein-protein interaction prediction methods. In summary, the residue and protein networks of calcium signaling pathway are studied to understand the details, working principles of the network and ultimately determine the drug targets.

In Chapter Two, the theoretical information about the Calcium signaling in yeast, worm and human is given. Chapter Three gives information about the methods, databases and tools used for Network Reconstruction as well as utilization procedure of these methods and sources. In the fourth, fifth and sixth Chapters, the core protein selection, analyses of the core proteins in sequentially and structurally, reconstruction of network, the

graph theoretical properties of these networks in comparison with other networks' topological properties as well as function assignment to proteins with unknown functions, hub proteins of the interaction network for yeast, worm and human are studied and presented in detail.

2. THEORY

2.1. Calcium Signaling in Yeast

The resting level of cytosolic free Ca^{2+} ($[Ca^{2+}]_{cyt}$) is kept very low (typically 50–100 nM) and this is maintained by active Ca^{2+}-pumps and - transporters, and the Ca^{2+}-buffering capacity of the cytoplasm. However, $[Ca^{2+}]_{cyt}$ becomes an intracellular signal when its concentration is transiently increased as a result of the activation of Ca^{2+}-permeable channels allowing Ca^{2+} to flow down a concentration gradient into the cytoplasm from Ca^{2+}-storage organelles or from the external medium. Following the $[Ca^{2+}]_{cyt}$ increase, the $[Ca^{2+}]_{cyt}$ recovers to its original resting level by the increased activity of Ca^{2+}-pumps and -transporters (Zelter *et al.*, 2004).

2.1.1. Under Normal (Resting) Conditions

Like mammalian cells, budding yeast maintains cytosolic Ca^{2+} concentration at very low levels in growing cells (Cui *et al.*, 2009). This homeostatic process has two basic characteristics:

- "zero" steady-state error to extracellular disturbance, which means that cytosolic Ca^{2+} concentration is tightly controlled.

- Robustness of such "perfect" adaptation, which means that the system is relatively insensitive to specific kinetic parameters (Cui and Kaandorp, 2006).

The underlying mechanism of this calcium homeostatic process involves the regulated internalization and sequestration of Ca^{2+} into a variety of intracellular compartments such as the vacuole, ER (endoplasmic reticulum) and Golgi apparatus through many membrane transport proteins (including carrier proteins and channel proteins) (Cui and Kaandorp, 2006).

A vacuolar calcium ATPase *Pmc1* and a vacuolar Ca^{2+}/H+ exchanger *Vcx1* independently transport cytosolic Ca^{2+} into large vacuoles, the major Ca^{2+} store and sink (Cui *et al.*, 2009). Cytosolic Ca^{2+} is also pumped into ER and Golgi through high affinity Ca^{2+}/Mn^{2+} ATPase termed *Pmr1*, which may also promote Ca^{2+} extrusion from yeast cells (Cui *et al.*, 2009)

Under normal conditions, extracellular Ca^{2+} enters the cytosol through an unknown Channel X on the plasma membrane whose encoded gene has not been identified yet (Cui and Kaandorp, 2006). Another transporter on the plasma membrane, transporter M, has been detected recently. This transporter transports Ca^{2+} into the yeast cell in response to very high extracellular Ca^{2+} (Cui *et al.*, 2009).

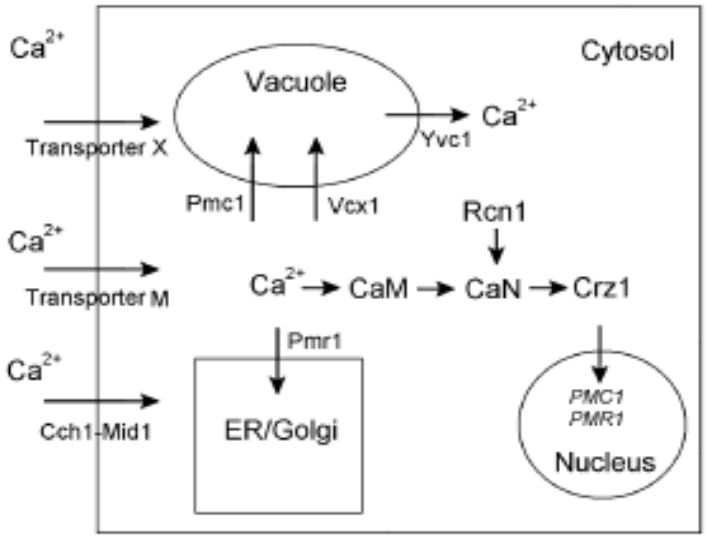


Figure 2.1. Calcium signaling in yeast in general

Both the expression and function of *Pmc1*, *Pmr1* and *Vcx1* are regulated by *calcineurin*, a highly conserved protein phosphatase that is activated by Ca^{2+}/*calmodulin.* Calmodulin is a small Ca^{2+}-binding protein. In resting cells, calmodulin exists in the Ca^{2+}-free (also named as apocalmodulin) form (Cui and Kaandorp, 2006).

2.1.2. Under Stress

In budding yeast, calcium signals promote cell survival under important adaptations in response to mating pheromones, membrane damaging compounds, and a variety of environmental stresses such as high salt, high pH, high osmolarity, and others (Kafadar and Cyert 2004). In response to the elevation of cytosolic Ca^{2+} arising from these changes, the universal Ca^{2+} sensor protein *calmodulin (Cmd1)* can bind and activate the protein phosphatase *calcineurin (Cna1, Cna2, Cnb1, Rcn1)*, which inhibits the function of *Vcx1* and induces the expression of *Pmc1* and *Pmr1* via activation of the *Crz1* transcription factor. These calcineurin sensitive adaptations appear to be crucial for proliferation of yeast cells in high calcium environments (Cui *et al.*, 2009).

Activated calcineurin increases the sequestering rates of Pmc1p and Pmr1p through a transcription factor Crz1p. Activated calcineurin also decreases the sequestering rate of Vcx1p by regulating its synthesis.

When yeast cells experience a severe hypertonic shock, vacuolar Ca^{2+} can be released into the cytosol through the mechanosensitive Ca^{2+} channel *Yvc1*, a distant homolog of mammalian TRPC-type Ca^{2+} channels, which results in transient elevation of cytosolic Ca^{2+} concentration. After Ca^{2+} release, the elevated cytosolic Ca^{2+} is reduced by the rapid sequestration of Ca^{2+} via *Vcx1* (Denis and Cyert, 2002).

Hypotonic shock also induced a transient increase in cytosolic Ca^{2+} (Batiza *et al.*, 1996); however, *Yvc1* is not required for these Ca^{2+} peaks. A plasmamembrane voltage-gated Ca^{2+} channel (VGCC) composed of Cch1 and Mid1 becomes activated resulting in the flow of external Ca^{2+} into the cytoplasm in response to hypotonic shock as well as depolarization, depletion of secretory Ca^{2+} and pheromone stimulation (Cui and Kaandorp, 2006).

In yeast, hyper- and hypotonic shocks induce two independent mitogen-activated protein (MAP) kinase signal transduction cascades. While hypertonic shock stimulates the HOG (High Osmolarity Glycerol) pathway in yeast, hypotonic shock stimulates the PKC1

(Protein Kinase C) signal transduction cascade. Components of both MAP kinase pathways are functionally conserved in mammalian cells (Batiza *et al.*, 1996).

2.2. Calcium Signaling in Worm

Nonmammalian model organisms such as Escherichia coli, Saccharomyces, Caenorhabditis elegans, Drosophila, zebrafish, and the plant Arabidopsis provide numerous experimental advantages for defining the molecular bases of complex physiological processes. C. elegans provides a particularly striking example of the experimental utility of nonmammalian model organisms.

Worms have a short life cycle, produce large numbers of offspring by sexual reproduction, and can be cultured easily and inexpensively in the laboratory. Sexual reproduction occurs by self-fertilization in hermaphrodites or by mating with males.

The reproductive and laboratory culture characteristics of C. elegans make it an exceptionally powerful model system for forward genetic analysis. In addition to forward genetic tractability, C. elegans also has a fully sequenced and well-annotated genome. Genomic sequence and virtually all other biological data on this organism are assembled in readily accessible public databases (e.g., WormBase; http://www.wormbase.org).

Like all nematodes, C. elegans has an unsegmented, cylindrical body that tapers at both ends (Figure 2.2). The body wall consists of a tough collagenous cuticle underlain by hypodermis, muscles, and nerves. A fluid-filled body cavity or pseudocoel separates the body wall from internal organs. Body shape is maintained by hydrostatic pressure in the pseudocoel.

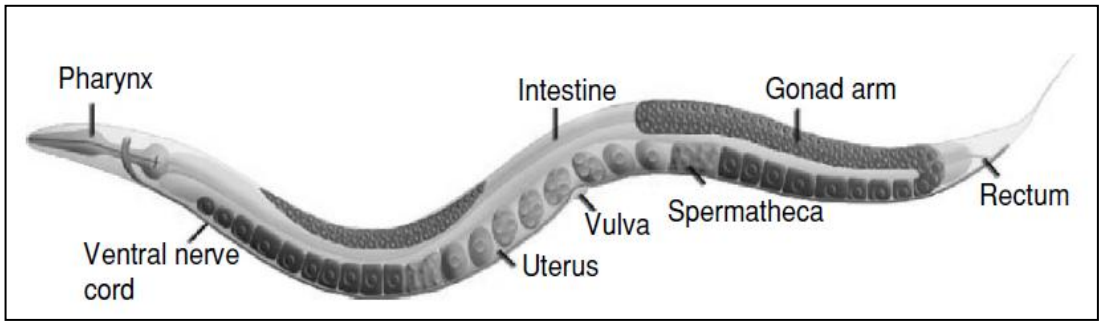


Figure 2.2. Schematic diagram showing anatomical features of an adult C. elegans hermaphrodite. (From www.wormatlas.org)

Despite the small cell number, C. elegans exhibits a striking degree of differentiation. Many physiological functions found in mammals have nematode analogs. Worms possess striated and nonstriated muscle cells; a digestive tract comprised of a muscular pharynx, intestine, and rectum; and excretory and reproductive systems.

The nervous system of adult hermaphrodites contains 302 neurons and 56 glial and support cells. Males have 381 neurons and 92 glial and support cells. White *et al.* have reconstructed and mapped the connectivity of the entire hermaphrodite nervous system using serial electron microscopy.

2.2.1. Calcium IP3 Signaling in the C. Elegans Hermaphrodite Gonad

The involvement of IP3 signaling in C. elegans reproduction was first demonstrated by Clandinin *et al.* (1998).

Rutledge *et al.* (2001) postulated that IP3/Ca^{2+} signaling events regulate ovulation by regulating the contractile activity of gonad sheath cells. Sheath cells are smooth muscle-like myoepithelial cells that surround and are coupled to oocytes via gap junctions. Prior to ovulation, the sheath cells contract weakly at a basal rate of ~8 contractions per minute. Basal contractile activity is triggered by release of major sperm protein (MSP) from sperm, which activates the VAB-1 Eph receptor tyrosine kinase. Oocytes undergoing meiotic maturation secrete *LIN-3*, which increases both the force and rate of sheath contraction and

induces dilation of the spermatheca-gonad valve. This combination of increased sheath contraction and spermatheca dilation induces ovulation.

Yin *et al.* (2004) utilized reverse genetic strategies and DIC imaging to test the hypothesis that IP3/Ca^{2+} signaling regulates sheath cell contraction. *itr-1* reduction- of-function mutations or knockdown of *ITR-1* expression by RNAi dramatically inhibit both basal and ovulatory sheath cell contractions. Gain-of-function *itr-1* mutations increase sheath contractile activity and suppress the inhibitory effects of *lin-3* and *let-23* mutations indicating that *ITR-1* functions downstream of *LIN-3/LET-23* signaling (Figure 2.3).

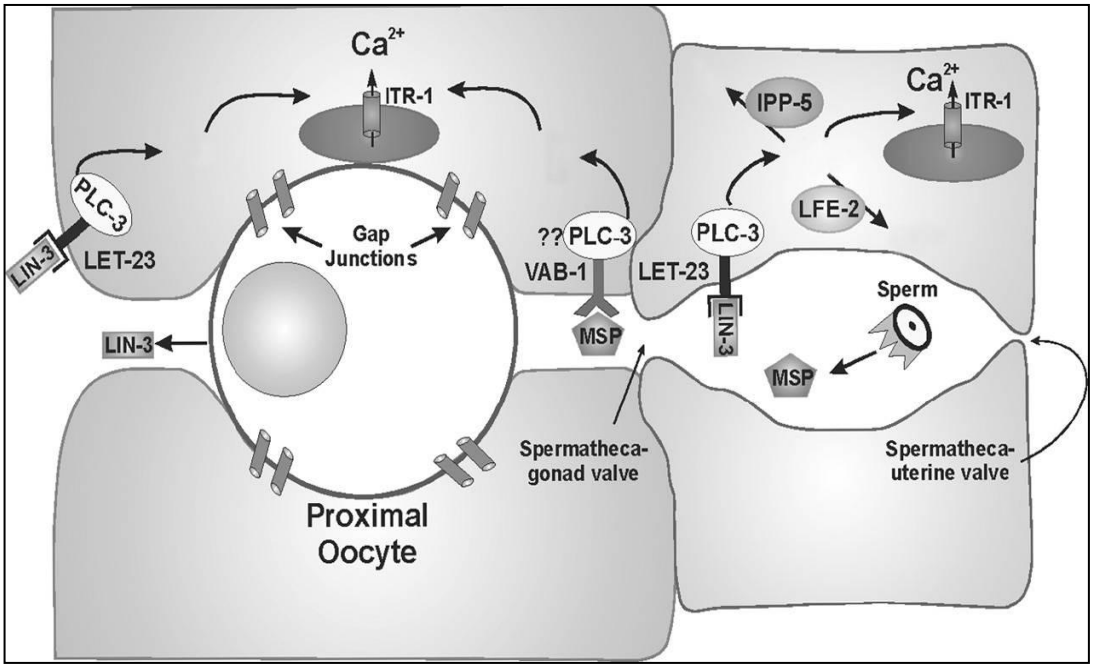


Figure 2.3. Schematic diagram of IP3 signaling pathways identified in the adult hermaphrodite gonad.

Major sperm protein (MSP) is released from sperm and activates the VAB-1 Eph receptor tyrosine kinase on the sheath cells to regulate basal contractility. *LIN-3* is an EGF-like factor secreted by maturing oocytes. Binding of *LIN-3* to its cognate EGF-like receptor, *LET-23*, increases both the force and rate of sheath contraction and induces dilation of the spermatheca–gonad valve. *PLC-3* is a predicted PLC-g homolog and generates IP3 upon activation of *LET-23* and possibly VAB-1. Binding of IP3 to the IP3R, *ITR-1* induces Ca^{2+} release from the ER. LFE-2 is an inositol 1,4,5-trisphosphate-3-kinase

(IP3K) which phosphorylates IP3 and converts it into IP4. *IPP-5* is an ortholog of the human type I inositol polyphosphate 5-phosphatase, which dephosphorylates IP3 and converts it into IP2.

2.2.2. Oscillatory Calcium Signaling in the C. elegans Intestine

The C. elegans defecation cycle is an ultradian rhythm that consists of sequential contraction of posterior body wall muscles, anterior body wall muscles, and enteric muscles every 50 s when the animals are feeding. Cycle is initiated by posterior body wall muscle contraction (pBoc). After relaxation of these muscles, the anterior body wall muscles contract (aBoc) and then expulsion occurs by enteric muscle contraction (Emc). The cycle repeats itself every 45–50 s (Figure 2.4).

Reduction-of-function mutations in *itr-1* slow or eliminate the defecation cycle whereas overexpression of the receptor increases cycle frequency. Rescue of *itr-1* mutants with the wild-type gene in the intestinal cells is sufficient to restore normal cycle activity, demonstrating that *ITR-1* is acting in the intestine to regulate defecation (Dal Santo *et al.*, 1999).

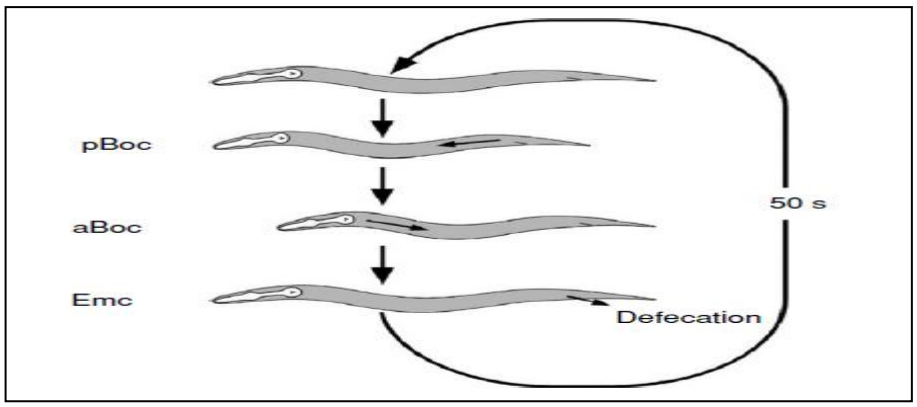


Figure 2.4. C. elegans defecation cycle (Based on Dal Santo *et al.*, 1999)

Cyclical elevation of cytoplasmic Ca^{2+} levels, driven in part by IP3-dependent release of Ca^{2+} from intracellular stores, triggers exocytic secretion from intestinal cells of an unidentified messenger that induces pBoc (Figure 2.5).

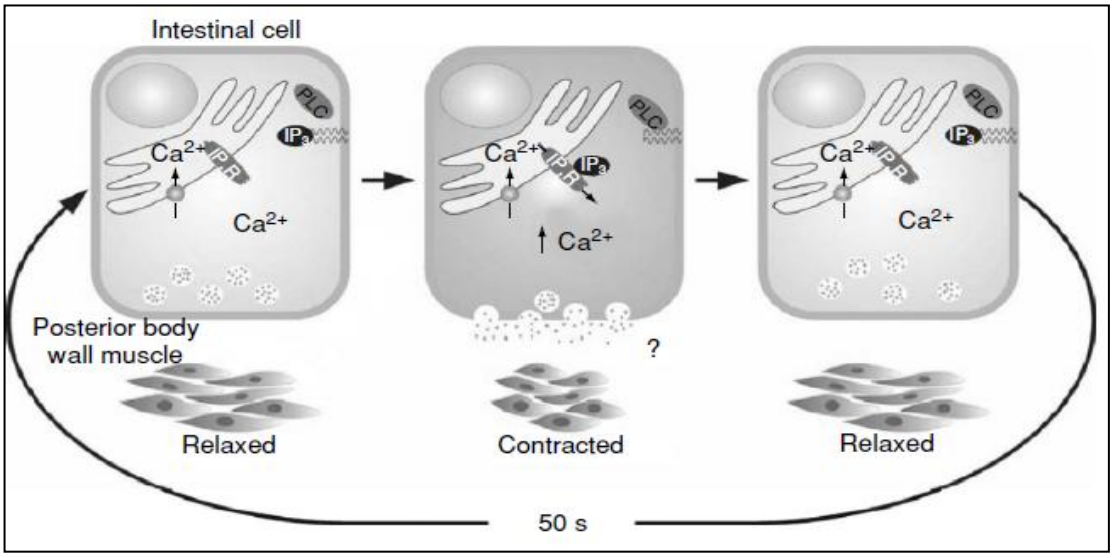


Figure 2.5. Model illustrating possible role of intracellular Ca in regulating defecation cycle. (Taken from Strange, 2003)

2.2.3. Calcium Signaling in C. Elegans Neurons and Muscle

Genetically encoded Ca probes have made it possible to characterize the role of Ca signaling in excitable cell function in C. elegans. In the following section, three examples are given of how a combination of in vivo Ca imaging and genetic analysis has provided insights into the molecular mechanisms of muscle contraction, mechanotransduction, and neurodegeneration.

Regulation of Pharyngeal Muscle Contraction: C. elegans is a filter feeder and the pharynx is a muscular organ that pumps food into the pharyngeal lumen, grinds it up, and then moves it into the intestine. Pharyngeal pumping is essential for feeding and Avery and Horvitz (1990) identified pumping mutants their starved appearance. This screen characterized 52 mutants including egl-19, which is a homolog of the a1 subunit of L-type voltage-gated Ca channels.

Kerr *et al.* (2000) showed that gain-of-function mutations in *egl-19* prolong the duration of Ca transients at the onset of pharyngeal muscle contraction. Kerr *et al.* (2000) also demonstrated that the rate of Ca increase during action potential firing was greatly increased in *unc-36* loss-of-function mutants. *unc-36* encodes a C. elegans homolog of the a2 subunit of L-type voltage-gated Ca channels. The findings of Kerr *et al.* (2000) suggest that UNC-36 negatively regulates Ca influx and are in accordance with heterologous expression studies, which suggest that vertebrate a2 subunits regulate L-type Ca channel inactivation.

The studies of Shimozono *et al.* (2004) suggest that Ca release from intracellular stores via the ryanodine receptor *UNC-68* may play a role in regulating intracellular Ca dynamics and plasma membrane Ca channel activity during pharyngeal contractions.

Mechanotransduction: C. elegans has been an invaluable model system for identifying genes involved in mechanosensory processes. Genes required for mechanosensory neuron function were identified by mutagenizing worms and assessing their ability to move away from gentle body touch. Touch insensitive mutant worms are mechanosensory abnormal and the genes responsible for this abnormality are termed mec. However, touch activated Ca transients are detected in wild-type animals but not in *mec-4* null mutants. The Ca response is also dependent on the L-type Ca channel *EGL-19*. Loss-of-function mutations in *egl-19* significantly reduce the magnitude of touch evoked Ca influx. Taken together, these data suggest that activation of MEC-4 upon mechanical stimulation depolarizes the touch neurons and this depolarization subsequently leads to Ca influx via *EGL-19*.

Cameleons have been used to track Ca changes in sensory neurons in response to chemical, osmotic, and mechanical stimuli. Evoked Ca transients are dependent on *OSM-9*, the founding member of the TRPV channel subfamily. *OSM-9* plays important roles in osmosensation, chemosensation, mechanosensation, and olfaction in C. elegans.

Neurodegeneration: The first DEG/ENaC gene identified in C. elegans was *deg-1* (degeneration of certain neurons). Gain-of-function mutations in deg-1 cause various neurons to swell and degenerate and the protein encoded by the gene was thus termed a

‘‘degenerin.’’ Expression of a hyperactive form of the *MEC-4* channel, termed MEC-4(d) also induces neuronal death. The cell death induced by MEC-4(d) and other degenerins is via necrosis rather than apoptosis.

MEC-4(d) is expressed ectopically in the ventral nerve cord, which resulted in the necrotic death and consequent paralysis of the animals. Paralyzed worms were then mutagenized and mutants in whom movement was restored were isolated. This screen identified the calreticulin encoding gene *crt-1* as a suppressor of MEC-4(d) induced neurodegeneration. *CRT-1* is a calcium binding protein normally expressed in the ER where it functions to modulate free-Ca levels and also acts as a molecular chaperone. The suppressing effect of loss-of-function *crt-1* mutations is dependent on Ca release from intracellular stores and can be blocked by gain-of-function mutations in the IP3 and ryanodine receptors *ITR-1* and *UNC-68*, respectively, or by RNAi of another store Ca binding protein, calnexin.

How does hyperactivation of MEC-4(d) initiate necrotic cell death? One possibility is that Na influx via MEC-4(d) depolarizes the cell and activates voltage-gated Ca channels, leading to an increase in intracellular Ca that initiates necrosis. However, loss-of-function mutations in any of the six voltage-gated Ca channels expressed in C. elegans do not inhibit MEC-4(d) induced necrosis. Calcium measurements in cultured touch neurons expressing cameleons demonstrated that MEC-4(d) induces an amiloride sensitive Ca influx. It is proposed that Ca entry via the MEC-4(d) channel triggers Ca release from intracellular stores via IP3 and ryanodine receptors. Store Ca release in turn activates proteases that initiate cell death.

2.3. Calcium Signaling in Human

The divalent cation calcium (Ca) is used by cells as a second messenger to control many cellular processes including muscle contraction, secretion, metabolism, neuronal excitability, cell proliferation, and cell death. The cytosolic level of Ca in resting cells is kept low (10 - 100 nM), but stimulation results in the level increasing to 500 -1000 nM range required to activate sensors such as calmodulin and troponin C. The cell has access to two sources of signal, Ca entry from the external medium and Ca release from internal

stores. These Ca ON mechanisms are balanced by Ca pumps which constitute the OFF mechanisms responsible for removing the Ca signal. These ON and OFF mechanisms are often organized to produce brief spikes and waves of calcium. Cells may avoid the cytotoxic effects of calcium by employing this oscillatory mode of calcium signaling (Clapham, 1995).

2.3.1. Ca On Mechanisms

Plasma Membrane Ca Channels: The plasma membrane has a variety of Ca entry channels that are characterized by their mechanisms of activation.

Voltage-operated channels (VOCs): A family of channels that are opened in response to membrane depolarization to mediate the selective entry of Ca. The multiple types (L, T, N, P/Q and R) are classified on the basis of their kinetics and pharmacological properties.

Receptor-operated channels (ROCs): Ca channels that are opened by the binding of specific agonists usually neurotransmitters such as glutamate (the NMDA receptor) or ATP.

Store-operated channels (SOCs): Many cells have SOCs in the plasma membrane that are opened by the depletion of the internal stores. Just how empty stores communicate with the membrane SOCs is still unclear, current proposals include a calcium influx factor (CIF) or information transfer through a direct protein-protein interaction.

Intracellular Ca Channels: The endoplasmic reticulum/sarcoplasmic reticulum (ER/SR) has two families of intracellular channels responsible for releasing Ca from this internal store.

Inositol 1,4,5-trisphosphate (IP3) receptors: Agonists acting through G-protein-linked or tyrosine kinase-linked receptors stimulate phospholipase C (PLC) to generate the second messenger IP3 which then diffuses into the cytoplasm to release stored Ca by binding to IP3 receptors.

Ryanodine receptors (RYR): Originally described in muscle cells, these RYRs are also found in neurons and other cell types. Cyclic ADP ribose is a putative second messenger for regulating the activity of RYRs. Ryanodine also binds to these receptors and can initiate release by locking the Ca channel in an open configuration.

Calcium-Induced Calcium Release (CICR): The IP3Rs and the RYRs are sensitive to Ca through both positive and negative feedback effects. A positive feedback process of Ca-induced Ca release (CICR) is of critical importance in generating and shaping Ca signals. Once the channels open, a microdomain of Ca begins to build up and the channels then close as the positive feedback effect is replaced by a negative feedback response. An important function of IP3 and cyclic ADP ribose is to increase the Ca sensitivity of the intracellular channels such that the cytoplasm becomes an excitable medium capable of generating Ca spikes and waves.

2.3.2. Ca Off Mechanisms

The Ca OFF mechanisms depend on pumps that remove the Ca signal during the recovery from stimulation. The surface membrane has a Na+/Ca exchanger (found mainly in excitable cells) and the ubiquitous plasma membrane Ca ATPase (PMCA). The latter is regulated by a variety of factors including calmodulin, acidic phos pholipids, and protein kinases A and C (PKA and PKC). The internal stores have a sarco/endoplasmic reticulum Ca ATPase (SERCA) (Berridge, 1998).

The mitochondria also play an important role in Ca signaling in that they sequester Ca rapidly during the recovery phase and then release it back slowly when the cell is at rest. They are particularly effective when they lie close to the Ca release channels. If the mitochondrion accumulates too much Ca it forms a permeability transition pore, which collapses the transmembrane potential leading to the release of cytochrome c and the initiation of apoptosis.

Ca-Binding Proteins: Cells have a variety of Ca-binding proteins functioning either as buffers to shape the cellular response or as sensors to carry out the messenger role of Ca.

Ca Buffers: Both the cytoplasm and the lumen of the ER/SR have proteins capable of buffering Ca. As Ca is pumped into the lumen of the ER/SR, it is buffered by storage proteins such as calsequestrin and calreticulin which have low affinity (Kd in the mM range) but high capacity (approximately 50 Ca ions bound/molecule). When Ca enters the cytosol it is rapidly buffered by proteins such as calbindin, calretinin, and parvalbumin. More than 90% of the Ca entering the cell is bound to these buffers with the small remainder representing the stimulus-evoked elevation of Ca responsible for activating the Ca sensors.

Ca Sensors: Ca sensors such as the EF-hand proteins, annexins and the S100 proteins mediate the intracellular effects of Ca. The EF hand proteins such as troponin C (TnC) and calmodulin (CaM), which have a characteristic Ca-binding domain between two helices (named after the E and F α-helices of parvalbumin), are the major Ca sensors. TnC functions in skeletal and cardiac muscle whereas CaM has a much more general role. CaM exerts some of its effects by stimulating a multifunctional CaM-dependent protein kinase II (CaMKII) or a CaM-dependent protein phosphatase known as calcineurin.

2.3.3. Signaling Functions of Calcium

Calcium functions as a second messenger to regulate a great variety of cellular processes (Clapham, 1995).

Contraction: Excitation-contraction coupling in skeletal and cardiac muscle depends upon the release of calcium by RYRs located on the sarcoplasmic reticulum. Pharmacomechanical coupling in smooth muscle is controlled by the release of calcium from either IP3Rs or RYRs depending on the muscle type.

Secretion: During stimulus-secretion coupling, calcium acts either to release preformed materials by exocytosis (e.g. transmitter release at synaptic endings) or to stimulate the ionic mechanisms responsible for fluid secretion (e.g. in exocrine glands).

Metabolism: Glycogen breakdown in liver cells is controlled by a calcium-dependent activation of phosphorylase.

Neuronal excitability: The excitability of neurons can be modulated through calcium-dependent effects on ion channels (e.g. potassium channels) and ionotropic receptors (e.g. AMPA receptors). Some of these effects of calcium are long lasting, such as long-term potentiation (LTP) or long-term depression (LTD), and have been implicated in learning and memory.

Cell proliferation: Calcium plays an important role both in fertilization and in controlling cell proliferation. In the case of lymphocyte activation, the immunosuppressant drug cyclosporin A acts by inhibiting the calcium-dependent transfer of information from the T cell receptor to the nucleus.

Cell death: Elevated levels of calcium, especially if maintained for long periods, can be cytotoxic. Calcium has been implicated in both necrosis and apoptosis.

2.3.4. Neuronal Calcium Signaling

Calcium plays an important role in regulating a great variety of neuronal processes including excitability, associativity, neurotransmitter release, synaptic plasticity, and gene transcription. Like other cells, neurons use both extracellular and intracellular sources of calcium. The mechanisms responsible for regulating the influx of external calcium are well established. For example, voltage-operated channels are used to trigger the release of neurotransmitter at synaptic junctions and they contribute to dendritic action potentials. In addition, neurotransmitters can induce an influx of the NMDA receptors located primarily at postsynaptic sites.

While much is known about these influx pathways, there is less information on the mechanism and role of the intracellular supply of calcium stored within the endoplasmic reticulum of neurons. Inositol 1,4,5-triphosphate receptors (IP3Rs) or ryanodine receptors (RYRs) distributed throughout the endoplasmic reticulum are responsible for releasing Ca from this internal store (Figure 2.6). Since the endoplasmic reticulum is a network distributed throughout the cell, it may be considered as a neuron-within-a-neuron, a concept that becomes all the more interesting because the endoplasmic reticulum, like the

plasma membrane, has integrative and regenerative properties that could play important roles in neural signaling.

Neuronal activity can lead to marked increases in the concentration of cytosolic calcium, which then functions as a second messenger that mediates a wide range of cellular responses. Calcium binds to calmodulin and stimulates the activity of a variety of enzymes, including calcium-calmodulin kinases and calcium-sensitive adenylate cyclases. These enzymes transduce the calcium signal and effect short-term biological responses, such as the modification of synaptic proteins and long-lasting neuronal responses that require changes in gene expression. Recent studies of calcium signal transduction mechanisms have revealed that, depending on the route of entry into a neuron, calcium differentially affects processes that are central to the development and plasticity of the nervous system, including activity-dependent cell survival, modulation of synaptic strength, and calcium-mediated cell death (Ghosh *et al*.., 1995).

The concentration of intracellular free Ca in most neurons is about 100 nM. This low resting concentration is maintained by the action of membrane-associated Ca adenosine triphosphatases (Ca pumps), which must act against a fairly steep concentration gradient to pump Ca into the two major sinks-the extracellular space and the internal Ca stores of the endoplasmic reticulum (ER), where the Ca concentration is typically about 2 mM (Berridge, 1998).

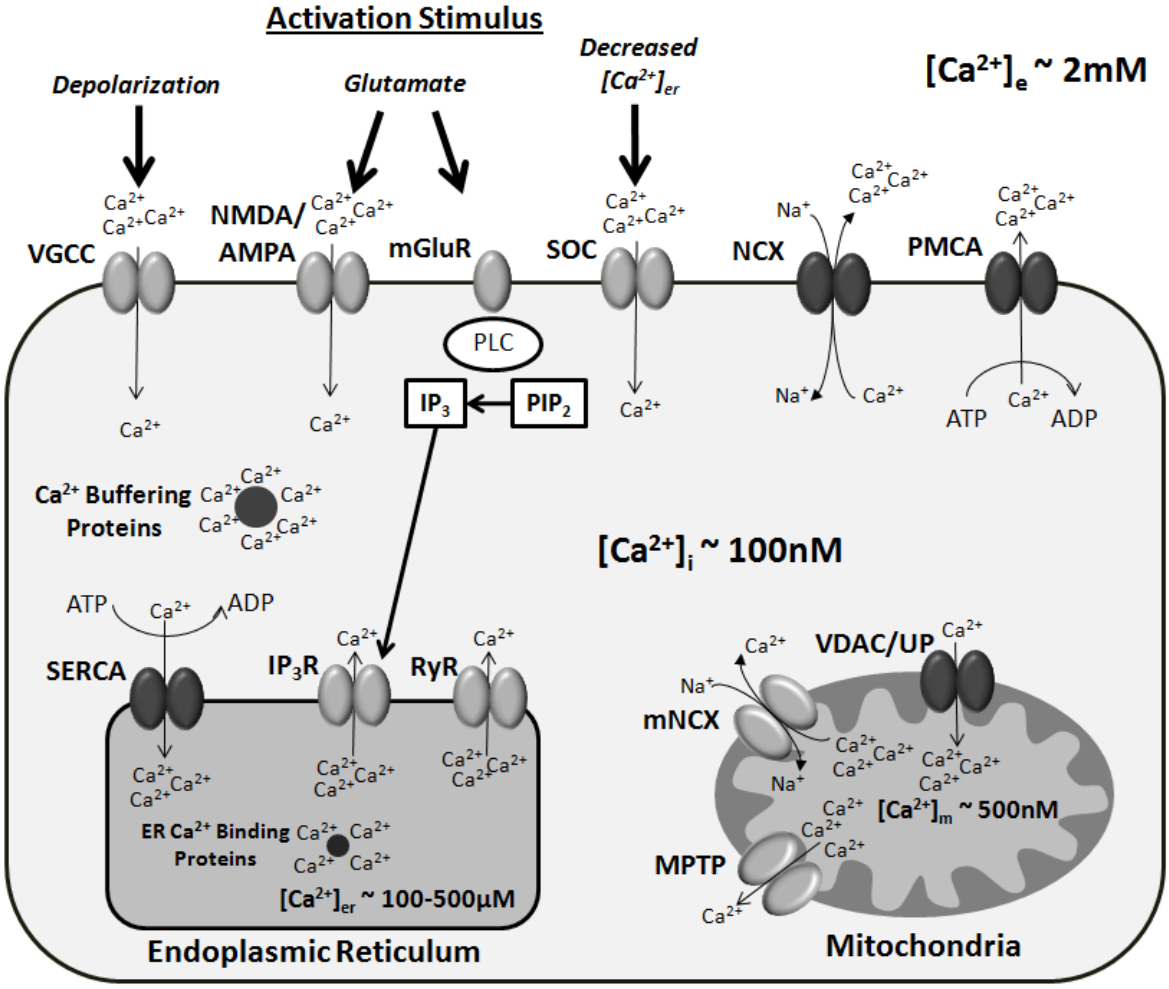


Figure 2.6. Neuronal Calcium Signaling Mechanism (Grant, 2008)

There are two major mechanisms by which Ca enters the cytosol from these Ca sinks. The first is by influx of extracellular Ca through various Ca channels embedded within the plasma membrane. The second is by release of Ca from the internal stores. The entry of Ca into the cytosol can be regulated by multiple mechanisms, allowing for exquisite spatial and temporal control of Ca by signals such as neurotransmitters and growth factors that act at the cell surface (Klipp *et al.*., 2006).

Calcium influx from the extracellular space: In most neurons of the central nervous system (CNS) there are at least two major classes of Ca channels: the voltage-operated Ca channels (VOCCs) and the NMDArs. In addition, subsets of central neurons express other Ca-permeable channels including the neuronal acetylcholine receptor (nAChR), the type 3 serotonin receptor (5HT3R), and the Ca-permeable AMPA and glutamate receptors.

Although much of our understanding of Ca channels comes from the study of VOCCs, in recent years the NMDAr has been intensively studied because of its involvement in events such as synaptic plasticity and cell death. A striking feature of the NMDAr is that, in contrast to most other ion channels, it is a ligand- and voltage- gated channel.

Opening of the channel requires the binding of the neurotransmitter glutamate as well as a depolarization of the membrane. The voltage dependence of the channel arises from the fact that at resting potential the channel is blocked by extracellular magnesium ions (Mg). Depolarization of the membrane allows the Mg block to be relieved, and in the open state the NMDAr allows the entry of both Ca and sodium (Na) ions. Since the induction of LTP is completely blocked by the inhibition of NMDArs as well as by chelating intracellular Ca, the prevailing hypothesis is that Ca influx through the NMDAr leads to the potentiation of active synapses.

Calcium release from internal calcium stores: Although a major route of Ca entry into the cytosol is through plasma membrane Ca channels, release from internal Ca stores also contributes significantly to the concentration of intracellular free Ca. There appear to be at least two pools of internal Ca stores in the ER, and the release of Ca from these two pools is mediated by distinct mechanisms.

The first mechanism is by the binding of inositol triphosphate (IP3) to its receptor on the ER membrane. IP3 is a central regulator of intracellular free Ca that is generated by exposure of cells to a variety of stimuli, including growth factors and neurotransmitters that act through guanosine triphosphate (GTP)-binding protein (G protein)-linked receptors.

Another mechanism by which Ca is released from the internal stores is by Ca-mediated Ca release. This mechanism is regulated by the activation of the ryanodine receptor (RyR, named after its specific agonist ryanodine). RyRs expressed in brain appear to be more closely related to cardiac muscle RyRs. Cardiac (and presumably neuronal) RyR activation has a biphasic Ca response. The release of Ca from RyR-activated internal stores may be acutely sensitive to the magnitude of Ca influx after synaptic stimulation.

As Ca enters the cytosol, it encounters a number of proteins that regulate its biochemical effects. Central among them is calmodulin, a small Ca-binding protein, which can act as an intracellular Ca sensor. Ca-calmodulin (CaM) binds to a number of enzymes and modulates their activity. These include CaM-dependent protein kinases (CaM kinases), protein phosphatases, and adenylate cyclases. These enzymes can either act locally at the synapse by inducing the modifications of preexisting synaptic proteins or can mediate more general cellular responses by activating molecules involved in the regulation of gene expression.

Calcium-calmodulin-dependent protein kinases: Of the five CaM kinases that have been identified, CaM kinases II and IV have been best characterized. Both have broad substrate specificities and, therefore, may regulate a number of Ca-dependent biochemical events. Whereas CaM kinase II has been extensively studied as a potential mediator of Ca-dependent synaptic changes, CaM kinase IV and certain isoforms of CaM kinase II may be specifically involved in mediating transcriptional activation of gene expression in response to changes in intracellular Ca.

A number of studies on the localization, activation, and perturbation of CaM kinase II present a compelling case that CaM kinase II is directly involved in the regulation of synaptic plasticity. CaM kinase II is enriched in the cortex and hippocampus and is concentrated both pre- and postsynaptically. CaM kinase II isolated from brain contains alpha and beta subunits in a ratio of about 4:1. Unlike protein kinase A (PKA) and PKC, in which there are separate catalytic and regulatory subunits, each subunit of CaM kinase II has catalytic activity. In the resting state, an autoinhibitory domain of the kinase keeps the catalytic site inaccessible. This inhibition can be relieved by the binding of CaM to the autoinhibitory domain. The kinase thus activated can not only phosphorylate substrates at serine residues but also undergoes autophosphorylation at a number of sites. The autophosphorylation of CaM kinase II causes the enzyme to switch to a Ca independent form, that is, the kinase maintains catalytic activity even after intracellular Ca concentrations have returned to basal levels. This is the property of CaM kinase II that initially led to the proposal that it could function as a "memory molecule" and might contribute to long-term modulation of synaptic function (Ghosh *et al.*., 1995).

2.3.5. Ca Signaling in KEGG

Ca that enters the cell from the outside is a principal source of signal Ca. Entry of Ca is driven by the presence of a large electrochemical gradient across the plasma membrane. Cells use this external source of signal Ca by activating various entry channels with widely different properties. The voltage-operated channels (VOCs) are found in excitable cells and generate the rapid Ca fluxes that control fast cellular processes. There are many other Ca-entry channels, such as the receptor-operated channels (ROCs), for example the NMDA (N-methyl-D-aspartate) receptors (NMDARs) that respond to glutamate. There also are second-messenger-operated channels (SMOCs) and store-operated channels (SOCs).

The other principal source of Ca for signalling is the internal stores that are located primarily in the endoplasmic/sarcoplasmic reticulum (ER/SR), in which inositol-1,4,5-trisphosphate receptors (IP3Rs) or ryanodine receptors (RYRs) regulate the release of Ca. The principal activator of these channels is Ca itself and this process of Ca-induced Ca release is central to the mechanism of Ca signalling. Various second messengers or modulators also control the release of Ca. IP3, which is generated by pathways using different isoforms of phospholipase C (PLCbeta, delta, epsilon, gamma and zeta), regulates the IP3Rs. Cyclic ADP-ribose (cADPR) releases Ca via RYRs. Nicotinic acid adenine dinucleotide phosphate (NAADP) may activate a distinct Ca release mechanism on separate acidic Ca stores. Ca release via the NAADP-sensitive mechanism may also feedback onto either RYRs or IP3Rs. cADPR and NAADP are generated by CD38. This enzyme might be sensitive to the cellular metabolism, as ATP and NADH inhibit it.

The influx of Ca from the environment or release from internal stores causes a very rapid and dramatic increase in cytoplasmic calcium concentration, which has been widely exploited for signal transduction. Some proteins, such as troponin C (TnC) involved in muscle contraction, directly bind to and sense Ca. However, in other cases Ca is sensed through intermediate calcium sensors such as calmodulin (CALM).

2.3.6. Brain Regions and Functions in General

There are three different parts of brain - the forebrain, midbrain and hindbrain. The forebrain is made up of the *cerebrum, thalamus and hypothalamus*. The midbrain consists of the *tectmum and tegmentum*. The hindbrain is made up of the *cerebellum, pons and medulla*. The midbrain, pons and medulla form the brain stem.

Cerebrum: The cerebrum is the largest part of the brain that is located in the forebrain. The cerebrum is divided into two equal halves called the hemispheres and consists of four lobes. Cerebrum is responsible for controlling our thoughts and actions. The four lobes of cerebrum are as follows:

- Parietal Lobe: This area of the brain responsible for recognition, movement, orientation and perception of stimuli.

- Frontal Lobe: The frontal lobe is associated with reasoning, planning, parts of speech, movement, emotions and problem solving.

- Occipital Lobe: The occipital lobe is associated with visual processing.

- Temporal Lobe: The temporal lobe is associated with the perception and recognition of memory, speech and auditory stimuli.

Cerebellum: The cerebellum is also divided into two hemispheres and is a part of the hind brain. The cerebellum contains more than half of the brain neurons. The cerebellum is one of the sensory areas of the brain that is responsible for motor control, sensory perception and coordination. The cerebellum is also associated for movement, posture and balance regulation and coordination.

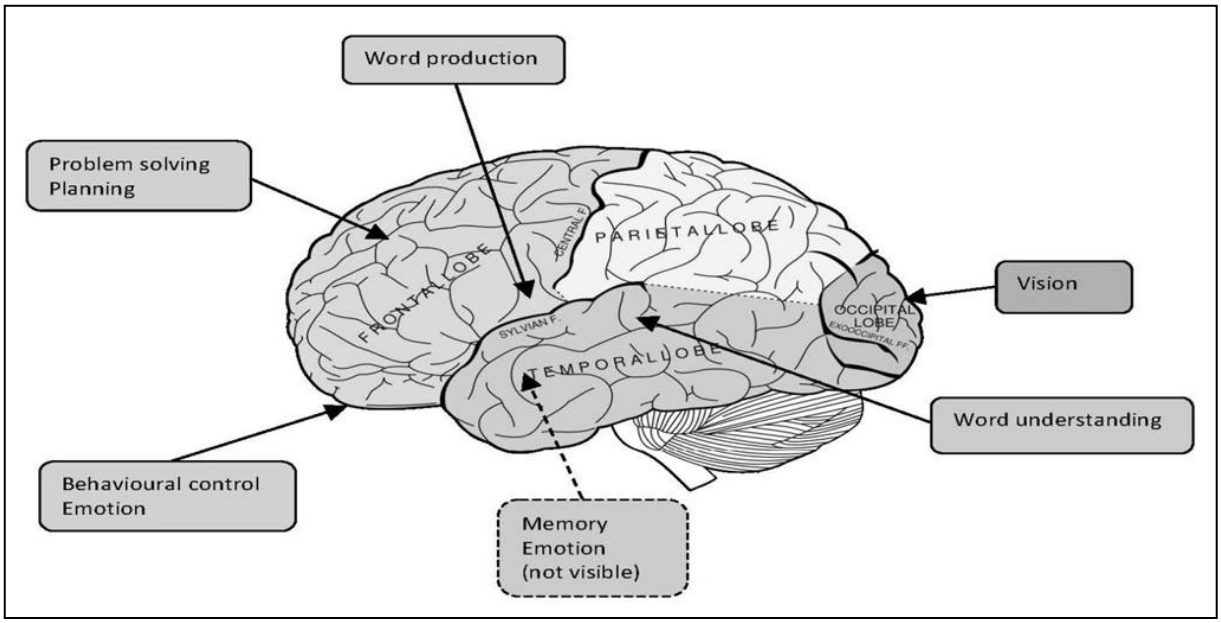


Figure 2.7. Lobes of Cerebrum

Limbic System: The limbic system is the 'emotional brain' that is deep within the *cerebrum*. The limbic system consists of the *hypothalamus, thalamus, hippocampus and amygdala*. The limbic system is one of the sensory areas of the brain controls our emotions and memories. The body temperature is regulated by the *hypothalamus* and controls emotions and our cranial urges. The messages to the *cerebral cortex* are transmitted by the *thalamus*. The *amyglyda* is located in the *temporal lobe* and is involved with memory, emotions and fear. The *hypocampus* is one of the major areas of the brain and their functions include learning and memory. The *hypocampus* converts short term memory into permanent memory and spatial navigation.

Brain Stem: The brain stem is located under the limbic system and is responsible for the basic life functions that include breathing, heart beat and blood pressure. The brain stem is the simplest part of the human brain connects the brain to the spinal cord. The midbrain a part of the brain stem is associated with vision, hearing, movement of the eye and other body movements. The *pons* is associated with motor control and sensory analysis. Certain *pons* areas are involved with consciousness and some for sleep. Certain structures of the *pons* are linked to the *cerebellum* and associated with functions involving posture and movement. The *medulla* is the brain structure that is between the *pons* and

spinal cord. The *medulla* is important for conducting vital body functions like breathing and heart rate.

2.3.7. Neurodegenerative Disorders and Calcium Signaling

In neurodegenerative disorders, such as Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), Huntington's disease (HD) and spinocerebellar ataxias (SCAs), neuronal calcium (Ca^{2+}) signaling is abnormal. Similar, but less severe, changes in neuronal Ca^{2+} signaling occur as a result of the normal aging process. For the treatment of these disorders, Ca^{2+} blockers will be most beneficial clinically when used in combination with other disease-specific therapeutic approaches (Bezprozvanny, 2009).

Most of the scientific effort is focused on identification of the major causes of these diseases and on developing ways to target them.

2.3.7.1. Neuronal Ca^{2+} signaling and aging. The risk of neurodegenerative disorders increases with age. Comparative studies performed with neurons from young and old rodents have indicated that neuronal Ca^{2+} signaling machinery undergoes significant age-dependent changes (Tuesco et.al, 2007).

Increased resting cytosolic Ca^{2+} concentration (Kirischuk *et al.*, 1996) and increased Ca^{2+} entry through VGCCs (Thibault *et al.*, 1996) have been observed in neuronal models of aging. Reductions in the Ca^{2+}-buffering capacity of the mitochondria (Murchison *et al.*, 2004) and the ER (Toescu *et al.*, 2003) have also been observed in aging models. In addition, several studies have shown that the rate by which Ca^{2+} is returned to resting concentration following a Ca^{2+} response is decreased in aged neurons (Toescu *et al.*, 2004). This may be due to decreased expression and/or activity of the neuronal membrane PMCA. Decreased Ca^{2+} uptake through the mitochondrial uniporter due to loss of mitochondrial membrane potential has also been implicated in decreased Ca^{2+} removal from the cytoplasm (Grant, J., 2008).

The resulting changes in neuronal Ca^{2+} dynamics lead to augmented susceptibility to induction of long-term depression (LTD) and an increase in the threshold frequency for induction of long-term potentiation (LTP) in aging neurons. LTD and LTP refer to activity-dependent and persistent changes in synaptic strength, which are widely considered to form a basis for formation and storage of memories in the brain.

These age-related changes in Ca^{2+} handling, along with reduced mitochondrial function, energy production and increased levels of oxidative stress, may serve to sensitize aged neurons to further disruptions in Ca^{2+} signalling due to toxic stimuli. Indeed, disruption in Ca^{2+} signalling is observed in several age-related neurodegenerative disorders.

The mechanisms responsible for age-related alterations in neuronal Ca^{2+} signaling are not clearly understood. One potential explanation is related to age-induced defects in mitochondrial function due to cumulative oxidative damage to mitochondria. The mitochondria from aged neurons are depolarized and less efficient in handling Ca^{2+} load. Age-related changes in the transcription of Ca^{2+} signaling genes have been observed in microarray studies. Some of these changes are directly caused by aging and some are compensatory, however, the overall picture is consistent with age-related alterations in neuronal Ca^{2+} signaling at multiple levels (Bezprozvanny, 2009).

Aging is associated with disruptions in learning and memory, which are common across numerous species. At a cellular level, changes have been observed in terms of reduced neuronal excitability and reduction of LTP coupled with increases in LTD. Subtle changes in neuronal Ca^{2+} signaling may at least in part mediate these age-related changes in neuronal function.

2.3.7.2. Neuronal Ca^{2+} signaling and Alzheimer's Disease (AD). AD is an age-related neurodegenerative disorder characterized by progressive impairment of memory, mood, and coordination. Postmortem brains from patients who had AD exhibit striking

deterioration in regions responsible for learning and memory, including the hippocampus, entorhinal cortex, basal forebrain, amygdala, frontal cortex, and inferior parietal cortex.

The dominant model for pathogenesis in the AD field is the ‘amyloid hypothesis’, which states that increased production of amyloid-beta peptide is a major cause of neuronal and synaptic loss in AD. The experimental support for the ‘amyloid hypothesis’ comes from:

- accumulation of amyloid plaques in the brains of AD patients;
- the familial AD (FAD) cases that result from missense mutations in amyloid precursor protein (APP); and
- the FAD cases that result from missense mutations in presenilins, which form a catalytic subunit of the APP-cleaving enzyme g-secretase. The ‘amyloid-targeting’ therapies have been the main focus of AD drug development.

Recent clinical trial results have suggested that additional targets beyond amyloid need to be seriously considered for AD treatment. A body of evidence has also suggested that neuronal Ca^{2+} dyshomeostasis has an important role in AD. One potential connection between AD pathogenesis and Ca^{2+} comes from the observation that Ab oligomers can form Ca^{2+}-permeable channels in membranes.

Exposure of phosphatidylserine (PtdS) on the cell surface, which is usually indicative of cells in conditions of energy deficit, enhances the ability of Ab to associate with the membrane. Age-related mitochondrial impairments might increase surface PtdS levels in affected neurons and set them up for Ab-mediated pore formation, Ca^{2+} influx and cell death (Figure 2.8). Indeed, neurons with reduced cytosolic ATP levels and elevated surface PtdS levels are particularly vulnerable to Ab toxicity. The probable explanation is that a high local concentration of Ab oligomers in the area surrounding amyloid plaques causes the formation of Ca^{2+}-permeable ion channels in the neuronal plasma membrane.

In addition to the direct effects of Ab on plasma membrane Ca^{2+} permeability, Ab oligomers also affect neuronal Ca^{2+} homeostasis by modulating the activity of NMDARs, alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors (AMPARs) and P/Qtype VGCCs (Figure 2.8).

Another potential connection between Ca^{2+} signaling and AD comes from the observation that many FAD mutations in presenilins result in abnormal Ca^{2+} signaling. To explain these observations, it has been suggested that mutant presenilins affect store-operated Ca^{2+} influx, increase activity and/or expression of intracellular Ca^{2+}-release channels, such as RyR and IP3R, or modulate the function of the sarcoplasmic and endoplasmic reticulum calcium ATPase (SERCA) ER Ca^{2+} pump.

Presenilins themselves have been reported to function as ER Ca^{2+}-leak channels (Figure 2.8) and many FAD mutations in presenilins result in loss of ER Ca^{2+}-leak function, leading to ER Ca^{2+} overload and supranormal Ca^{2+} release from the ER. Although they differ in the details of the proposed mechanisms, the majority of these studies concluded that many FAD mutations in presenilins result in excessive Ca^{2+} release from the ER through IP3R and RYR.

There are several potentially toxic downstream effects resulting from Ca^{2+} influx through Ab channels and excessive Ca^{2+} release from the ER. As discussed earlier, elevated cytosolic Ca^{2+} might lead to activation of CaN and hence to neurite atrophy (Figure 2.8). Excessive Ca^{2+} levels activate calpains, which degrade signaling enzymes involved in learning and memory (Figure 2.8). Old neurons are sensitized to cytosolic Ca^{2+} toxicity because Ca^{2+}-buffering capacity declines with advancing age. Indeed, a tight correlation has been observed between the reduction in expression of Ca^{2+}-binding proteins (CaBPs) in the dentate gyrus region of the hippocampus and onset of cognitive symptoms in AD. The supranormal cytosolic Ca^{2+} signals might cause excessive Ca^{2+} handling by mitochondria and induction of apoptotic cell death (Figure 2.8).

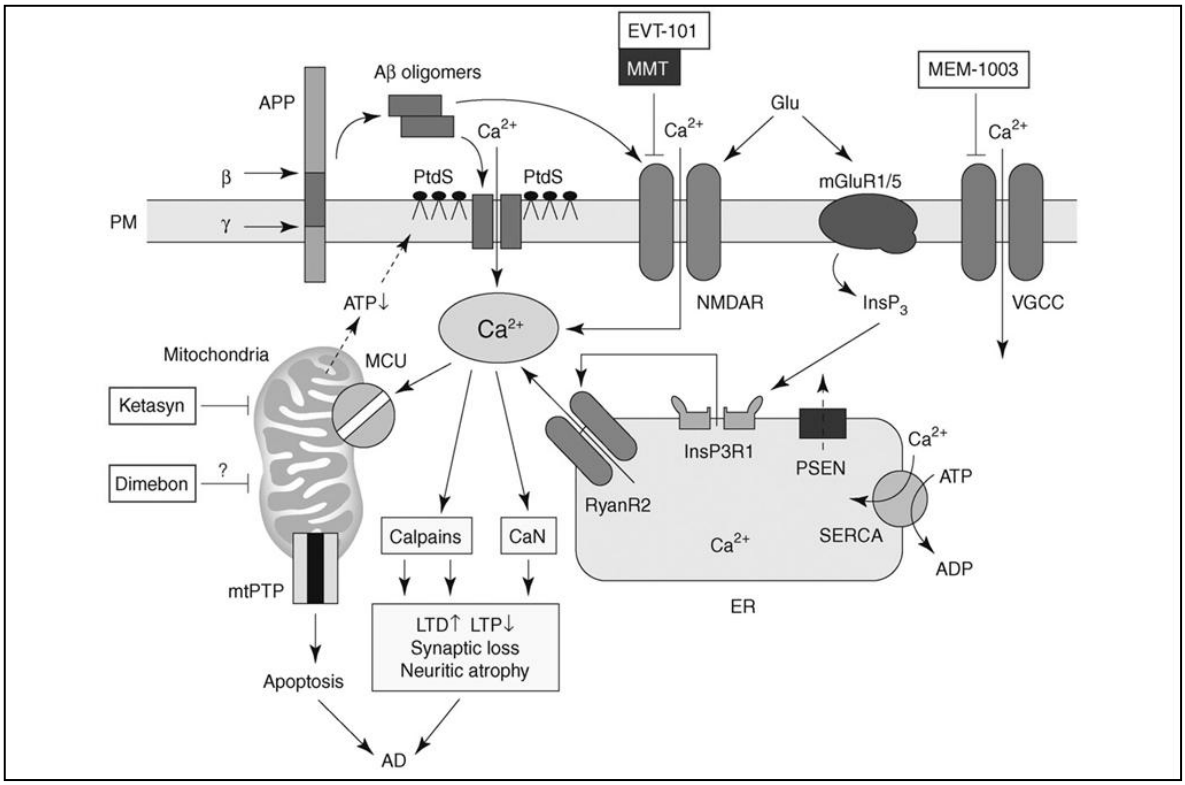


Figure 2.8. The model of Ca^{2+} dysregulation in AD (Bezprozvanny, 2009)

In Figure 2.8, sequential cleavages of b-amyloid precursor protein (APP) by β-secretase (β) and γ-secretase (γ) generate amyloid βpeptide (Aβ). Aβ forms oligomers, which can insert into the plasma membrane and form Ca^{2+}-permeable pores. The association of Aβ oligomers with the plasma membrane is facilitated by binding to surface phosphatidylserine (PtdS); age and Ca^{2+}-related mitochondrial impairment leads to ATP depletion and might trigger flipping of PtdS from the inner portion of the plasma membrane to the cell surface. Reduction in ATP levels and loss of membrane integrity causes membrane depolarization, which leads to facilitation of Ca^{2+} influx through NMDAR and VGCC. Ab oligomers can also affect activity of NMDAR, AMPAR and VGCC directly. Glutamate stimulates activation of mGluR1/5 receptors, production of InsP3 and InsP3-mediated Ca^{2+} release from the ER. Presenilins (PS) function as an ER Ca^{2+}-leak channels and many FAD mutations impair Ca^{2+}-leak-channel function of PS, resulting in excessive accumulation of Ca^{2+} in the ER. Increased ER Ca^{2+} levels result in enhanced Ca^{2+} release through InsP3-gated IP3R1 and Ca^{2+}-gated RYR2. PS might also modulate activity of IP3R, RYR and SERCA pump directly. Elevated cytosolic Ca^{2+} levels result in the activation of calcineurin (CaN) and calpains and lead to facilitation of LTD,

inhibition of LTP, modification of neuronal cytoskeleton, synaptic loss and neuritic atrophy. Excessive Ca^{2+} is taken up by mitochondria through mitochondrial Ca^{2+} uniporter (MCU), eventually leading to opening of mitochondrial permeability-transition pore (mtPTP) and apoptosis.

The NMDAR inhibitor memantine (MMT) is approved for the treatment of AD and the NR2B-specific antagonist EVT-101 was recently developed for AD treatment. 'CNSoptimized' L-type VGCC inhibitor MEM-1003, putative 'mitochondrial agent' Dimebon and 'mitochondrial energizer' Ketasyn are in clinical trials for AD (Bezprozvanny, 2009).

2.3.7.3. Neuronal Ca^{2+} signaling and Parkinson Disease (PD). PD is characterized by motor disturbances, loss of autonomic function, cognitive impairment, and depression. These symptoms are associated with the selective loss of dopaminergic neurons in the substantia nigra pars compacta (SNc) region of the brain as well as less characterized loss of neurons from other regions of the brain. The mechanisms of neuronal loss in PD have yet to be fully elucidated, but important aspects of PD appear to be oxidative stress and mitochondrial dysfunction which may be in part due to environmental toxin exposure (Grant, 2008).

Most of the genes implicated in familial PD (e.g. PINK1, DJ-1, LRRK2 and Parkin) encode proteins associated with certain aspects of mitochondrial function, which points to mitochondria as a crucial locus in PD pathogenesis (Figure 2.9).

The prevalent idea in the PD field is the 'dopamine hypothesis', which states that dopamine (DA) acts as a natural toxin and oxidation of cytosolic DA to 6- hydroxy-DA and other metabolites damages mitochondria (Figure 2.9) and causes cell death of SNc neurons. Consistent with the continuous oxidative damage, affected SNc neurons accumulate large amounts of neuromelanin (NM), which is a lysosome composed of oxyradical DA derivatives of lipids and proteins (Figure 2.9).

The US FDA-approved treatment for PD is administration of levodopa (L-dopa), which is converted to DA and leads to elevation of DA levels in the cytosol and synaptic vesicles of remaining SNc neurons (Figure 2.9). In an apparent contradiction to the 'dopamine hypothesis', administration of L-dopa does not accelerate disease progression in PD patients despite increasing levels of DA in their brain.

Also, PD has low levels of penetrance and most people do not develop PD despite having similar levels of DA in their SNc neurons. These observations point to a 'multi-hit' hypothesis of PD, which states that SNc neurons in PD succumb to a combined effect of DA-related oxidative stress and an additional 'factor X'.

Can Ca^{2+} play the part of 'factor X' in this scenario? Midbrain dopaminergic neurons that express high levels of the CaBP calbindin are relatively spared in PD patients and in animal models. Calpain activation has been observed in sporadic PD and in animal models. A-synuclein is a major component of Lewy bodies observed in the brains of sporadic PD patients and mutations in asynuclein cause autosomal-dominant hereditary PD. The probable mechanism of a-synuclein toxicity is related to the formation of small a-synuclein aggregates (protofibrils). Biophysical studies indicated that synuclein protofibrils generate ion pores in synthetic lipid membranes and induce Ca^{2+} influx in neurons (Figure 2.9). The proposed mechanism of synuclein-mediated Ca^{2+} influx might be similar to that proposed for Ab oligomer-formed Ca^{2+} channels.

Additional support for the 'Ca^{2+} hypothesis' of PD was provided by physiological studies. In contrast to most other neurons in the nervous system, the SNc dopaminergic neurons use CaV1.3 L-type Ca^{2+} channels to drive spontaneous pacemaking activity in the 2–4 Hz range. The continuous Ca^{2+} influx creates an excessive metabolic load on SNc neurons, making them particularly vulnerable to secondary insults on mitochondrial function (Figure 2.9). The reliance of SNc neurons on L-type Ca^{2+} channels to control pacemaking increases with age, which might explain why age is such a significant risk factor for developing PD. Pharmacological inhibition of CaV1.3 L-type Ca^{2+} channels with dihydropyridine isradipine restored Ca^{2+}-independent 'juvenile' pacemaking activity in SNc neurons. Subcutaneous delivery of isradipine significantly protected SNc neurons in

animal models of PD. In support of this linkage to PD, a recent retrospective epidemiological study found that treatment of hypertension with Ca^{2+}-channel antagonists significantly diminished the risk of developing PD. These observations prompted a controlled clinical trial of isradipine in PD patients.

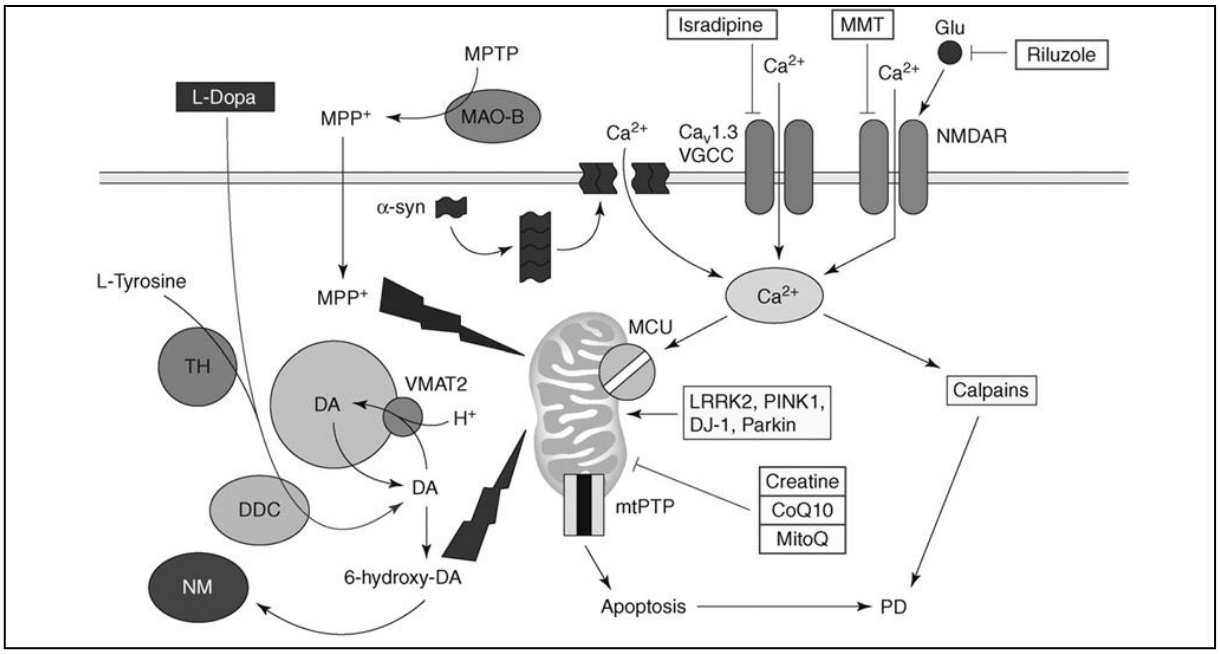


Figure 2.9. The model of Ca^{2+} dysregulation in PD (Bezprozvanny, 2009)

In Figure 2.9, continuous Ca^{2+} influx to SNc neurons is mediated by CaV1.3 L-type voltage-gated Ca^{2+} channels (CaV1.3). In response to glutamate, Ca^{2+} influx is mediated by NMDA receptors (NMDAR). Alpha-synuclein forms aggregates (protofibrils) which may form Ca^{2+}-permeable channels in the plasma membrane. Elevated cytosolic Ca^{2+} is transported into mitochondria through the activity of mitochondrial Ca^{2+} uniporter (MCU). Dopamine (DA) is generated from Ltyrosine by the action of tyrosine hydroxylase (TH) and is loaded into the synaptic vesicles by the activity of the DA/H+ cotransporter (VMAT2). Cytosolic DA is oxidized to 6- hydroxy-DA, which causes damage to proteins and mitochondria by oxidative stress. The products of DA oxidation accumulate as neuromelanin (NM). Cumulative damage to mitochondria resulting from Ca^{2+} overload and DA-mediated oxidative stress leads to an opening of mtPTP and apoptotic cell death of SNc neurons in PD. An importance of mitochondria is highlighted by several mitochondria-related genes (e.g. LRRK2, PINK1, DJ-1, Parkin), which are mutated in familial PD. In a chemical model of PD, the toxin 1-methyl-4-phenyl-1,2,3,6-

tetrahydropyridine (MPTP) is converted to the 1-methyl-4-phenylpyridinium (MPP+) by the glial enzyme monoamine oxidase B (MAO-B). MPP+ enters SNc neurons and potently inhibits mitochondrial complex I, causing selective cell death of SNc neurons. The US FDA-approved treatment for PD is levodopa (L-dopa), which is converted to DA by aromatic L-amino acid decarboxylase (DCC) inside SNc neurons. Generated DA is loaded to synaptic vesicles and alleviates symptoms of PD temporarily. The drugs tested or in PD clinical trials currently are 'mitochondrial stabilizers' (creatine, CoQ10, MitoQ), NMDAR antagonist memantine (MMT), antiglutamate agent riluzole and L-type VGCC inhibitor isradipine.

2.3.7.4. Neuronal Ca^{2+} signaling and ALS. Amyotrophic lateral sclerosis (ALS) is characterized by progressive loss of muscle force, loss of breathing capacity, and limb spasticity. The symptoms result from progressive loss of cortical, spinal, and brainstem motor neurons. Roughly 5-10% of ALS cases are familial in nature, while 20% of cases of the familial form of ALS are caused by mutations in the enzyme superoxide dismutase 1 (SOD-1).

It results from selective degeneration of motor neurons (MNs). The mutations in SOD1 appear to cause disease not by affecting the dismutase activity of SOD1 but rather by inducing aggregation of mutant SOD1 and/or by causing pathological association of mutant SOD1 with mitochondria.

Genetic experiments provide strong evidence in support of non-cell autonomous mechanism of neuronal toxicity in ALS. The emerging model suggests thatMN degeneration in ALS is caused by neuroinflammatory activation of microglia, which attack MNs and neighboring astrocytes (Figure 2.10). The activated microglia release proinflammatory agents, such as tumor necrosis factor-alpha (TNF-a), NO and $O2^-$, which damage MN and neighboring astrocytes (Figure 2.10). Activated microglia also release large amounts of glutamate (Figure 2.10) and elevated levels of glutamate are found in the cerebrospinal fluid (CSF) of ALS patients.

There are several lines of independent evidence supporting an important role for Ca^{2+} signaling in ALS pathology. Calpain activation was observed in MNs from sporadic ALS

patients and in SOD1-mutant mice. Excessive mitochondrial Ca^{2+} accumulation and mitochondrial swelling has been observed in motor nerve terminals from sporadic ALS patients. The spinal and hypoglossal MNs expressing low levels of CaBPs are most vulnerable in ALS, whereas oculomotor neurons expressing high levels of CaBPs are spared. Cross-breeding of SOD1-mutant mice with the mice over-expressing parvalbumin in spinal MNs delayed onset of the symptoms and prolonged survival of these mice. MNs express a high proportion of Ca^{2+}-permeable AMPA receptors, which lack the GluR2 subunit, and pharmacological inhibition of AMPA receptors protected MNs from damage induced by activated microglia in co-culture experiments. D-serine secreted by activated microglia sensitizes NMDAR to glutamate activation and promotes NMDAR-mediated excitotoxicty in ALS. Thus, both AMPAR and NMDAR play a role in glutamate-induced Ca^{2+} overload in ALS (Figure 2.10). In contrast to activated microglia, astrocytes have a largely protective role in ALS excitotoxicity. Astrocytes appear to upregulate the expression of GluR2 subunits in MNs, which reduces AMPAR mediated Ca^{2+} influx. Astrocytes express the glutamate-uptake transporter excitatory amino acid transporter type 2 (EAAT2), which has a major role in clearing glutamate from the extracellular space.

Reduction in the levels of EAAT2 transporting activity was observed in the spinal cord of ALS patients (Figure 2.10). A patient has been identified with a partial loss-of-function mutation in the EAAT2 gene, further supporting an important role of glutamate clearance in ALS pathogenesis

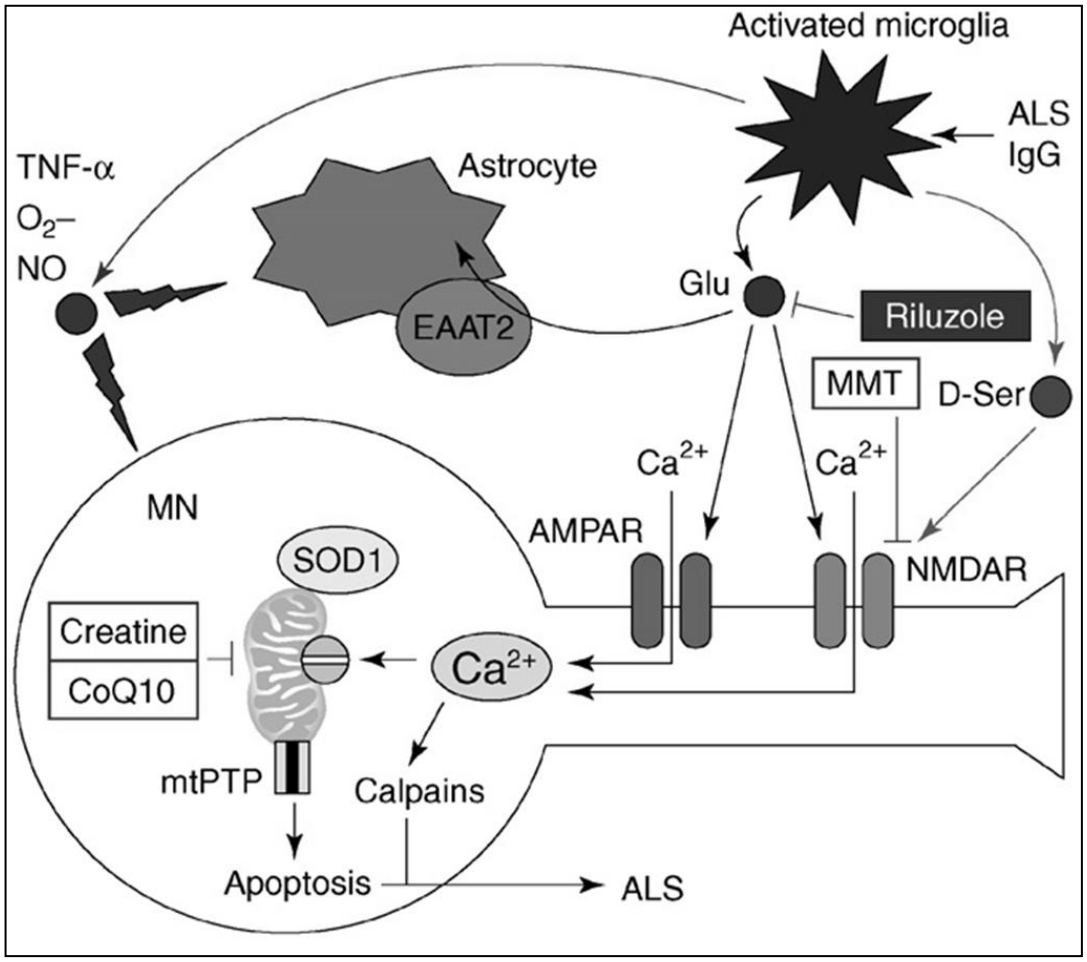


Figure 2.10. The model of excitotoxicity and Ca^{2+} dysregulation in ALS (Bezprozvanny, 2009)

In Figure 2.10, the pathogenic cascade in ALS involves interactions among activated microglia, astrocytes and motor neurons (MNs). In an experimental situation, microglia can be activated by antiserum collected from ALS patients (ALS IgG). Activated microglia release pro-inflammatory factors TNF-a, NO and O2 _. Activated microglia also release large amounts of glutamate, which causes activation of AMPA and NMDA receptors on MNs. Activated microglia also release D-serine, which further sensitizes NMDAR to glutamate activation. Astrocytes express glutamate-uptake transporter EAAT2, which is involved in clearing glutamate from the extracellular space. Ca^{2+} influx by Ca^{2+}-permeable AMPA receptors and NMDA receptors results in mitochondrial Ca^{2+} overload, mitochondrial swelling, opening of mitochondrial permeability-transition pore (mtPTP) and apoptosis of MNs. Mutant SOD1 binds to MN mitochondria and further impairs their ability to handle Ca^{2+} load. Antiglutamate agent riluzole is approved by the US FDA for

treatment of ALS. NMDAR antagonist memantine (MMT) and ‘mitochondrial stabilizers’ creatine and CoQ10 are in ALS clinical trials currently.

2.3.7.5. Neuronal Ca^{2+} signaling and HD. Huntington’s disease (HD) is an autosomal-dominant progressive neurodegenerative disorder characterized by involuntary movement, loss of coordination, cognitive decline and behavioral difficulties such as depression and psychosis. These symptoms are caused by selective loss of GABAergic medium spiny neurons (MSNs) in the striatum.

The disorders discussed already – AD, PD and ALS – are mostly sporadic with rare familial forms. There are some common themes that emerge from the analysis of the role of neuronal Ca^{2+} signaling in these sporadic disorders. These disorders are ‘multi-hit’ and will probably require a combination therapy, with Ca^{2+} inhibitors included as a part of the treatment regimen.

In contrast to these disorders, HD is a purely genetic disorder that is caused by a single mutation – CAG repeat (polyglutamine) expansion in the huntingtin (Htt) gene. The medium spiny striatal neurons (MSNs) are most affected in HD. Most researchers agree that mutant Httexp protein acquires a ‘toxic gain of function’. Destabilization of neuronal Ca^{2+} signaling is one of the toxic functions of the Httexp protein. Consistent changes in the expression levels of many Ca^{2+} signaling proteins were observed in microarray studies of the brains from HD patients and also from HD mouse models.

There are several points of Httexp interference with MSN Ca^{2+} signaling (Figure 2.11). Httexp binds directly and specifically to the IP3R1 C-terminal region, and the association of Httexp with IP3R1 was confirmed independently in an unbiased screen. Binding to Httexp increases sensitivity of IP3R1 to activation by InsP3. The importance of IP3R1 activation for Httexp neurotoxicity was validated in pharmacological experiments with MSN cultures from a HD mouse model and in genetic experiments with a Drosophila HD model.

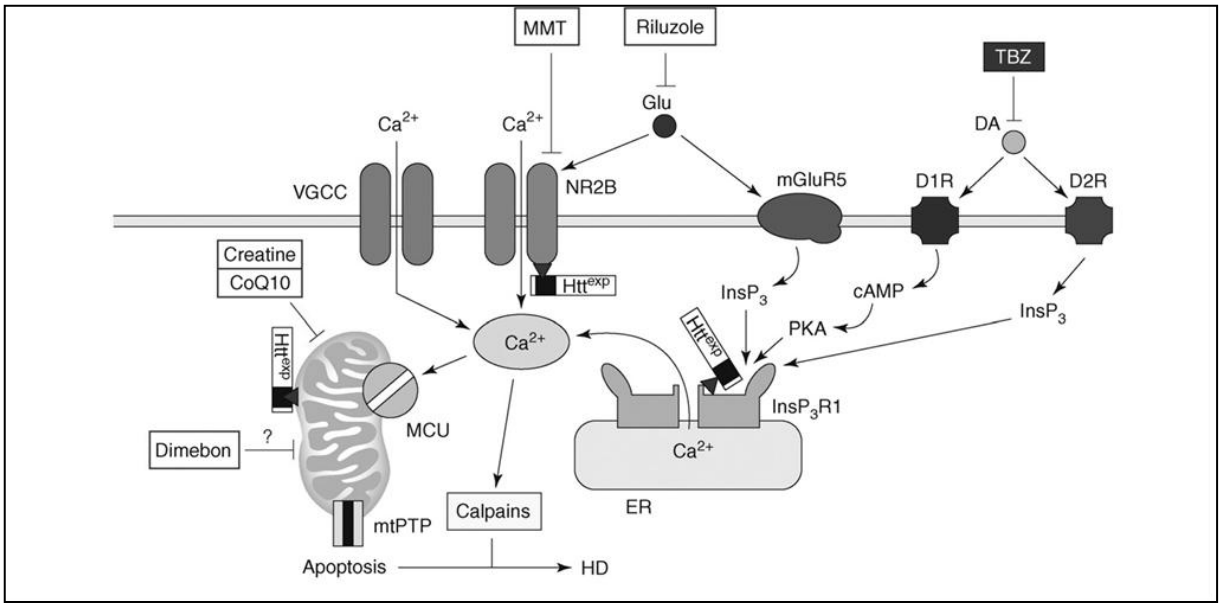


Figure 2.11. The model of Ca^{2+} dysregulation in HD (Bezprozvanny, 2009)

In HD MSN (Figure 2.11), the Httexp perturbs Ca^{2+} signaling through several synergistic mechanisms. Httexp enhances function of NR2B-containing MDAR, probably by promoting trafficking to the plasma membrane. Httexp binds strongly to the IP3R1C terminus and sensitizes the IP3R1 to activation by InsP3. The low levels of glutamate released from corticostriatal projection neurons lead to supranormal Ca^{2+} influx by NMDAR and Ca^{2+} release through the IP3R1. Additional Ca^{2+} influx to MSN is mediated by voltage-gated Ca^{2+} channels (VGCCs). Dopamine released from midbrain dopaminergic neurons stimulates D1-class and D2-class DARs, which are expressed abundantly in MSNs. D1-class DARs are coupled to activation of adenyl cyclase, increase in cAMP levels and activation of PKA. PKA potentiates glutamate-induced Ca^{2+} signals by facilitating the activity of NMDAR and IP3R1. D2 receptors are coupled directly to InsP3 production and activation of IP3R1. Supranormal Ca^{2+} signals activate calpain, which cleave Httexp and other substrates. Excessive cytosolic Ca^{2+} signals result in mitochondrial Ca^{2+} uptake by MCU, which eventually triggers mtPTP opening and apoptosis. The mitochondrial Ca^{2+} handling is further destabilized by direct association of Httexp with mitochondria. Antidopamine agent tetrabenazine (TBZ) is approved by the US FDA for symptomatic treatment of HD. NMDAR antagonist memantine (MMT), putative ‘mitochondrial agent’ Dimebon and ‘mitochondrial stabilizers’ creatine and CoQ10 are tested in HD clinical trials. Antiglutamate agent riluzole was tested and failed.

2.4. Calcium Related Domains in General

The EF hand Domain: is a helix-loop-helix structural domain found in a large family of calcium-binding proteins. The EF-hand motif contains a helix-loop-helix topology, much like the spread thumb and forefinger of the human hand, in which the Ca^{2+} ions are coordinated by ligands within the loop. It consists of two alpha helices positioned roughly perpendicular to one another and linked by a short loop region (usually about 12 amino acids) that usually binds calcium ions.

Upon binding to Ca^{2+}, this motif may undergo conformational changes that enable Ca^{2+}-regulated functions as seen in Ca^{2+} effectors such as calmodulin (CaM) and troponin C (TnC) and Ca^{2+} buffers such as calreticulin and calbindin D9k. The majority of the known EF-hand Calcium-binding proteins (CaBPs) contain paired EF-hand motifs. CaBP's with single EF hands have also been discovered in both bacteria and eukaryotes. In addition, "EF-hand-like motifs" have been found in a number of bacteria.

The EH Domain: The EH (for Eps15 Homology) domain is a protein-protein interaction module of approximately 95 residues which was originally identified as a repeated sequence present in three copies at the N-terminus of the tyrosine kinase substrates Eps15 and Eps15R. The EH domain was subsequently found in several proteins implicated in endocytosis, vesicle transport and signal transduction in organisms ranging from yeast to mammals. EH domains are present in one to three copies and they may include calcium-binding domains of the EF-hand type.

The structure of several EH domains has been solved by NMR spectroscopy. The fold consists of two helix-loop-helix characteristic of EF-hand domains, connected by a short antiparallel beta-sheet. The target peptide is bound in a hydrophobic pocket between two alpha helices. Sequence analysis and structural data indicate that not all the EF-hands in the EH domain are capable of binding calcium because of substitutions of the calcium-liganding residues in the loop. This domain is often implicated in the regulation of protein transport/sorting and membrane trafficking.

The EGF_Ca Domain: A sequence of about forty amino-acid residues found in epidermal growth factor (EGF) has been shown to be present in a large number of membrane-bound and extracellular, mostly animal, proteins. Many of these proteins require calcium for their biological function and a calcium-binding site has been found at the N-terminus of some EGF-like domains. Calcium-binding may be crucial for numerous protein-protein interactions.

The C2 Domain: The C2 domain is a Ca^{2+}-dependent membrane-targeting module found in many cellular proteins involved in signal transduction or membrane trafficking. C2 domains are unique among membrane targeting domains in that they show wide range of lipid selectivity for the major components of cell membranes, including phosphatidylserine and phosphatidylcholine.

This C2 domain is about 116 amino-acid residues and is located between the two copies of the C1 domain in Protein Kinase C (that bind phorbol esters and diacylglycerol) and the protein kinase catalytic domain. Regions with significant homology to the C2-domain have been found in many proteins. The C2 domain is thought to be involved in calcium-dependent phospholipid binding and in membrane targetting processes such as subcellular localisation.

Na-Ca exchanger Domain: This is a family of sodium/calcium exchanger integral membrane proteins. This family covers the integral membrane regions of the proteins. Sodium/calcium exchangers regulate intracellular Ca^{2+} concentrations in many cells; cardiac myocytes, epithelial cells, neurons retinal rod photoreceptors and smooth muscle cells. Ca^{2+} is moved into or out of the cytosol depending on Na+ concentration. In humans and rats there are 3 isoforms; NCX1 NCX2 and NCX3.

Calcipressin: Calcipressin is also known as calcineurin-binding protein, since it inhibits calcineurin-mediated transcriptional modulation by binding to calcineurin's catalytic domain.

Calcipressin 1 negatively regulates calcineurin by direct binding and is essential for the survival of T helper type 1 cells. Calcipressin 1 is a phosphoprotein that increases its

capacity to inhibit calcineurin when phosphorylated, and this phosphorylation also controls the half-life of calcipressin 1 by accelerating its degradation.

Calcineurin is a calcium-responsive enzyme that dephosphorylates the nuclear factor of activated T cells (NFAT). In so doing it promotes its nuclear translocation and uniquely links calcium signalling to transcriptional regulation. Calcipressins are a family of proteins derived from three genes. Calcipressin 1 is also known as modulatory calcineurin-interacting protein 1 (MCIP1), Adapt78 and Down syndrome critical region 1 (DSCR1). Calcipressin 2 is variously known as MCIP2, ZAKI-4 and DSCR1-like 1. Calcipressin 3 is also called MCIP3 and DSCR1-like 2. DSCR1 (Adapt78) is associated with successful adaptation to oxidative stress and calcium stress as well as with diseases like Alzheimer's and Down syndrome.

The DSCR1 (Adapt78) isoform 1 protein, calcipressin 1, inhibits calcineurin and protects against acute calcium-mediated stress damage, including transient oxidative stress. Calcipressin 1 is encoded by DSCR1, a gene on human chromosome 21. Calcipressin 1 isoform 1 has an N-terminal coding region, which generates a new polypeptide of 252 amino acids. Endogenous calcipressin 1 exists as a complex together with the calcineurin A and B heterodimer.

Calreticulin: Calreticulin is a high-capacity calcium-binding protein which is present in most tissues and located at the periphery of the endoplasmic (ER) and the sarcoplamic reticulum (SR) membranes. It probably plays a role in the storage of calcium in the lumen of the ER and SR and it may well have other important functions.

Structurally, calreticulin is a protein of about 400 amino acid residues consisting of three domains:

- An N-terminal, probably globular, domain of about 180 amino acid residues (N-domain). A central domain of about 70 residues (P-domain) which contains three repeats of an acidic 17 amino acid motif. This region binds calcium with a low-capacity, but a high-affinity.

- A C-terminal domain rich in acidic residues and in lysine (C-domain). This region binds calcium with a high-capacity but a low-affinity.

Calreticulin is evolutionarily related to several other calcium-binding proteins, including Onchocerca volvulus antigen RAL-1, calnexin and calmegin.

Neutral trehalase Ca^{2+} binding domain: Neutral trehalases mobilise trehalose accumulated by fungal cells as a protective and storage carbohydrate. This family represents a calcium-binding domain similar to EF hand. Residues 97 and 108 in O42893 have been implicated in this interaction. It is thought that this domain may provide a general mechanism for regulating neutral trehalase activity in yeasts and filamentous fungi.

To conclude, presence of these domains in the protein sequence indicates the probable binding of calcium to those domain sites implying that the protein may be involved in a calcium-related function or process.

3. MATERIALS & METHODS

For the reconstruction and macro/micro scale analyses of Calcium signaling network, various databases and tools are used besides the algorithms compiled throughout the thesis.

3.1. Databases

3.1.1. Interactome Sources

BIOGRID: The Biological General Repository for Interaction Datasets (BioGRID) database is developed to house and distribute collections of protein and genetic interactions from major model organism species. BioGRID (Stark *et al.*, 2006) currently contains over 330,000 interactions from 17 different species, as derived from both high-throughput studies and conventional focused studies.

BioGRID uses a wide range of gene identifiers to describe all biological interactors. For each experiment supporting a particular interaction, the interacting partners are recorded as a pair of gene identifiers. For physical interaction evidence this means that the article describes an experiment which supports an interaction between a product of the bait gene and a product of the prey (hit) gene. Genetic interaction evidence codes record experiments where the phenotypes of two strains, each having a perturbation of one of the interacting genes, are compared with the phenotype of a strain with both genes perturbed. Self-interactions are recorded, as are reciprocal interactions if the bait-prey directionality is clear.

As well as small-scale curated experiments, BioGRID contains data from papers reporting high throughput screens. Since the experimental setup of high throughput screens for interactions can be adjusted to give variable numbers of “hits”, it is common for publications to report a confidence score for each interaction. Before loading any high throughput data into BioGRID, the data is assessed and interactions considered low confidence by the authors are not loaded. Generally, the confidence score used by the author is loaded, and on BioGRID gene pages containing interactions from that publication

the scores can be displayed by choosing the corresponding publication. In addition, the number of interactions in BioGRID for each publication is displayed next to that publication. (See App A.1 for Genetic and Physical Experimental Methods)

DIP: The DIP – Database of Interacting Proteins database (Xenarios *et al.*, 2000) catalogs experimentally determined interactions between proteins. It combines information from a variety of sources to create a single, consistent set of protein-protein interactions. The data stored within the DIP database were curated, both, manually by expert curators and also automatically using computational approaches that utilize the knowledge about the protein-protein interaction networks extracted from the most reliable, core subset of the DIP data.

The key tables - PROTEIN, SOURCE and EVIDENCE – store information on individual proteins, sources of experimental information and information on individual experiments respectively. The information on protein–protein interactions is stored in two tables - INTERACTION and INT_PRT. Such arrangement of the tables enables description of binary interactions but also of multi-protein complexes. The METHOD table provides a list of controlled vocabulary terms, together with references to the corresponding PSI ontology entries, which are used to annotate the experiments.

3.1.2. Sequence, Structure & Annotation Sources

SGD: SGD – Saccharomyces Genome Database (Cherry *et al.*, 1998) is an organized collection of genetic and molecular biological information about *Saccharomyces cerevisiae*, bakers' and brewers' yeast. It contains the sequences of yeast genes and proteins; descriptions and classifications of their biological roles, molecular functions, and subcellular localizations; links to literature information; links to functional genomics datasets; and tools for analysis and comparison of sequences.

The basic unit of SGD is the Locus Page. Each yeast gene or open reading frame has an individual Locus Page. Genetic loci that are not tied to a DNA sequence and chromosomal features (for example, centromeres) also have Locus Pages. All information

relevant to a particular locus is either presented on or linked to the Locus Page. Tools for analysis and comparison of the locus are also accessible from the Locus Page.

Gene products in SGD are classified according to the Gene Ontology (GO) system of classification. Each gene product is assigned GO terms that describe its molecular function and subcellular location, and the biological process in which it is involved. The terms may be searched to find the set of proteins that have been annotated to a particular term.

UniProt: UniProt - The Universal Protein Resource (Apweiler *et al.*, 2004) is a comprehensive resource for protein sequence and annotation data. The UniProt databases are the UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef), and the UniProt Archive (UniParc).

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added. This includes widely accepted biological ontologies, classifications and cross-references, and clear indications of the quality of annotation in the form of evidence attribution of experimental and computational data.

The UniProt Knowledgebase consists of two sections: a section containing manually-annotated records with information extracted from literature and curator-evaluated computational analysis, and a section with computationally analyzed records that await full manual annotation. For the sake of continuity and name recognition, the two sections are referred to as "UniProtKB/Swiss-Prot" (reviewed, manually annotated) and "UniProtKB/TrEMBL" (unreviewed, automatically annotated), respectively.

CYGD: MIPS - Munich Information Center for Protein Sequences mainly focuses on the genome-oriented bioinformatics, in particular the systematic analysis of genome information including the development and application of bioinformatics methods in genome annotation, expression analysis and proteomics. MIPS supports and maintains a

set of generic databases as well as the systematic comparative analysis of microbial, fungal, and plant genomes.

CYGD - The MIPS Comprehensive Yeast Genome Database (Güldener *et al.*, 2005) aims to present information on the molecular structure and functional network of the entirely sequenced, well-studied model eukaryote, the budding yeast *Saccharomyces cerevisiae*. In addition the data of various projects on related yeasts are used for comparative analysis.

The Gene Ontology: The Gene Ontology (GO) project (Blake and Harris, 2002) is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The GO collaborators are developing three structured, controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes i.e. operations or sets of molecular events, cellular components i.e. the parts of a cell or its extracellular environment and molecular functions i.e. the elemental activities of a gene product at the molecular level, such as binding or catalysis in a species-independent manner.

A GO annotation consists of a GO term associated with a specific reference that describes the work or analysis upon which the association between a specific GO term and gene product is based. Each annotation must also include an evidence code to indicate how the annotation to a particular term is supported. (See App A.2 for Evidence Codes)

Out of all the evidence codes available, only Inferred from Electronic Annotation (IEA) is not assigned by a curator. Manually-assigned evidence codes fall into four general categories: experimental, computational analysis, author statements, and curatorial statements.

Use of an experimental evidence code in a GO annotation indicates that the cited paper displayed results from a physical characterization of a gene or gene product that has supported the association of a GO term.

PDB: The PDB –Protein Data Bank archive (Laskowski *et al.,* 1997) is a repository of atomic coordinates and other information describing proteins and other important biological macromolecules. Structural biologists use methods such as X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy to determine the location of each atom relative to each other in the molecule. They then deposit this information, which is then annotated and publicly released into the archive by the PDB.

The primary information stored in the PDB archive consists of coordinate files for biological molecules. These files list the atoms in each protein, and their 3D location in space. These files are available in several formats (PDB, mmCIF, XML). A typical PDB formatted file includes a large "header" section of text that summarizes the protein, citation information, and the details of the structure solution, followed by the sequence and a long list of the atoms and their coordinates. The archive also contains the experimental observations that are used to determine these atomic coordinates.

The PDBsum is a pictorial database that provides an at-a-glance overview of the contents of each 3D structure deposited in the Protein Data Bank (PDB).

It shows the molecule(s) that make up the structure (i.e. protein chains, DNA, ligands and metal ions) and schematic diagrams of their interactions. Extensive use is made of the freely available RasMol molecular graphics program to view the molecules and their interactions in 3D.

PFAM: Proteins are generally comprised of one or more functional regions, commonly termed domains. The presence of different domains in varying combinations in different proteins gives rise to the diverse repertoire of proteins found in nature. Identifying the domains present in a protein can provide insights into the function of that protein.

The Pfam database (Sonnhammer *et al.*, 1997) is a large collection of protein domain families. Each family is represented by multiple sequence alignments and hidden Markov models (HMMs).

There are two levels of quality to Pfam families: Pfam-A and Pfam-B. Pfam-A entries are derived from the underlying sequence database, known as Pfamseq, which is built from the most recent release of UniProtKB at a given time-point. Each Pfam-A family consists of a curated seed alignment containing a small set of representative members of the family, profile hidden Markov models (profile HMMs) built from the seed alignment, and an automatically generated full alignment, which contains all detectable protein sequences belonging to the family, as defined by profile HMM searches of primary sequence databases.

Pfam-B families are un-annotated and of lower quality as they are generated automatically from the non-redundant clusters of the latest release. Although of lower quality, Pfam-B families can be useful for identifying functionally conserved regions when no Pfam-A entries are found.

Pfam entries are classified in one of four ways:

Family: A collection of related proteins

Domain: A structural unit which can be found in multiple protein contexts

Repeat: A short unit which is unstable in isolation but forms a stable structure when multiple copies are present

Motifs: A short unit found outside globular domains

Related Pfam entries are grouped together into clans; the relationship may be defined by similarity of sequence, structure or profile-HMM.

Human Protein Reference Database – HPRD: HPRD (Peri *et al.*, 2004) contains annotations pertaining to human proteins based on experimental evidence from the literature. This includes PPIs as well as information about post-translational modifications, sub cellular localization, protein domain architecture, tissue expression and association with human diseases. In addition to interactions of proteins with other proteins, HPRD also reports interactions of proteins with nucleic acids and small molecules. The PPI data is sub classified as binary or complex interactions based on topology and the number of participants. Binary PPIs are direct interactions between two proteins while complexes

represent interactions with more than 2 participants and the topology of interaction is unknown. Relevant publications are cited for each interaction.

The type of experiment is also indicated as *in vivo* (e.g. coimmunoprecipitation), *in vitro* (e.g. GST pull-down assays) or yeast two-hybrid. They consider two proteins to be interacting *in vivo* when, through the experiment performed, it can be assured that the proteins interacting are truly interacting in a mammalian cell. An interaction occurs *in vitro* when the experiment was performed in such a way that one cannot conclude it is happening in the context of a mammalian cell. Because Yeast 2-hybrid and similar experiments test interactions in the context of a yeast cell, they are technically *in vivo*, but HPRD have annotated them specifically as Yeast 2-hybrid because of the greater degree of false positive interactions seen with this method.

Human ProteinPedia – HUPA: Human Proteinpedia (Prasad *et al.*, 2009) is a community portal for sharing and integration of human protein data. It allows research laboratories to contribute and maintain protein annotations. Human Protein Reference Database (HPRD) integrates data that is deposited in Human Proteinpedia along with the existing literature curated information in the *context* of an individual protein.

Protein annotations present in Human Proteinpedia are derived from a number of platforms such as

- Co-immunoprecipitation and mass spectrometry-based protein-protein interaction
- Co-immunoprecipitation and Western blotting based protein-protein interaction
- Fluorescence based experiments
- Immunohistochemistry
- Mass Spectrometric Analysis
- Protein and peptide microarray
- Western blotting
- Yeast two-hybrid based protein-protein interaction

3.2. Tools

3.2.1. Sequence & Structure Related Tools

BLAST & BLASTp: BLAST - The Basic Local Alignment Search Tool (Altschul *et al.*, 1997) is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

To run, BLAST requires a query sequence to search for, and a sequence to search against (also called the target sequence) or a sequence database containing multiple such sequences. BLAST will find subsequences in the database which are similar to subsequences in the query.

The main idea of BLAST is that there are often high-scoring segment pairs (HSP) contained in a statistically significant alignment. BLAST searches for high scoring sequence alignments between the query sequence and sequences in the database using a heuristic approach.

Standard protein-protein BLAST (blastp) is used for both identifying a query amino acid sequence and for finding similar sequences in protein databases. Like other BLAST programs, blastp is designed to find local regions of similarity. When sequence similarity spans the whole sequence, blastp will also report a global alignment, which is the preferred result for protein identification purposes.

Conseq & Consurf: The degree to which an amino (or nucleic) acid position is evolutionarily conserved is strongly dependent on its structural and functional importance. Thus, conservation analysis of positions among members from the same family can often reveal the importance of each position for the protein (or nucleic acid)’s structure or function.

ConSeq is a web server for the identification of structurally and functionally important residues in protein sequences where the structure of the protein is not known. Given a set of homologous proteins in the form of multiple sequence alignment (MSA), the evolutionary rate at each amino acid site in the MSA is calculated.

The slowly evolving sites are often biologically important. To determine which of these sites are important for maintaining the protein structure and which are functionally important, the MSA is used to predict the relative solvent accessibility state of each site (i.e. buried vs. exposed). The basic assumption is that functionally important residues, which for example take part in ligand binding, DNA binding and protein-protein interactions, are often evolutionarily conserved and are most likely to be solvent accessible, whereas highly conserved residues within the protein core are likely to have an important structural role in maintaining the protein's fold. However it should be emphasized that the discrimination between the functional and structural residues ("functional" = highly conserved and exposed, whereas "structural" = highly conserved and buried) might be problematic in cases where a certain residue has both functional role and structural roles.

The conservation scores calculated by ConSeq are a relative measure of evolutionary conservation at each sequence site of the query sequence. The lowest score represents the most conserved position in a protein. It does not necessarily indicate 100% conservation (e.g. no mutations at all), but rather indicates that this position is the most conserved in this specific protein calculated by the use of a specific MSA.

ConSeq is now integrated in ConSurf server. The ConSurf server (Ashkenazy *et al.*, 2010) is a tool that enables the identification of functionally important regions on the surface of a protein or domain, of known three-dimensional (3D) structure, based on the phylogenetic relations between its close sequence homologues.

3.2.2. Topology & Visualization Related Tools

Cytoscape is an open source bioinformatics software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. Although Cytoscape was originally designed for biological research, now it is a general platform for complex network analysis and visualization. Cytoscape core distribution provides a basic set of features for data integration and visualization. Additional features are available as plugins. Plugins are available for network and molecular profiling analyses, new layouts, additional file format support, scripting, and connection with databases.

Network Analyzer: Network Analyzer is a Cytoscape plugin performing analysis of biological networks and calculating network topology parameters including the diameter of a network, the average number of neighbors, and the number of connected pairs of nodes. It also computes the distributions of more complex network parameters such as node degrees, average clustering coefficients, topological coefficients, and shortest path lengths. It displays the results in diagrams, which can be saved as images or text files.

Network Analyzer can perform topological analysis on directed networks (containing only directed edges) as well as on undirected networks (containing only undirected edges).

BINGO: BiNGO (Maere *et al.*, 2005) is a Java-based tool to determine which Gene Ontology (GO) categories are statistically overrepresented in a set of genes or a subgraph of a biological network. BiNGO is implemented as a plugin for Cytoscape, which is an open source bioinformatics software platform for visualizing and integrating molecular interaction networks.

BiNGO maps the predominant functional themes of a given gene set on the GO hierarchy, and outputs this mapping as a Cytoscape graph. Gene sets can either be selected or computed from a Cytoscape network (as subgraphs) or compiled from sources other than Cytoscape (e.g. a list of genes that are significantly upregulated in a microarray experiment). The main advantage of BiNGO over other GO tools is the fact that it can be used directly and interactively on molecular interaction graphs. Another plus is that

BiNGO takes full advantage of Cytoscape's versatile visualization environment. This allows the user to produce customized high-quality figures.

CytoHubba: Cyto-Hubba is a Java plugin for Cytoscape, a facilitated platform for the analysis and visualization of molecular interaction networks (Lin *et al.*, 2008).

Cyto-Hubba provides a user-friendly interface to analyze topology of protein-protein interaction networks, such as human, yeast, fly etc. Users can obtain not only the identified nodes with scores from calculation but the subnetwork composed of these nodes by utilizing cyto-Hubba in Cytoscape.

In this plugin, nodes are explored by different analysis methods on protein-protein interaction network, including: Degree, the Maximum Neighborhood Component (MNC), which is composed of a target node and its neighbors, and also the Density of Maximum Neighborhood Component (DMNC).

- Degree (Jeong *et al.*,2001): In this method, the score of a node is its degree, the number of links incident to this node.

- Maximum Neighborhood Component (MNC): The neighborhood of a node v, nodes adjacent to v, induce a subnetwork N(v). The score of node v, MNC(v), is defined to be the size of the maximum connected component of N(v).

- Density of Maximum Neighborhood Component (DMNC): For a node v, let N be the node number and E be the edge number of MNC(v) respectively. The score of node v, DMNC(v), is defined in Eq. 3.1 for some e between 1 and 2. It may be assumed that the maximum neighborhood component has a strong community structure.

$$DMNC(v) = E / N^{e} \tag{3.1}$$

- Double Screening Scheme (DSS - DMNC||MNC): Each scoring method catches certain postulated topological characteristic of essential proteins. Therefore, a DSS is proposed. That is to say, two scoring methods A and B, are used to extract mixed characteristic of essential proteins. For *n*, most possible essential proteins are expected in the output, the 2*n* top ranked proteins by method A are selected firstly. The selected 2*n* proteins are further ranked by method B and the *n* top ranked proteins are output. The number 2*n* is an empirical value for this double screening method.

Among all methods, DMNC is found to be the one that shares the least proteins with the others. Accordingly, the topological characters extracted by DMNC may differ from those by the other methods. Second, Lin *et al.* (2008) evaluate the performance of the six scoring method by the coverage of yeast essential proteins. They found out that DMNC has the highest hit rate on the essential protein list. Therefore, they choose DMNC as the first method in the DSS. The second method of the double screen scheme is chosen on the same criteria. Among the five methods, MNC is the best mate of DMNC.

The results Lin *et al.* obtained are tabulated in Table 3.1.

Table 3.1. The percentage of top n ranked proteins, identified from the yeast protein interaction dataset (yeast-2007/01/07) by three scoring methods and by the double screening scheme (DMNC || MNC, Hubba algorithm), to be yeast essential proteins.

	Scoring methods			DSS
	Degree (%)	MNC (%)	DMNC (%)	DMNC || MNC (%)
Top10	60.00	40.00	80.00	100.00
Top20	35.00	50.00	80.00	95.00
Top30	43.33	46.67	76.67	83.33
Top40	45.00	45.00	70.00	72.50
Top50	50.00	48.00	66.00	76.00
Top60	50.00	46.67	61.67	70.00
Top70	48.57	48.57	57.14	67.14
Top80	47.50	48.75	55.50	65.50
Top90	47.78	47.78	56.67	63.33
Top100	48.00	48.00	58.00	63.00

MCODE: MCODE – Molecular Complex Detection plugin for Cytoscape finds clusters (highly interconnected regions) in a network. Clusters mean different things in different types of networks. For instance, clusters in a protein-protein interaction network are often protein complexes and parts of pathways, while clusters in a protein similarity network represent protein families. MCODE also makes the visualization of large networks manageable by extracting the dense regions around a protein of interest.

MCODE (Bader and Hogue, 2003) consists of three stages: vertex weighting, complex prediction and an optional post-processing step. The weighting of nodes is based on the core clustering coefficient. Bader and Hogue (2003) propose the use of this coefficient instead of the standard clustering coefficient, as it increases the weights of heavily interconnected graph regions while giving small weights to the less connected vertices, which are abundant in the scale-free protein interaction networks. Once the weights are computed, the algorithm traverses the weighted graph in a greedy fashion to isolate densely connected regions. The post-processing step filters or adds proteins based on connectivity criteria. Complexes are filtered if they do not contain at least a 2-core (graph of minimum degree 2). The algorithm may be run with the 'fluff' option, which increases the size of the complex according to a given 'fluff' parameter between 0.0 and 1.0. For every vertex in the complex, *v*, its neighbors are added to the complex if they have not yet been seen and if the neighborhood density (including *v*) is higher than the given fluff parameter. Vertices that are added by the fluff parameter are not marked as seen, so there can be overlap among predicted complexes with the fluff parameter set. If the algorithm is run using the 'haircut' option, the resulting complexes are 2-cored, thereby removing the vertices that are singly connected to the core complex. If both options are specified, fluff is run first, then haircut.

MCODE has been used in several recent publications describing mapping of large-scale interaction networks. MCODE performed best on random data, by detecting the fewest false positives, owing to its ability to report complexes covering only part of the network, rather than partitioning all the nodes into complexes.

3.3. Selective Permissibility Algorithm (SPA)

SPA is a novel algorithm (Arga *et al..*, 2007) that integrates protein–protein interaction data with GO annotations (www.geneontology.org) to reconstruct a protein interaction network, which is composed of candidate proteins for signal transduction mechanisms in yeast S. cerevisiae. GO annotations are used to curate data from protein–protein interaction databases resulting in a protein interaction network that is composed of the candidate proteins for the yeast signaling pathways.

The algorithm (Figure 3.1) has two distinctive elements, the input protein(s) and the selection criterion. The input to the algorithm is a protein (or a set of proteins) known to have a certain function in the yeast signaling pathways. In the first step, all interactions of the protein(s) are extracted from interactome data sources. In the next step, the relevance of these interacting proteins is tested through a selection criterion which is designed to include the GO annotations.

A candidate protein is included into the network if all of the three GO annotations (component/function/process) of the protein are present in the Annotation Collection. Interacting proteins are either accepted or rejected on this basis, and in the next step of the algorithm the accepted proteins are used as input proteins and the cycle is repeated until the whole interaction data source is scanned, that is, no new interacting partners are identified.

This approach can be used to (i) identify unknown members of documented pathways, (ii) identify network modules consisting of a group of related components that often incorporate the same functional mechanism, or (iii) identify smaller correlated clusters of interacting proteins; and (iv) with further extension, to predict new signaling pathways.

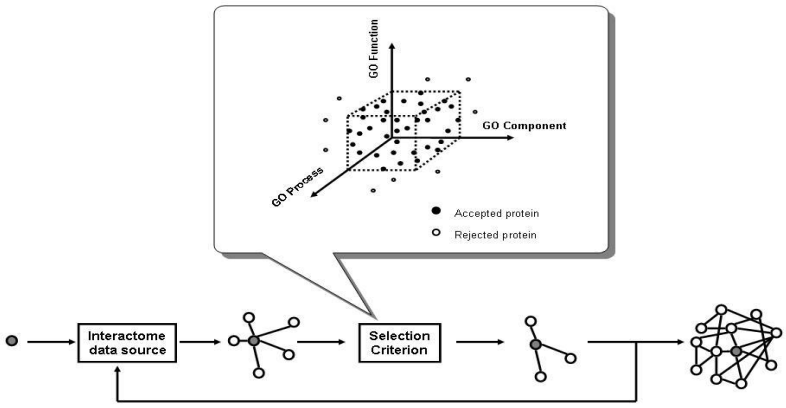


Figure 3.1. The geometrical representation of the selection step of SPA. The Annotation Collection represents a rectangular subspace, S (X, Y, Z), where the coordinates are the GO annotations in terms of component, function and process, respectively. Any protein is described as a point in the 3-D space, p(x,y,z). If the point p is inside the rectangular subspace (S), the corresponding protein is selected. Else, the protein is rejected. (Arga *et al.*., 2007)

3.4. Graph Theory

A major motivation for the study of biological networks is the need for tailored analysis methods which can extract meaningful biological information from the data becoming available through the efforts of experimentalists. Knowledge of the topologies of biological networks, and of their impact on biological processes is needed if one is to fully understand, and develop more sophisticated treatment strategies for, complex diseases such as cancer. The mathematical discipline which underpins the study of complex networks in Biology is graph theory (Grindrod *et al.*., 2004).

3.4.1. Degree distribution and scale-free property

Degree *k* is the elementary character of a node, accounting for the number of other nodes linked to this node. The mean degree of a network, $\langle k \rangle$, is defined as the average number of edges that a node has. For an undirected network with N nodes and L edges, the mean degree can be characterized by Equation 3.2.

$$\langle k \rangle = \frac{2L}{N} \tag{3.2}$$

A normalized version of this parameter is the network density. The density is a value between 0 and 1. It shows how densely the network is populated with edges (self-loops and duplicated edges are ignored). A network which contains no edges and solely isolated nodes has a density of 0. In contrast, the density of a clique is 1.

The degree distribution *P(k)* represents probability of the nodes that have exact *k* links, which may define the type of the network. $P(k)$ can be obtained by counting the number of nodes n(k) with k=1,2,… edges and dividing by the total number of nodes, N.

$$P(k) = \frac{n_k}{N} \tag{3.3}$$

In many real life networks, the degree distribution has no well-defined peak but has a power-law distribution, $P(k) \sim k^{-\gamma}$, where γ is a constant. Such networks are known as scale-free network. (Figure 3.3) The power-law form of the degree distribution implies that the networks are extremely inhomogeneous.

In the scale-free network, there are many nodes with few links and a few nodes with many links. The highly connected proteins (hubs) could possibly play a key role in the functionality of the network.

Biological networks tend to be very heterogeneous: while some 'hub' nodes are highly connected, the majority of nodes tend to have very few connections. Heterogeneity is another measure of scale-free network characteristics reflecting the tendency of a network to contain hub nodes.

It has been noted that the scale-free structure has implications for the robustness and vulnerability of networks to failure and attack. Specially, while removing most of the nodes in a scale-free network will have little effect on the network's connectivity, the targeted removal of hub nodes can disconnect the network relatively easily. This has led to

the suggestion that genes or proteins which are involved in a large number of interactions, corresponding to hub nodes, may be more important for an organism's survival than those of low degree.

The degree exponent γ being lower when compared to other networks implies that this network's property is mostly dominated by the highly connected proteins (Przulj N., 2005).

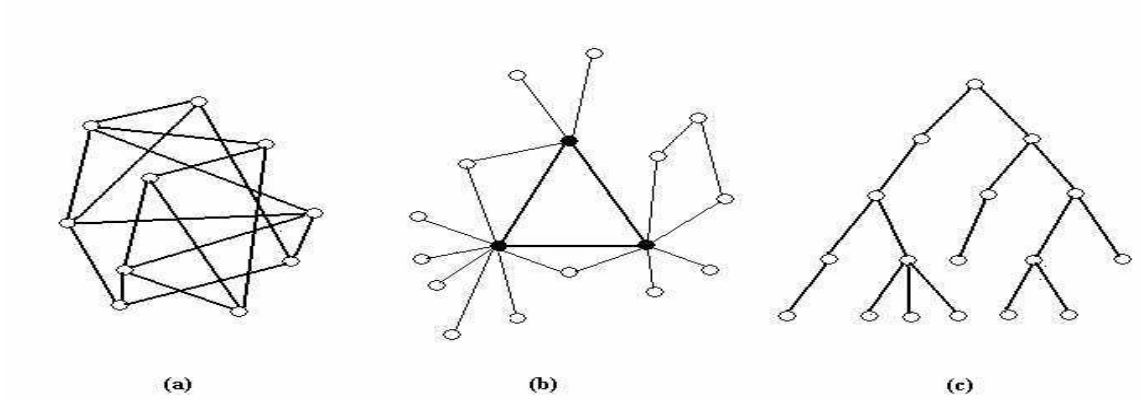


Figure 3.2. Representation of network structures: (a) random, (b) scale-free and, (c) hierarchical (Stocker *et al.*, 2001)

3.4.2. Shortest path length and small-world property

The second topological measurement is the distance between two nodes, which is given by the number of edges along the shortest path connecting them. As there are many possible paths between two nodes, the shortest path plays a special role. The number of links by which a node is connected to the other nodes varies from node to node.

The diameter of the network, d, also known as the average path length, is the average of the distances between all pairs of nodes.

$$d = \frac{\sum_j jL(j)}{\sum_j L(j)} \tag{3.4}$$

where j is the shortest path length and L(j) is the frequency of nodes have path length j.

The diameter of a network essentially characterizes the network's interconnectivity. The shorter the diameter is, the more interconnectivity the network has. In a sense, the average path length in a network is an indicator of how readily "information" can be transmitted through it. Thus, the small world property observed in biological networks suggests that such networks are efficient in the transfer of biological information: only a small number of intermediate reactions are necessary for any one protein/gene/metabolite to influence the characteristics or behaviour of another. If the size of the network is n, the average path length and diameter are of the same order of magnitude as log(n)**.** This property is referred to as small-world property. Compared to small diameters, a larger one is supposed to undermine the organism's ability to respond efficiently to internal or external perturbations. The "conserved diameter" phenomena may result from selection over a long period of evolution to response to internal defects and robust environmental fluctuations (Vishveshwara *et al.*, 2002; Przulj, 2005).

3.4.3. Clustering coefficient and hierarchical modularity

The final aspect of network structure is concerned with how densely clustered the edges in a network are. In a highly clustered network, the neighbours of a given node are very likely to be themselves linked by an edge. Typically, the first step in studying the clustering and modular properties of a network is to calculate its average clustering coefficient, C, and the related function, C(k), which gives the average clustering coefficient of nodes of degree k in the network.

Suppose *n* is a node of degree *k* in an undirected graph *G* and that there are *e* edges between the *k* neighbours of *n* in *G*. Then the clustering coefficient of *n* in *G* is given by

$$C_n = \frac{2e}{k\left(k-1\right)} \tag{3.5}$$

Thus, C_u measures the ratio of the number of edges between the neighbours of *n* to the total possible number of such edges, which is k(k-1)/2. The average clustering coefficient of a graph G is defined in the obvious manner.

The function C(k) takes the form $C(k) \sim k^{-\beta}$. Thus, as the degree of a node increases, its clustering coefficient decreases. This suggests that the neighborhoods of low-degree nodes are densely clustered while those of hub nodes are quite sparsely connected. Nodes of high degree are significantly more likely to connect to nodes of low degree than to other "hubs". In order to account for this, a hierarchical modular structure was suggested (Przulj N., 2005) for metabolic networks in which:

(i) Individual modules are comprised of densely clustered nodes of relatively low degree;

(ii) Different modules are linked by hub nodes of high degree.

3.4.4. Centrality

Several researchers have applied centrality measures to identify structurally important genes or proteins in interaction networks and investigated the biological significance of the genes or proteins identified in this way. Particular attention has been given to the relationship between centrality and essentiality, where a gene or protein is said to be essential for an organism if the organism cannot survive without it. The use of centrality measures to predict essentiality based on network topology has potentially significant applications to drug target identification (Przulj, 2005; Vishveshwara *et al.*, 2002).

There are four classical concepts of centrality which have recently been applied to biological interaction networks:

(i) Degree centrality: An important node is involved in a large number of interactions.

Formally, for an undirected graph *G*, the degree centrality of a node *n* is

$$C_d(n) = \deg(n) \tag{3.6}$$

(ii) Closeness centrality: An important node is typically "close" to, and can communicate quickly with, the other nodes in the network.

$$C_c(n) = \frac{1}{avg(L(n,m))} \tag{3.7}$$

where $L(n,m)$ is the length of the shortest path between two nodes *n* and *m*. The closeness centrality of each node is a number between 0 and 1. The closeness centrality of isolated nodes is equal to 0.

(iii) Betweenness centrality: An important node will lie on a high proportion of paths between other nodes in the network.

$$C_b(n) = \sum_{s \neq t \neq n} \frac{\sigma_{st}(n)}{\sigma_{st}} \tag{3.8}$$

where *s* and *t* are nodes in the network different from *n*, σ_{st} denotes the number of shortest paths from *s* to *t*, and $\sigma_{st}(n)$ is the number of shortest paths from *s* to *t* that *n* lies on.

Betweenness centrality is computed only for networks that do not contain multiple edges. The betweenness value for each node *n* is normalized by dividing by the number of node pairs excluding *n*: $(N-1)(N-2)/2$, where *N* is the total number of nodes in the connected component that *n* belongs to. Thus, the betweenness centrality of each node is a number between 0 and 1.

(iv) Eigenvector centrality: An important node is connected to important neighbours.

Highly connected nodes causing network failures and essential genes (i.e. genes whose knockout leads to unviable cells) were observed to be correlated, a property known as lethality–centrality. The degree centrality has been used to characterize the node importance (Gürsoy et al, 2008; Zotenko *et al.*, 2008).

Betweenness is based on the frequency with which a node lies on the shortest path between all other nodes; proteins with high betweenness control the flow of information across a network and, thus, can be important for minimizing response (Hahn and Kern, 2005). Nodes that bridge gaps separating clusters in a network (Ravasz *et al.*, 2002) are likely to be proteins with high betweenness.

Closeness centrality marks how far or close a certain node is to all the others, so that it can be interpreted as how central the position of the protein is in the network (Ortutay and Vihinen, 2009). In other words it is a measure of how quickly information can be transferred in a network (Gürsoy *et al.*, 2008).

In the aspect of these definitions and statements, the nodes with a high betweenness value are expected to have a high degree centrality also. However Joy and colleagues (Joy *et al.*, 2005) found out that there are proteins that have high betweenness, but low connectivity. This finding is not explained by algorithms proposed to explain the scale-free property of protein interaction networks, where low-connectivity proteins also have low betweenness. These data suggest the existence of some modular organization of the network, and that the high-betweenness, low-connectivity proteins may act as important links between these modules.

3.5. Prediction of Protein Function

Originally, proteins identified from genome sequencing projects were mostly annotated through homology, inferred with the aid of pair-wise alignment tools such as BLAST (Altschul *et al.*, 1997), and followed by manual intervention (Gotoh, 1999; Pearson, 1995; Procter *et al.*, 2010). Researchers would have an idea of the function of a protein by finding a significant sequence similarity to another protein whose function had

been experimentally characterized. This homology-based annotation transfer is essentially the most widely used form of computational function prediction. The rationale behind homology-based annotation is that, if two sequences have a high degree of similarity, then they have evolved from a common ancestor, and thus they should have similar, if not identical, functions.

However, with increasing number of sequences as well as the effects of gene duplications, which might be followed by divergence of function, the power of homology-based annotation is being challenged. Adding to this is the problem of errors in annotation even in human curated databases, which spread misannotations when homology- based approaches are used. In addition, most of the newly identified proteins do not show significant sequence similarity with experimentally characterized proteins worsening the problem for manual curators to keep up with the influx of data (Yooseph *et al.*, 2007).

The newly available large-scale networks of molecular interactions within the cell have made it possible to go beyond these one-dimensional approaches, and study protein function in the context of a network. The common principle underlying methods for functional annotation is that proteins that lie closer to one another in the PPI network are more likely to have similar function.

3.5.1. Neighborhood Counting

The simplest and most direct method for function prediction determines the function of a protein based on the known function of proteins lying in its immediate neighborhood. Schwikowski *et al.* (2000) predict for a given protein up to three functions that are most common among its neighbors. Although simple and effective, the obvious caveats of this approach are that associations are not assigned any significance values and the full topology of the network is not taken into account in the annotation process.

3.5.2. Module- Assisted Method

The prevalence of the modularity paradigm in molecular cell biology has led to an extensive use of modules in prediction of molecular functions. By functional module one

typically means a group of cellular components and their interactions that can be attributed to a specific biological function (Hartwell *et al.,* 1999). Instead of predicting functions for individual genes, a module-assisted approach attempts to first identify coherent groups of genes and then assign functions to all the genes in each group. The module-assisted methods differ mainly in their module detection technique. Once a module is obtained, simple methods are usually used for function prediction within the module. For example, every function shared by the majority of the module's genes is assigned to all the genes in the module.

Several works focused on the detection of functional modules based solely on protein interaction data. Most of the works decompose the PPI network into subnetworks based on some topological properties. MCODE has been used in several recent publications describing mapping of large-scale interaction networks (LaCount *et al*., 2005; Rual *et al*., 2005). It is available as a plug-in for the Cytoscape network visualization software (Shannon et al, 2003).

The scheme in Figure 3.3 shows a network in which the functions of some proteins are known (top), where each function is indicated by black color. Unannotated proteins are in white. In the direct methods (left), these proteins are assigned a color that is unusually prevalent among their neighbors. The direction of the edges indicates the influence of the annotated proteins on the unannotated ones. In the module-assisted methods (right), modules are first identified based on their density. Then, within each module, unannotated proteins are assigned a function that is unusually prevalent in the module. In both methods, proteins may be assigned with several functions.

No systematic comparison of network-based function prediction, covering both direct and module-assisted methods, has been undertaken to date. In a simplistic comparison of two basic methods, it is found that a simple neighbor-counting method has a higher specificity in predicting functions when compared to the more involved module-assisted MCODE algorithm. This may be explained by the focus of MCODE on processes imposing subgraphs with a dense interaction pattern (Sharan *et al.,* 2007).

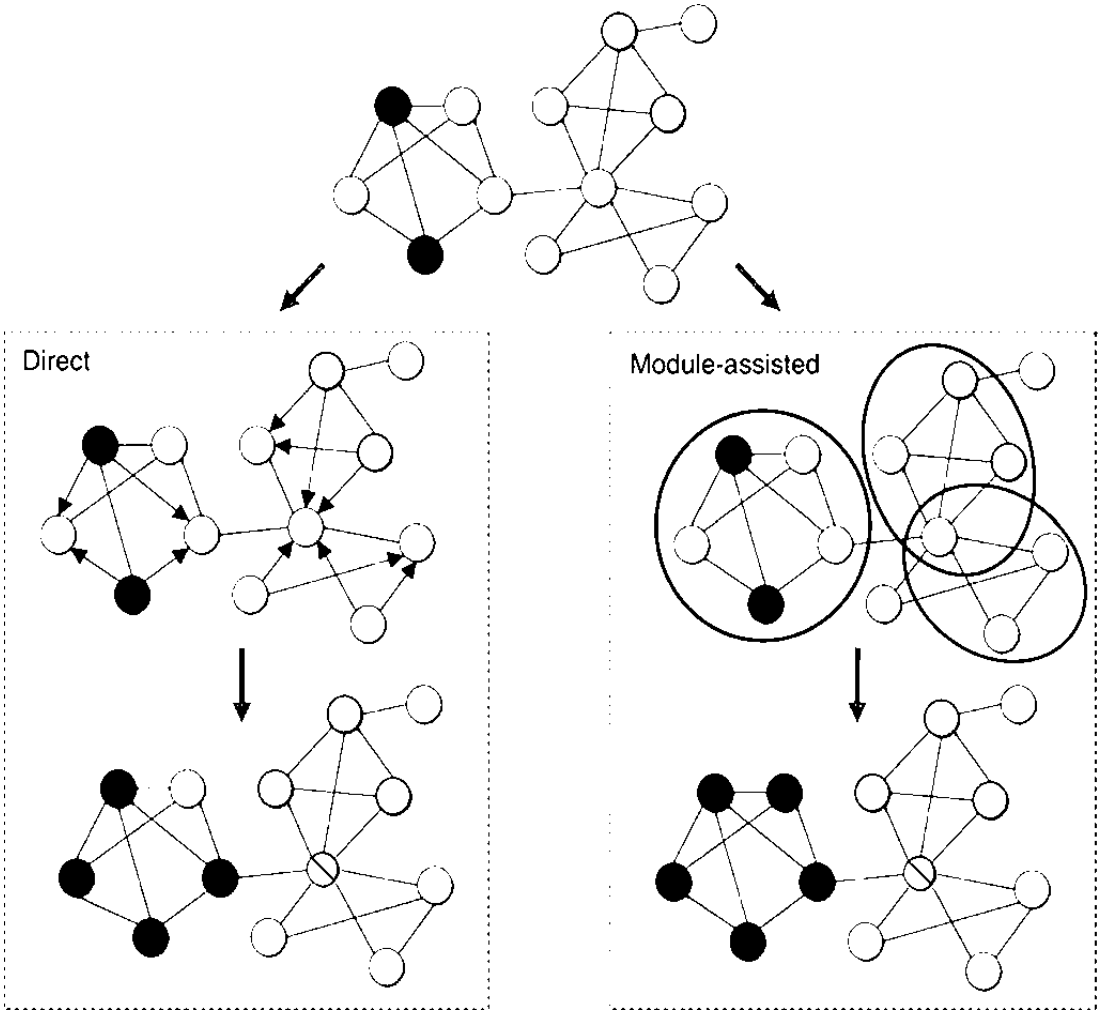


Figure 3.3. Direct versus module-assisted approaches for functional annotation. (Sharan *et al*., 2007)

In this work, the functional assignments made by modular approach are compared with the neighborhood counting method for verification purposes.

3.6. Utilization of Methods, Databases and Tools

The above mentioned methods and sources (summarized in Table 3.2) are utilized at various steps of the thesis work.

The initial step is to conduct necessary data, i.e. up-to-date interactome and gene ontology records for SPA analysis. Gene ontology data (downloaded on 16.05.2010) is obtained from GO database. The interactome data are downloaded from BioGRID and

DIP. Further elimination, according to the method of interaction determination (See App A1), is required for BioGRID data since it collects both physical and genetic interactions. Only physical interactions are taken into the constructed PPI table. Data obtained from DIP, being experimentally curated physical data, is included in the constructed PPI data directly.

As the second step, core protein selection is carried out by identifying the calcium related proteins from various sources like UniProt, GO, SGD, CYGD and WormBase together with literature search for yeast and worm. The determined core proteins are investigated in terms of sequence, structure and domain characteristics using PDB database and ConSeq. Also, using BLASTp, the evolutionary history of these proteins is obtained by evaluating the functional and sequential conservation scores throughout the evolution.

Selection of core proteins is of high importance since the selection criteria of SPA depends on the GO annotations of these proteins. Protein found to be not conserved by means of sequence or function/domain is eliminated from the core protein list since this protein will not be helpful for the development and comparison for human signaling network.

Next step is to construct the Annotation Table of the core proteins via a program written in MATLAB (See App A.3 for algorithm and App B for the Annotation Tables). This is a table representing the GO annotations of each core protein.

Having the tabulated core protein list and corresponding annotation table, the SPA method is applied to construct the Protein Interaction Network (PIN) using a program written in MATLAB (See App A.4 for the algorithm). Firstly, the interactions of core proteins are determined from the constructed PPI data. The GO annotations of the interacting partners of core proteins are tabulated and checked via the code in MATLAB whether they match with the SPA selection criteria or not. The accepted ones, then, are the new query proteins for the SPA. Determining interaction partners of the accepted proteins and criteria check procedure continues until no new accepted protein is obtained.

The next step, after the determination of accepted proteins, is to construct the PPI by obtaining the interactome of these proteins and to feed it to Cytoscape for visualization and analyses purposes. Then, using Network Analyzer plugin, the topological properties of the reconstructed PIN are determined. Using the connectivity data, the network is evaluated with respect to Graph Theory in terms of scale free properties and small world characteristics.

Another plugin, cytohubba, is compiled for the detection of hub proteins via Degree Method. The detected hub proteins' structural and sequential features are explored using UniProt, GO and SGD databases in addition to ConSeq, Pfam and BLASTp tools.

Furthermore, using BINGO plugin, the annotational enrichment analysis is carried out to be able to visualize the predominant GO terms in PIN and for hub proteins.

For module detection, the initial step is to reveal the proteins with unknown function. The function of a protein is said to be unknown if the "molecular function" term in GO is "molecular_function" with the GO id: 3674 and no other molecular function term. There are a number of proteins in the GO database with the GO id 3674, but some of these proteins also have other molecular function terms determined via IEA (Inferred from Electronic Assay – see App A.2). Since all function terms are inferred from Electronic Assay without any curation, the "molecular_function" term is assigned implying "there is no curated functional annotation for this protein". In order to detect the above-mentioned functionally unknown proteins (Go id 3674 being the only assigned term – will be mentioned as functionally unknown protein hereafter), elimination is required. Using the MATLAB code written for Annotation Table, the GO annotations of the nodes of PIN are tabulated and the ones with the GO id 3674 are detected. The next step is to determine the functionally unknown proteins using EXCEL's "match" command.

For function assignment to the functionally unknown proteins, the modules containing those proteins are detected using MCODE plugin. The functional enrichment in the modules is analyzed via BINGO plugin and function assignment procedure is established.

Table 3.2. Databases and tools used in network reconstruction and analyses

Databases	***Description***
BIOGRID	curated biological database of protein-protein and genetic interactions
DIP	database cataloging experimentally determined interactions between proteins
SGD	database of the molecular biology and genetics of the yeast Saccharomyces cerevisiae
UNIPROT	central repository of protein data created by combining Swiss-Prot, TrEMBL and PIR
PDB	repository for the 3-D structural data of large biological molecules, such as proteins and nucleic acids
GENE ONTOLOGY	bioinformatics initiative to unify the representation of gene and gene product attributes across all species
PFAM	collection of multiple sequence alignments and hidden Markov models covering protein domains and families
Tools	
BLASTp	used for both identifying a query amino acid sequence and for finding similar sequences in protein databases
CONSURF	tool that enables the identification of functionally important regions on the surface of a protein or domain
CYTOSCAPE	bioinformatics software platform for visualizing and analyzing molecular interaction networks
NETSEARCH	linear path analysis tool

4. RESULTS & DISCUSSION FOR CALCIUM SIGNALING IN YEAST

4.1. Core Proteins and Resultant Protein Interaction Network

4.1.1. Determination of Core Proteins for SPA

The choice of core proteins for SPA is one of the most crucial steps in SPA algorithm, since the annotation table (see B.2) is constructed from the GO terms of core proteins. The core proteins are directly related in accepting and rejecting process of SPA. In order to determine the core proteins, four different approaches are applied.

As the initial step, the proteins belonging to the "30.01.09.03 Ca Mediated Signal Transduction" category in the MIPS Comprehensive Yeast Genome Database (Güldener *et al.*, 2005) are chosen as the core proteins. 12 proteins are listed under the mentioned category, namely: FRT2, DPL1, CMK1, LCB3, RCN1, LCB5, RSF1, CRZ1, CMK2, LCB4, FRT1, and PLC1.

The second approach is to search the Gene Ontology for calcium related terms. In GO Database downloaded on 16.05.2010, there are a total of 127 annotations which include the term "calcium", 77 "biological process", 47 "molecular function" and 3 "cellular component" annotations (see B.1) for yeast. However, 8 of them include the term "calcium-independent"; so they are excluded from the list.

Within the group of 119 calcium-related annotations for all species, 53 proteins are found to be expressed in yeast. Since 8 of them were also listed in the MIPS Database, there are 57 unique core proteins so far.

Third approach is to find the proteins containing calcium-related domains. From the PFAM database; 11 yeast proteins having calcium-related domains (EF hand, Na_Ca_ex, calcipressin, calreticulin, trehalase_ca_bi and C2) are collected. These 11 proteins were already in the list obtained by GO search.

The last approach is to carry out a literature survey in order to find experimental evidence for the proteins determined so far by database searches and to count in the ca-related proteins not found by database searches. 6 novel proteins (CDC24, INM1, INM2, KEX2, NEO1 and STE12) are reported as calcium-related. In addition, the literature survey yielded experimental evidence for 39 existing proteins (see Table 4.1 for references).

In conclusion, a set of 63 unique probable calcium-related proteins is generated using 4 different methods.

The schematic decomposition and details of core proteins can be found in Figure 4.1 and Table 4.1 respectively.

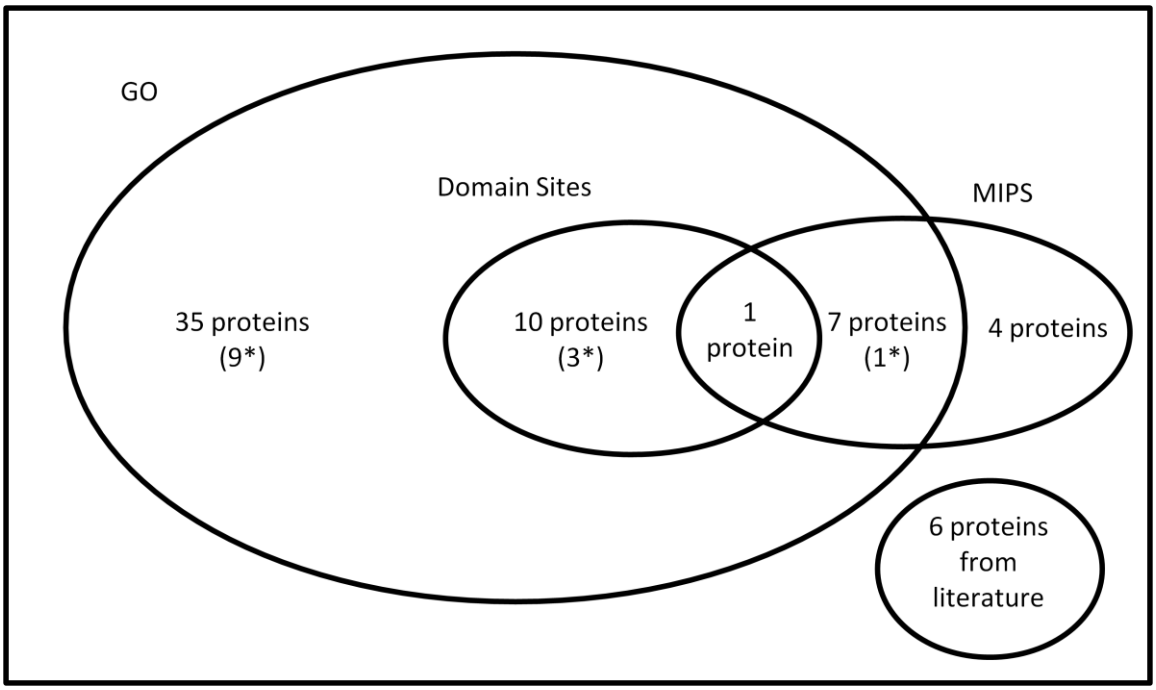


Figure 4.1. Core Protein Sources

***** Literature Evidence not found

Table 4.1. Yeast Core Protein List

ORF	NAME	AA	UNIPROT ID	DOMAINS	SOURCE	LITERATURE
YAL028W	FRT2	528	P39734		MIPS	Heath *et al.*, 2004
YBR036C	CSG2	410	P35206		GO	Beeler *et al.*, 1994
YBR109C	**CMD1**	**147**	**P06787**	**2 EF-Hand**	**GO, PFAM**	**Cunningham and Fink, 1996**
YBR127C	VMA2	517	P16140		GO	Fu *et al.*, 2000
YDL128W	**VCX1**	**411**	**Q99385**	**2 Na_Ca_ex**	**GO, PFAM**	**Cunningham and Fink, 1996**
YDR294C	DPL1	589	Q05567	ATP_synt ab N ATP_synt ab ATP_synt ab C	MIPS, GO, PFAM	Sengupta *et al.*, 2001
YDR373W	FRQ1	190	Q06389	3 EF-hand	GO, PFAM	Ames *et al.*, 2000
YEL031W	SPF1	1215	P39986	E1_E2 ATPase Hydrolase	GO, PFAM	Cronin *et al.*, 2002
YFR014C	CMK1	446	P27466	Pkinase	MIPS, PFAM	Cyert, 2001
YGL006W	**PMC1**	**1173**	**P38929**	**Cation_ATPase_N E1-E2_ATPase Hydrolase Cation_ATPase_C**	**GO, PFAM**	**Degand et al., 1999**
YGL020C	GET1	235	P53192	CHD5	GO, PFAM	Ando and Suzuki, 2005
YGL155W	CDC43	376	P18898	4 Prenyltrans	GO, PFAM	Mayer *et al.*, 1992
YGL167C	**PMR1**	**950**	**P13586**	**Cation_ATPase_N E1-E2_ATPase Hydrolase Cation_ATPase_C**	**GO, PFAM**	**Degand et al., 1999**
YGR217W	**CCH1**	**2039**	**P50077**	**4 Ion_trans**	**GO, PFAM**	**Locke et al., 2000**
YJL134W	LCB3	409	P47013	PAP2	MIPS, GO, PFAM	Birchwood *et al.*, 2001
YKL159C	**RCN1**	**211**	**P36054**	**Calcipressin**	**MIPS, GO, PFAM**	**Cui et al., 2009**
YKL190W	**CNB1**	**175**	**P25296**		**GO, SMART**	**Cyert and Thorner, 1992**
YLR220W	CCC1	322	P47818	VIT1	GO, PFAM	Pozos *et al.*, 1996
YLR260W	LCB5	687	Q06147	DAGK_cat	MIPS, GO, PFAM	Birchwood *et al.*, 2001
YLR433C	**CNA1**	**553**	**P23287**	**Metallophos**	**GO, PFAM**	**Cyert et al., 1991**
YML057W	**CMP2**	**604**	**P14747**	**Metallophos**	**GO, PFAM**	**Cyert et al., 1991**
YML099C	ARG81	880	P05085	Zn_clus DUF3468	GO, PFAM	
YMR030W	RSF1	376	Q05043		MIPS, GO	

Table 4.2. Yeast Core Protein List - cont.

YNL027W	**CRZ1**	**678**	**P53968**	**Zf_C2H2**	**MIPS, GO, PFAM**	**Matheos et al., 1997**
YNL083W	SAL1	545	P48233	3 Mito_carr	GO, PFAM	Chen, 2004
YNL291C	**MID1**	**548**	**P41821**		**GO**	**Locke et al., 2000**
YOL016C	CMK2	447	P22517	Pkinase	MIPS, PFAM	Cyert, 2001
YOR087W	**YVC1**	**675**	**Q12324**		**GO**	**Denis and Cyert, 2002**
YOR171C	LCB4	624	Q12246	DAGK_cat	MIPS, GO, PFAM	Birchwood *et al.*, 2001
YOR324C	FRT1	602	Q99332		MIPS	Heath *et al.*, 2004
YPL268W	PLC1	869	P32383	PI-PLC-X PI-PLC-Y	MIPS, GO, PFAM	Flick and Thorner, 1993
YAL058W	CNE1	502	P27825	Calreticulin	GO, PFAM	
YDR001C	NTH1	751	P32356	Trehalase Ca_bi Trehalase	GO, PFAM	
YGL106W	MLC1	149	P53141		GO	
YGR058W	PEF1	335	P53238		GO	Vernarecci *et al.*, 2007
YNL084C	END3	349	P39013		GO	
YOR257W	CDC31	161	P06704	2 EF-hand	GO, PFAM	Sullivan *et al.*, 1998
YPL048W	CAM1	415	P29547	GST N GST C EF1G	GO, PFAM	Kambouris *et al.*, 1993
YAL041W	CDC24	854	P11433	CDC24 RhoGEF PB1	LIT. SURVEY, PFAM	Miyamoto *et al.*, 1991
YDR287W	INM2	292	Q05533	Inositol_P	LIT. SURVEY, PFAM	Lopez *et al.*, 1999
YHR046C	INM1	295	P38710	Inositol_P	LIT. SURVEY, PFAM	Lopez *et al.*, 1999
YHR084W	STE12	688	P13574	STE	LIT. SURVEY, PFAM	Muller, 2001
YIL048	NEO1	1151	P40527	E1-E2_ATPase Hydrolase	LIT. SURVEY, PFAM	Prezant *et al.*, 1996
YNL238W	KEX2	814	P13134	Peptidase_S8 P_proprotein	LIT. SURVEY, PFAM	Fuller *et al.*, 1989
YAL048C	GEM1	662	P39722	2 Miro 2 EF-hand associated	GO, PFAM	Wolff *et al.*, 1999
YBL047C	EDE1	1381	P34216	UBA	GO, PFAM	

Table 4.3. Yeast Core Protein List - cont.

YBR001C	NTH2	780	P35172	Trehalase Ca_bi Trehalase	GO, PFAM	
YDR129C	SAC6	642	P32599	4 CH	GO, PFAM	
YGR100W	MDR1	950	P53258	GRAM TBC	GO, PFAM	
YHR204W	MNL1	796	P38888	Glyco_hydro_47	GO, PFAM	
YIL140W	AXL2	823	P38928	HE_PIG SKG6	GO, PFAM	
YIR006C	PAN1	1480	P32521	3 DUF1720	GO, PFAM	Tang and Cai, 1996
YJL083W	TAX4	604	P47030		GO	
YJR131W	MNS1	549	P32906	Glyco_hydro_47	GO, PFAM	
YKR019C	IRS4	615	P36115		GO	
YLR057W		849	Q12205	Glyco_hydro_47	GO, PFAM	
YLR250W	SSP120	234	P39931		GO	
YML072C	TCB3	1545	Q03640	4 C2	GO, PFAM	Schulz and Creutz, 2004
YMR105C	PGM2	569	P37012	PGM_PMM_I PGM_PMM_II PGM_PMM_III PGM_PMM_IV	GO, PFAM	Fu *et al.*, 2000
YOL001W	PHO80	293	P20052	Cyclin	GO, PFAM	Sopko *et al.*, 2006
YOR086C	TCB1	1186	Q12466	4 C2	GO, PFAM	Schulz and Creutz, 2004
YPL031C	PHO85	305	P17157	Pkinase	GO, PFAM	Sopko *et al.*, 2006
YPR188C	MLC2	163	Q06580		GO	

* The proteins in grey shaded boxes are the members of 12 core protein group.

4.1.2. Resultant Protein Interaction Networks

Starting with 63 proteins (Figure 4.2 and Table 4.2) described as calcium-related protein in MIPS-CYGD, GO, PROSITE, SMART and literature; SPA resulted in a protein interaction network (PIN) consisting of *3226* proteins and *20167* undirected physical interactions among these proteins.

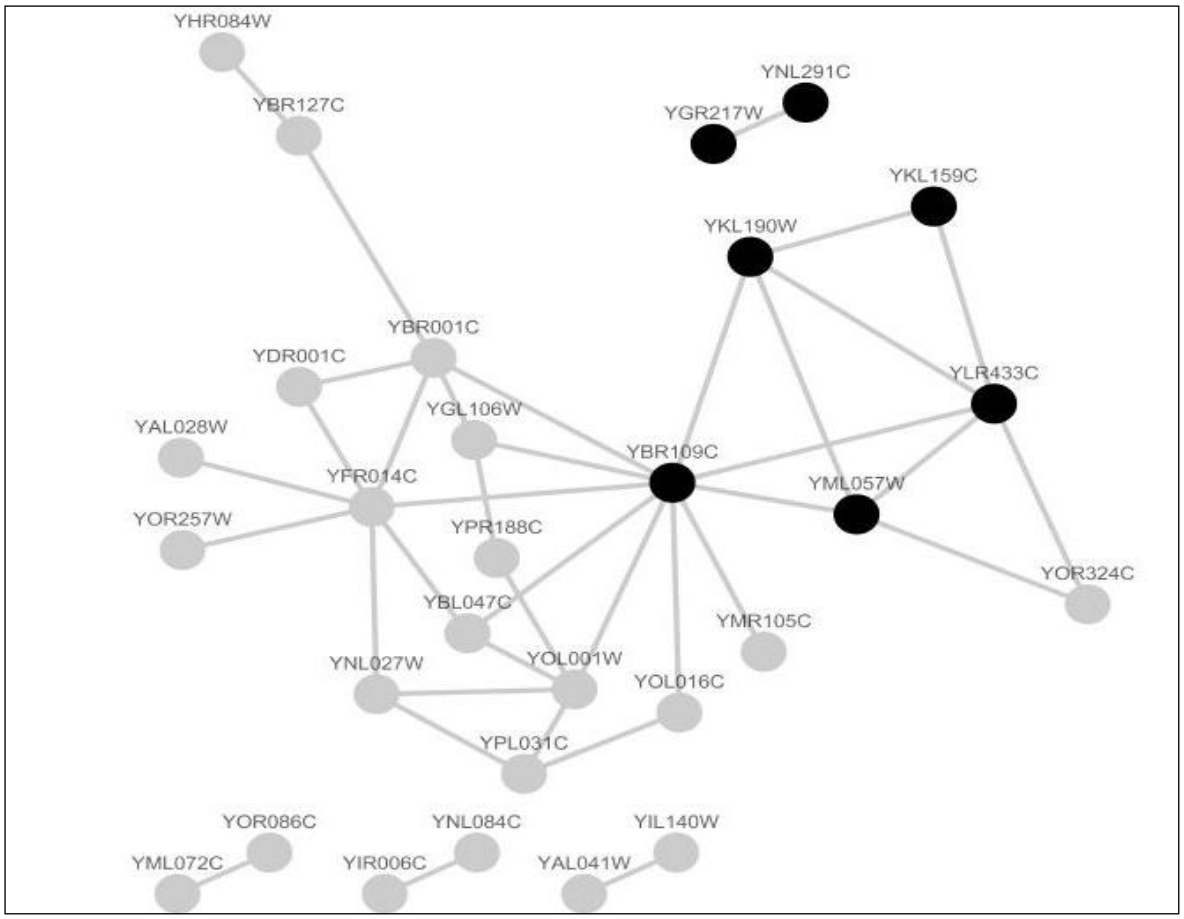


Figure 4.2. Yeast Core protein network consisting of 63 & 12 proteins. Produced by Cytoscape 2.6.

For 63 core; 8 proteins form isolated groups of two, 21 proteins have 34 interactions among themselves and 34 proteins have no interactions at all (not shown). For 12 core (bold ones); 2 proteins form isolated groups of two, 5 proteins have 8 interactions among themselves and 5 proteins have no interactions at all (not shown).

The resultant PIN from 63 core proteins covered 49 % of the proteins (inner circle in Figure 4.3) and also it covered 21 % of the whole genome interactome (outer circle in Figure 4.3) in Yeast Genome (SGD – 18.06.2010 data). As a result, the constructed Calcium Signaling protein interaction network became huge and cumbersome.

In order to attain a more "Calcium-Signaling-specific" network, the core set should be narrowed down. By doing this, it would be easier to dissect the constructed network and to draw healthy conclusions relevant to Calcium Signaling.

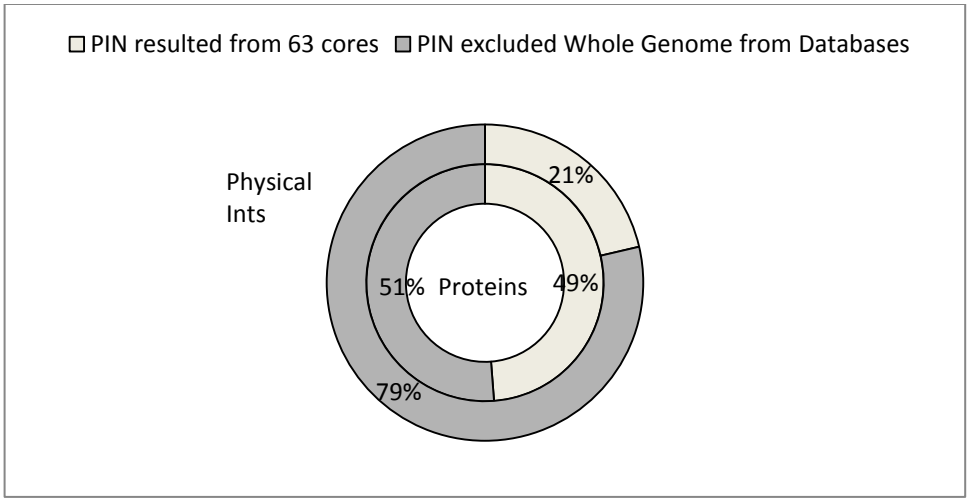


Figure 4.3. PIN with 63 cores vs. Whole Genome for Yeast

The basic principle is to include the proteins that are *proven to be directly related* to calcium by *experimental* techniques. The core set of 63 proteins is examined thoroughly and the indirectly related proteins are eliminated from the core list. Also since Gene Ontology includes the GO terms inferred by electronic annotations (IEA) as well as experimental annotations; the ones having GO term related to calcium determined with IEA only are excluded from the core list as well. After those eliminations, the resultant set of core proteins comprises of *12 proteins* (Figure 4.2 and Table 4.2) which are experimentally proven to be directly assigned to the calcium signaling in the cell (details in Table 4.1).

Table 4.4. Comparison of 63 & 12 Core Proteins for Yeast and Their Resultant PINs

	63 CORE		*12 CORE*	
	Core PIN	*Resultant PIN*	*Core PIN*	*Resultant PIN*
# of nodes	21	3226	5	1826
# of edges	34	20167	8	10078
# of isolated & lonely nodes	42	2	7	-
Total # of proteins	63	3228	12	1826

When SPA with 63 core proteins yielded in a large and complex network (Figure 4.4a) ,12 core proteins resulted in a PIN consisting of ***1826*** proteins and ***10078*** undirected physical interactions (Figure 4.4b).

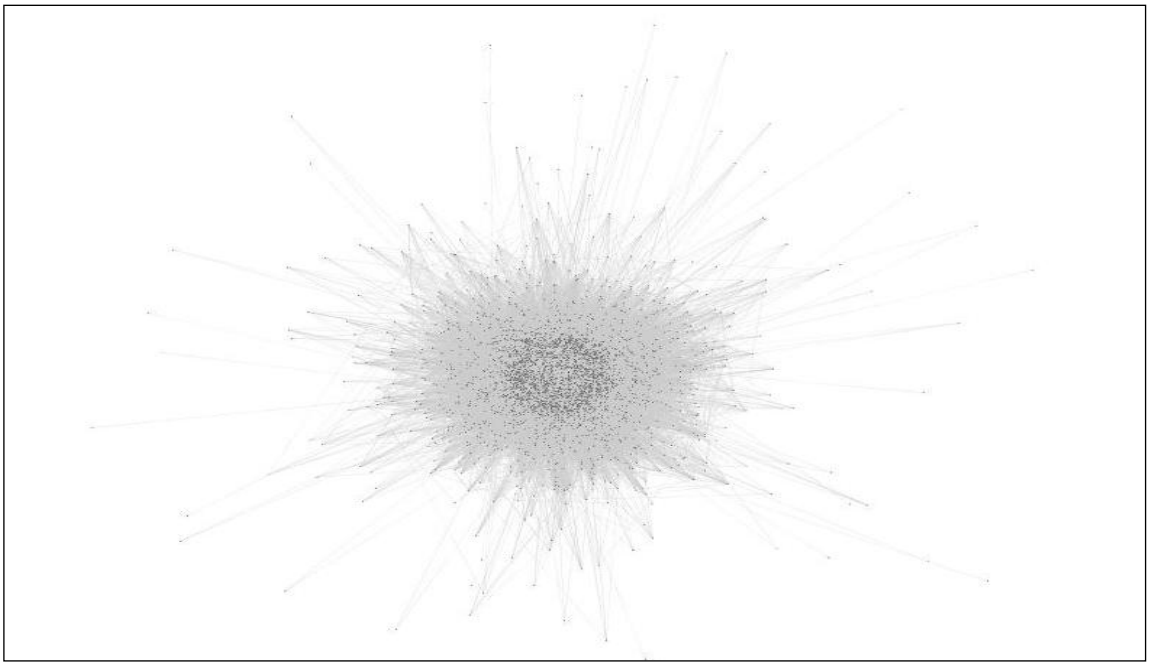


Figure 4.4a

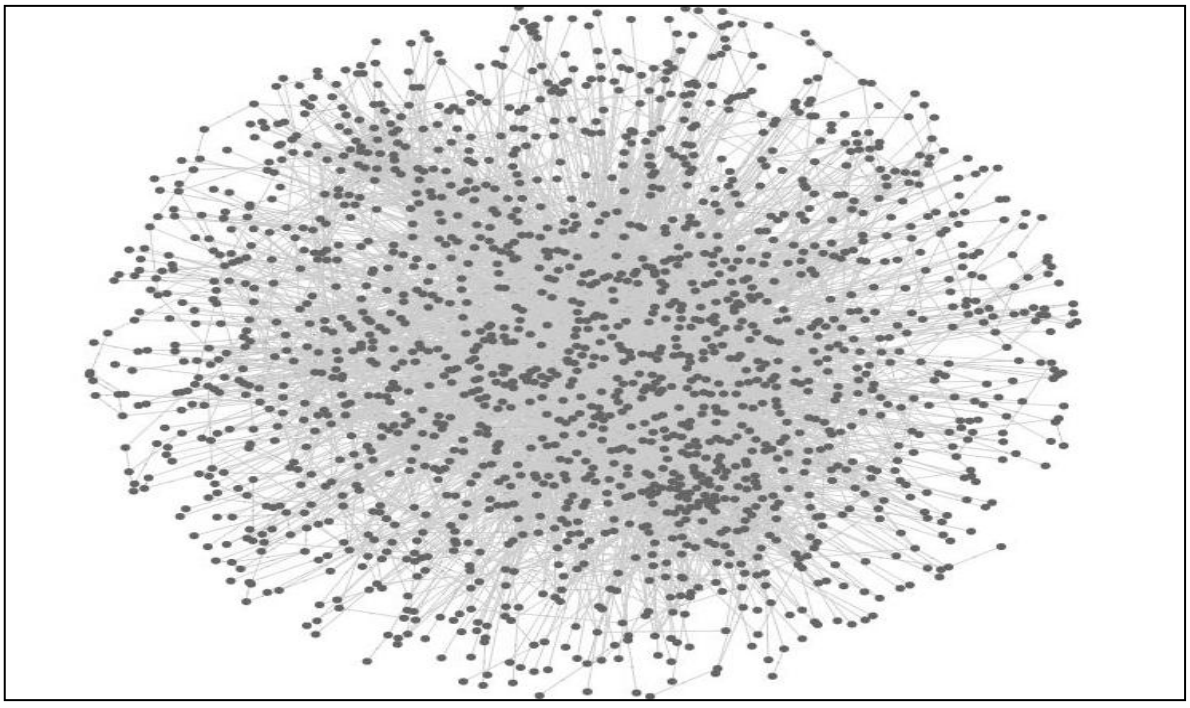


Figure 4.4b

Figure 4.4. a) SPA resultant Yeast PIN obtained from 63 core proteins with 3226 nodes and 20167 edges. b) SPA resultant Yeast PIN obtained from 12 core proteins with 1826 nodes and 10078 edges. Produced by Cytoscape 2.6.

In Figure 4.5, the steps (i.e. the number of the accepted proteins in each step) of SPA starting with both 63 and 12 core proteins are depicted. As it can be seen, in the 2nd step, the accepted proteins of 63 core proteins cover most of the members of the resultant PIN. As a result, the algorithm initiated with 63 core proteins converges earlier than that of 12 core proteins.

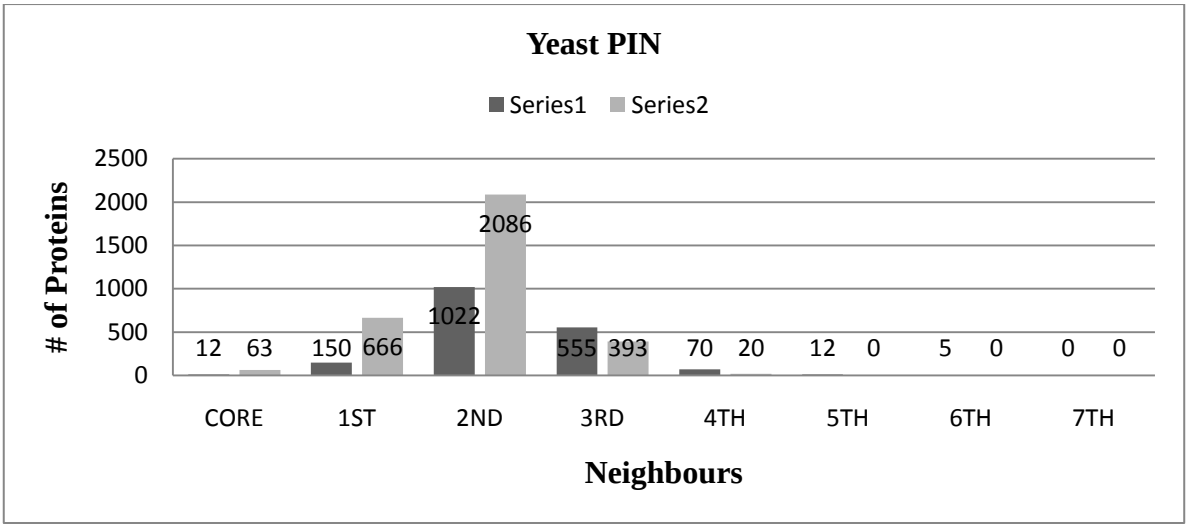


Figure 4.4. Network reconstruction steps in SPA for 63 & 12 core Proteins

4.2. Analysis of Reconstructed Network

4.2.1. Evolutionary History of Core Proteins

Protein sequences determine protein function. Thus proteins with high sequence identity and high structural similarity tend to possess functional similarity and evolutionary relationship. In order to validate the choice of core proteins, the core proteins' homologues through different species are investigated.

The blastp search uses the core protein as the query sequence and looks for the homologues. *C.elegans* (worm), *D. melanogaster* (fruit fly), *D.rerio* (zebra fish), *X.laevis* (frog), *M.musculus* (mouse) and finally *H.sapiens* (human) are chosen as model organisms according to their order of evolutionary time span. The reason these organisms are chosen is that these organisms are the model organisms that have been extensively studied.

In Table 4.3, the beginning (yeast) and the corresponding homologues of the last organism (human) are given only. Having a high identity score in the last organism implies that the protein also has high identity scores in-between organisms. The whole evolutionary homology search results can be found in Table B.3.1.

Out of 12 core proteins, 6 of them, namely calmodulin (CMD1), calcineurin subunits (CNB1, CNA1, CMP2) and calcium transporting ATPases (PMC1, PMR1), have high identity scores with a high significance (low e-value).

Among the others left, MID1 is the only protein which has low identity score and high e-value. Also, the homologue in human to this protein has, as expected, an irrelevant function when compared to that of MID1. This protein becomes active in yeast when there is mating pheromone stimulation. So, this protein is mainly specific only to yeast which explains the low identity score and irrelevant function. However, in yeast, it is known that MID1 has an important role in external Ca transport into the cytoplasm by activating CCH1. So it should stay in the core protein list.

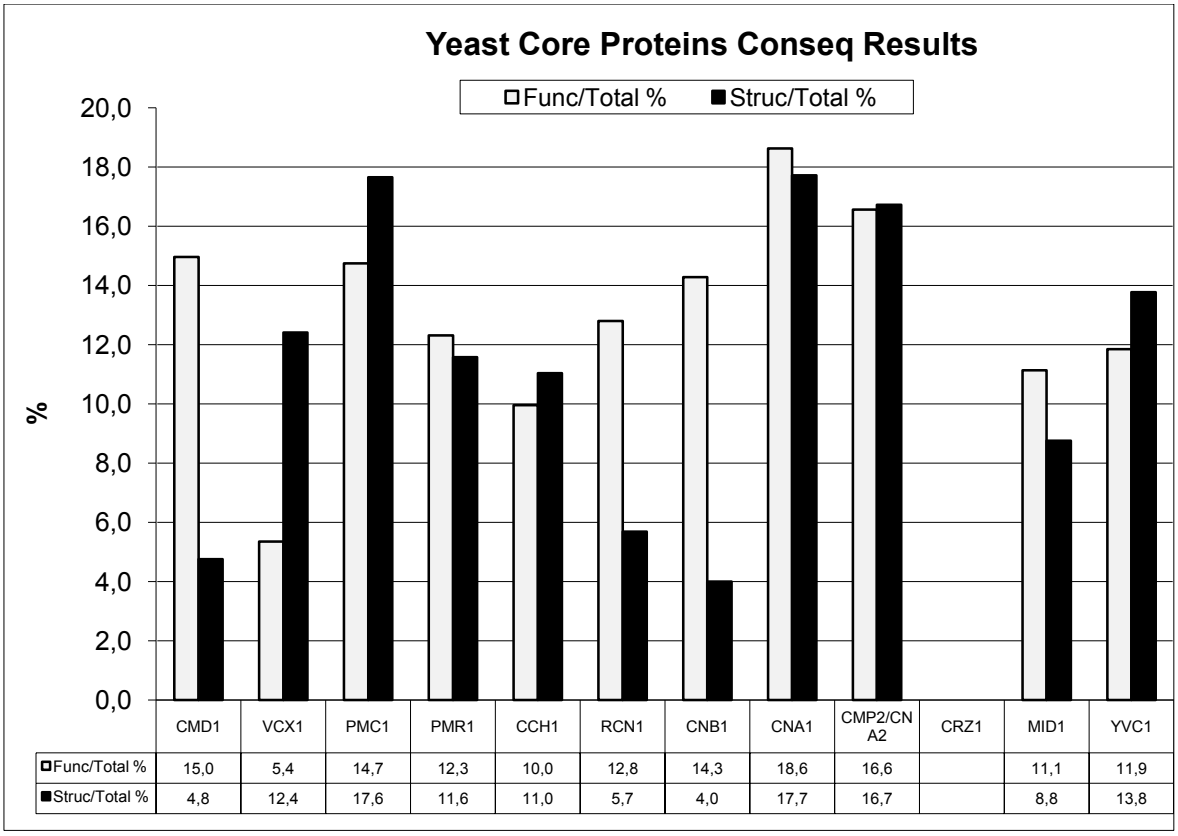
The CRZ1, VCX1, RCN1 and finally YVC1 have low identity scores but having low identity score does not always imply irrelevance. The homologues found share the similar functions with the query protein. This result can be explained with the fact that, the functionally important region of these proteins that is conserved may be small when compared to the size of the total a.a sequence of these proteins.

In addition there is no structure information for any of the core proteins except CMD1 with PDB id:1LKJ

Table 4.5. Homologues of yeast core proteins in human

S. CEREVISIAE (Yeast)	***prot name***	CMD1	CNB1	CNA1	CMP2	CRZ1	VCX1	PMR1	MID1	PMC1	RCN1	YVC1	CCH1
		Calmodulin	Calcineurin subunit B	Calcineurin A1	Calcineurin A2	Calcineurin-Responsive Zinc finger	VaCuolar $H+/Ca^{2+}$ eXchanger	Ca transporting ATPase 1	Mating pheromone-induced death protein 2	Ca transporting ATPase 2	Regulator of CalciNeurin	Yeast vacuolar conductance protein 1	Ca channel protein
H.SAPIENS (Human)	***% id.***	59,9	52,6	47	45,2	8	8	47,8	3,1	34,9	12,3	17,5	10,84
	e-val	7,07E-50	9,95E-49	9,90E-149	1,47E-154	1,48E-13	1,25E+00	0,00E+00	1,48E+00	1,89E-154	7,35E-01	1,49E-01	3,09E-39
	prot name	calmodulin	calcineurin subunit B type 1	serine - threonine-protein phosphatase 2B	serine - threonine-protein phosphatase 2B catalytic subunit alpha isoform isoform 1	zinc finger protein 397OS	Na K Ca exchanger 3 precursor	Ca transporting ATPase type 2C member 1 isoform	TPA_inf: olfactory receptor OR14-7	plasma membrane Ca transporting ATPase 2 isoform 2	regulator of calcineurin 3 isoform 2,3,4b,5	short transient receptor potential channel 7 isoform 1	L-type Ca channel (HFCC)

The parallel results are also observed by ConSeq program (Figure 4.6). The conserved residue analyses are carried out using ConSeq, which a web server for the identification of biologically important residues and estimation of their solvent accessible area in protein sequences by collecting homologues and aligning them by multiple alignment method.



	CMD1	VCX1	PMC1	PMR1	CCH1	RCN1	CNB1	CNA1	CMP2/CNA2	CRZ1	MID1	YVC1
Func/Total %	15,0	5,4	14,7	12,3	10,0	12,8	14,3	18,6	16,6		11,1	11,9
Struc/Total %	4,8	12,4	17,6	11,6	11,0	5,7	4,0	17,7	16,7		8,8	13,8

Figure 4.5. The Conserved Residue Analysis of 12 yeast core proteins by Conseq

The functionally important residues are the conserved residues on the surface of the protein. These sites are possibly taking part in ligand binding and protein–protein interactions etc. The structurally important residues however, are the conserved ones buried inside the core and may be crucial on protein folding.

In accordance with the inference of homology search carried out, calcineurin subunits as well as calcium transporting ATPases have high percentages in both functional and structural conservation although their exposed residues are more than their buried residues (Figure 4.6 and Table B.3.2).

Calmodulin having the highest identity score in homology search, however, has a low structural conservation percentage. But it has the third highest functional conservation score in ConSeq results. So it can be inferenced that, calmodulin's binding attributes are more dominant than its folding characteristics.

All proteins, with the exception of VCX1, have more exposed residues than buried residues; implying that their folding residue sites are not abundant as much as their possible functional residue sites. VCX1 is the only protein with a structurally important residue percentage much higher than functionally important residue percentage; so the dominance of buried residue in this protein is logical (Table 4.4).

On the other hand, the domain conservation is the highest in VCX1, PMC1, CCH1, CNA1 and CMP2 which are also found to be conserved by homology search with significant e-values except VCX1. This protein, having two highly conserved Na-Ca exchanger domains, may have a low identity % in homology search but the domain conservation explains the conservation of the function (Na-Ca exchanger) in human homolog.

Another feature of the domain conservation search is that, almost all domains' conserved residues are the main constituents of the proteins' conserved residues (Table 4.4); i.e. the domain conservation in conserved residues of the proteins is generally above 50%.

For the proteins having high ConSeq conservation scores both functionally and structurally, the domain-base buried residues are more than the exposed ones. This explains the reason of structural conservation although having more exposed residues in whole protein.

Also, the functionally and structurally important residues are mostly found in the domain sites of the proteins.

For CRZ1, ConSeq could not come up with enough number of homologues to run the analysis; so no result is given for that protein.

Table 4.6. ConSeq results of yeast core proteins and their domains

ORF name	Protein name	Tot. Res.	Cons	Domain Cons. in	Exposed	Buried	Func/conserved*	Domain F/Cons F**	Struc/conserved*	Domain S/Cons S**
			%	Cons. Residues	%	%	%	%	%	%
YBR109C	CMD1	147	25,2		68,7	31,3	59,5		18,9	
	EF hand	29	37,9	29,7	62,1	37,9	63,6	31,8	27,3	42,9
	EF hand	29	20,7	16,2	58,6	41,4	33,3	9,1	33,3	28,6
YDL128W	VCX1	411	26,5		37,0	63,0	20,2		46,8	
	Na Ca ex	150	35,3	48,6	33,3	66,7	24,5	59,1	43,4	45,1
	Na Ca ex	134	36,6	45,0	19,4	80,6	18,4	40,9	51,0	49,0
YGL006W	PMC1	1173	39,6		55,6	44,4	37,3		44,6	
	Cation_ATPase_N	71	32,4	5,0	63,4	36,6	39,1	5,2	34,8	3,9
	E1-E2_ATPase	277	48,0	28,7	49,1	50,9	36,8	28,3	42,1	27,1
	Hydrolase	135	52,6	15,3	54,8	45,2	45,1	18,5	50,7	17,4
	Cation_ATPase_C	196	42,9	18,1	35,7	64,3	26,2	12,7	51,2	20,8
YGL167C	PMR1	950	29,2		51,4	48,6	42,2		39,7	
	Cation_ATPase_N	71	14,1	3,6	59,2	40,8	50,0	4,3	50,0	4,5
	E1-E2_ATPase	243	37,4	32,9	48,1	51,9	39,6	30,8	40,7	33,6
	Hydrolase	197	34,0	24,2	56,9	43,1	37,3	21,4	43,3	26,4
	Cation_ATPase_C	174	21,3	13,4	37,9	62,1	43,2	13,7	29,7	10,0
YGR217W	CCH1	2039	28,9		53,3	46,7	34,4		38,1	
	Ion trans	245	49,0	20,3	42,4	57,6	32,5	19,2	42,5	22,7
	Ion trans	194	35,6	11,7	24,7	75,3	17,4	5,9	43,5	13,3
	Ion trans	229	48,9	19,0	34,1	65,9	25,9	14,3	44,6	22,2
	Ion trans	205	44,9	15,6	32,2	67,8	34,8	15,8	46,7	19,1
YKL159C	RCN1	211	20,4		69,2	30,8	62,8		27,9	
	Calcipressin	168	19,0	74,4	67,3	32,7	68,8	81,5	21,9	58,3
YLR433C	CNA1	553	40,0		59,0	41,0	46,6		44,3	
	Metallophos	210	52,4	49,8	46,7	53,3	43,6	46,6	50,0	56,1
YML057W	CMP2/CNA2	604	36,3		60,8	39,2	45,7		46,1	
	Metallophos	227	49,3	51,1	48,5	51,5	43,8	49,0	50,0	55,4

*Ratio of functional/structural residues to that of conserved residues in whole protein ** Ratio of functional/structural residues in domain to that of in whole protein

4.2.2. Reliability of GO Annotations

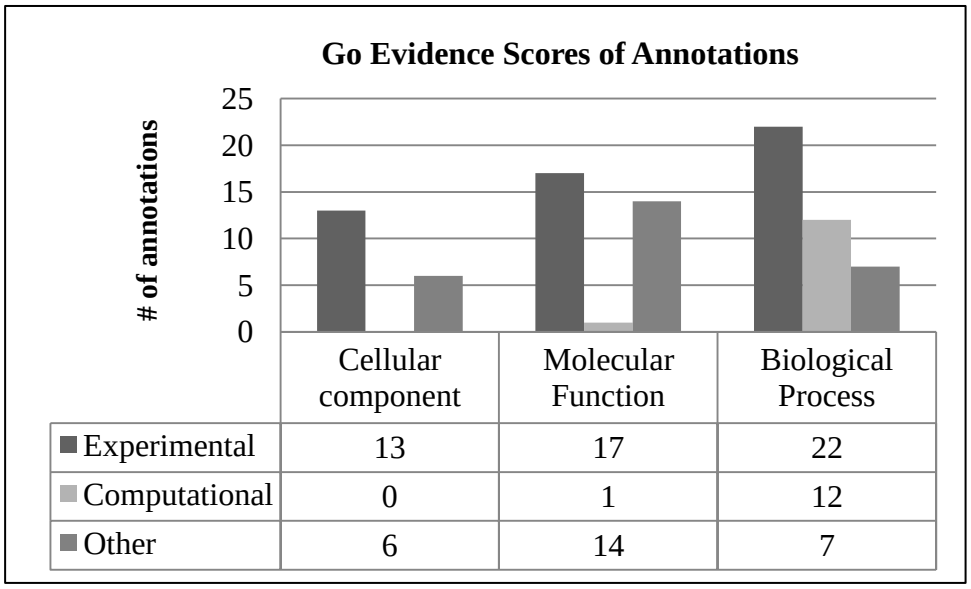
The selection criterion for the SPA is the collection of GO annotations of core proteins. As mentioned, proteins are accepted to the network if all three of their GO annotations match the GO Annotation Table of core proteins. Therefore, the reliability of GO annotations is important to include the proper proteins into the network.

A GO annotation consists of a GO term associated with a specific reference that describes the work or analysis upon which the association between a specific GO term and gene product is based. Each annotation must also include an evidence code to indicate how the annotation to a particular term is supported.

Manually-assigned evidence codes fall into four general categories: experimental, computational analysis, author statements, and curatorial statements. (See App A.2)

The use of an *experimental evidence code* in a GO annotation indicates that the cited paper displayed results from a physical characterization of a gene or gene product that has supported the association of a GO term. The use of the *computational analysis evidence codes* indicates that the annotation is based on an in silico analysis of the gene sequence and/or other data as described in the cited reference. The evidence codes in this category also indicate a varying degree of curatorial input. Use of the *curatorial statement evidence codes* indicates an annotation made on the basis of a curatorial judgement that does not fit into one of the other evidence code classifications. Out of all the evidence codes available, only *Inferred from Electronic Annotation* (IEA) is not assigned by a curator.

The Annotation Table in this work contains 20 cellular component terms, 33 molecular function terms and 42 biological process terms with a total number of 95 GO terms.



	Cellular component	Molecular Function	Biological Process
Experimental	13	17	22
Computational	0	1	12
Other	6	14	7

Figure 4.6. Reliability of Annotation Table for 12 core proteins of yeast

All three groups of GO annotations are mostly composed of reliable sources (experimental evidence codes) with a percentage of higher than 50%. The contribution of computational evidence codes to the annotations is the least in cellular component (none) and most in biological process (29%).

The evidence code of annotations under the group “others” are all inferred from Electronical Annotation (IEA) which is the only evidence code without a curator and probably the least reliable code. The highest percentage of IEA stemmed annotation is in molecular function by 44%. The cellular component and biological process have lower IEA derived annotations (32 and 17% respectively).

In conclusion, the most reliable GO annotations are for cellular component term. Biological process term also has a high reliability score, since the sum of experimental and rather reliable code computational annotations is 75%. The molecular functions term, indeed, has a reliability of 53%; the rest is IEA (44%). So the molecular function annotations can also be accepted as reliable.

4.2.3. Annotation Table Enrichment in Yeast PIN

The annotation table, set as the main criteria for SPA method, is constructed from the GO terms of the 12 core proteins. With the intention of assigning new functions to the proteins not having any known function, the GO terms of "cellular component unknown" (GO id: 5575), "molecular function unknown" (GO id: 3674) and "biological process unknown" (GO id: 8150) are also included in the Annotation Table.

The Annotation Table of 12 core proteins contains 20 cellular component terms, 33 molecular function terms and 42 biological process terms with a total number of 95 GO terms. On the other hand, the proteins involved in the reconstructed Ca signaling PIN (1826 proteins) enriched the annotation table to a total number of 2320 GO terms; 448 being cellular component, 686 being molecular function and 1186 being biological process. (Figure 4.8)

The GO term enrichment in the annotation table resulted in 25 fold increase compared to the initial annotation table. The most enriched annotation is the biological function with an increase of 28 fold. The detailed list of annotation tables and the amount of enrichment in the GO terms are given in App B.

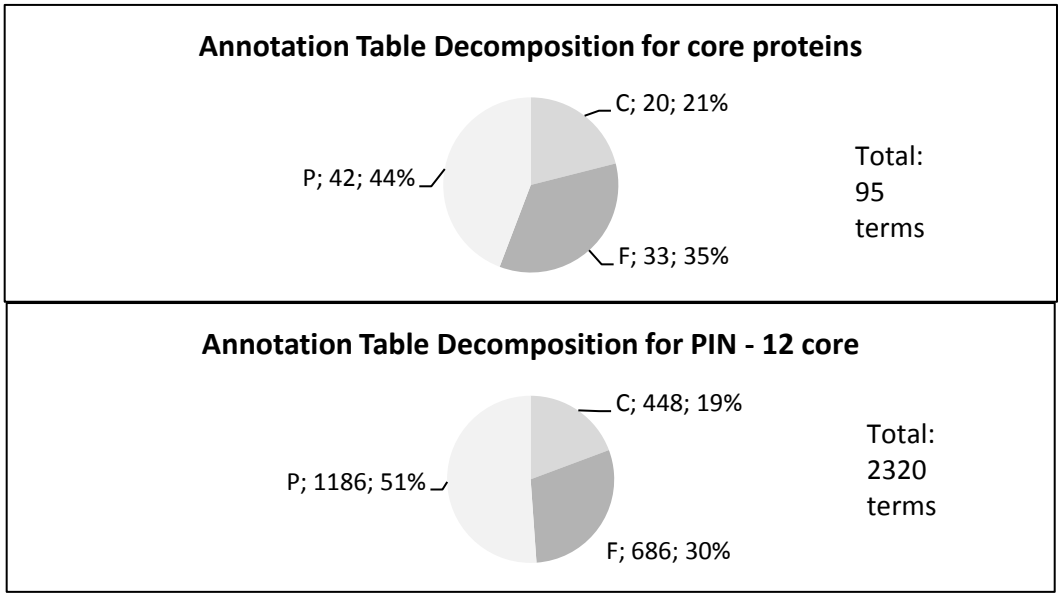


Figure 4.7. Annotation Table Comparison of core proteins and PIN for yeast

The dominant biological process terms in the core protein set are *ion transport* and *cellular ion homeostasis* (Figure 4.9) as expected; since the core proteins chosen are mainly the calcium channels on membranes and the calcium binding proteins in cytoplasm in order to sustain calcium ion homeostasis in the cell. Hence, *cytoplasm* and *membrane* being the dominant cellular component terms (Figure 4.13) is consistent with the biological process dominance.

Also the most frequent molecular function terms are *ion binding* and *channel activity* (Figure 4.11) which are the main functions of the core proteins to keep the calcium level low in the cytosol by binding the free cytosolic calcium and sequestering it to the organelles through channels on membranes or to respond to the environmental stress conditions and by releasing calcium to the cytoplasm.

On the other hand; *transcription, protein transport* and *metabolic process* are the abundant biological process terms (Figure 4.10) in the reconstructed PIN. Accordingly, the dominant molecular function terms are *nucleotide binding, DNA binding* in addition to *ion binding* (Figure 4.12) which takes place in *nucleus, cytoplasm* and *membranes* (Figure 4.14).

In the mentioned figures (4.9 – 4.14), the percentages do not give the decomposition of the annotations but the abundance of that term. The term "% Freq in Network" implies the predominance percentage of the related GO term in the reconstructed PIN; whereas "% Freq in Whole Annotation" term gives the percentage of the frequency of the related term in the whole yeast genome.

A protein can have more than one annotation term at the same time, so the same protein may contribute to the percentage of any term in the figures. As an example, for the biological term, the core protein PMC1 (YGL006W) makes contribution to ion transport, transport, calcium ion transport, transmembrane transport, cellular calcium ion homeostasis, metabolic process, ATP biosynthetic process and vacuolar protein catabolic process all at once.

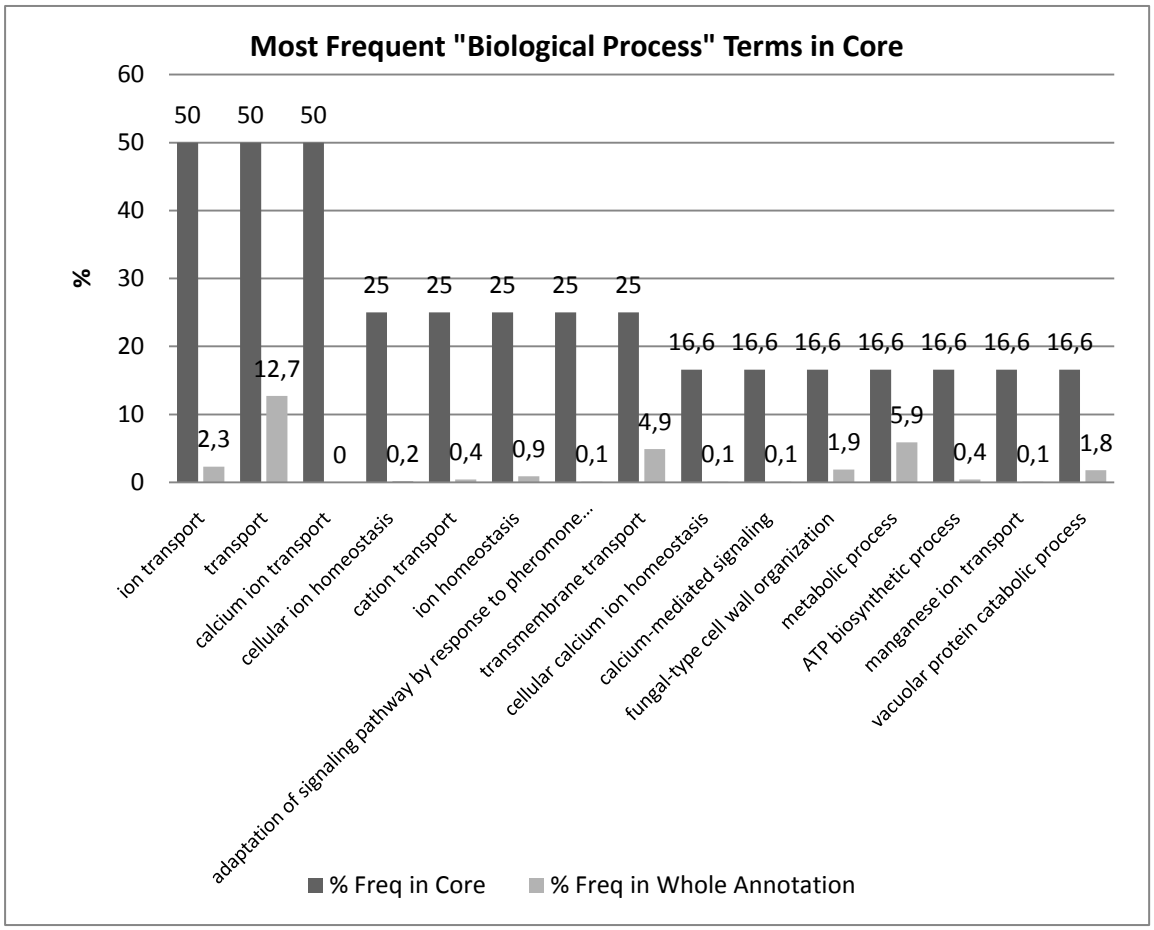


Figure 4.8. Predominant “Biological Process” terms in yeast Core Proteins

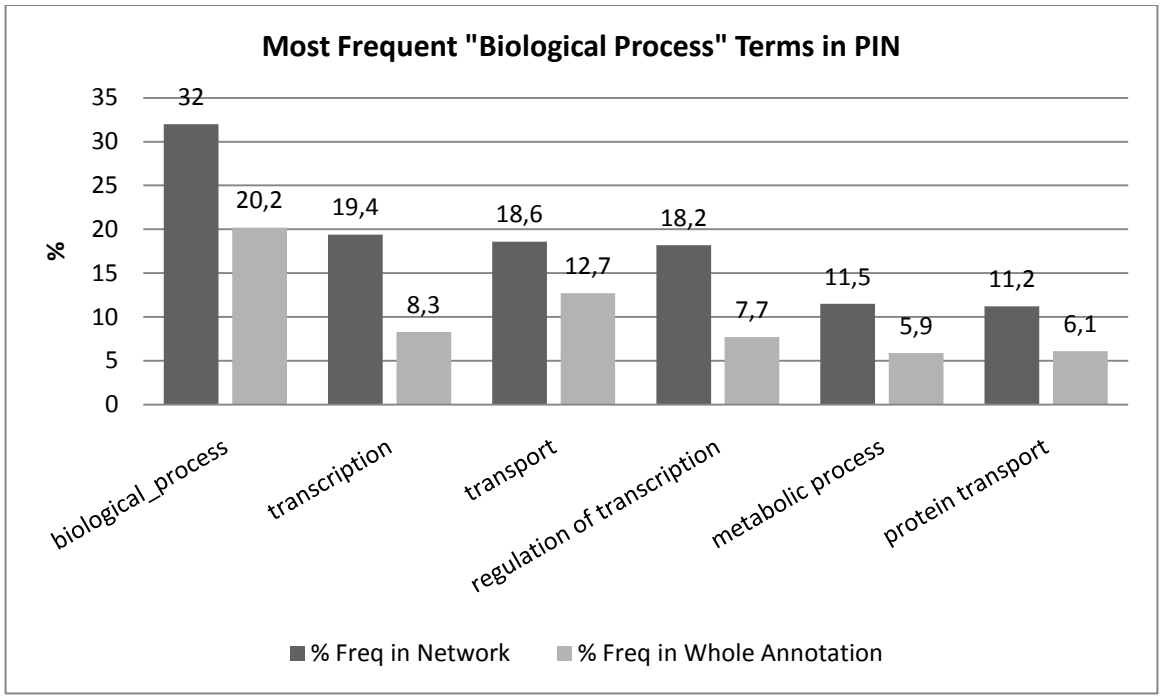


Figure 4.9. Predominant “Biological Process” terms in yeast PIN

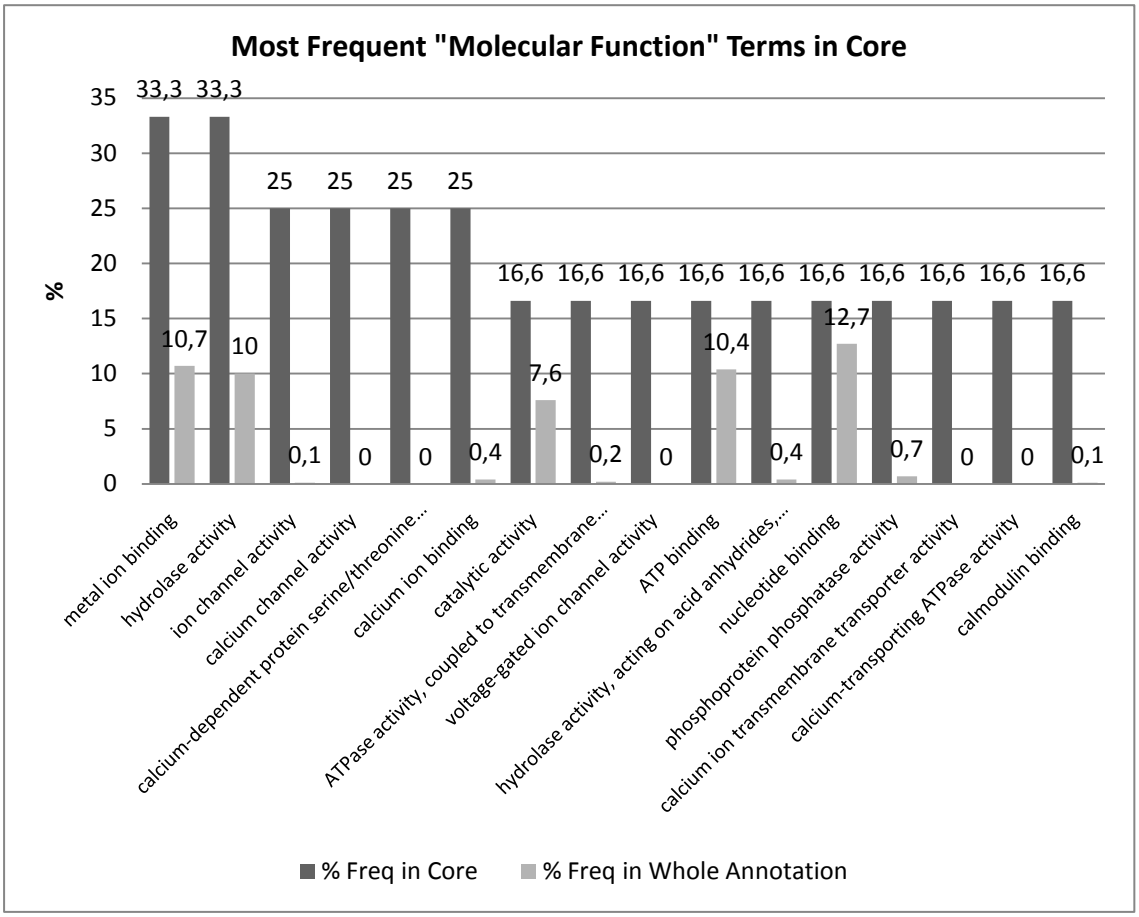


Figure 4.10. Predominant “Molecular Function” terms in yeast Core Proteins

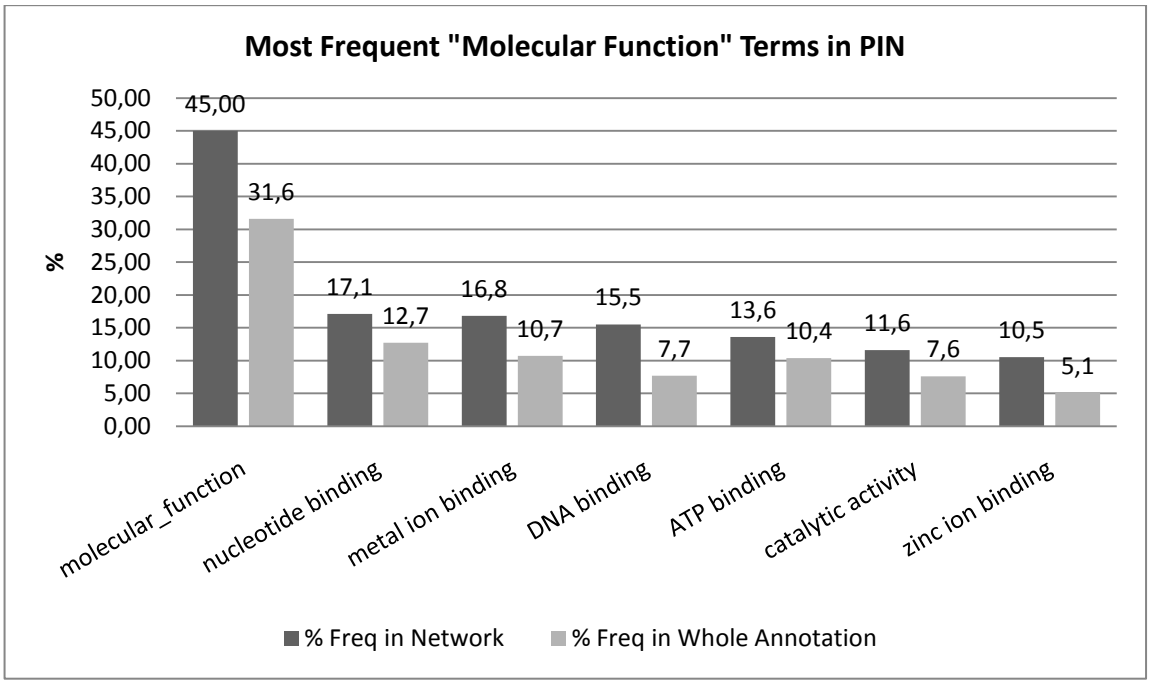


Figure 4.11. Predominant “Molecular Function” terms in yeast PIN

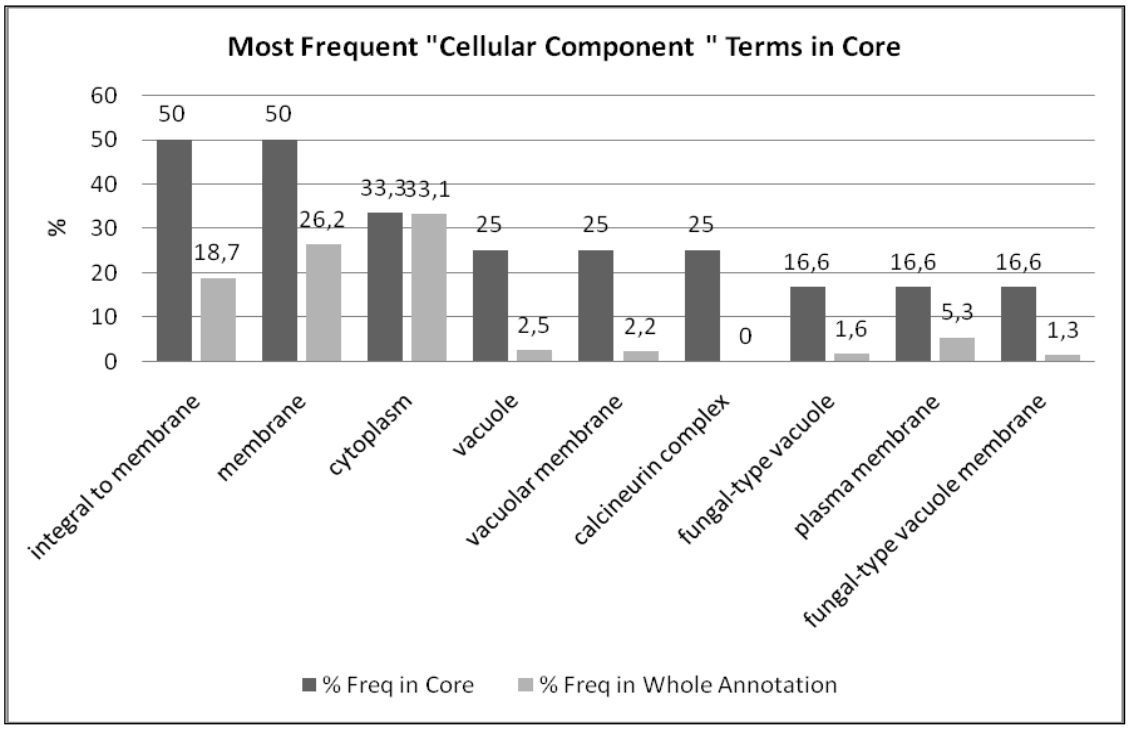


Figure 4.12. Predominant "Cellular Component" terms in yeast Core Proteins

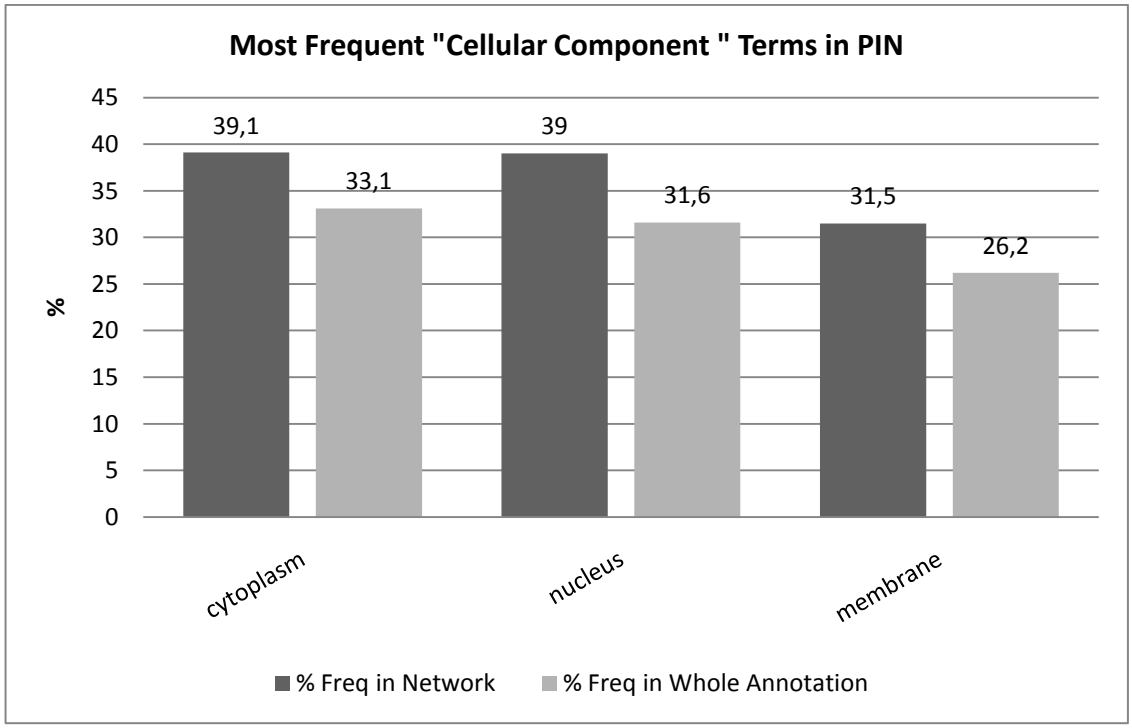


Figure 4.13. Predominant "Cellular Component" terms in yeast PIN

Consequently, it can be concluded that, while the core proteins mainly deal with ion transport and binding for cellular homeostasis especially concerning Calcium; the resultant

network's key processes are related to transcription and transport of proteins for required metabolic or signaling processes.

4.3. Topology Based Analysis of Yeast PIN

4.3.1. Topological Analysis of Yeast Network

The undirected protein interaction graph of the yeast signaling network was analyzed using the Network Analysis algorithm, a Cytoscape plugin. (Table 4.5)

Table 4.7. Topological properties of the yeast signaling networks

YEAST	63 core		12 core		Whole Genome
	Core	*PPI Network*	*Core*	*PPI Network*	*PPI Network*
# of nodes	21	3226	5	1826	5685
# of edges	34	20167	8	10078	52141
Mean connectivity (degree)	3,24	12,5	3,2	11,0	18,3
Mean path length	2,48	3,43	1,2	3,56	2,78
Network diameter	5	9	2	11	6
Network density	0,16	0,004	0,8	0,006	0,003
Network heterogenity	0,67	1,43	0,23	1,46	2,61
Clustering coefficient	0,35	0,15	0,87	0,15	0,24
Network centralization	0,37	0,08	0,33	0,112	0,45
Shortest paths	420	10403850	20	3332450	32313540
Power Law Constant (γ)		1.67 (R^2:0.90)		1.49 (R^2:0.90)	1.41 (R^2:0.87)
# of isolated & lonely nodes	42	2	7	-	922
Total # of proteins	63	3228	12	1826	6607

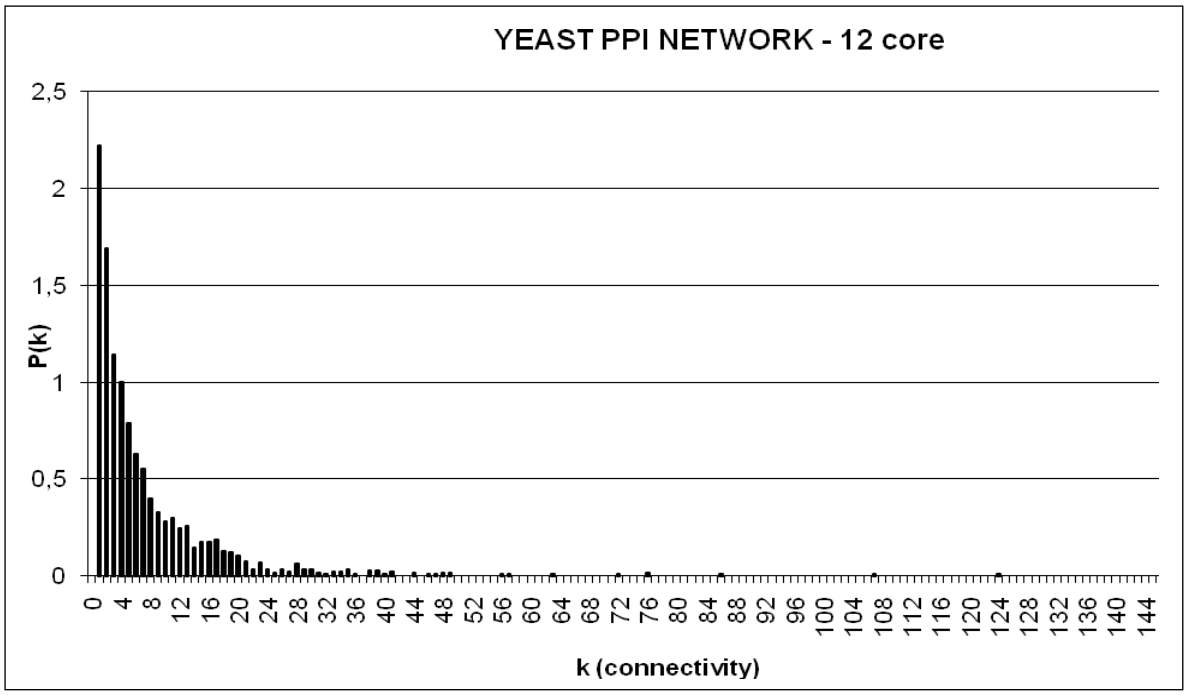


Figure 4.14. Degree distribution in the yeast signaling network

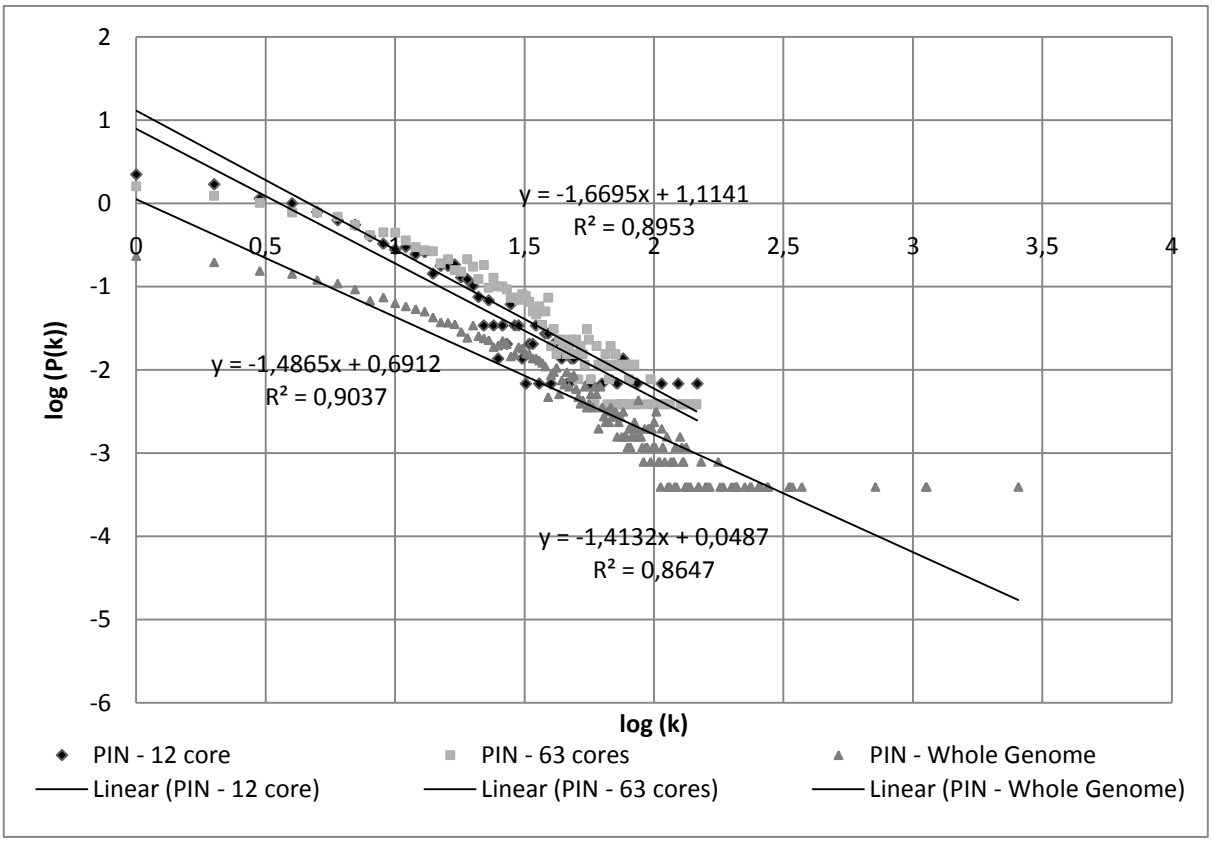


Figure 4.15. Log transform of degrees versus log transform of P(k) (probability of having connectivity, k) in yeast signaling network

Scale-free property & power law: Both three networks follow nearly a power law model ($P(k) \approx k^{-\gamma}$, γ =1.67 with R^2=0.90, γ =1.49 with R^2=0.90 and γ =1.41 with R^2=0.87 for PINs resulted from 63, 12 cores and whole genome respectively), implying that networks can be characterized by a scale-free degree distribution with a general emergence of hubs (Figures 4.12 and 4.13), as many other biological networks. The exponent γ is smaller for 12 core network than that of 63 core network indicating that the 12 core network is more hub dominant than 63 core network. However, whole genome PIN, with the smallest γ value, is the one that is mostly dominated by hubs among those three. When compared to other γ value in literature (1.53 and 1.86 for insulin and EFGR signaling in H. sapiens respectively), the whole genome PIN is preserves its high hub dominance. In addition, 63 and 12 core PINs are more hub dominant than the human's EFGR signaling network.

The network density is the normalized version of the degree and varies between 0 - 1. It shows how densely the network is populated with edges. Since, for both three networks, it is close to 0, one can conclude that not every node has many edges, but some hubs. The network heterogeneity reflects the tendency of a network to contain hub nodes. Network heterogeneity of 1.43, 1.46 and 2.61 supports the scale-free distribution, claiming that, the networks especially the PIN from whole genome have tendency to contain hubs instead of a constant node/edge ratio.

Small-world property, mean path length & diameter: The magnitude of log(nodes) for PINs of 63 and 12 cores and whole genome respectively are 4.3, 3.3 and 3.75 which are of the same order of magnitude as avg. path lengths (3.4, 3.6 and 2.8) and diameters of the PINs (9, 11 and 6) meeting the criteria for having small-world property.

Since the average path lengths of both three networks are almost the same, it can be concluded that these networks have the same efficiency in the transfer of biological information, whole genome network being the fastest among them. When compared to other networks found from literature, these three networks are also the most efficient ones in transfer of biological information, together with network of *M.musculus*.

The diameter of a network essentially characterizes the network's interconnectivity. The diameters of 63 and 12 core networks are near to each other; so, the degree of interconnectivity in the networks does not differ from each other. When compared to the diameters of networks found from literature, they have smaller diameters together with *S. cerevisiae* (signaling) constructed by Arga *et al.*., *H.pylori* and *M.musculus* from DIP. The diameter of the whole genome PIN, on the other hand, is smaller than these two and also the smallest among all given networks in Table 4.4; implying that, it has the highest interconnectivity.

Clustering coefficients: Both 63 and 12 core networks have the same values of clustering coefficients (0.15) suggesting that the neighborhoods of low-degree nodes are densely clustered while those of hub nodes are quite sparsely connected in these networks. Whole genome PIN, however, has higher CC (0.24), showing that it has a high level of redundancy and cohesiveness and is highly modular.

It is seen that, most of the scale free networks listed in Table 4.4 are also highly modular like the three networks studied.

Centrality: In yeast PIN, all the centrality parameters are consistent with the lethality-centrality rule and the scale-free property. In other words, there are many nodes with low connectivity and a few nodes with high connectivity, named hubs (Figure 4.17). Also the hub proteins tend to have the highest betweenness and closeness values within the network (Figure 4.18 and 4.19). There are no outliers observed as in the case of Joy *et al.*.

In conclusion, both 63 and 12 core networks have almost the same topological properties although number of nodes varies by a factor of two and number of edges by a factor of three between networks. The whole genome network is the most robust, highly modular and most dominant hub containing network among those three.

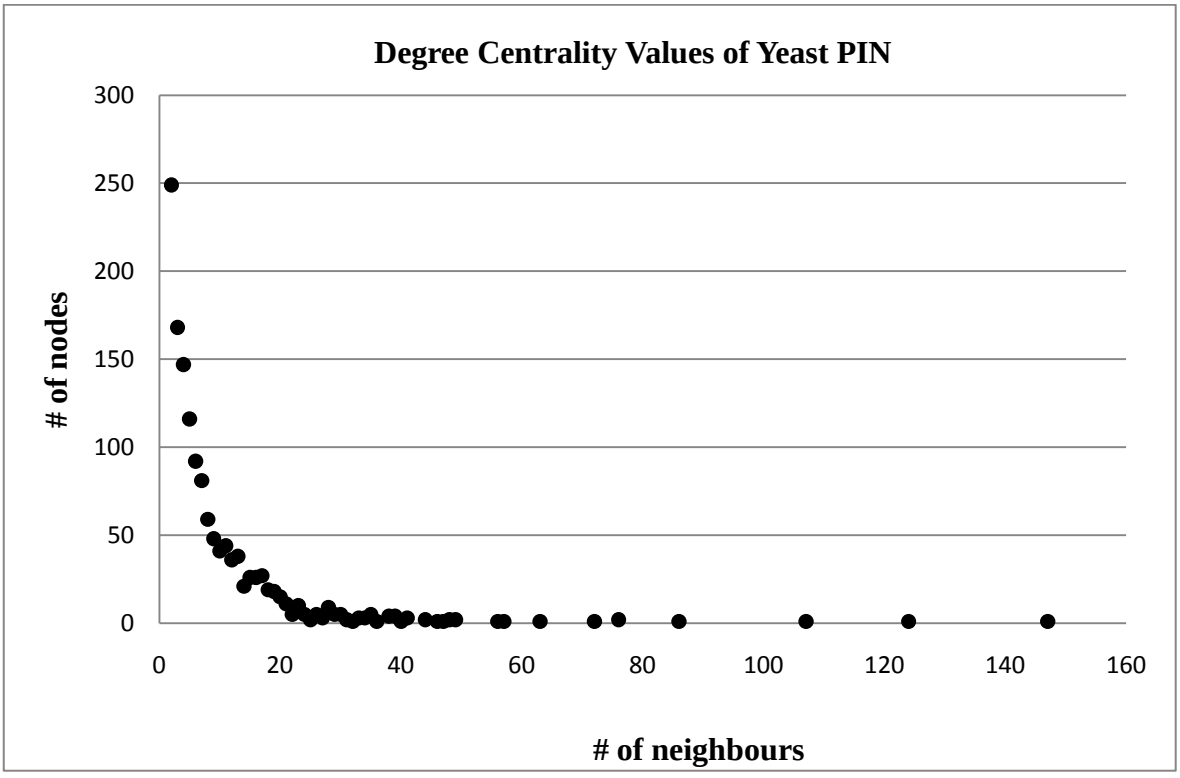


Figure 4.16. Degree centrality distribution of yeast PIN

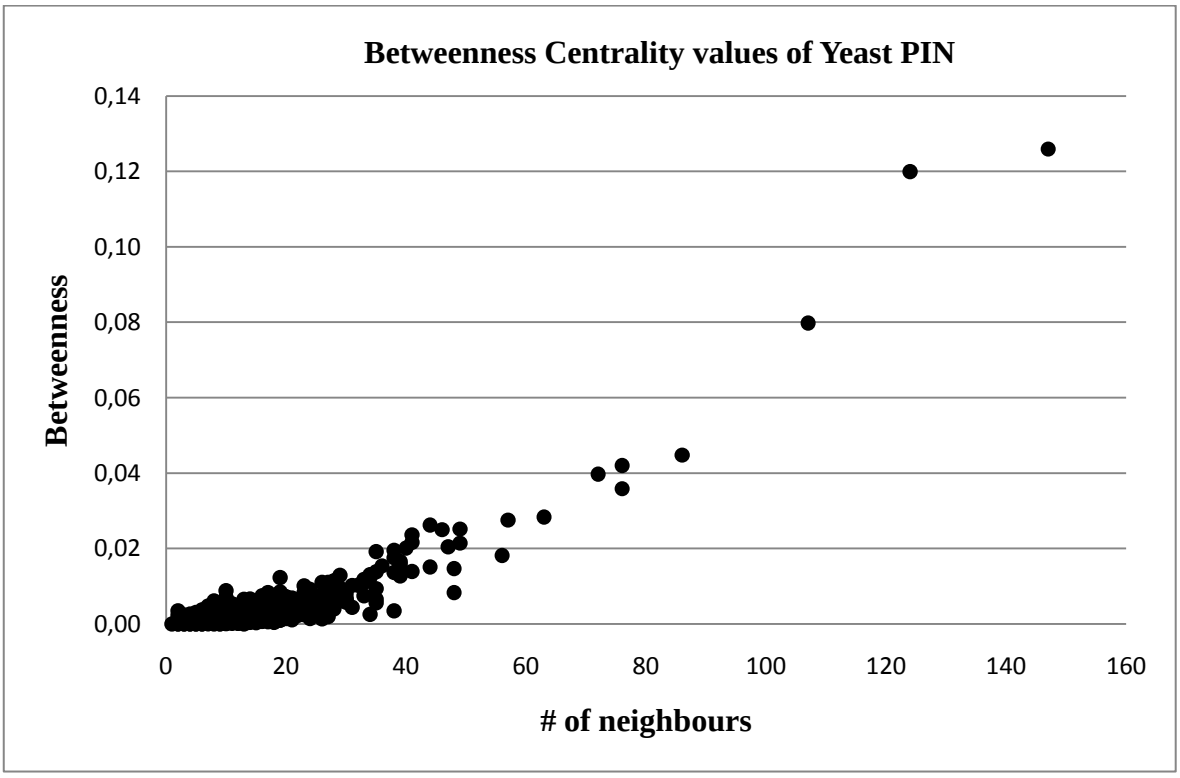


Figure 4.17. Betweenness centrality distribution of yeast PIN

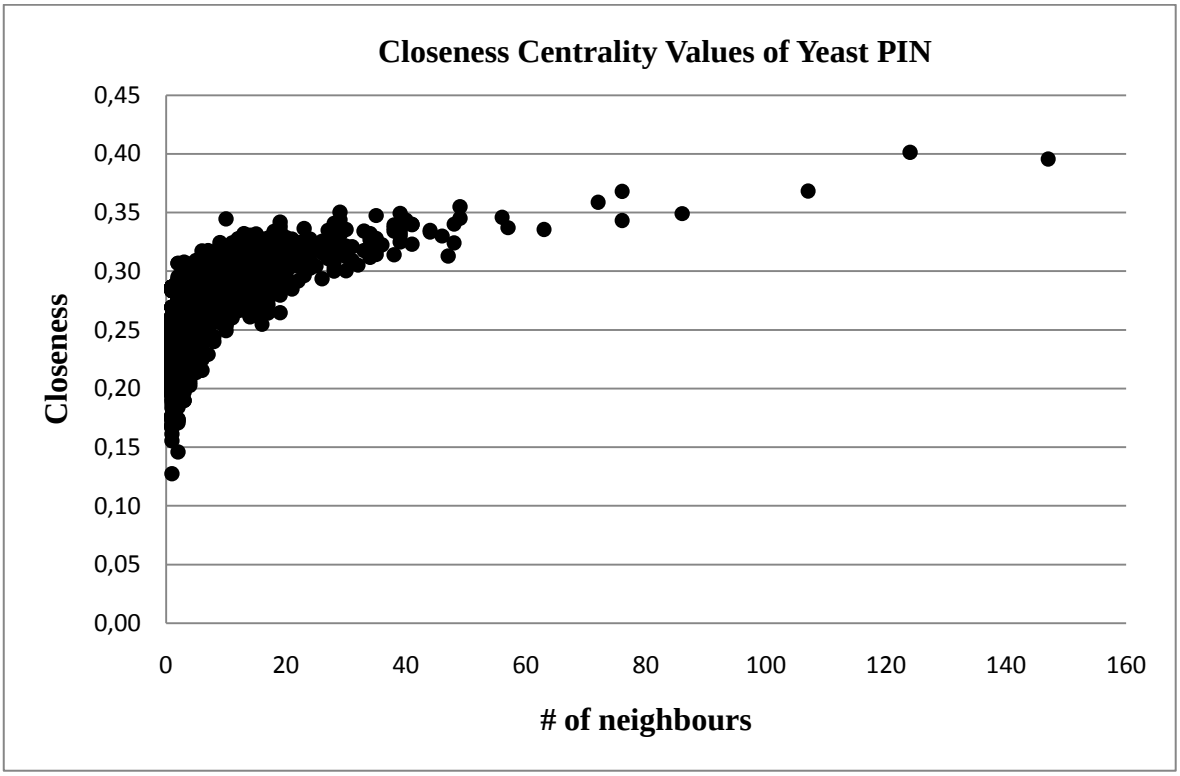
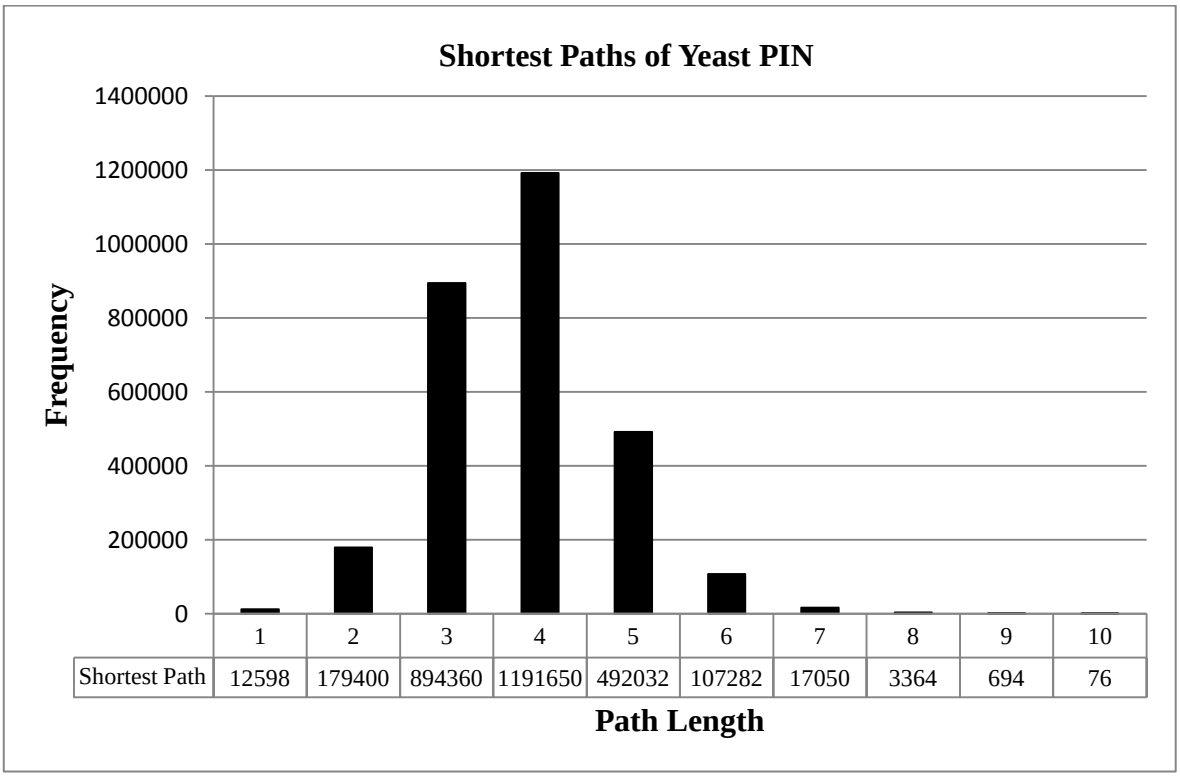


Figure 4.18. Closeness centrality distribution of yeast PIN



	1	2	3	4	5	6	7	8	9	10
Shortest Path	12598	179400	894360	1191650	492032	107282	17050	3364	694	76

Figure 4.19. Shortest path distribution of yeast PIN

Table 4.8. Topological properties of various networks found from literature

	# of nodes	# of edges	Mean connectivity (degree)	Mean path length	Network diameter	Clustering coefficient	Reference
S. cerevisiae (Ca signaling - 12 core)	1826	10078	11,04	3,56	11	0,15	Present study
S. cerevisiae (Ca signaling - 63 core)	3226	20167	12,5	3,43	9	0,15	Present study
S. cerevisiae - BIOGRID	5685	52141	18,3	2,78	6	0,24	Present study
S.cerevisiae - core - DIP	2640	6600	5	5	13	0,33	Li *et al.*
S. cerevisiae	2115	2240	2,12	6,8	14		Jeong et.al.
S. cerevisiae	5253	65673	25	5,21	11		Arga et.al.
S. cerevisiae (signaling)	1363	3649	5,35	6,81	9		Arga et.al.
S.cerevisiae - DIP	4773	15444	6,5	5,01	13		Li *et al.*
S.cerevisiae - MIPS	2043	5434	5,32	7,71	16		Yook et.al
S.cerevisiae - DIP	5798	20098	6,93	4,9	12		Yook et.al
S.cerevisiae	1870	4480	4,79	6,8	15		Uetz et.al.
S.cerevisiae	3280	8868	5,41	4,9	13		Ito et.al.
S.cerevisiae - core	797	1560	3,91	6,2	16		Ito et.al.
H. Sapiens (Insulin)	498	2887	11,6	2,9	5		Tekir et.al
H.Sapiens (EFGR signaling)	329	1795	10,91	4,7	11		Tekir et.al
D.melanogaster - DIP	7451	22819	6,2	4,4	11	0,02	Li *et al.*
C.elegans - DIP	2638	4030	3,1	4,8	14	0,06	Li *et al.*
C.elegans (Noncanonical Wnt)	667	922	2,76	5,39	21		Alazi et.al
H.pylori - DIP	710	1420	4	4,1	9	0,08	Li *et al.*
H.sapiens - DIP	1065	1369	2,6	6,8	21	0,21	Li *et al.*
E.coli - DIP	553	761	2,8	5,5	16	0,62	Li *et al.*
M.musculus - DIP	329	286	1,7	3,6	9	0,15	Li *et al.*

4.3.2. Hub Proteins of Yeast Network

The evolutionary rate of protein having more protein–protein interactions is much slower than those proteins having fewer interacting partners (Makino *et al.*, 2006). These highly connected proteins called hubs play an important role in biological processes of the cell.

In order to identify the hub proteins in the network, the centrality measures and clustering coefficient parameters of each node in the network are thoroughly analyzed. Hub nodes in a network are defined as nodes involved in high number of interactions - high degree (Przulj N., 2005), being able to spread information to other reachable nodes immediately - high closeness centrality (Ortutay and Vihinen, 2009, Gürsoy *et al.*, 2008) and traversed by high number of shortest paths - high stress and high betweenness centralities (Hahn and Kern, 2005, Ravasz *et al.* 2002). It is also reported that most neighbors of highly connected nodes have suppressed connectivity, implying that neighbors of hubs are sparsely connected – low neighborhood connectivity and low clustering coefficient (Maslov and Sneppen, 2002, Przulj N., 2005). In the light of these statements, the nodes sharing the high degree, high closeness, high stress, high betweenness, low clustering coefficient and low neighborhood connectivity measures are chosen to be hubs, namely HSP82, CDC28 and TPK1 (Table 4.7). These hubs have 418 interactions with 383 proteins in total covering 4.1 % of the interactions and 21.0 % of the proteins in PIN (Table 4.8). The detailed information on hub proteins is given in App. C.

HSP82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures (Borkowich *et al.*, 1989). CDC28, on the other hand, has an essential surveillance function required to maintain genetic stability through mitosis (Kitazono and Kron, 2002). For TPK1, it is stated that while no single TPK gene is essential and the Tpk isomers are functionally redundant for cell viability, the presence of at least one of the isomers is required for normal growth since more than 87% of all PKA substrates are uniquely phosphorylated by just one of the TPKs (Toda *et al.*., 1987).

Although no experimental evidence can be found in literature implying a direct relation of hub proteins with calcium, they are determined as the key proteins to the

processes like cell division, pheromone signaling or transcription known to be triggered by calcium signaling in the cell. Particularly, a probable hub protein LGE1, with no known molecular function annotation, is also proposed to have a novel function in promoting transcription elongation (Song *et al.*, 2010).

The hub proteins' dominant molecular function terms are *nucleotide binding, ATP binding* and *transferase activity* taking place in the biological processes of *protein amino acid phosphorylation, vacuolar protein catabolic activity* and *regulation of transport* which possibly occurs in *cytoplasm* and *nucleus* (Figure 4.21).

When compared to the frequent terms in PIN, it can be concluded that the key process in transcription (frequent biological process term in PIN), is amino acid phosphorylation (frequent biological process term in hubs); achieved by nucleotide binding and transferase activity taking place in nucleus (frequent function and component terms in hubs).

In addition; the crucial steps of other important processes in PIN, namely, metabolic process and protein transport occurring in cytoplasm, are vacuolar protein catabolic activity and regulation of transport taking place in cytoplasm (frequent terms in hubs).

Being a hub protein implies to be an essential protein for the survival of the cell via achieving an important function in a key process. It is expected to observe high conservation scores and functionally important domain sites in a hub protein.

The ConSeq results (Figure 4.22 and Table 4.9) show that functionally important residues are more than structurally important residues being consistent with the abundance of exposed residues. The conservation score of heat shock protein family (HSP82) is higher than the protein kinase family. HSPs are a class of functionally related proteins whose expression is increased when cells are exposed to elevated temperatures or other stress and their function is crucial for the cell to survive under stress. The highly conserved HSPs in PIN are ATPases responsible for transport in cytosol. Hsp90 domain (HSP82) plays an important role in the folding of newly synthesized proteins and stabilization and refolding of denatured proteins after stress.

Protein kinases (CDC28, TPK1) are a group of enzymes that possess a catalytic subunit which transfers the gamma phosphate from nucleotide triphosphates (often ATP) to one or more amino acid residues in a protein substrate side chain, resulting in a conformational change affecting protein function. Protein kinases play a role in a multitude of cellular processes, including division, proliferation, apoptosis, and differentiation. Phosphorylation usually results in a functional change of the target protein by changing enzyme activity, cellular location, or association with other proteins.

Protein kinases share a conserved catalytic core common with both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. In the N-terminal extremity of the catalytic domain there is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. In the central part of the catalytic domain there is a conserved aspartic acid residue which is important for the catalytic activity of the enzyme.

The domain conservation scores (Table 4.9) imply that all the conserved residues in the whole protein are placed in domain sites. Hence, the functionally and structurally important residues of the protein are also found in domain sites and the domains of the proteins characterizes the function of the concerning protein. HSP proteins contain HSP related domains and kinase proteins include Pkinase domains.

The dominant role the hubs have in the network can be seen in Figure 4.23. With a total of 418 interactions with 383 nodes (almost 21% of all nodes in the PIN), the hubs are the important connectors in the PIN.

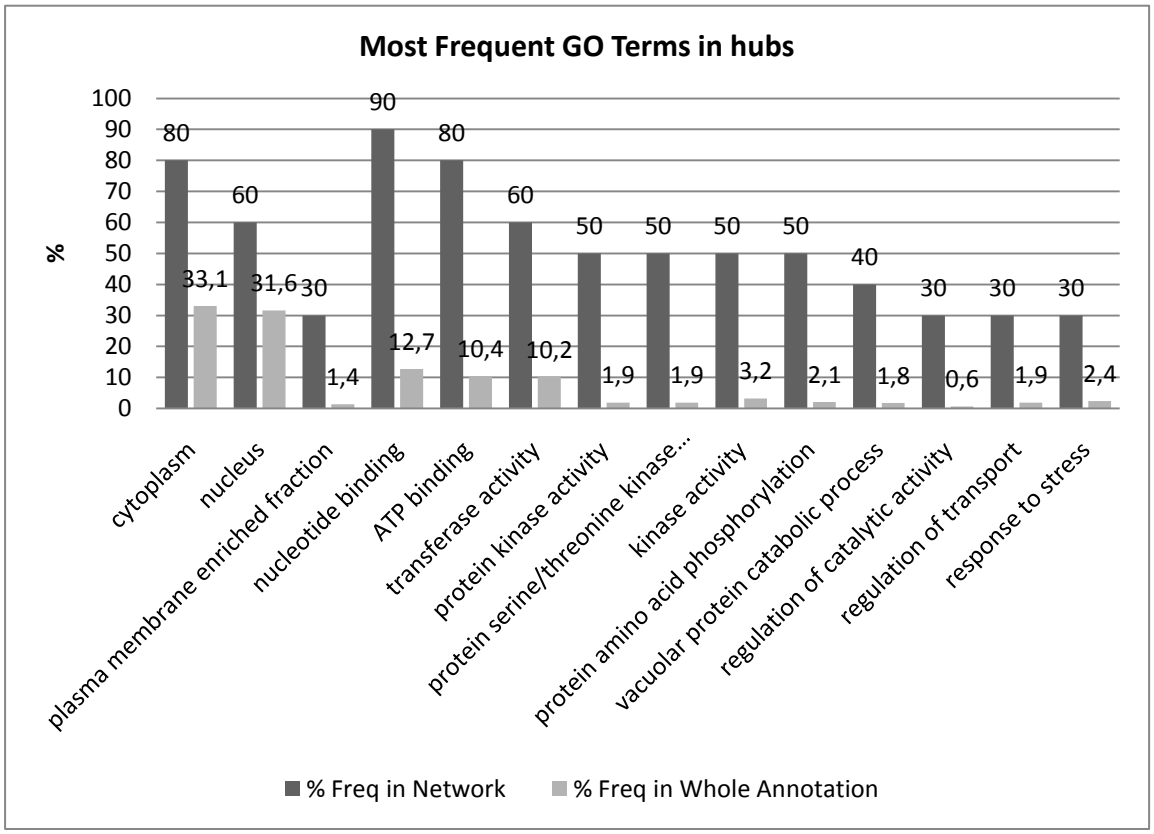
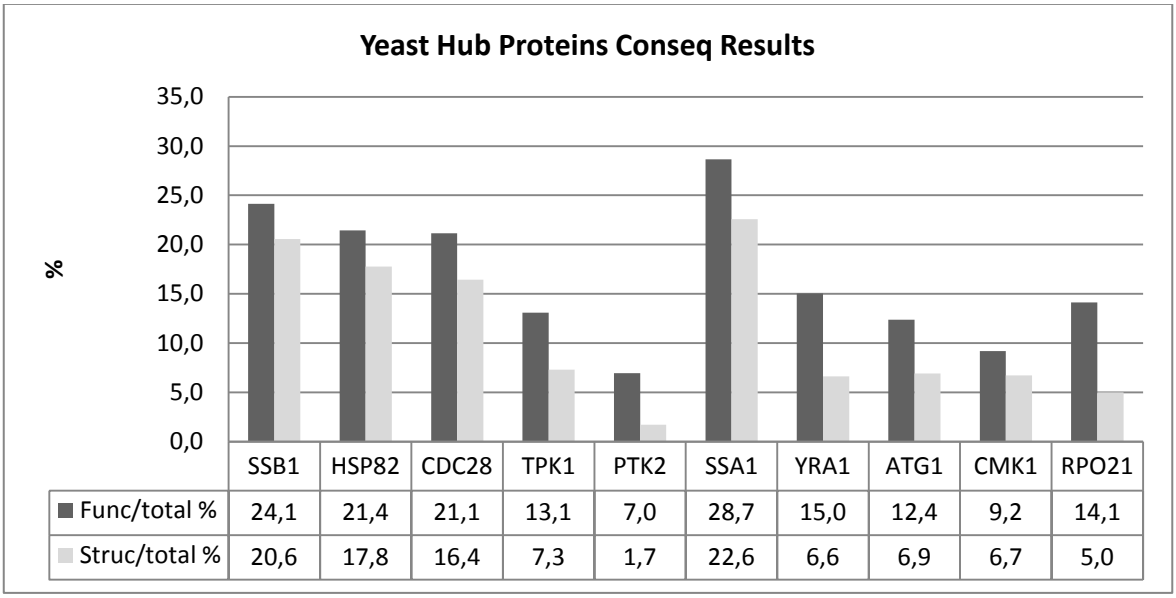


Figure 4.20. Predominant GO terms in yeast PIN hubs



	SSB1	HSP82	CDC28	TPK1	PTK2	SSA1	YRA1	ATG1	CMK1	RPO21
Func/total %	24,1	21,4	21,1	13,1	7,0	28,7	15,0	12,4	9,2	14,1
Struc/total %	20,6	17,8	16,4	7,3	1,7	22,6	6,6	6,9	6,7	5,0

Figure 4.21. The conserved residue analysis of yeast hub proteins

Table 4.9. Topological parameters of probable hub proteins of yeast signaling network

	Protein	Degree	Betweenness	Closeness	Stress	Neighborhood Connectivity	Clustering Coefficient
YPL240C	HSP82	216	0.14	0.43	4110884	19.71	0.02
YLR262C	YPT6	162	0.04	0.40	1659346	33.49	0.09
YBR160W	CDC28	112	0.06	0.39	1623734	19.99	0.03
YGL058W	RAD6	104	0.03	0.39	1083304	37.44	0.11
YPL055C	LGE1	101	0.01	0.38	613564	43.66	0.18
YOL012C	HTZ1	99	0.01	0.39	534148	46.68	0.21
YDL020C	RPN4	95	0.02	0.38	930160	35.56	0.09
YJL164C	TPK1	91	0.04	0.37	1158242	14.65	0.01
YDR159W	SAC3	91	0.01	0.37	606262	40.42	0.15
YDR392W	SPT3	86	0.01	0.38	419046	45.76	0.20

Table 4.10. Hub Proteins of the Yeast signaling network

ORF	name	ints in PIN	ints in whole genome PIN	Uniprot	aa	PDB	Domains	Description
YPL240C	HSP82	216	342 6th	P02829	709	2CG9 1-677	HATPase_c HSP90	Hsp90 chaperone required for pheromone signaling and negative regulation of Hsf1p
YBR160W	CDC28	112	253 12th	P00546	298		Pkinase	Catalytic subunit of the main cell cycle cyclin-dependent kinase (CDK)
YJL164C	TPK1	91	276 9th	P06244	397	1FOT 81-397	Pkinase	cAMP-dependent protein kinase catalytic subunit

Table 4.11. ConSeq results of yeast hub proteins and their domains

ORF name	Protein name	Tot. Res.	Cons	Domain Cons. in	Exposed	Buried	Func/conserved	Domain F/Cons F	Struc/conserved	Domain S/Cons S
			%	Cons. Residues	%	%	%	%	%	%
YPL240C	HSP82	29	45,7		58,3	41,7	46,9		38,9	
	HATPase_c	411	56,8	27,2	51,0	49,0	37,5	21,7	46,6	32,5
	HSP90	150	41,2	67,0	60,9	39,1	51,6	73,7	34,1	58,7
YBR160W	CDC28	134	45,0		54,7	45,3	47,0		36,6	
	Pkinase	1173	46,2	99,3	53,1	46,9	46,6	98,4	36,8	100,0
YJL164C	TPK1	71	27,2		67,8	32,2	48,1		26,9	
	Pkinase	277	38,4	90,7	55,7	44,3	45,9	86,5	28,6	96,6

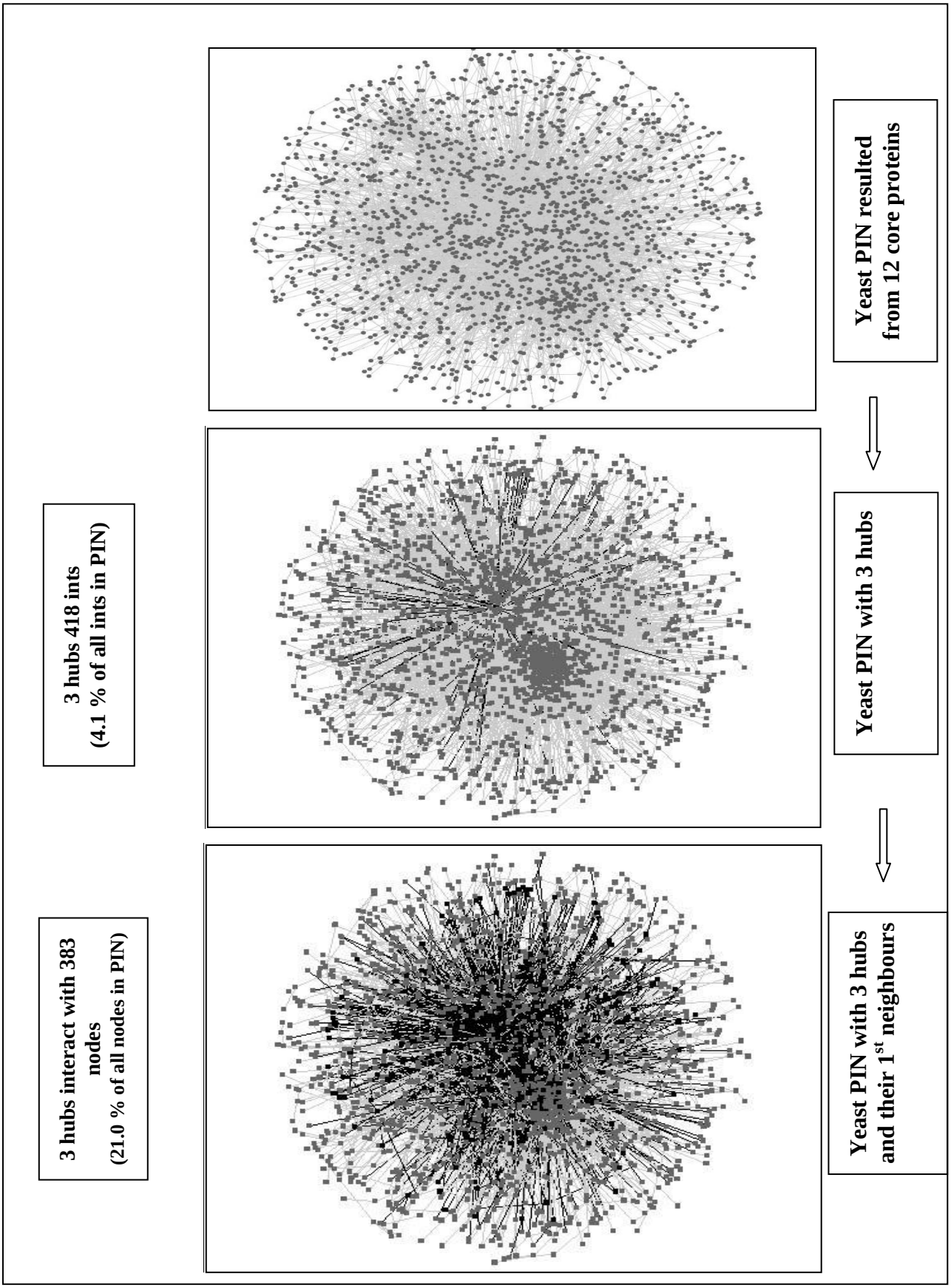


Figure 4.22. The hub proteins and their interactions in the yeast PIN

In order to determine the role of hub proteins in the reconstructed PIN, they are compared to the hubs of the PIN constructed from whole genome (5685 nodes, 52141 edges). 10 hubs in the first 40 hubs of whole yeast genome PIN (mostly in the first 20 hubs) are also the hubs of the reconstructed PIN for calcium. In addition to those, 5 hub proteins in the first 50 hubs of whole genome PIN are a member of calcium PIN.

These results show that, the hubs of calcium PIN do not only have important roles in calcium related paths, but also in the whole genome for yeast to survive implying the vitality of calcium signaling pathway for yeast.

4.3.3. Module Detection and Proteins with Unknown Function

The prevalence of the modularity paradigm in molecular cell biology has led to an extensive use of modules in prediction of molecular functions. By functional module, one typically means a group of cellular components and their interactions that can be attributed to a specific biological function (Hartwell et al, 1999). Instead of predicting functions for individual genes, a module-assisted approach attempts to first identify coherent groups of genes and then assign functions to all the genes in each group. Once a module is obtained, simple methods are usually used for function prediction within the module. For example, every function shared by the majority of the module's genes is assigned to all the genes in the module.

In Yeast genome, there are 582 proteins with unknown function, 569 of them being in the reconstructed network with 2223 interactions.

Using MCODE, a Cytoscape plugin, 30 modules containing 96 functionally unknown proteins are determined (details and figures can be found in Table D3.1.) Out of these 30 modules, 12 modules (Modules 1-14 except module 3&7) are chosen to be appropriate for function assignment to functionally unknown proteins. The module selection criterion depends on the number of functionally known protein dominance in the mentioned module/subgroup interacting with the functionally unknown protein. For example, in Module 1, there are 3 subgroups; one containing proteins with known functions, second one comprising of 3 functionally unknown proteins (dark nodes) with 7

known proteins, and finally the one in the circle with 1 functionally unknown protein interacting with 8 functionally known proteins out of 10. Since the criteria for functional annotation is to find the group containing the least unknown and most known proteins, only the circled subgroup is suitable for the purpose.

Another approach in function prediction is the simplest and most direct method for function prediction, neighborhood counting, which determines the function of a protein based on the known function of proteins lying in its immediate neighborhood. The module based assigned functions are checked whether they match with the findings of neighborhood counting method (Table D.3.2).

The assigned functions and properties of related proteins can be found in Table 4.19. More detailed information can be found in App D.

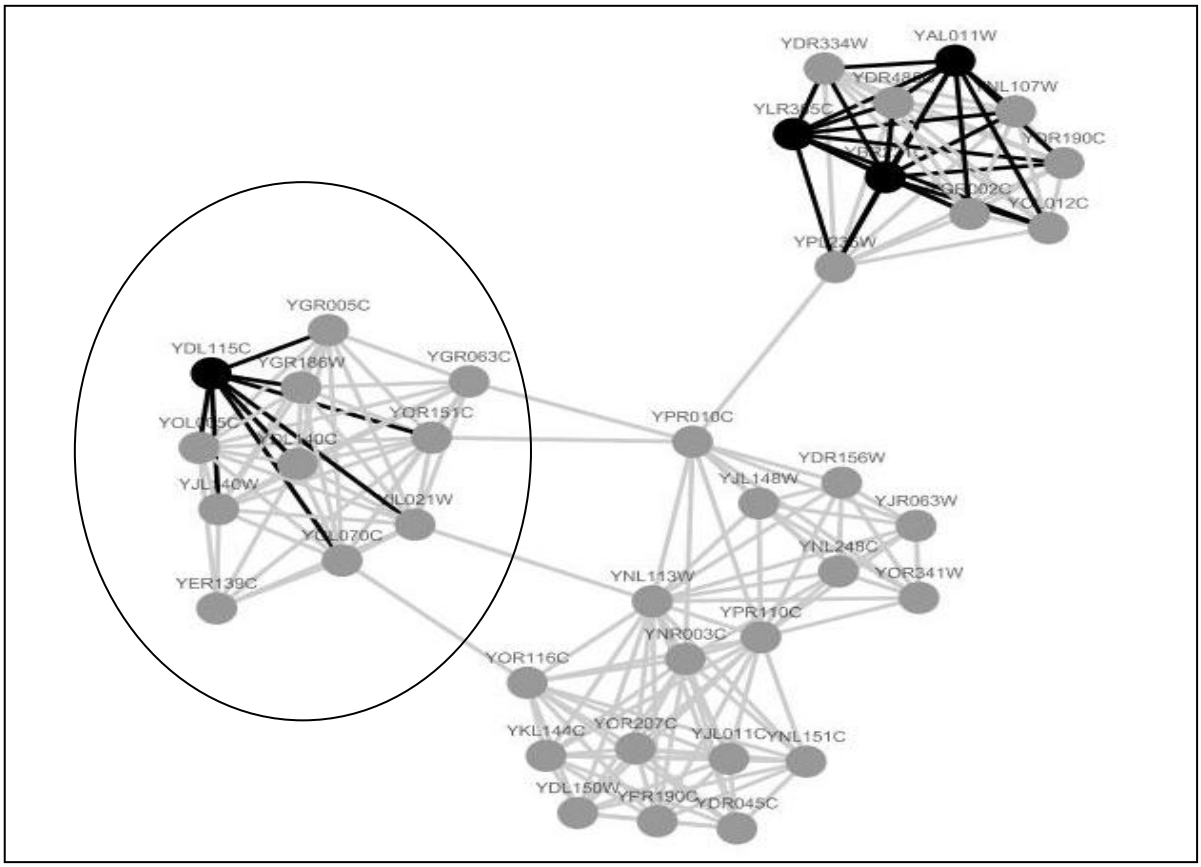


Figure 4.23. Module 1 used in function assignment to YDL115C

Table 4.12. Functional Enrichment in Module 1

Go Id	Go Name	Freq in Module	Freq in Genome	Nodes
3677	DNA binding	7/11 63.6	487/6310 7.7	YGR186W YGR005C YOR151C YDL140C YGL070C YIL021W YOL005C
46872	metal ion binding	6/11 54.5	679/6310 10.7	YOR151C YGR063C YDL140C YGL070C YIL021W YER139C
3899	DNA-directed RNA polymerase activity	6/11 54.5	36/6310 0.5	YOR151C YJL140W YDL140C YGL070C YIL021W YOL005C

In Module 1, the sub-module of 11 proteins in the circle is the suitable one for function assignment. The proteins involved in this module are mainly responsible for DNA binding (63.6 %), metal ion binding (54.5 %) and DNA-directed RNA polymerase activity (54.5 %). So it can be claimed that, YDL115C, the functionally unknown protein may also be responsible for these functions.

Function assignment to YDL115C (IWR1) from Module 1: One of the biological process terms for this protein given in Gene Ontology Database is “general transcription from RNA polymerase II promoter” and “gene-specific transcription from RNA polymerase II promoter” which are both inferred from experiment (Mutant Phenotype - IMP) (See GO Evidence Code list in App A.2). Also the cellular component term “DNA-directed RNA polymerase II, holoenzyme”, is inferred from experiment (Physical Interaction – IPI) to this protein. So, the assignment of “DNA binding” and “DNA-directed RNA polymerase activity” functions to this protein seems highly appropriate.

The third molecular function term assigned is “metal ion binding”. It is known that RNA polymerase II transcription factors require metal to bind specific DNA sequences. And Zn^{2+} is most likely the metal involved in this process (Dailey *et al.*., 1987).

This protein has a domain of unknown function. It is known to interact with RNA polymerase II and deletion of this protein results in hypersensitivity to the K1 killer toxin, which supports the DNA directed polymerase activity.

This protein has 70 physical interactions in various databases and the common functional term of 21 proteins out of the 70 interacting proteins is DNA binding (30 %) (Table D.3.2).

That is to say; these three functional annotation assignments for IWR1 are consistent within each other and also with the biological process and cellular component terms inferred by experiments.

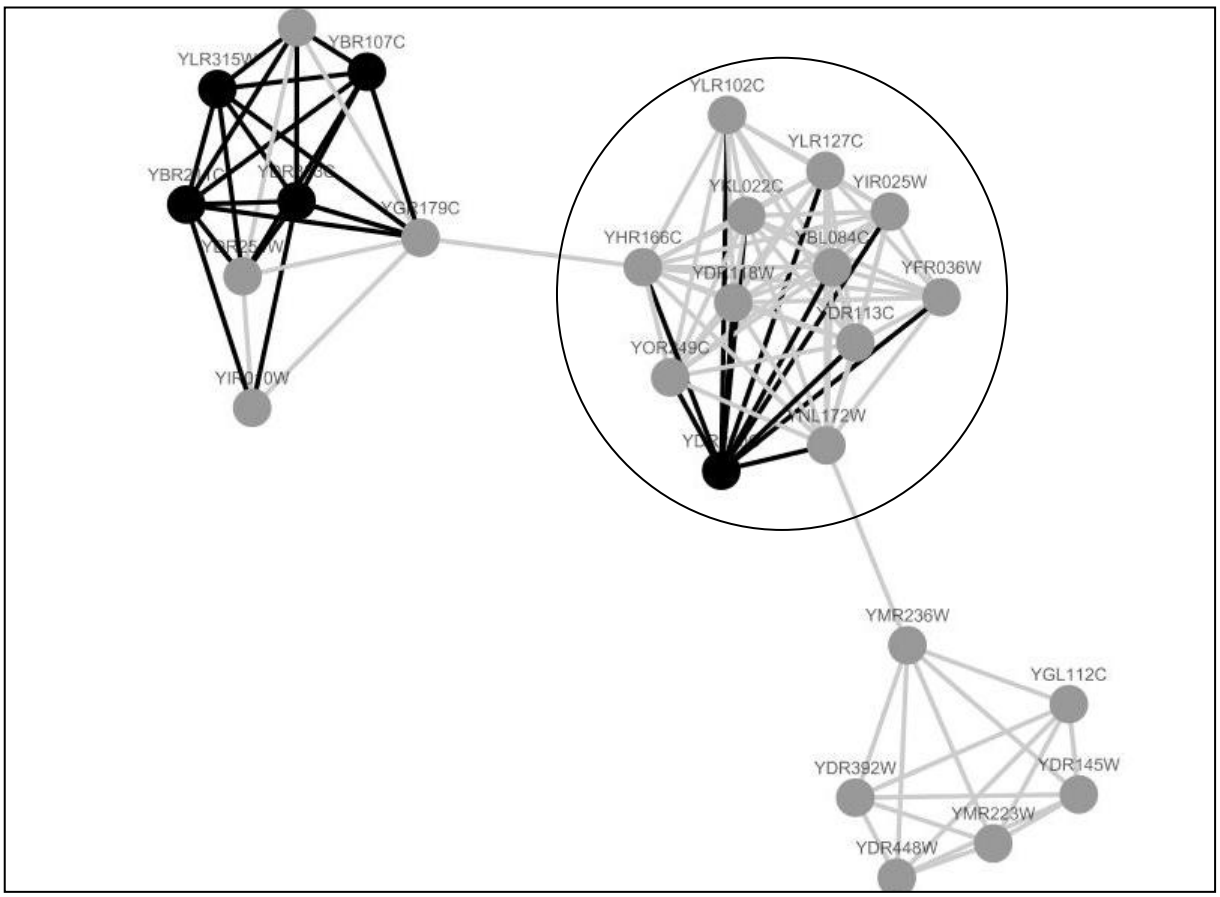


Figure 4.24. Module 2 used in function assignments to YIR025W and YDR383C

In Module 2, there are three sub-group of proteins; one having functionally known proteins, the other one having so many functionally unknown proteins and the last one in the circle having 2 functionally unknown proteins (YDR260C and YIR025W) out of 12. The dominant molecular function terms in this sub-group are protein binding (83.3 % - all proteins except the 2 unknowns) and ubiquitin-protein ligase activity (75% - 9 proteins out

of 10 functionally known one). Hence, the two functionally unknown proteins are most likely to take place in these functions in the cell.

Table 4.13. Functional Enrichment in Module 2

Go Id	Go Name	Freq in Module	Freq in Genome	Nodes
5515	protein binding	10/12 83.3	242/6310 3.8	YBL084C YNL172W YOR249C YLR102C YLR127C YDR118W YHR166C YDR113C YKL022C YFR036W
4842	ubiquitin-protein ligase activity	9/12 75.0	82/6310 1.2	YBL084C YNL172W YOR249C YLR102C YLR127C YDR118W YHR166C YKL022C YFR036W

Function assignment to YDR383C (NKP1) from Module 2: For this protein, there are only electronically annotated biological process terms in GO database related with cell cycle, mitosis and meiosis. Since there is no experimental evidence concerning the biological processes annotated to the protein; the “biological process unknown” term is also listed.

For cellular component term, however, there is the term “kinetochore” inferred from direct assay. Kinetochore is the protein structure on chromosomes where the spindle fibers attach during cell division to pull the chromosomes apart. It assembles on the centromere and links the chromosome to microtubule polymers from the mitotic spindle during mitosis and meiosis.

The molecular function annotations derived for NKP1 are “protein binding” and “ubiquitin-protein ligase activity”. In order to verify the validity of these assignments, the relation between kinetochores and ubiquitin-protein ligase should be unveiled.

An ubiquitin protein ligase complex called SCF is responsible for initiating S-phase. The ubiquitin protein ligase complex called anaphase-promoting complex (APC) is responsible for initiating Anaphase.

Kinetochores that lack stable attachments to the spindle microtubules serve as catalytic staging areas for the assembly of inhibitor complexes. These inhibitor complexes then leave the kinetochores and block activity of the APC throughout the cell. It is suggested that microtubule occupancy at kinetochores turns off the capability of the kinetochore to produce the APC/C inhibitor. Subsequently, the inhibitor concentration in the cell wanes and anaphase initiates (Gorbsky *et al*., 1999).

To conclude, "protein binding" and "ubiquitin-protein ligase activity" terms match with the experimentally inferred term "kinetochore" as well as supporting the accuracy of the electronically annotated biological process terms (cell cycle, mitosis and meiosis).

Function Assignment to YIR025W (MND2) from Module 2: The biological process and cellular component terms associated with this protein inferred from experimental methods are related to "protein ubiquitination" by the "anaphase promoting complex", inducing "mitotic metaphase/anaphase transition" leading to "chromatid segregation" at anaphase in "nucleus".

MND2 has Apc15p protein (PF05841) domain. The anaphase-promoting complex (APC) is a conserved multi-subunit ubiquitin ligase required for the degradation of key cell cycle regulators. Members of this family are components of the anaphase-promoting complex homologous to Apc15p. The presence of this domain supports the assigned function "ubiquitin-protein ligase activity".

Out of 26 proteins which MND2 physically interacts, almost half of them are responsible for "protein binding" and "ubiquitin-protein ligase activity" (46.1 % each) (Table D.3.2). IntAct also assigns "protein binding" term to this protein stemming from the interactions it has.

MND2, being in the same modular group (Module 2) with SWM1 is also found to function in "protein binding" and "ubiquitin-protein ligase activity" which are highly probable to be true.

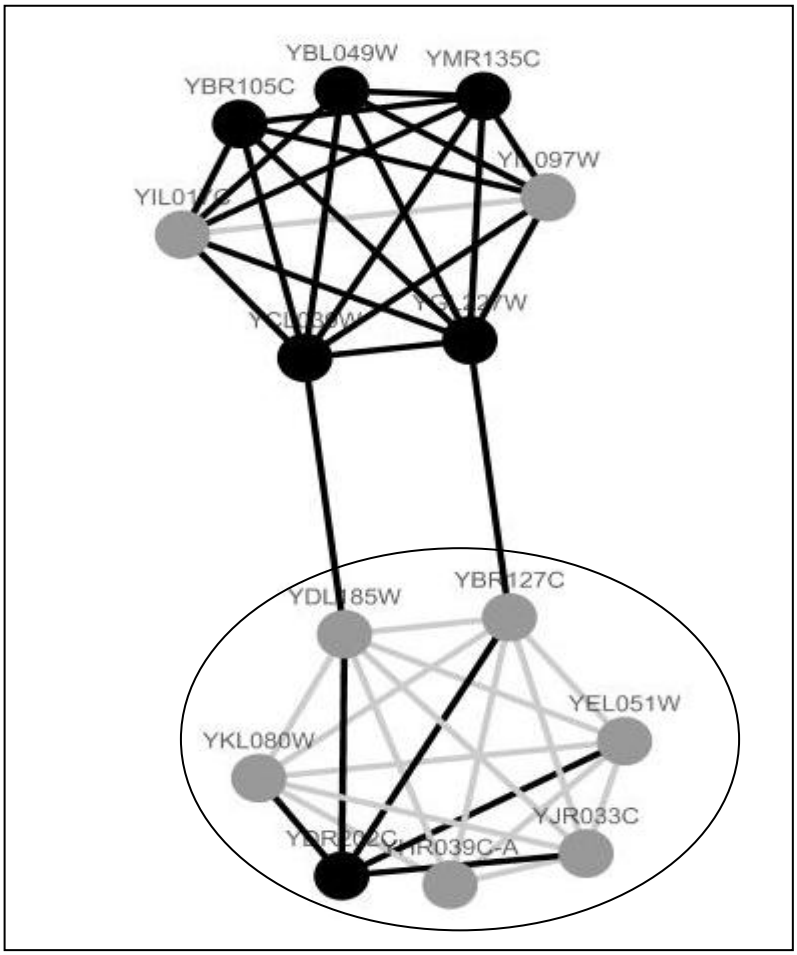


Figure 4.25. Module 4 used in function assignments to YDR202C

Module 4 has 2 subgroups; the upper one with only 2 known proteins and the lower one with one unknown protein, YDR202C, out of 7. For the lower one in circle, proton-transporting ATPase activity (85.7 % - all proteins except the unknown) and hydrolase activities (71.4 and 57.1 %) are in abundance. So, the unknown protein's function is highly possible to be related to the mentioned ATPase and hydrolase activities.

Function assignment to YDR202C (RAV2) from Module 4: It is experimentally determined and listed in GO database that RAV2 has a role in "vacuolar acidification" through "regulation of protein complex assembly" by being a part of the "RAVE complex".

Saccharomyces cerevisiae has a lysosome-like vacuole that is acidic and contains a variety of hydrolases. Vacuolar acidification has been postulated to play a central role in

such processes as Ca^{2+} homeostasis. In vitro studies have shown that there are Ca^{2+} transport systems in the yeast vacuole that require a proton electrochemical gradient across the vacuolar membrane. The vacuolar proton-translocating ATPases (V-ATPase) are multi-subunit enzymes composed of a peripheral complex (V_1) attached to a membrane-bound complex V_0) (Kane, 2006), responsible for the acidification of the vacuolar network (Yamashiro *et al*., 1990).

Table 4.14. Functional Enrichment in Module 4

Go Id	**Go Name**	**Freq in Module**	**Freq in Genome**	**Nodes**
46961	proton-transporting ATPase activity, rotational mechanism	6/7 85.7	20/6310 0.3	YKL080W YBR127C YHR039C-A YDL185W YJR033C YEL051W
16820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	5/7 71.4	26/6310 0.4	YKL080W YBR127C YHR039C-A YDL185W YJR033C
16787	hydrolase activity	4/7 57.1	637/6310 10.0	YBR127C YHR039C-A YDL185W YEL051W

The RAVE complex plays a role in both regulatory assembly and disassembly of the V-ATPase and in biosynthetic assembly of the V-ATPase. In the absence of the RAVE complex, V_0 complexes are assembled and targeted to the vacuole, and there is some attachment of V_1 to V_0, but the assembled complexes are unstable and there is very little ATPase activity in isolated vacuoles (Kane, 2006).

The protein's functions assigned in this work are "proton-transporting ATPase activity" and "hyrolase activity" especially "catalyzing transmembrane movement of substances". These annotations for the protein seem logical since vacuolar acidification is achieved via V-ATPase and in order for the V-ATPase pump to be active, ATP should be hydrolyzed into ADP and inorganic phosphate (which binds to the V-ATPase to make it active) via hyrolase activity.

Also, the neighborhood count method results match with the terms assigned. RAV2 has 11 interactions and most frequent function terms in the neighbors are “proton-transporting ATPase activity” (45.4 %) and “hyrolase activity” (27.2 %) especially “catalyzing transmembrane movement of substances” (27.2 %).

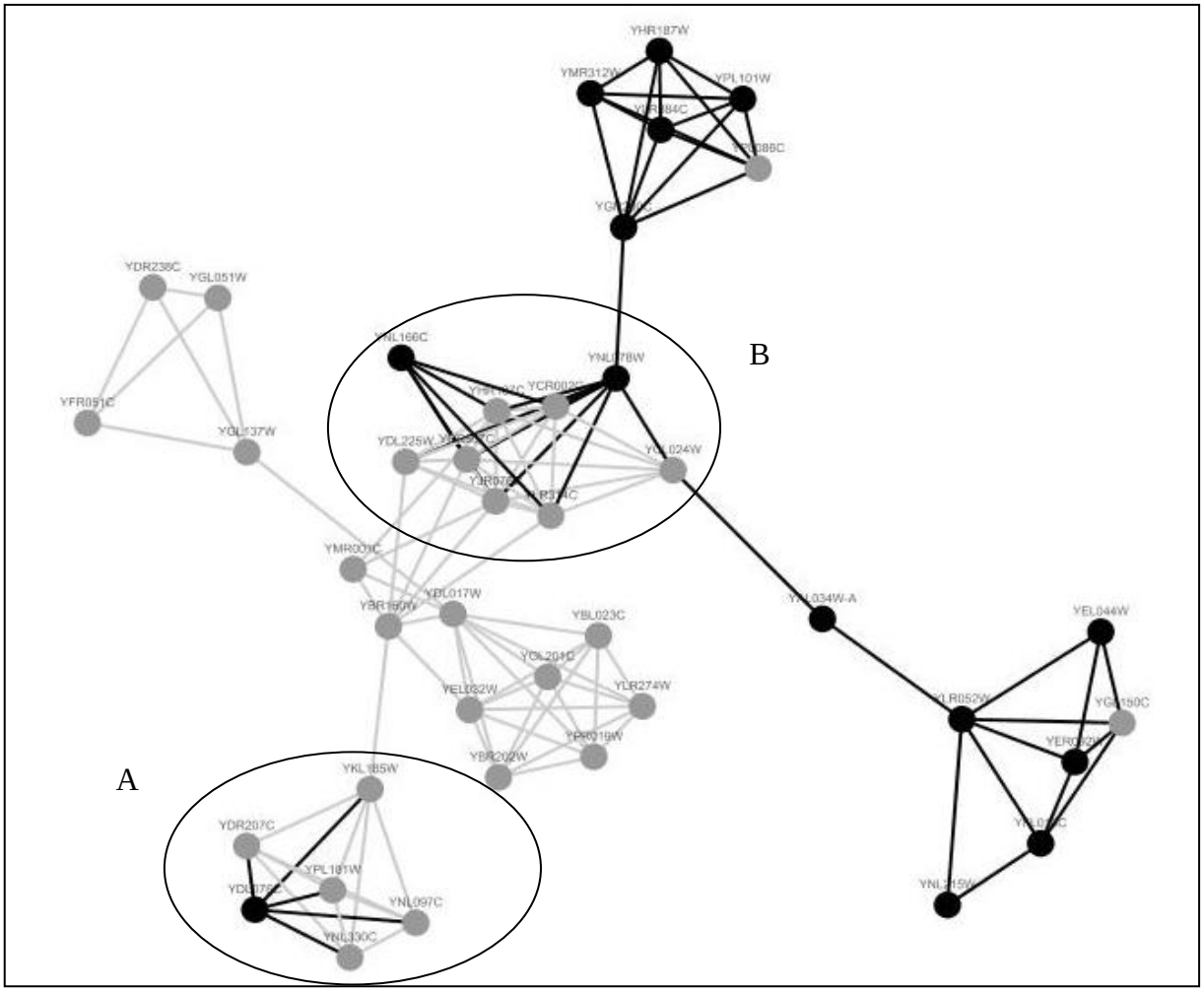


Figure 4.26. Module 5 used in function assignments to YDL076C (A), YNL166C (B) and YNL078W (C)

Out of six sub-groups in Module 5, the two (A and B) contain dominantly functionally known proteins with small number of functionally unknown proteins which makes them suitable for function assignment. For A, histone deacetylase activity (83.3 % - all proteins except the unknown), zinc ion binding (66.6 %), metal ion binding (66.6 %) and sequence specific DNA binding (50 %) are the most frequent function terms being possible candidates for the function of YDL076C. For B, on the other hand, nucleotide

binding (80 %), GTP binding, phosphatidylinositol binding and structural constituent of cytoskeleton (50 %) are in dominant. YNL166C and YNL078W are probably responsible for these functions in the cell.

Table 4.15. Functional Enrichment in Module 5

No	Go Id	Go Name	Freq in Module	Freq in Genome	Nodes
5A	4407	histone deacetylase activity	5/6 83.3	20/6310 0.3	YPL181W YDR207C YKL185W YNL330C YNL097C
	8270	zinc ion binding	4/6 66.6	323/6310 5.1	YPL181W YDR207C YKL185W YNL097C
	46872	metal ion binding	4/6 66.6	679/6310 10.7	YPL181W YDR207C YKL185W YNL097C
	43565	sequence-specific DNA binding	3/6 50.0	191/6310 3.0	YDR207C YKL185W YNL330C
5B	166	nucleotide binding	8/10 80.0	802/6310 12.7	YBR160W YCR002C YLR314C YJR076C YDR507C YHR107C YCL024W YDL225W
	5525	GTP binding	5/10 50.0	102/6310 1.6	YCR002C YLR314C YJR076C YHR107C YDL225W
	5545	phosphatidylinositol binding	5/10 50.0	12/6310 0.1	YCR002C YLR314C YJR076C YHR107C YDL225W
	5200	structural constituent of cytoskeleton	5/10 50.0	51/6310 0.8	YCR002C YLR314C YJR076C YHR107C YDL225W

Function assignment to YDL076C (RXT3) from Module 5: RXT3 is defined both experimentally and by electronic annotation in GO database, to involve in “regulation of transcription” through “chromatin modification” via “histone deacetylation” achieved by the histone deacetylases found in “Rpd3L complex” (Ruiz-Roig *et al.*, 2010; Zhang, 2003).

Histone tails are normally positively charged due to amine groups present on their lysine and arginine amino acids. These positive charges help the histone tails to interact with and bind to the negatively charged phosphate groups on the DNA backbone. Acetylation, which occurs normally in a cell, neutralizes the positive charges on the histone by changing amines into amides and decreases the ability of the histones to bind to DNA. This decreased binding allows chromatin expansion, permitting genetic transcription

to take place. Histone deacetylases remove those acetyl groups, increasing the positive charge of histone tails and encouraging high-affinity binding between the histones and DNA backbone. The increased DNA binding condenses DNA structure, preventing transcription.

Also, this protein has Rxt3 (PF08642) domain. Rxt3 has been shown in yeast to be required for histone deacetylation. Therefore, molecular functions “histone deacetylase activity” and “sequence specific DNA binding” assigned to the protein make satisfactory sense.

However, the relevance of the other two functions, namely “metal ion binding” and the sibling term “zinc ion binding”, to the above given biological processes, cannot be explicitly shown. Still it is experimentally proven that, a type of histone deacetylase has a zinc finger part in C terminus and it binds to ubiquitin through that region. In addition, it is well-documented that ubiquitinated histone H1, H2A, and H2B correlate with transcriptionally active chromatin regions, and that acetylated histones H3 and H4 correlate with increased transcriptional activation. The opposite case may also be valid; i.e. transcriptional repression may be achieved through both deacetylation and deubiquitination (Hook *et al.*, 2002).

RXT3 is listed as an interactor for 52 proteins, 14 of them sharing the function term “DNA binding” (26.9 %). Although the percentage is relatively small, histone deacetylase activity (19.2 %), metal ion binding (17.3 %), sequence specific DNA binding (15.3 %) and zinc ion binding (11.5 %) are the next commonly shared functions in these neighbors (Table D.3.2).

Function assignment to YNL166C (BNI5) from Module 5: The experimentally curated biological process terms given in the GO database focus on “cytokinesis” and “septin ring assembly” occurring in “cellular bud neck” and especially in “cellular bud neck septin ring”.

Septins are a highly conserved family of GTP-binding proteins with essential functions in cytokinesis. In yeast, septins maintain the compartmentalization of the yeast

plasma membrane during cell division by forming rings at the cortex of the bud neck. This event is achieved by the aggregation, arrangement and bonding together of septins and associated proteins to form an organized structure resembling a ring at the cell cortex (septin ring assembly).

Members of the septin family share common structural domains: a GTPase domain in the central region of the protein, a stretch of basic residues at the amino terminus, and a predicted coiled-coil domain at the carboxy terminus. The GTPase domain is important for self-association, binding other proteins, and targeting to the mother bud neck. The coiled-coil domain is not important for self-association but is important for association with other proteins (septin ring assembly) (Casamayor *et al*., 2003).

Also, yeast septins preferentially bind to phosphatidylinositol 4-phosphate PI(4)P. PI(4)P is likely to be present at the plasma membrane where it contributes to the septin organization, localization, or stability. The PI 4-kinases Stt4, which localizes to the plasma membrane, and required for the maintenance of vacuole morphology, cell wall integrity, and actin cytoskeleton organization synthesizes PI(4)P pool at the plasma membrane (Casamayor *et al*.., 2003).

Since septins have a GTPase domain, they bind guanine nucleotides and that nucleotide binding is important in septin-septin interactions and septin complex formation (Nagaraj *et al.*, 2008).

BNI5 has only 15 interactions listed in and almost 50 % of the interacting proteins are responsible for “nucleotide binding”. The other most frequent terms are “GTP binding”, “phosphatidylinositol binding” and “structural constituent of cytoskeleton” (33.3 % each) in accordance with the modular functional assignment results (Table D.3.2).

In conclusion, the functional terms assigned to BNI5, namely “nucleotide binding”, “GTP binding”, “phosphatidylinositol binding” and “structural constituent of cytoskeleton” are all coherently verified.

Function assignment to YNL078W (NIS1) from Module 5: NIS1 is proven to have a role in “regulation of mitosis” occurred in “cellular bud neck” controlled via “nucleus”.

Being in the same module with BNI5, NIS1 shares the common molecular function terms with BNI5. Those annotations seem logical since the biological term implies budding of yeast by forming a cellular bud neck.

In addition, out of 31 interactors 13 of them (41.9 %) has functional term “nucleotide binding” and 8 of them (25.8 %) share the term “structural constituent of cytoskeleton” (Table D.3.2.).

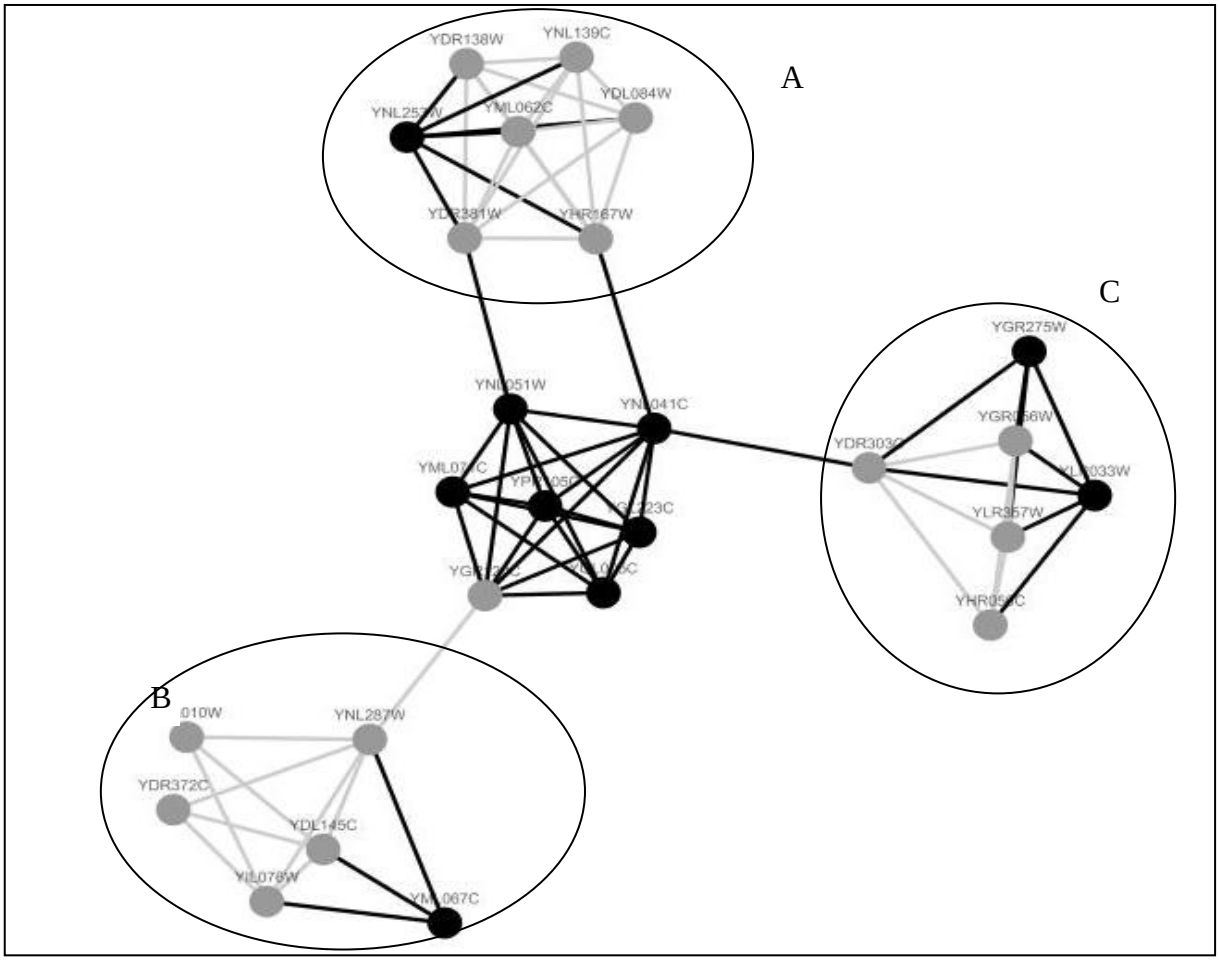


Figure 4.27. Module 6 used in function assignments to YNL253W (A), YML067C (B), YLR033W and YGR275W (C)

Table 4.16. Functional Enrichment in Module 6

No	Go Id	Go Name	Freq in Module	Freq in Genome	Nodes
6A	3676	nucleic acid binding	6/7 85.7	309/6310 4.8	YML062C YDR381W YDR138W YNL139C YHR167W YDL084W
6B	3674	molecular_function	4/6 66.6	1994/6310 31.6	YIL076W YDL145C YML067C YNL287W
	5198	structural molecule activity	3/6 50.0	64/6310 1.0	YIL076W YDL145C YNL287W
6C	3677	DNA binding	4/6 66.6	487/6310 7.7	YHR056C YGR056W YLR357W YDR303C
	8094	DNA-dependent ATPase activity	4/6 66.6	37/6310 0.5	YHR056C YGR056W YLR357W YDR303C

Module 6 has 3 suitable subgroups (A, B and C) for functional assignment; except the one in the middle with only one known protein. In A, the dominant function term; i.e. the assigned function for YNL253W is nucleic acid binding (85.7 % - all proteins except the unknown). In C, DNA binding and DNA-dependent ATPase activity (66.6 % - all proteins except the two unknowns) are the assigned functions for YGR275W and YLR033W.

In B, the most frequent function term is “molecular_function” implying that the proteins in the module have unknown function or their functions are determined by IEA evidence score only. “Structural molecule activity” term is shared by the group of proteins that are also listed in “molecular_function”. So the function terms of proteins in sub-group B are not curated and the functional assignment will not be reliable for YML067C.

Function assignment to YNL253W (TEX1) from Module 6: Eukaryotic mRNAs are generated and processed through a number of connected metabolic steps that include transcription, capping, splicing, poly(A)+ addition and RNA export.

Jimeno and Colleagues (2002) run various experiments, providing evidence for a functional link between transcription and mRNA metabolism, mediated by the THO complex and RNA export factors. In GO database, consistent with the literature findings,

biological process term “mRNA export from nucleus” and cellular component terms “transcription export complex”, “THO complex” are associated with TEX1.

However, the only molecular function term that can be assigned from the modularity search is “nucleic acid binding”; which in fact is valid since transcription implies nucleic acid binding.

TEX1’s physical interactors share “nucleic acid binding”, “nucleotide binding” and “RNA binding” as the common functional terms (19.1 % each) supporting the modularity search result and also adding a new coherent function term, “RNA binding”.

Function assignment to YGR275W (RTT102) from Module 6: “Nucleosome disassembly”, the breakdown of the beadlike structural unit of eukaryotic chromatin composed of histones and DNA, is the first biological process term annotated for RTT102; which is the outcome of “chromatin remodeling” and “chromosome segregation” processes. This chromatin alteration is carried out in the control of “SWI/SNF” (SWItch/Sucrose NonFermentable) and “Rsc” complexes.

An important element of eukaryotic gene regulation is the dynamic nature of chromatin because of the ability of various protein complexes to alter the structure of chromatin. Chromatin is altered by chemical modification of the core histone proteins through acetylation, phosphorylation, methylation, and ubiquitination (Sengupta *et al.*, 2001).

The first of the ATP-dependent chromatin-remodeling complexes is the SWI/SNF complex. The yeast SWI/SNF complex has 11 different subunits. The largest subunit of this complex, Swi2/Snf2, is a 194-kDa protein containing seven highly conserved subdomains corresponding to the DNA-dependent ATPase domain. Single amino acid changes in the DNA-dependent ATPase domain eliminate both the ATPase activity of Swi2/Snf2 and the chromatin-remodeling activity of SWI/SNF. The next largest subunit of SWI/SNF, Swi1 (148 kDa), had been shown to interact with DNA and contains an AT-rich interactive domain DNA-binding domain that may be involved in SWI/SNF binding of DNA (Sengupta *et al*., 2001).

The RSC complex is another ATP-dependent chromatin-remodeling complex found in yeast that is closely related to SWI/SNF. Rsc4 subunit has multiple transcriptional roles, including regulating interactions with RNA polymerases and histone H3, as well as regulating chromatin remodeling and gene selectivity. One of these roles can be the impact of RSC on RNA elongation from RNA polymerase II promoter; which is the extension of an RNA molecule after transcription initiation at an RNA polymerase II-specific promoter by the addition of ribonucleotides catalyzed by RNA polymerase II (Soutourina *et al.*., 2006).

This protein has Rtt102p-like transcription regulator protein (PF09510) domain. Rtt102p (Regulator of Ty1 Transposition Protein 102) is a transcription regulator protein found in fungi that appears to be integrally associated with both the Swi-Snf and the RSC chromatin remodelling complexes

RSC is involved in transcription regulation and nucleosome positioning, and is responsible for the transfer of a histone octamer from a nucleosome core particle to naked DNA. The reaction requires ATP and involves an activated RSC-nucleosome intermediate.

RTT102 has 51 interactions and 14 of these interacting partners (27.4 %) have a role in “DNA binding” and “DNA-dependent ATPase activity” (Table D.3.2).

In the light of this information, the assigned functions “DNA binding” and “DNA-dependent ATPase activity” seem valid but not very specific.

Function assignment to YLR033W (RSC58) from Module 6: The RSC58 protein is in the same modular group with RTT02; having the same biological process and cellular component terms. As expected, the assigned functional annotations are also common.

However, these proteins don’t have the same domain. RSC58 contains Bromodomain (PF00439) domain. The precise function of the domain is unclear, but it may be involved in protein-protein interactions and may play a role in assembly or activity of multi-component complexes involved in transcriptional activation.

Being involved in transcriptional activations may support the DNA related function assignments. Besides, like RTT102, RSC58's interacting partners share the terms "DNA binding" (39.6 %) and "DNA- dependent ATPase activity" (27.5 %) supporting the assigned functional terms for this protein.

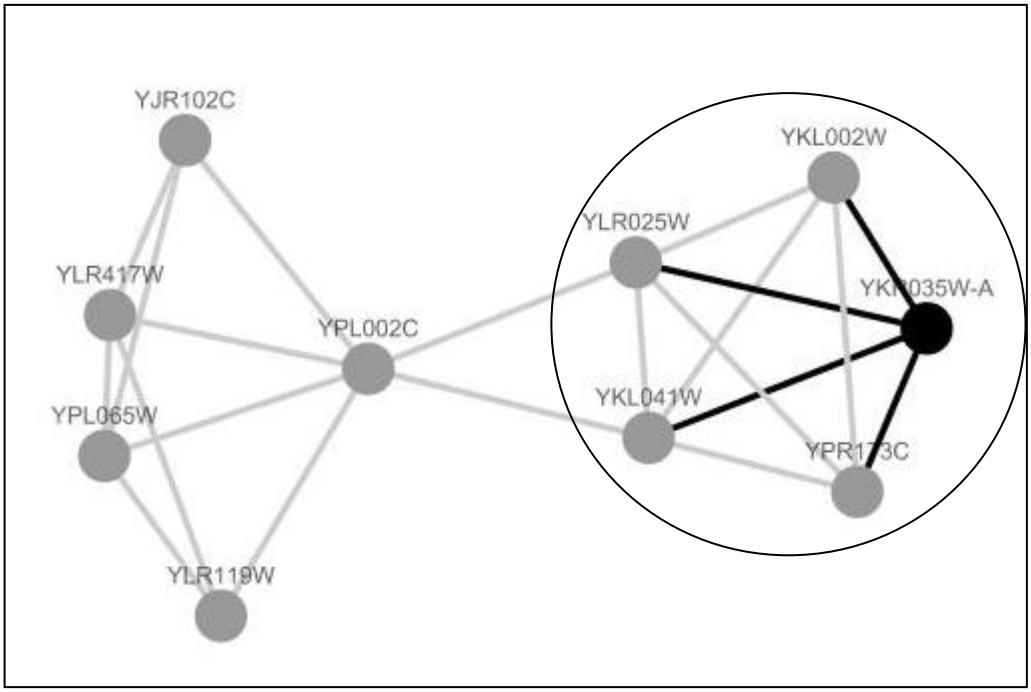


Figure 4.28. Module 8 used in function assignment to YKR035W-A

Table 4.17. Functional Enrichment in Module 8

Go Id	Go Name	Freq in Module	Freq in Genome	Nodes
5515	protein binding	3/5 60.0	242/6310 3.8	YLR025W YKL002W YKL041W

In Module 8, the only unknown protein is YKR035W-A. The dominant function term in the module is protein binding (60.0 %), making it the only choice for functional assignment.

Function assignment to YKR035W-A (DID2) from Module 8: Subcellular compartmentalization is an important feature of eukaryotic cells, preventing inappropriate meetings of certain intracellular components as well as facilitating efficient ordered

reactions. To maintain this compartmentalization, cells have evolved mechanisms to ensure that specific proteins are delivered to specific organelles (Bryant and Stevens, 1998).

The vacuole of *Saccharomyces cerevisiae* is central for much of the physiology of this organism.

A well-studied transport pathway via "endosome" in the yeast *Saccharomyces cerevisiae* is responsible for "delivery of proteins to the vacuole", the functional equivalent of the lysosome in mammalian cells (Babst *et al.*, 1997).

Multiple transport pathways depend on the endosomal membrane system. Endocytosed proteins are first delivered to an early endosome, where they are either recycled back to the plasma membrane or delivered to the vacuole via a late endosome.

There are two types of protein targeting via late endosome to vacuole:

"Late endosome to vacuole transport": The directed movement of proteins from late endosomes to the vacuole. In yeast, after transport to the prevacuolar compartment, endocytic content is delivered to the late endosome and on to the vacuole. This pathway is analogous to endosome to lysosome transport.

"Late endosome to vacuole transport via multivesicular body sorting pathway": The directed movement of proteins from endosomes to vacuoles by a pathway in which molecules are sorted into multivesicular bodies, which then fuse with the vacuole.

Those biological processes are the processes annotated in GO database, to DID2 protein. This protein involves especially in the late endosome to vacuole transport of proteins. Also DID2 contains Snf7 domain which are involved in protein sorting and transport from the endosome to the vacuole/lysosome in eukaryotic cells. Thus, the function "protein binding" is most likely to be valid for DID2.

Furthermore, 37.5 % of the interactors have the functional term "protein binding" supporting the assigned term and also 25 % has a role in "ATP binding" as well as

“nucleotide binding” which may be a more specific role for the protein in transport in endosomal membrane system.

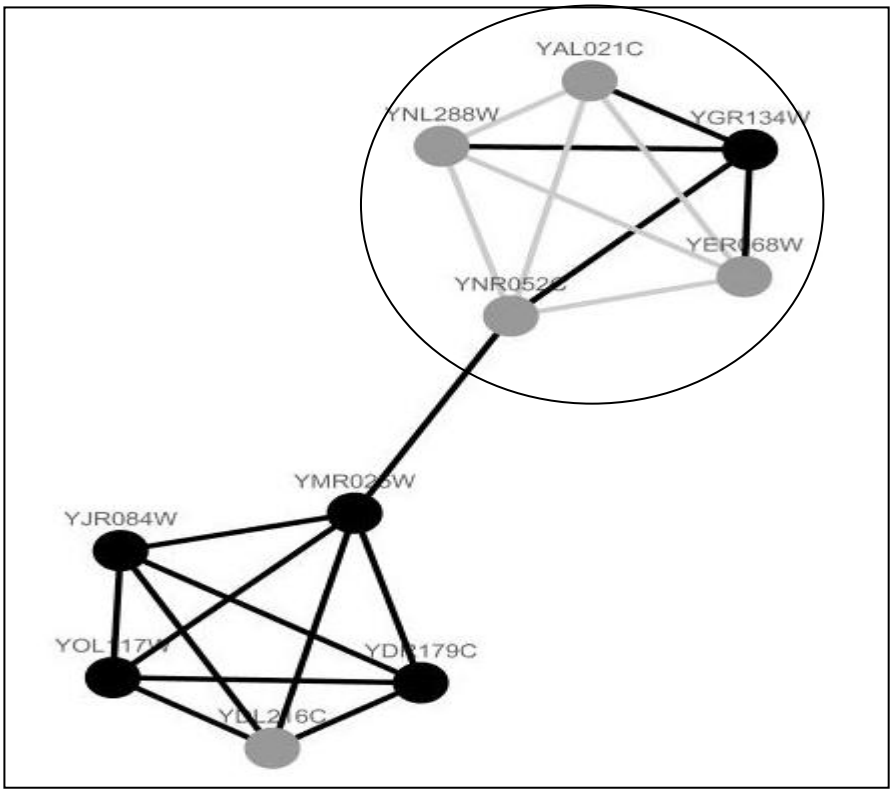


Figure 4.29. Module 9 used in function assignment to YGR134-W

Table 4.18. Functional Enrichment in Module 9

Go Id	Go Name	Freq in Module	Freq in Genome	Nodes
46872	metal ion binding	3/5 60.0	679/6310 10.7	YAL021C YNR052C YER068W
3723	RNA binding	3/5 60.0	446/6310 7.0	YAL021C YNR052C YER068W

Module 9 contains two sub-groups; lower one having only one known protein and upper one suitable for function assignment with one unknown protein, YGR134W. This protein may be responsible for metal ion binding (60 %) and RNA binding (60 %) since these are the most frequent terms in the sub-group.

Function assignment to YGR134W (CAF130) from Module 9: The “Ccr4-Not complex” is a global regulator of gene expression that is conserved from yeast to human. It affects genes positively and negatively and is thought to “regulate transcription” factor IID function. The Ccr4-Not complex regulates many different cellular functions, including RNA degradation and “transcription” initiation. It may be a regulatory platform that senses nutrient levels and stress (Collart, 2003).

Ccr4-Not has been connected functionally in the “nucleus” to “transcription by RNA polymerase II” and in the “cytoplasm” to mRNA degradation (Azzouz *et al.*, 2009).

The RING finger protein CNOT4 is a component of the CCR4–NOT complex whose N-terminus harbors a conserved domain with eight cysteine residues. This domain adopts a RING finger fold. RING fingers are a specialized type of zinc fingers, which bind two atoms of zinc required for proper folding of RING domains and for their subsequent biological activities (Albert *et al.*, 2002).

“Regulation of transcription from RNA polymerase II promoter” via “CCR4-NOT complex” is manually curated and listed in GO database. “RNA binding” function is reasonable when taking the process and component terms into consideration. Also “metal ion binding” term can be verified by the fact that CCR4-NOT complex has a zinc finger domain.

The neighborhood count method sustains the “RNA binding” term since 20 % of the neighbors share this common function term.

Module 10 seems to be suitable for functional assignment at first glance, since out of five proteins, only one protein is functionally unknown. It can be stated that, YBL097W may have a role in ATP or DNA binding however the predominance of molecular function terms is not very clear (each has 40 % frequency). So this module is eliminated and evaluated as unsuitable for functional assignment.

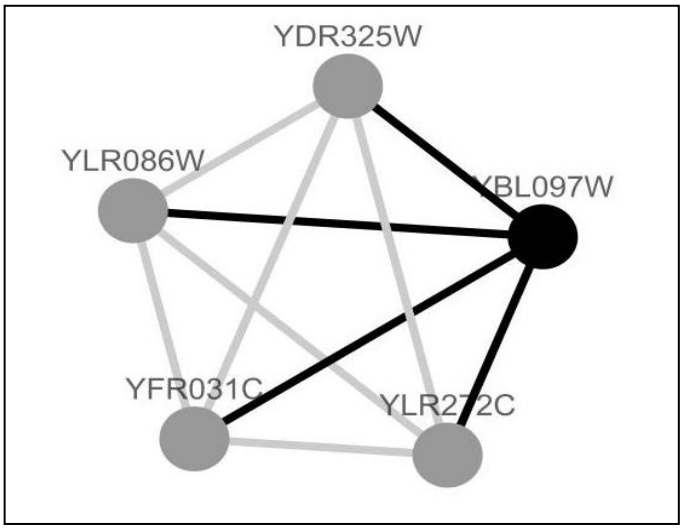


Figure 4.30. Module 10 used in function assignment to YBL097W

Table 4.19. Functional Enrichment in Module 10

Go Id	Go Name	Freq in Module	Freq in Genome	Nodes
3674	molecular_function	3/5 60.0	1994/6310 31.6	YBL097W YDR325W YLR272C
217	DNA secondary structure binding	2/5 40.0	6/6310 0.0	YLR086W YFR031C
5524	ATP binding	2/5 40.0	657/6310 10.4	YLR086W YFR031C
3680	AT DNA binding	2/5 40.0	5/6310 0.0	YLR086W YFR031C
5488	binding	2/5 40.0	274/6310 4.3	YDR325W YLR272C
166	nucleotide binding	2/5 40.0	802/6310 12.7	YLR086W YFR031C
16887	ATPase activity	2/5 40.0	139/6310 2.2	YLR086W YFR031C
3690	double-stranded DNA binding	1/5 20.0	28/6310 0.4	YFR031C

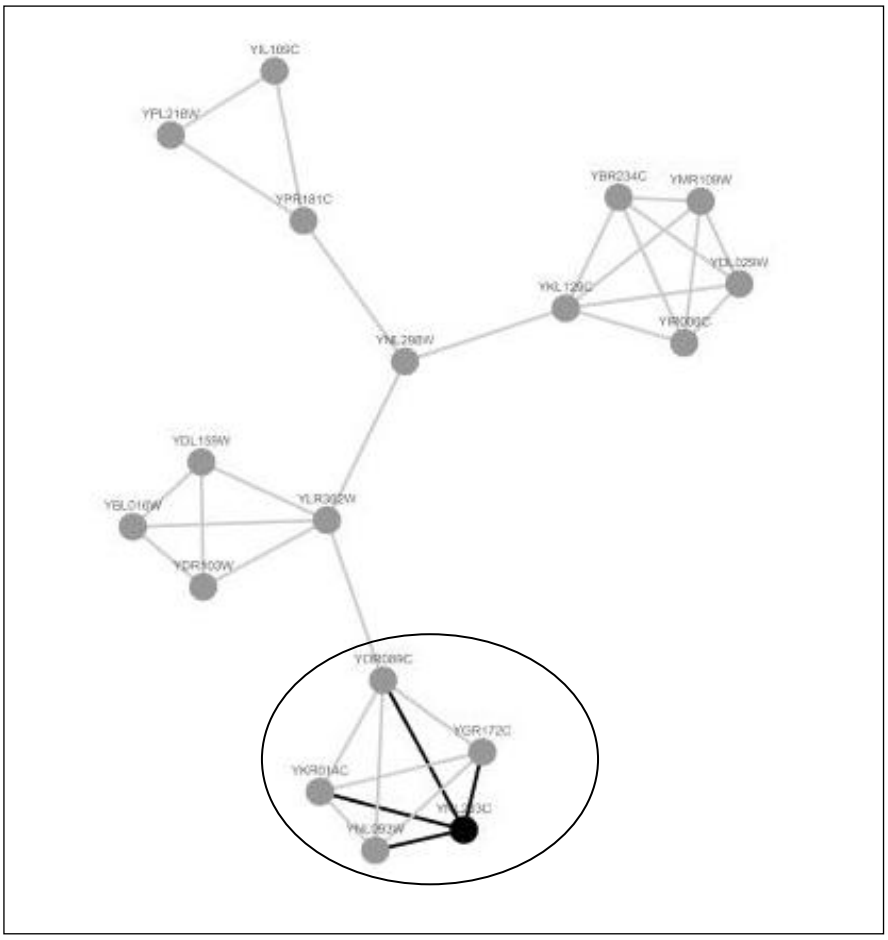


Figure 4.31. Module 12 used in function assignment to YNL263C

Table 4.20. Functional Enrichment in Module 12

Go Id	Go Name	Freq in Module	Freq in Genome	Nodes
5525	GTP binding	3/5 60.0	102/6310 1.6	YOR089C YKR014C YNL093W
166	nucleotide binding	3/5 60.0	802/6310 12.7	YOR089C YKR014C YNL093W
3924	GTPase activity	3/5 60.0	72/6310 1.1	YOR089C YKR014C YNL093W

Module 12 has a sub-group in circle with one unknown protein, YNL263C. GTP binding, nucleotide binding and GTPase binding are the most probable function terms for this protein, each having a 60 % frequency percentage.

Function assignment to YNL263C (YIF1) from Module 12: The directed movement of substances from the “endoplasmic reticulum (ER) to the Golgi” is mediated by “COP II vesicles”.

Proteins synthesized in the ER that are destined to travel via the exocytic pathway to the Golgi apparatus, also termed cargo, are sorted from ER-resident proteins and packaged into COPII-coated vesicles.

The formation of the COPII coat on ER membranes drives vesicle budding. Budding from the ER involves “activation of the Sar1p GTPase” by the ER-resident protein Sec12p, a Sar1p-specific guaninenucleotide exchange factor (GEF). The molecular events that trigger the activation of Sar1p to the “GTP-bound” form to initiate budding are presently unknown. Soon after formation, COPII vesicles shed their coat, and then fuse with the cis-Golgi compartment. Conversion to SAR1-GDP triggers coat release and recycles COPII subunits (Sato and Nakano, 2001).

YIF1 protein is experimentally proven to take place in the “ER to Golgi vesicle mediated transport” via “ER to Golgi transport vesicle”. This protein may aid in “GTPase activity” and “GTP binding” functions as a guanine nucleotide exchange factor like Sec12.

In addition, it contains YIF1domain (PF03878) required for membrane fusion of ER derived vesicles. It also plays a role in the biogenesis of ER derived COPII transport vesicles which are strictly related to the assigned functions.

The neighborhood count method agrees with the results obtained from modularity method since 11 of the 33 interacting partners have roles in “GTPase activity”, “GTP binding” and 14 of them having “nucleotide binding” function (Table D.3.2).

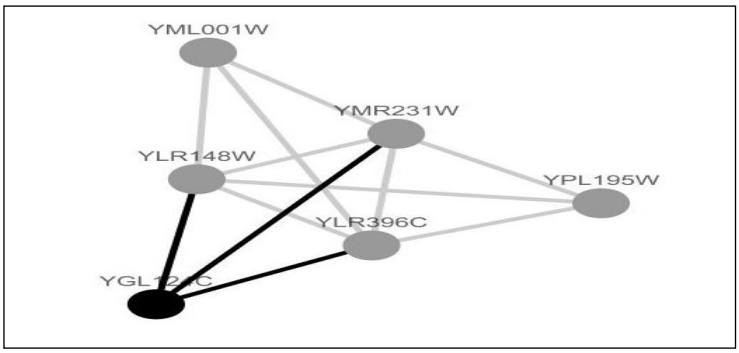


Figure 4.32. Module 11 used in function assignment to YGL124C

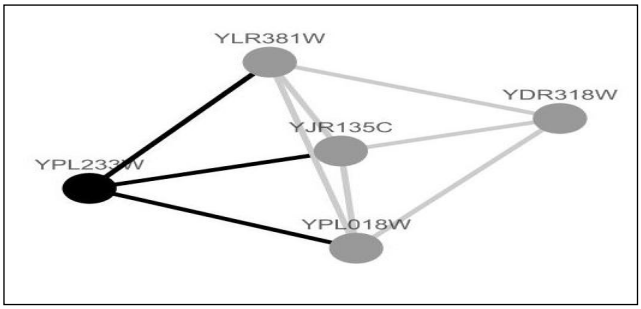


Figure 4.33. Module 13 used in function assignment to YPL233W

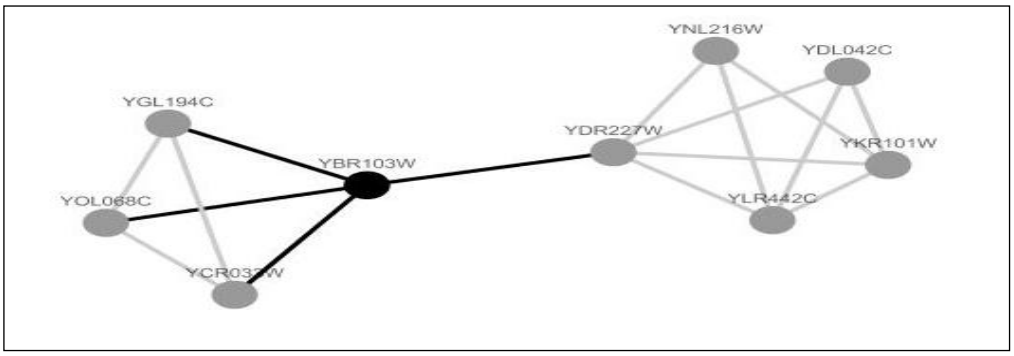


Figure 4.34. Module 14 used in function assignment to YBR103W

Table 4.21. Functional Enrichment in Modules 11, 13 and 14

No	Go Id	Go Name	Freq in Module	Freq in Genome	Nodes
11	35091	phosphoinositide binding	3/6 50.0	32/6310 0.5	YLR396C YMR231W YLR148W
13	5515	protein binding	4/5 80.0	242/6310 3.8	YLR381W YPL018W YDR318W YJR135C
14	3677	DNA binding	4/9 44.4	487/6310 7.7	YCR033W YDR227W YLR442C YNL216W

The three modules, Module 11, 13 and 14, do not have sub-groups. Each has one functionally unknown protein. For Module 11, phosphoinositide binding (50 %) is the frequent term and assigned function for YGL124C. Protein binding (80 % - all proteins except the unknown) is the highly possible functional role for YPL233W in Module 13. Finally, YBR103W may function in DNA binding (44.4 %).

Function assignment to YGL124C (MON1) from Module 11: Any model of vesicular traffic must include mechanisms to guarantee temporal and spatial specificity, because without such regulators, vesicular traffic would result in the rapid homogenization of all cellular compartments. Both proteins and lipids participate in this regulation. Among proteins, a major role is played by guanosine triphosphatases (GTPases). Among lipids, growing evidence suggests a key function for phosphatidylinositol (Ptdlns) and its phosphorylated derivatives, collectively referred to as phosphoinositides (PIs).

Genetic studies in yeast provided initial support for a role of PIs not only in Ca^{2+}-regulated exocytosis, but more generally in fundamental aspects of vesicular transport. Direct evidence for an essential role of PIs in vesicle-mediated transport was provided when the yeast VPS34 gene was found to encode a Ptdlns-3 kinase. Yeast vps mutants define a set of nearly 50 genes whose functions are required for the sorting and delivery of soluble vacuolar hydrolases from the late Golgi to the vacuole (Camille *et al.*, 1996).

It contains Mon1 (PF03164) domain. In Saccharomyces cerevisiae a protein complex of Mon1 and Ccz1 functions with the small GTPase Ypt7 to mediate vesicle trafficking to the vacuole. The Mon1/Ccz1 complex is conserved in eukaryotic evolution and members of this family (previously known as DUF254) are distant homologues to domains of known structure that assemble into cargo vesicle adapter (AP) complexes.

MON1 protein is experimentally proven to take place in the “protein targeting to vacuole”, “autophagy”, “vesicle docking” in “fungal type vacuolar membrane” and “cytosol”. This protein may be responsible for “phosphoinositide binding” and aiding vesicular traffic.

Table 4.22. Functions Assigned to the Proteins with Unknown Function

GO id	Assigned Molecular Function Term	Name	ORF & UNIPROT ID	DOMAIN
3677	DNA binding	**IWR1**	**YDL115C**	**DUF1762 (PF08574)**
46872	metal ion binding		**Q07532**	
3899	DNA-directed RNA polymerase activity			
5515	protein binding	**NKP1**	**YDR383C**	
4842	ubiquitin-protein ligase activity		**Q12493**	
5515	protein binding	**MND2**	**YIR025W**	**Apc15p (PF05841)**
4842	ubiquitin-protein ligase activity		**P40577**	
46961	proton-transporting ATPase activity, rotational mechanism	**RAV2**	**YDR202C**	**Rogdi_Iz (PF10259)**
16820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances		**Q03956**	
16787	hydrolase activity			
4407	histone deacetylase activity	**RXT3**	**YDL076C**	**Rxt3 (PF08642)**
8270	zinc ion binding		**Q07458**	
46872	metal ion binding			
43565	sequence-specific DNA binding			
166	nucleotide binding	**BNI5**	**YNL166C**	
5525	GTP binding		**P53890**	
5545	phosphatidylinositol binding			
5200	structural constituent of cytoskeleton			
166	nucleotide binding	**NIS1**	**YNL078W**	
5525	GTP binding		**P53939**	
5545	phosphatidylinositol binding			
5200	structural constituent of cytoskeleton			
3676	nucleic acid binding	**TEX1**	**YNL253W**	**WD40 (PF00400)**
			P53851	

Table 4.23. Functions Assigned to the Proteins with Unknown Function - cont.

3677 8094	DNA binding DNA-dependent ATPase activity	**RTT102**	**YGR275W** **P53330**	**Rtt102p (PF09510)**
3677 8094	DNA binding DNA-dependent ATPase activity	**RSC58**	**YLR033W** **Q07979**	**Bromodomain (PF00439)**
5515	protein binding	**DID2**	**YKR035W-A** **P69771**	**Snf7 (PF03357)**
46872 3723	metal ion binding RNA binding	**CAF130**	**YGR134W** **P53280**	
35091	phosphoinositide binding	**MON1**	**YGL124C** **P53129**	**Mon1 (PF03164)**
5525 166 3924	GTP binding nucleotide binding GTPase activity	**YIF1**	**YNL263C** **P53845**	**Yif1 (PF03878)**
3677	DNA binding	**BRN1**	**YBL097W** **P38170**	**Barren (PF05786)**
5515	protein binding	**NSL1**	**YPL233W** **Q12143**	**Mis14 (PF08641)**
3677	DNA binding	**SIF2**	**YBR103W** **P38262**	**LisH (PF08513)** **WD40 (PF00400)**

4.3.4. Calcium Related Pathways in Yeast Under Stress

Under stress conditions like hypo/hypertonic shock and mating, levels of calcium change drastically followed by rapid sequestration of calcium and maintaining homeostatic conditions again.

After the yeast PIN is reconstructed, the proteins involved in these three pathways, namely HOG (High Osmolarity Glycerol), mating pheromone and PKC pathways, are investigated with respect to their presence in the PIN.

Applying Netsearch for the reconstructed yeast PIN, the alternative paths, starting from the membrane protein and ending with transcription factors, are investigated. The maximum pathway length for the Netsearch algorithm is chosen to be 5 (Table 4.20) since higher lengths resulted in many numbers of linear paths and made it hard to visualize or study. The results including all 5 step-paths are shown in App. In this section only the shortest paths (3 and 4 step paths) are shown and discussed.

In the shortest path graphs, the color blue represents the proteins present in the pathway constructed according to the literature findings. The red ones are the hub proteins of yeast PIN. The colorless proteins are the ones found in the yeast PIN, but not in the pathway constructed according to the literature findings.

HOG pathway: The HOG pathway appears to be only activated by a specific stress signal: the increase of osmolarity through addition of salt or glucose to the medium (Saito and Tatebayashi, 2004). The pathway consists of two input branches, the Sln1 branch and the Sho1 branch (Figure 4.36).

Sln1 is a membrane protein that senses the osmotic changes. It has two transmembrane regions and an intracellular histidine kinase domain that signals to two other proteins, Ypd1 and Ssk1. Under normal conditions, it is continuously phosphorylated and transmits its phosphate group to Ypd1, which in turn passes it on to Ssk1. In this way, Ssk1 is kept phosphorylated and inactive. This module consisting of Sln1, Ypd1, and Ssk1 is a so-called phosphorelay system (Klipp and Liebermeister, 2006). Such phosphorylation

of Ssk1 is thought to prevent interaction of Ssk1 with the MAPKKKs Ssk2 and Ssk22 and, consequently, the downstream components of the pathway remain inactive.

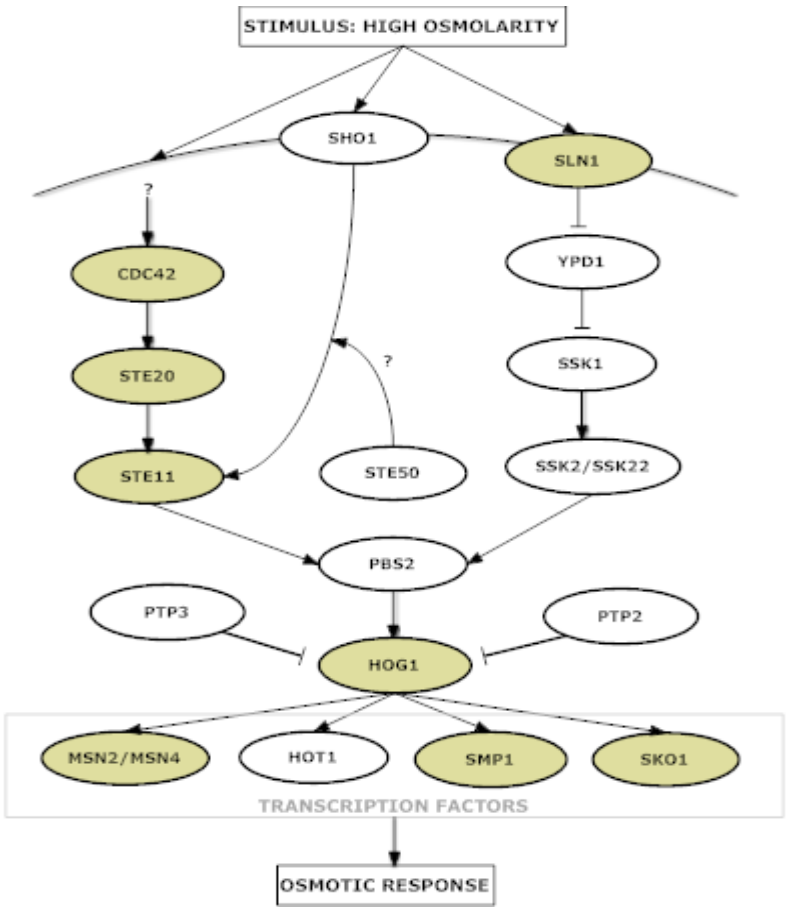


Figure 4.35. HOG (High Osmolarity Glycerol) Pathway of Yeast (Igual and Estruch (2000), Qiang Zhao PHD Thesis (2001), Nadal *et al.* (2002), Saito and Tatebayashi (2004), Klipp and Liebermeister (2006)). Shaded ones are present in the PIN constructed by SPA.

Upon osmotic stress, phosphorylation of Sln1 is interrupted by the inhibition of histidine kinase activity of the protein and Ssk1 switches to a nonphosphorylated, active state. In this form, it triggers the HOG MAP kinase cascade, which involves the redundant proteins Ssk2 and Ssk22 as well as Pbs2 and Hog1. Phosphorylated Hog1 can enter the nucleus and regulate the transcription of a series of genes that are involved in glycerol production (Nadal *et al.*, 2002).

The Sho1 osmosensor contains four transmembrane segments and a cytoplasmic SH3 domain that binds to the N-terminal region of Pbs2. Sho1 is predominantly localized to the

cytoplasmic membrane at regions of polarized growth, such as the emerging bud and the bud neck (Raitt *et al.*, 2000; Reiser *et al.*, 2000, 2003). Sho1 is absolutely required for activation of Hog1 via the Sho1 branch. Plasma membrane localization of the SH3 domain is required, and Sho1 might not sense osmolarity directly but might instead provide a docking site (via its SH3 domain) for downstream proteins.

In addition, Cdc42, Ste20, Ste50, and Ste11 have also been implicated in the Sho1 pathway (Tatebayashi *et al.*, 2006). Cdc42 is a Rho-type small G-protein, and binds and activates Ste20, a prototype of the PAK family protein kinases. Activated Ste20 phosphorylates the MAPKK kinase (MAPKKK) Ste11. Ste50, an SAM (Sterile Alpha Motif) domain-containing protein, binds Ste11, itself another SAM domain-containing protein, via a heterotypic SAM–SAM interaction (Grimshaw *et al.*, 2004).

The reconstructed PIN contains the key elements of HOG pathway; the membrane protein Sln1, the transcription factors Msn2/Msn4, Smp1 and Sko1 as well as the mid-proteins Cdc42, Ste20 and Ste11. The absence of Sho1, phosphorelay elements (Ypd1 and Ssk1) and MAPKKKs Ssk2/Ssk22 as well as Pbs2 may imply that there are some alternative pathways for this signaling cascade.

Starting with Sln1 as the membrane protein and Cdc42, the Netsearch algorithm resulted in three 4-step path for Sln1 and 93 5-step path for Cdc42 (Table 4.20 and Figure 4.37).

These paths are expected to contain Hog1 protein since it is said to be the key protein for activation of transcription factors. However, for Sln1, the shortest paths do not pass through Hog1; instead Sln1 is connected to Msn2/Msn4 by Skn7 (nuclear response regulator and transcription factor required for optimal induction of heat-shock genes in response to oxidative stress; involved in osmoregulation), Hsf1 (heat shock transcription factor, activates multiple genes in response to stresses) and Ssn3 (protein kinase involved in phosphorylation of the RNA polymerase II C-terminal domain; involved in glucose repression). This result may be explained by the findings of Igual and Estruch. In their work, they stated that although osmotic stress induces nuclear accumulation of both Hog1 and Msn2/4, the increase in nuclear concentration of these transcription factors can also be

observed in a *hog1* mutant and vice versa (Igual and Estruch, 2000), implying that there should be an alternate to this path. Also SKN7 has been proven to act as a phosphorelay system member that turns on the HOG pathway until yeast suffers from cell shrinking (Wei and Chen, 2010).

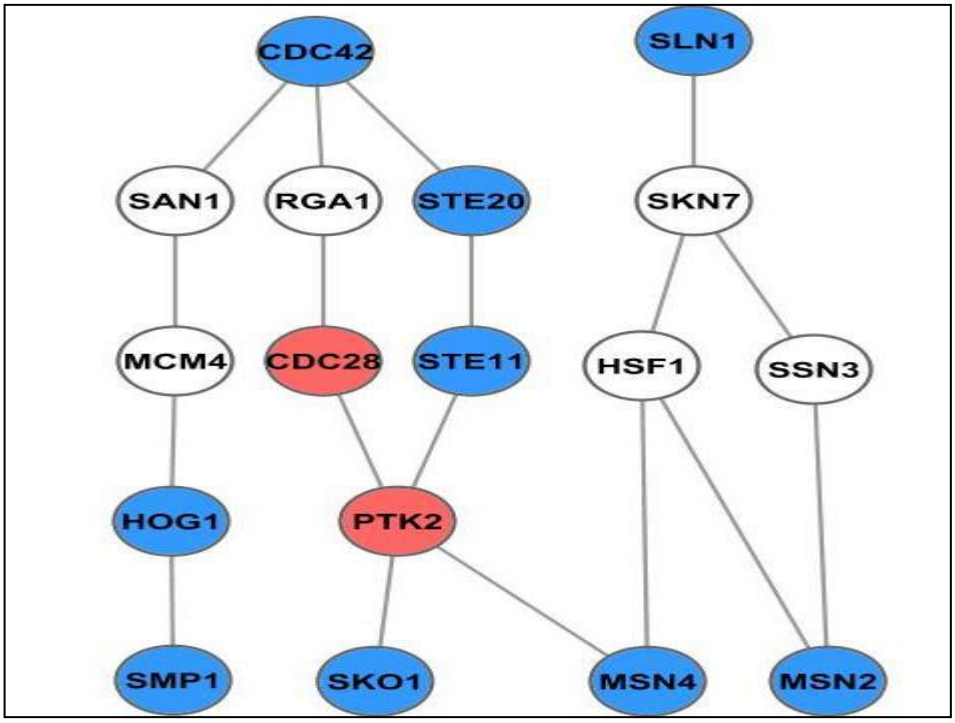


Figure 4.36. The shortest paths of Sln1 and Cdc42 found by Netsearch.

The linear paths starting with Cdc42 pass through Hog1 as well as the yeast PIN hub proteins Cdc28 and Ptk2. Ptk2 seems to function for both Pbs2 and Hog1, since it connects Ste11 with Sko1 and Msn4.

Mating pheromone pathway: The signal transduction pathway that senses the presence of extracellular pheromone and orchestrates the cellular responses to it is known as the yeast mating pheromone response pathway. Mating is initiated by the binding of the mating pheromone to a seven-transmembrane, G-protein-coupled receptor, Ste2/Ste3, on the cell-surface. Receptor occupancy stimulates the Gα (Gpa1) subunit of the G protein to exchange GDP for GTP; GTP-bound Gα then releases the Gγ (Ste18) and Gβ (Ste4) heterodimer (Figure 4.38).

Following release from Gα, the membrane-bound Gβ and Gγ complex transmits the signal by binding to three different effectors: (1) a Ste5/Ste11 complex; (2) the Ste20 protein kinase, and; (3) a Far1/Cdc24 complex. Activated Ste11 phosphorylates target residues in the activation loop of Ste7. As a result, Ste7 is activated. Activated Ste7 then phosphorylates, and thereby activates, its targets, the MAPKs Kss1 and Fus3 on a threonine and a tyrosine residue in their activation loop (Barwell, 2004). Key substrates of Fus3 and Kss1 are the Ste12/Dig1/Dig2 transcription factor complex and the Far1 protein (Figure 4.38).

The shortest paths found by Netsearch (Figure 4.39) are mostly comprised of the pathway proteins with the additional proteins Gpa2 (nucleotide binding alpha subunit of the heterotrimeric G protein), Tec1 (member of Ste12/Dig1/Tec1 complex) (Chou *et al.*,2006) and hub protein of yeast PIN, Ssb1.

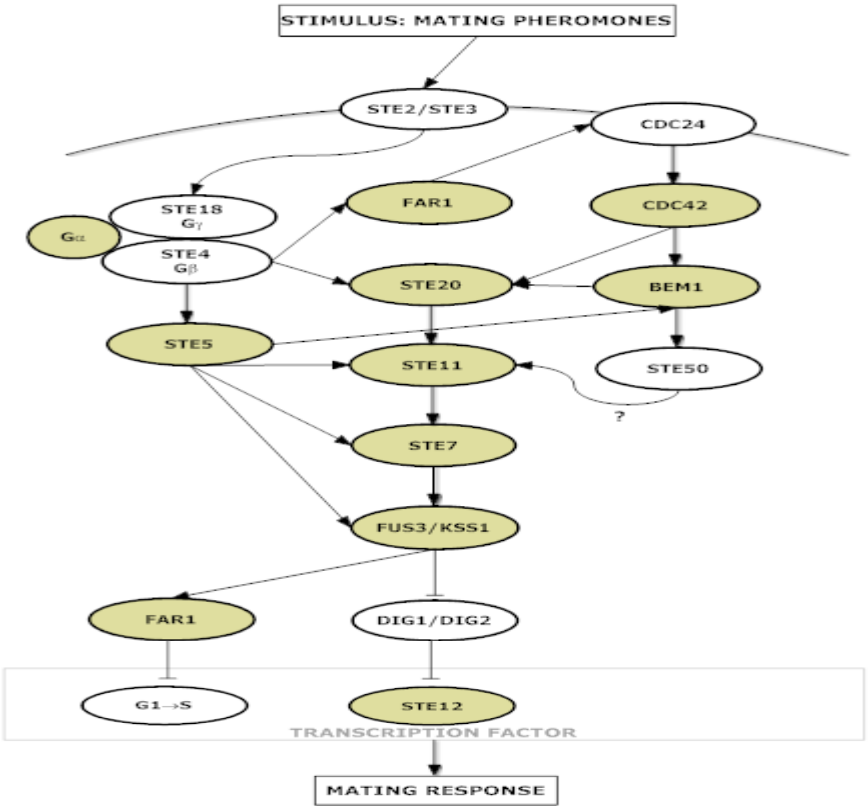


Figure 4.37. Mating Pheromone Pathway of Yeast (Igual and Estruch (2000), Saito and Tatebayashi (2004), Klipp and Liebermeister (2006)). Shaded ones are present in the PIN constructed by SPA.

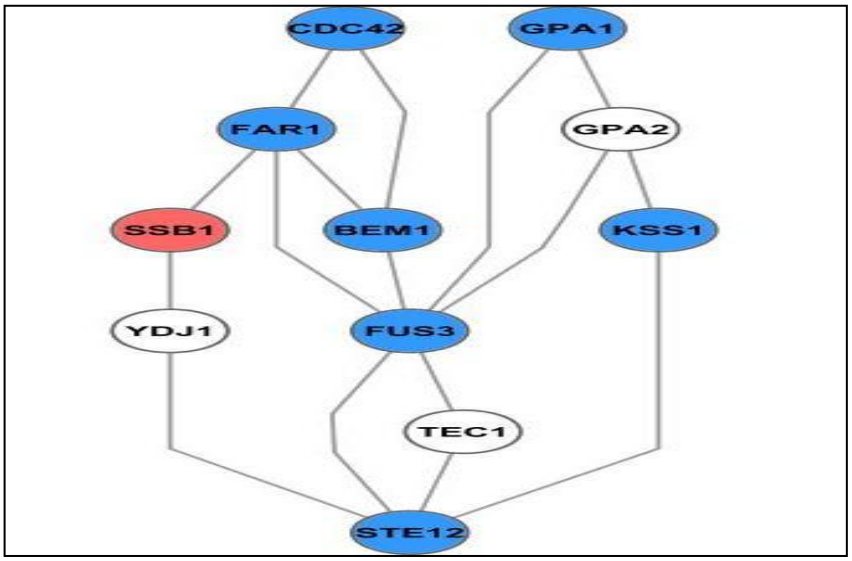


Figure 4.38. The shortest paths of Gpa1, Far1 and Cdc42 found by Netsearch.

PKC pathway: In its natural environment, yeast cells are exposed to not only increases but also decreases in osmolarity. The intracellular protein kinase C (PKC) signal transduction pathway is essential for sensing cell integrity under a variety of environmental conditions or morphogenetic events. The PKC response regulates cell wall and actin cytoskeleton dynamics, and it is activated during polarized growth such as budding and mating. In addition, PKC signaling is activated by environmental conditions that jeopardize cell wall stability, including high temperature, hypotonic shock or impaired cell wall synthesis (Reinoso Martin *et al.*, 2003).

Accordingly, the absence of PKC signaling causes cell lysis when yeast is exposed to any of these inducing conditions. Osmotic stabilization can prevent cell lysis, which also indicates defective maintenance of a functional cell wall.

Sensing of cell wall perturbations requires dedicated surface sensors as Mid2 and Hsc77, a member of Wsc family. The activation of the PKC pathway proceeds through the small G protein Rho1, via Pkc1, and a downstream MAP kinase cascade.

Although the molecular mechanisms by which sensors transmit the signal to downstream effectors remain ill defined, the Rho1-GDP/GTP exchange factor Rom2 may mediate Rho1 activation. Rho1 is a small GTPase upregulated by the GDP/GTP exchange factors Rom1 and Rom2 and downregulated by the GTPase-activating proteins Sac7 and

Bem2. Among other functions, Rho1 binds and activates Pkc1, which in turn activates the MAP kinase kinase kinase Bck1/Slk1, the functionally redundant MAPKK kinases Mkk1 and Mkk2, and the MAP kinase Slt2/Mpk1. PKC signaling is constantly guarding cell integrity, and the expression of many cell wall biosynthesis genes requires PKC.

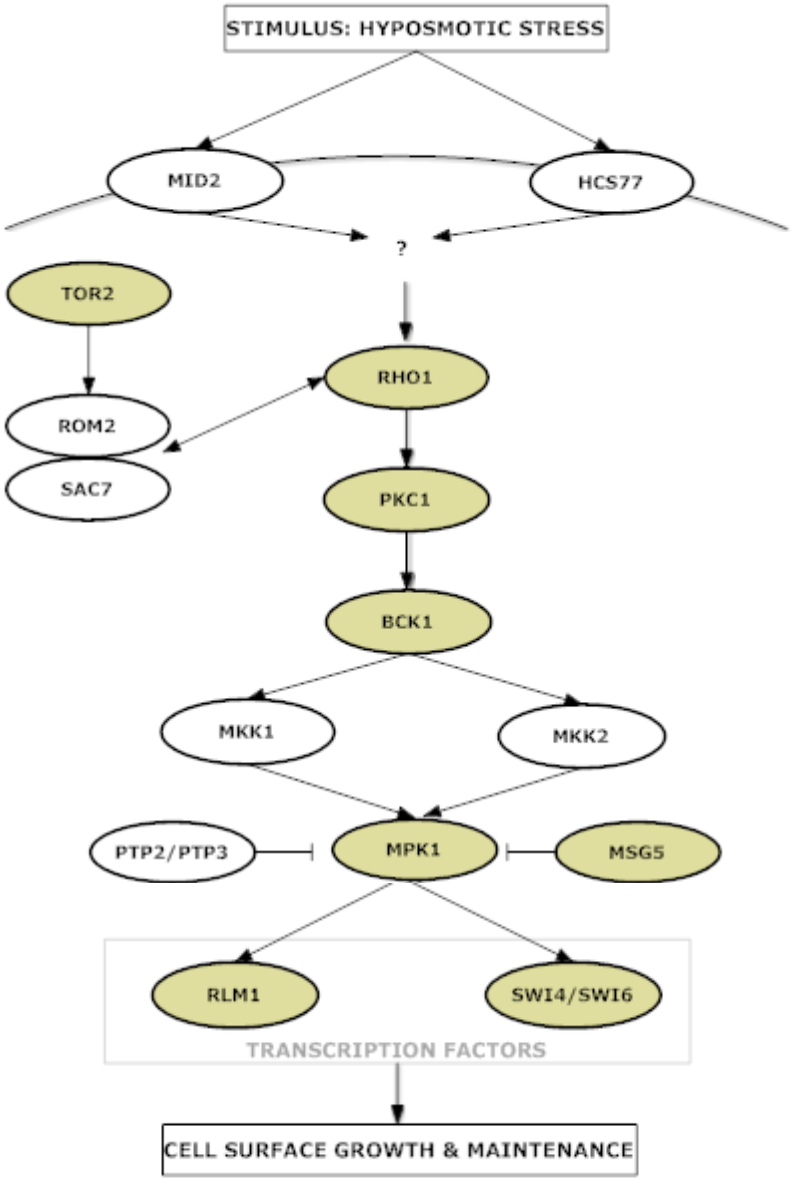


Figure 4.39. PKC Pathway of Yeast (Igual and Estruch (2000), Saito and Tatebayashi (2004), Klipp and Liebermeister (2006)). Shaded ones are present in the PIN constructed by SPA

The shortest paths found for PKC signaling are completely alternative to the present PKC pathway (Figure 4.40 and 4.41). Only the starting proteins and the transcription factors as well as Pkc1 are common with the PKC pathway implying this important mechanism has many alternative pathways for cell robustness and fast response. Slt2

(Reinoso-Martin *et al.*, 2003), Smi1 (Markovich *et al.*, 2004), Skn7, Cdc28 and Mbp1 (Bouquin *et al.*, 1999) are some of the shortest path proteins that are presented to have a role in PKC pathways.

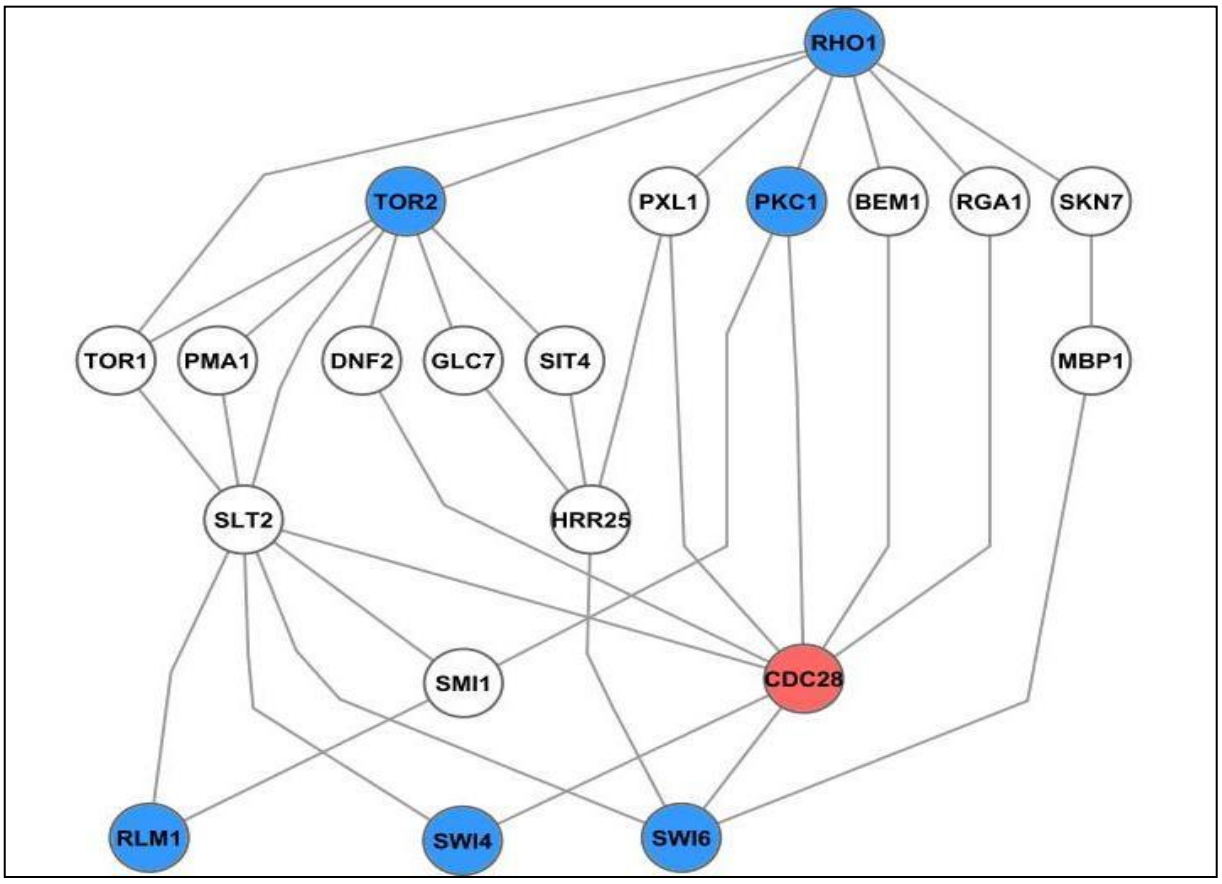


Figure 4.40. The shortest paths of Tor2 and Rho1 found by Netsearch.

Table 4.24. The Number of Shortest Paths for HOG, Mating Pheromone and PKC Pathways

	HOG PATHWAY		MATING PHEROMONE PATHWAY			PKC PATHWAY	
Shortest path length	SLN1	CDC42	CDC42	FAR1	Gα	RHO1	TOR2
3	-	-	-	1	1	-	3
4	3	-	2	3	3	17	13
5	13	93	44	78	16	125	177
6	237	1600	644	1453	186	1393	1991
SUM	253	1693	690	1535	206	1535	2184

5. RESULTS & DISCUSSION FOR CALCIUM SIGNALING IN WORM

5.1. Core Proteins & Resultants Protein Interaction Network

5.1.1. Determination of Core Proteins in SPA

The first step is to find the proteins including the term “calcium” in Gene Ontology in terms of component, function and process with experimental evidence scores (GO Database downloaded on 16.05.2010). In GO Database, for worm, *21* proteins having experimental evidence scores, paired off within the group of calcium-related annotations.

The second approach is to search the Wormbase, which is a website for nematodes. In Wormbase, *94* proteins are determined, 13 being the same with GO findings.

Also the literature is searched for two purposes; to find experimental evidence for the proteins determined so far by database searches and to count in the ca-related proteins unfound by database searches. *10* novel proteins are revealed as calcium-related.

In conclusion, a set of 112 unique probable calcium-related proteins is generated using 3 different methods. (Table A.2)

The basic principle is to include the proteins that are *proven to be directly related* to calcium by *experimental* techniques. The core set of 112 proteins is examined thoroughly and the indirectly related proteins are eliminated from the core list. After those eliminations, the set of core proteins is down to *21* proteins. According to the evolutionary history of core proteins *4* of them are also eliminated since their homologs are not related to calcium. The resultant set of core proteins comprises of *17 proteins* which are experimentally proven to be directly assigned to the calcium signaling in the cell (details in Appendix F.1).

5.1.2. Resultant Protein Interaction Network

SPA with 17 core proteins resulted in a PIN consisting of *1714* proteins and *3174* undirected physical interactions after 8 runs. 11 core proteins have no interactions in the interactome data so they are not included in the resultant PIN.

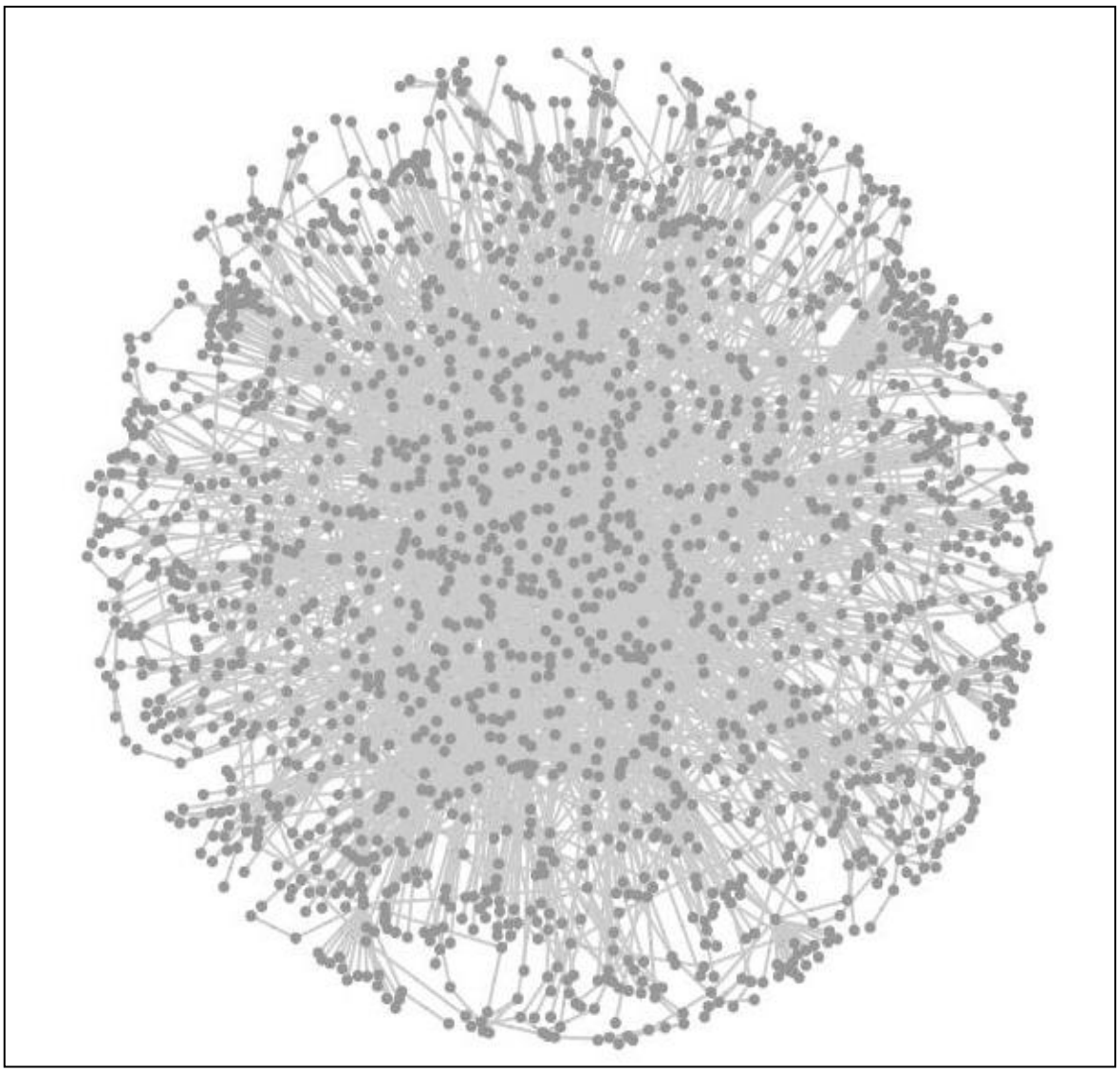


Figure 5.1. SPA resultant Worm PIN obtained from 17 core proteins with 1714 nodes and 3174edges. Produced by Cytoscape 2.6.

In order to connect the lonely core proteins to the PIN, String server is used. STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four

sources: genomic context, high throughput experiments, conserved coexpression and previous knowledge. STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable with a confidence score. The database currently covers 2,590,259 proteins from 630 organisms.

However, the lonely proteins neither have relevant interactions to the PIN nor the interactions yield high confidence scores. So the lonely proteins stay unconnected to the PIN as before.

A challenging task is to gather the data obtained from different databases (BIOGRID, Intact, Wormbase and GO) under the same language. All these databases define worm genes/proteins with different IDs. Also the proteins have synonyms making the gathering process more difficult and time consuming. To be able to use the data in MATLAB, all data is converted to the annotation "WBGenexxxxxx" format.

Also, there is an unexpected situation with worm gene ontology data. Many proteins/genes have no cellular component or molecular function term at all. In fact, for those proteins, it is expected to be assigned with the terms "cellular component unknown" or "molecular function unknown" in this case. So when running MATLAB, the "unknown" terms are assigned to the proteins with no annotation at all.

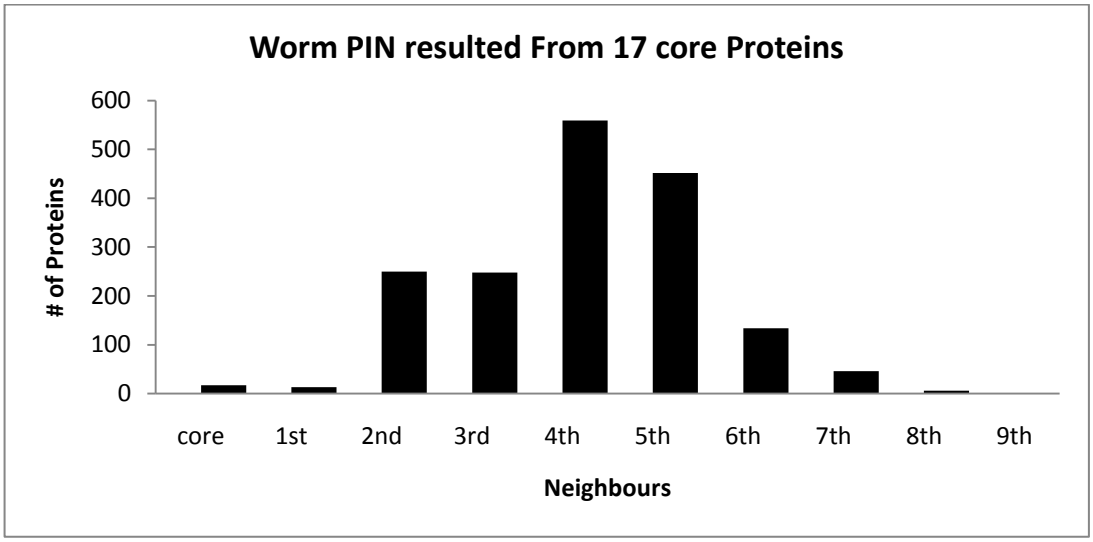


Figure 5.2. Details of network reconstruction steps in SPA for 17 core Proteins for Worm

5.2. Analysis of Reconstructed Network

5.2.1. Evolutionary History of Core Proteins

Sequence homology search is performed using “blastp” (protein-protein BLAST), for both identifying a query amino acid sequence and for finding similar sequences in protein databases.

In Table 5.1, the beginning (worm) and the corresponding homologues of the last organism (human) is given only. Having a high identity score in the last organism implies that the protein also has high identity scores in-between organisms. The whole evolutionary homology search results can be found in App F.2.

Table 5.1. Homologues of worm core proteins in human

C.ELEGANS	**H.SAPIENS (Human)**			
Worm	% identity	e-val	prot name	ID
ckk-1	48,15	1,76E-105	calcium/calmodulin-dependent protein kinase kinase 2 isoform 2	NP_705719
cmk-1	61,49	5,50E-123	calcium/calmodulin-dependent protein kinase type 1	NP_003647
cnb-1	78,36	1,06E-74	calcineurin subunit B type 1	NP_000936
crt-1	57,72	2,78E-135	calreticulin precursor	NP_004334
csq-1	29,02	4,72E-44	calsequestrin-1	NP_001222
egl-19	57,49	0,00E+00	voltage-dependent L-type calcium channel subunit alpha-1D isoform c	NP_001122311
itr-1	37,76	0,00E+00	inositol 1,4,5-trisphosphate receptor type 3	NP_002215
mca-1	50,48	0,00E+00	plasma membrane calcium-transporting ATPase 3 isoform 3b	NP_001001344
mca-3	61,47	0,00E+00	plasma membrane calcium-transporting ATPase 3 isoform 3b	NP_001001344
pmr-1	58,27	0,00E+00	calcium-transporting ATPase type 2C member 1 isoform 1a	NP_055197
rcn-1	34,30	2,92E-32	calcipressin-3	NP_038469
sca-1	66,95	0,00E+00	sarcoplasmic/endoplasmic reticulum calcium ATPase 2 isoform 1	NP_733765
ser-1	19,62	2,07E-44	D(2) dopamine receptor isoform long	NP_000786
slo-1	56,14	0,00E+00	calcium-activated potassium channel subunit alpha-1 isoform c	NP_001154824
spe-41	17,80	9,08E-33	short transient receptor potential channel 3 isoform a	NP_001124170
tax-6	68,60	0,00E+00	serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform isoform 2	NP_001124163
unc-36	21,86	2,42E-57	voltage-dependent calcium channel subunit alpha-2/delta-1	NP_000713

Out of 17 core proteins, 11 of them, namely calmodulin dependent protein kinases (ckk-1, cmk-1), calcineurin related and calcium binding proteins (cnb-1, crt-1), calcium transporting ATPases (mca-1, mca-3, pmr-1, sca-1), calcium channels (slo-1, egl-19) and finally phosphatase (tax-6) have high identity scores with a high significance (low e-value).

The csq-1, itr-1, rcn-1, ser-1, spe-41 and finally unc-36 have low identity scores but having low identity score does not always imply irrelevance. The homologues found share the similar functions with the query protein. This result can be explained with the fact that, the functionally important region of these proteins that is conserved may be small when compared to the size of the a.a sequence.

5.2.2. Annotation Table Enrichment in Worm PIN

The annotation table, set as the main criteria for SPA method, is constructed from the GO terms of the 17 core proteins. With the intention of assigning new functions to the proteins not having any known function, the GO terms of “cellular component unknown” (GO id: 5575), “molecular function unknown” (GO id: 3674) and “biological process unknown” (GO id: 8150) are also included in the Annotation Table.

The Annotation Table of core proteins contains 25 cellular component terms, 31 molecular function terms and 81 biological process terms with a total number of 137 GO terms.

On the other hand, the proteins involved in the reconstructed Ca signaling PIN enriched the annotation table to a total number of 1115 GO terms; 152 being cellular component, 319 being molecular function and 644 being biological process.

The GO term enrichment in the annotation table resulted in 12 fold increase compared to the initial annotation table. The most enriched annotation is the cellular process with an increase of 16 fold (Figure 5.3).

The detailed list of annotation tables and the amount of enrichment in the GO terms are given in App F.3.

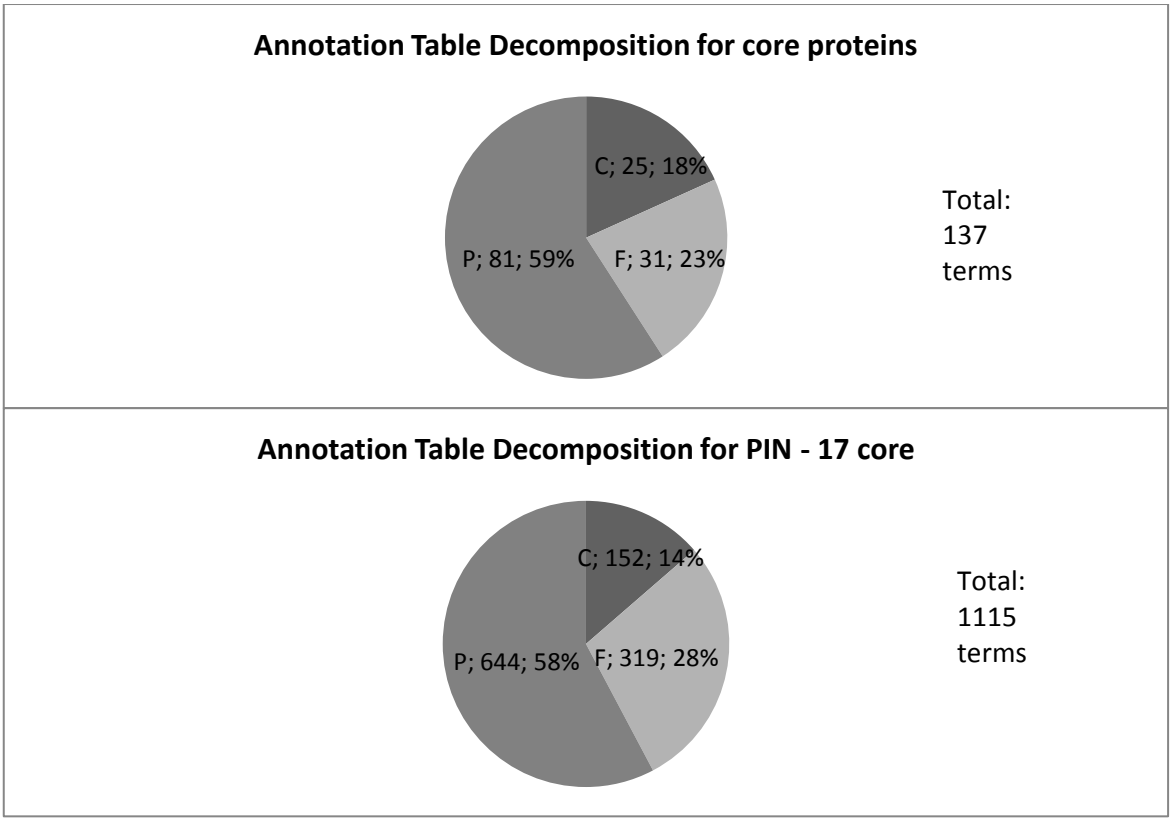


Figure 5.3. Annotation Table Comparison of core proteins and PIN for worm

The dominant biological process terms in the core protein set are *calcium mediated signaling*, *thermosensory behaviour* and *larval development* in general as expected; since the core proteins chosen are mainly dealing with calcium signaling and also since calcium has vital role in gonads of worm in cytoplasm or on membranes in order to sustain calcium ion homeostasis in the cell. Hence, *cytoplasm* and *integral to membrane* being the dominant cellular component terms is consistent with the biological process dominance. Also the most frequent molecular function terms are *protein binding* and *calcium ion binding*.

On the other hand; *embryonic development ending in birth or egg hatching, reproduction, nematode larval development, positive regulation of growth rate, growth* and *locomotion* are the abundant biological process terms in the reconstructed PIN (Figure 5.4).

Accordingly, the dominant molecular function terms are *protein binding and ATP binding* which take place in *nucleus*, and *membranes* (Figure 5.5 and 5.6).

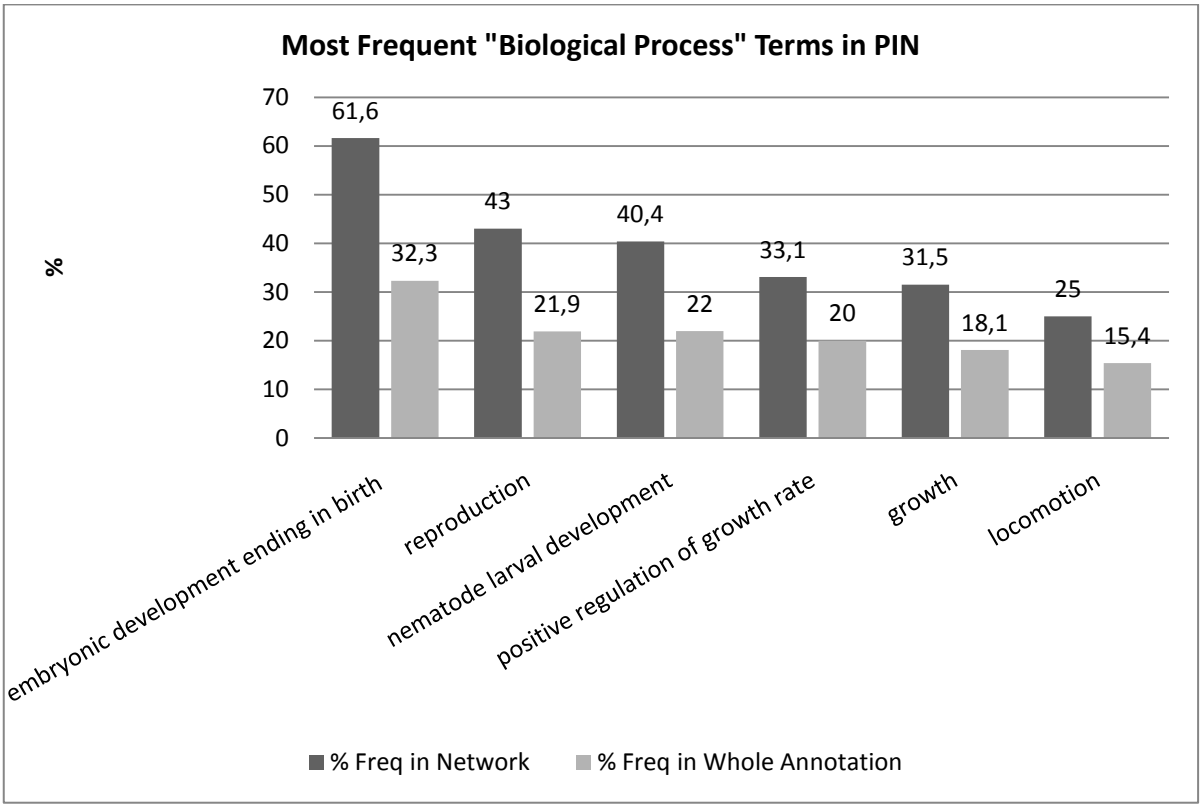


Figure 5.4. Predominant “Biological Process” terms in worm PIN

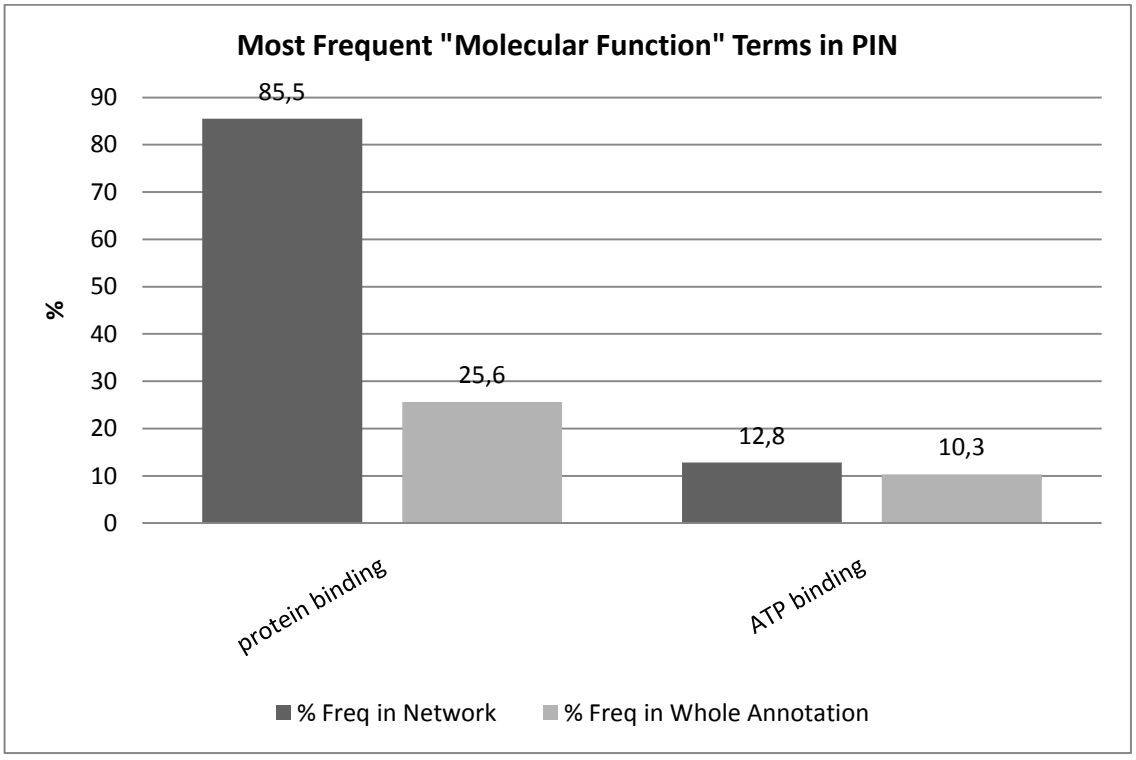


Figure 5.5. Predominant "Molecular Function" terms in worm PIN

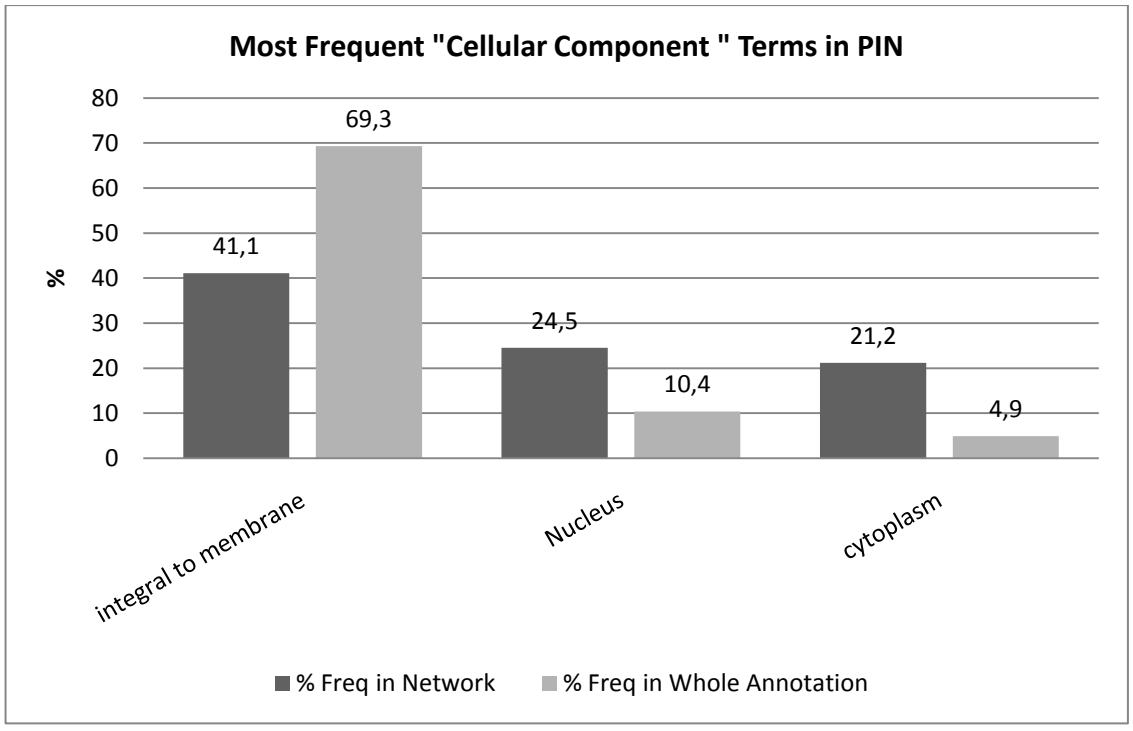


Figure 5.6. Predominant "Cellular Component" terms in worm PIN

5.3. Topology Based Analysis of Worm PIN

5.3.1. Topological Analysis of Worm PIN

The undirected protein interaction graph of the worm signaling network was analyzed using the Network Analysis algorithm, a Cytoscape plugin. The reconstructed signaling network includes 1714 proteins (represented as nodes) and 3174 interactions (represented as edges) among these proteins. (Table 5.2)

Table 5.2. Topological properties of the worm signaling networks and whole genome

WORM	17 core *PPI Network*	Whole Genome *PPI Network*
# of nodes	1714	3741
# of edges	3174	7118
Mean connectivity (degree)	3,70	3,81
Mean path length	4,78	4,79
Network diameter	12	13
Network density	0,002	0,001
Network heterogenity	1,80	2,11
Clustering coefficient	0,03	0,03
Network centralization	0,08	0,05
Shortest paths	2936082,00	13991340
Power Law Constant (γ)	1.71 (R^2:0.89)	1.73 (R^2:0.90)
# of isolated & lonely nodes	11	-
Total # of proteins	1725	3741

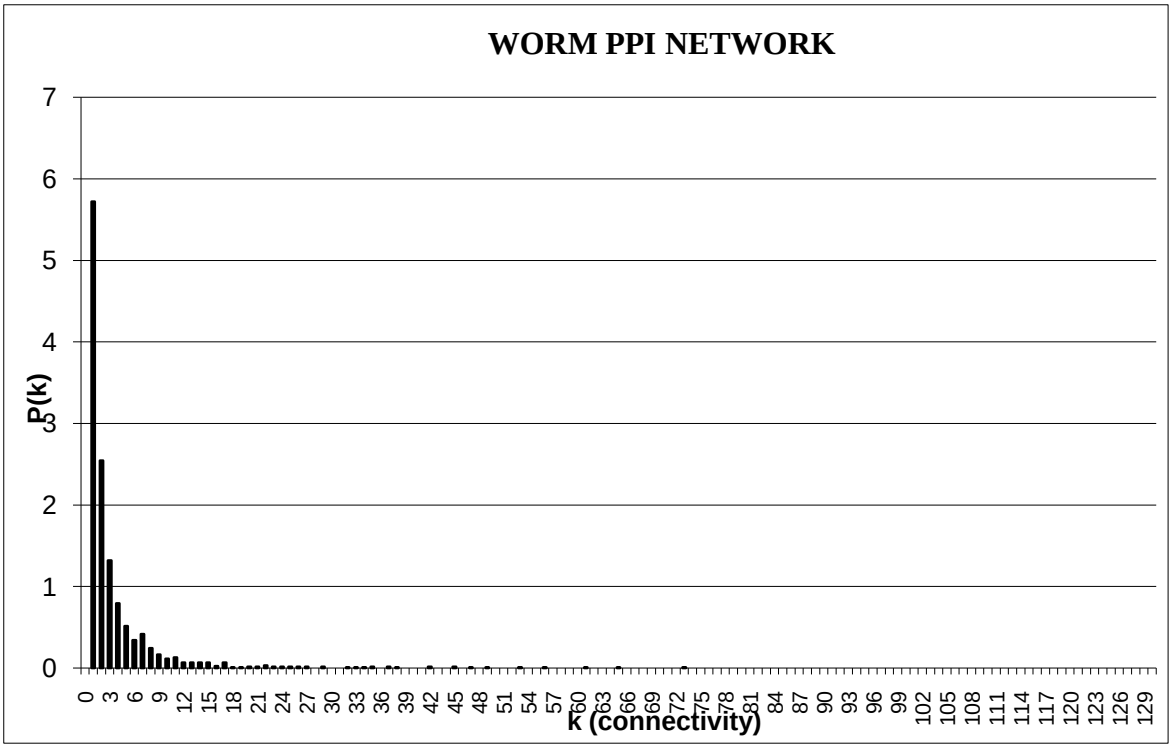


Figure 5.7. Degree distribution in the worm signaling network

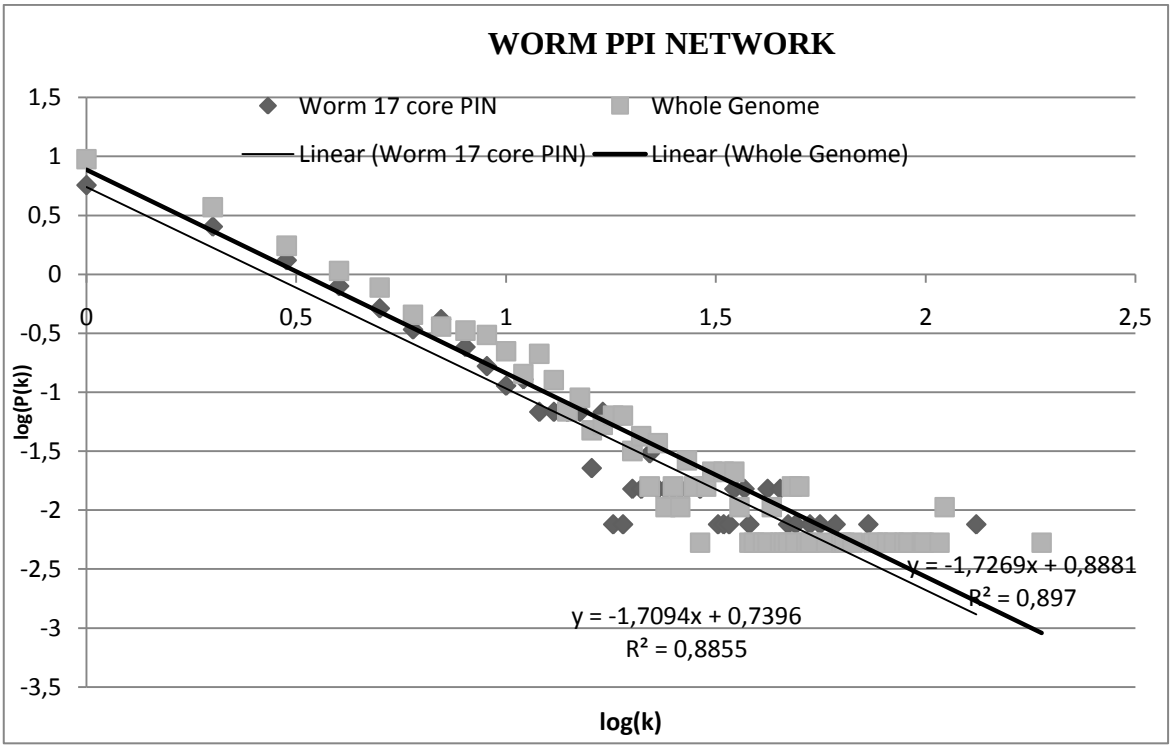


Figure 5.8. Log transform of degrees versus log transform of P(k) (probability of having connectivity, k) in worm signaling network and whole interactome

Scale-free property & power law: Both networks follow nearly a power law model ($P(k) \approx k^{-\gamma}$, γ=1.71 with R^2=0.89, γ=1.73 with R^2=0.90 for PINs resulted from 17 cores and whole genome respectively), implying that networks can be characterized by a scale-free degree distribution with a general emergence of hubs, as many other biological networks. The exponent γ is almost the same between two networks indicating that the role of hubs does not differ considerably.

The network density is the normalized version of the degree and varies between 0 - 1. It shows how densely the network is populated with edges. Since, for both networks, it is close to 0, one can conclude that not every node has many edges, but some hubs. The network heterogeneity reflects the tendency of a network to contain hub nodes. Network heterogeneity of 1.80 and 2.11 supports the scale-free distribution, claiming that, the networks especially the PIN from whole genome have tendency to contain hubs instead of a constant node/edge ratio.

Small-world property, mean path length & diameter: The magnitude of log(nodes) for PINs of 17 cores and whole genome respectively are 3.23 and 3.57 which are of the same order of magnitude as avg. path lengths (4.78 and 4.79) and diameters of the PINs (12 and 13) meeting the criteria for having small-world property.

Since the average path lengths of both networks are almost the same, it can be concluded that these networks have the same efficiency in the transfer of biological information. When compared to other networks found from literature, these networks are not the most/least efficient ones in transfer of biological information like C.elegans – DIP network (Li *et al.*, 2004). Still yeast whole genome network is the fastest, together with network of M.musculus.

The diameter of a network essentially characterizes the network's interconnectivity. The diameters of core network and whole genome are near to each other; so, the degree of interconnectivity in the networks does not differ from each other. When compared to the diameters of networks found from literature, they stay in between and also consistent with C.elegans – DIP network (Li *et al.*, 2004)

Clustering coefficients & modularity: Both networks have the same values of clustering coefficients as well as C.elegans – DIP network (Li et.al.). These three have the smallest clustering coefficient values together with D.melanogaster and H.pylori networks both obtained from DIP interactions.

It is seen that, most of the scale free networks listed in Table 5.3 are also highly modular like the networks studied.

Centrality: In network constructed for worm, all the centrality parameters are consistent with the lethality-centrality rule and the scale-free property. In other words, there are hubs in the work with the highest betweenness and closeness values within the network. There are no outliers observed as in the case of Joy *et al.* (2005).

In conclusion, both networks have almost the same topological properties although number of nodes varies by a factor of two and number of edges by a factor of three between networks.

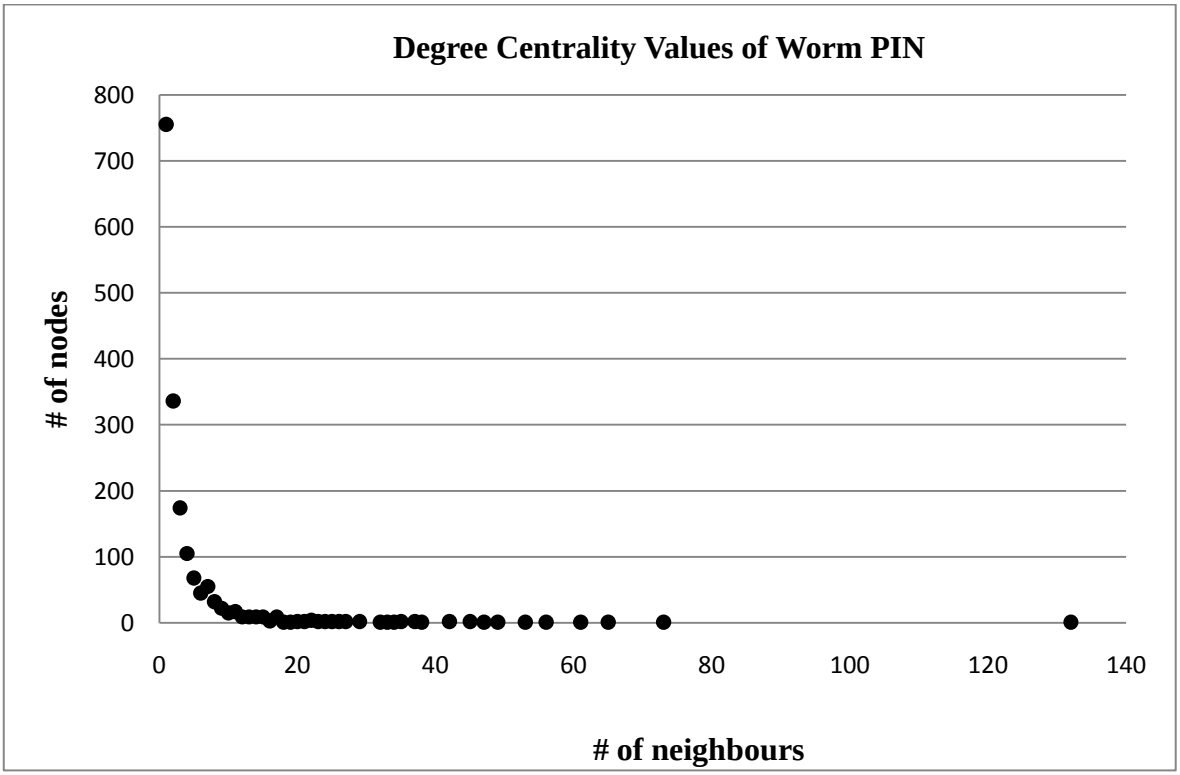


Figure 5.9. Degree distribution of worm PIN

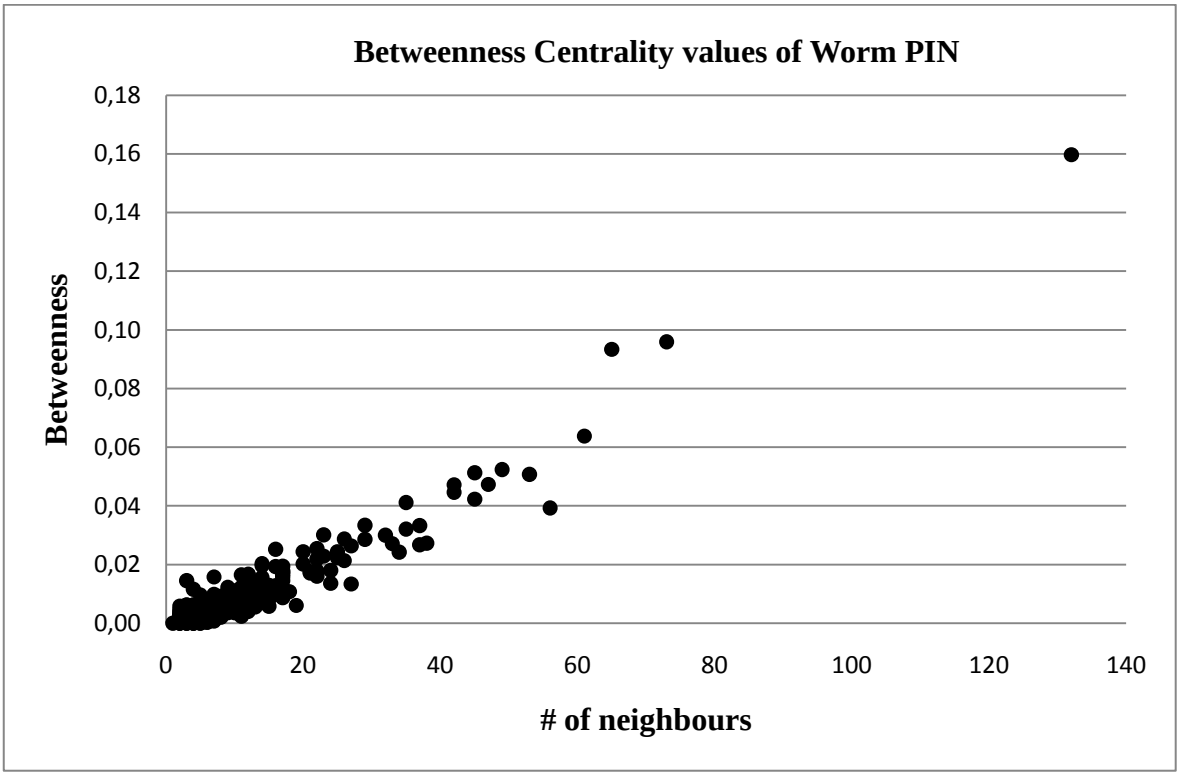


Figure 5.10. Betweenness centrality of worm PIN

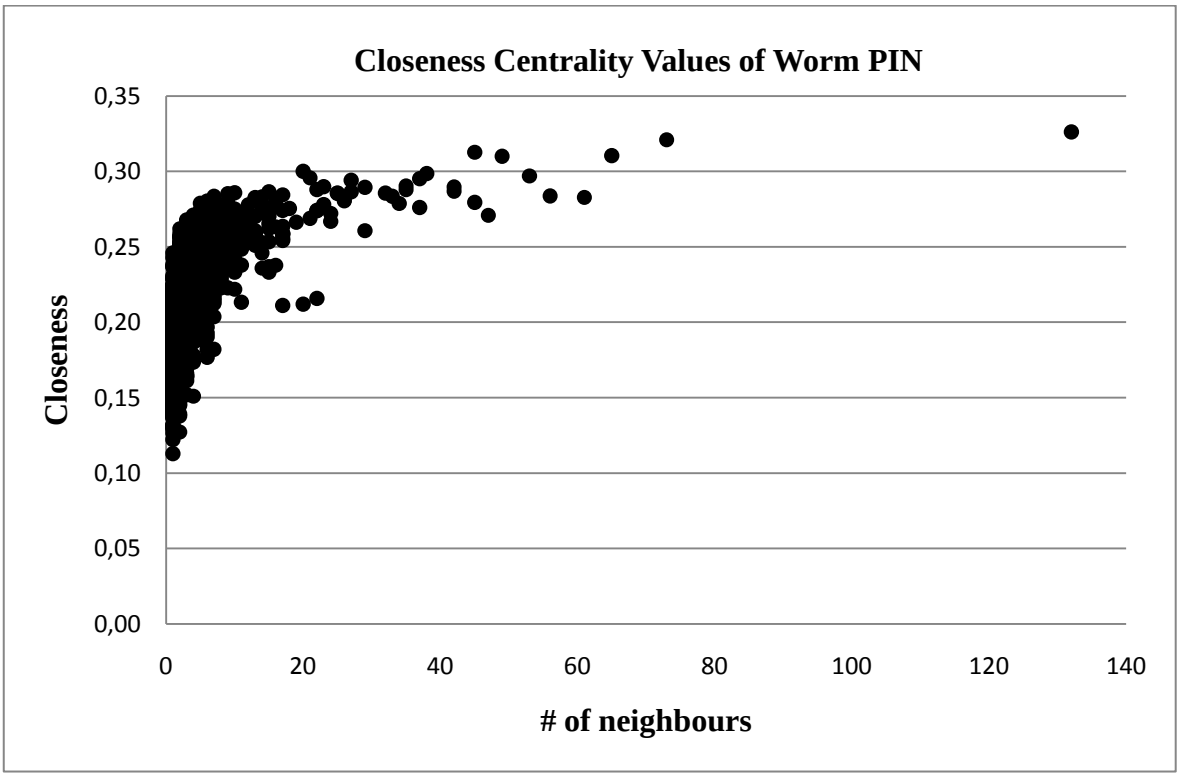
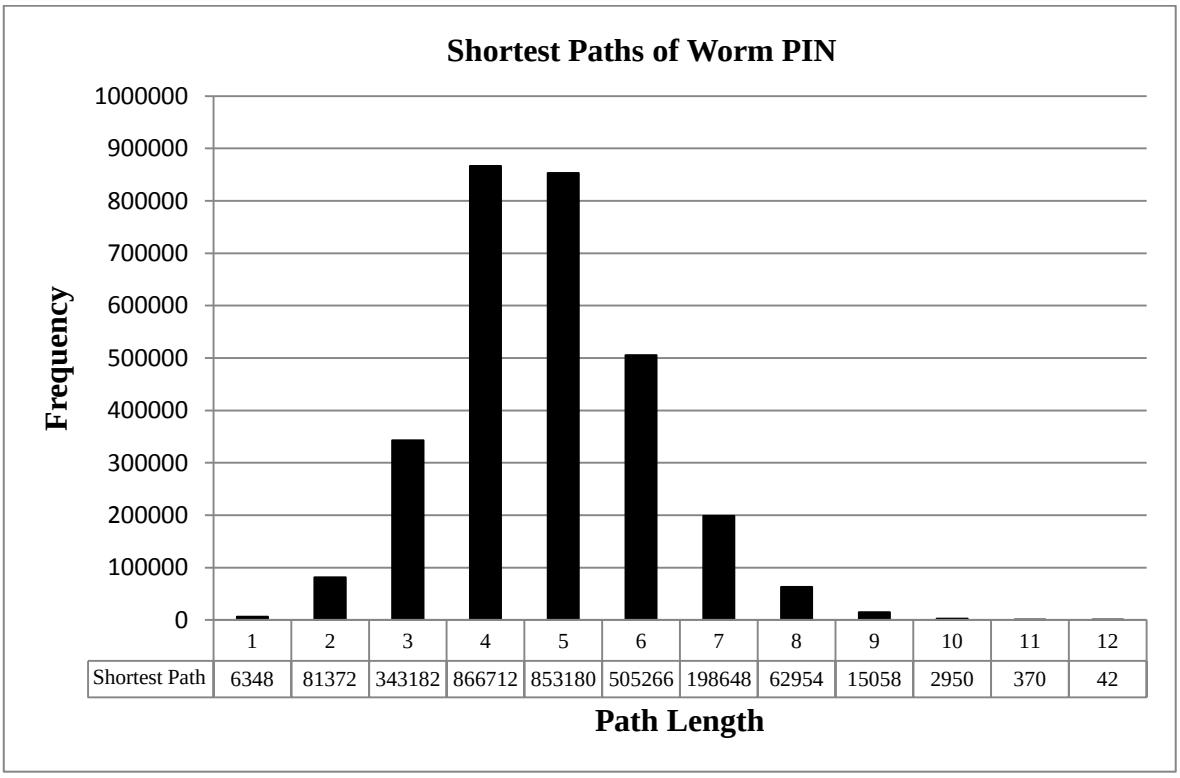


Figure 5.11. Closeness centrality of worm PIN



	1	2	3	4	5	6	7	8	9	10	11	12
Shortest Path	6348	81372	343182	866712	853180	505266	198648	62954	15058	2950	370	42

Figure 5.12. Shortest path distribution of worm PIN

Table 5.3. Topological properties of various networks found from literature

	# of nodes	# of edges	Mean connectivity (degree)	Mean path length	Network diameter	Clustering coefficient	Reference
S. cerevisiae (Ca signaling - 12 core)	1703	6299	7,4	3,82	10	0,16	Present study
S. cerevisiae (Ca signaling - 63 core)	3226	20167	12,5	3,43	9	0,15	Present study
S. cerevisiae - BIOGRID	5685	52141	18,3	2,78	6	0,24	Present study
C. elegans (Ca Signaling - 17 core)	1714	3174	3,70	4,78	12	0,03	Present study
C.elegans -BIOGRID & Intact	3741	7118	3,81	4,79	13	0,03	Present study
S.cerevisiae - core - DIP	2640	6600	5	5	13	0,33	Li *et al.*
S. cerevisiae	2115	2240	2,12	6,8	14		Jeong et.al.
S. cerevisiae	5253	65673	25	5,21	11		Arga et.al.
S. cerevisiae (signaling)	1363	3649	5,35	6,81	9		Arga et.al.
S.cerevisiae - DIP	4773	15444	6,5	5,01	13		Li *et al.*
S.cerevisiae - MIPS	2043	5434	5,32	7,71	16		Yook et.al
S.cerevisiae - DIP	5798	20098	6,93	4,9	12		Yook et.al
S.cerevisiae	1870	4480	4,79	6,8	15		Uetz et.al.
S.cerevisiae	3280	8868	5,41	4,9	13		Ito et.al.
S.cerevisiae - core	797	1560	3,91	6,2	16		Ito et.al.
D.melanogaster - DIP	7451	22819	6,2	4,4	11	0,02	Li *et al.*
C.elegans - DIP	2638	4030	3,1	4,8	14	0,06	Li *et al.*
H.pylori - DIP	710	1420	4	4,1	9	0,08	Li *et al.*
H.sapiens - DIP	1065	1369	2,6	6,8	21	0,21	Li *et al.*
E.coli - DIP	553	761	2,8	5,5	16	0,62	Li *et al.*
M.musculus - DIP	329	286	1,7	3,6	9	0,15	Li *et al.*

5.3.2. Hub Proteins of the Worm Network

In worm calcium signaling network constructed from 17 core proteins by SPA method, the top six hubs (the nodes with the highest degree) are identified using the Cytoscape plugin cytohub. Those hubs have 438 interactions with 356 proteins in total covering 13.8 % of the interactions and 20.8 % of the proteins in PIN.

Table 5.4. Hub Proteins of the Worm signaling network

	name	**ints in PIN**	**ints in whole genome PIN**	**Uniprot**	**aa**	**Description**
WBGene00001561	gei-4	132	189 1st	**Q9TZ39**	545	Gex interacting protein protein 4, isoform a; required for embryonic viability, fertility, and vulval morphogenesis
WBGene00003936	pat-12	73	111 2nd	**Q95QA6**	663	Gex interacting protein protein 16, isoform a
WBGene00000228	atn-1	65	98 3rd	**Q23158**	920	Ca^{2+}-binding actin-bundling protein (actinin), alpha chain (EF-Hand protein superfamily)
WBGene00001090	drh-1	61	89 5th	**O44165**	1037	Dicer-related helicase required for both germline and somatic RNA interference (RNAi)
WBGene00001377	eya-1	56	77 7th	**O17670**	503	involved in embryonic and larval tissue morphogenesis as well as larval viability
WBGene00011449	T04H1.2	53	91 4th	**Q22174**	588	Predicted translation elongation factor

In order to determine the role of hub proteins in the reconstructed PIN, they are compared to the hubs of the PIN constructed from whole genome (3741 nodes, 7118 edges). The top ten hubs of whole genome PIN are the hubs of the reconstructed PIN for calcium.

These results show that, the hubs of calcium PIN do not only have important roles in calcium related paths, but also in the whole genome for worm to survive implying the vitality of calcium signaling pathway for worm.

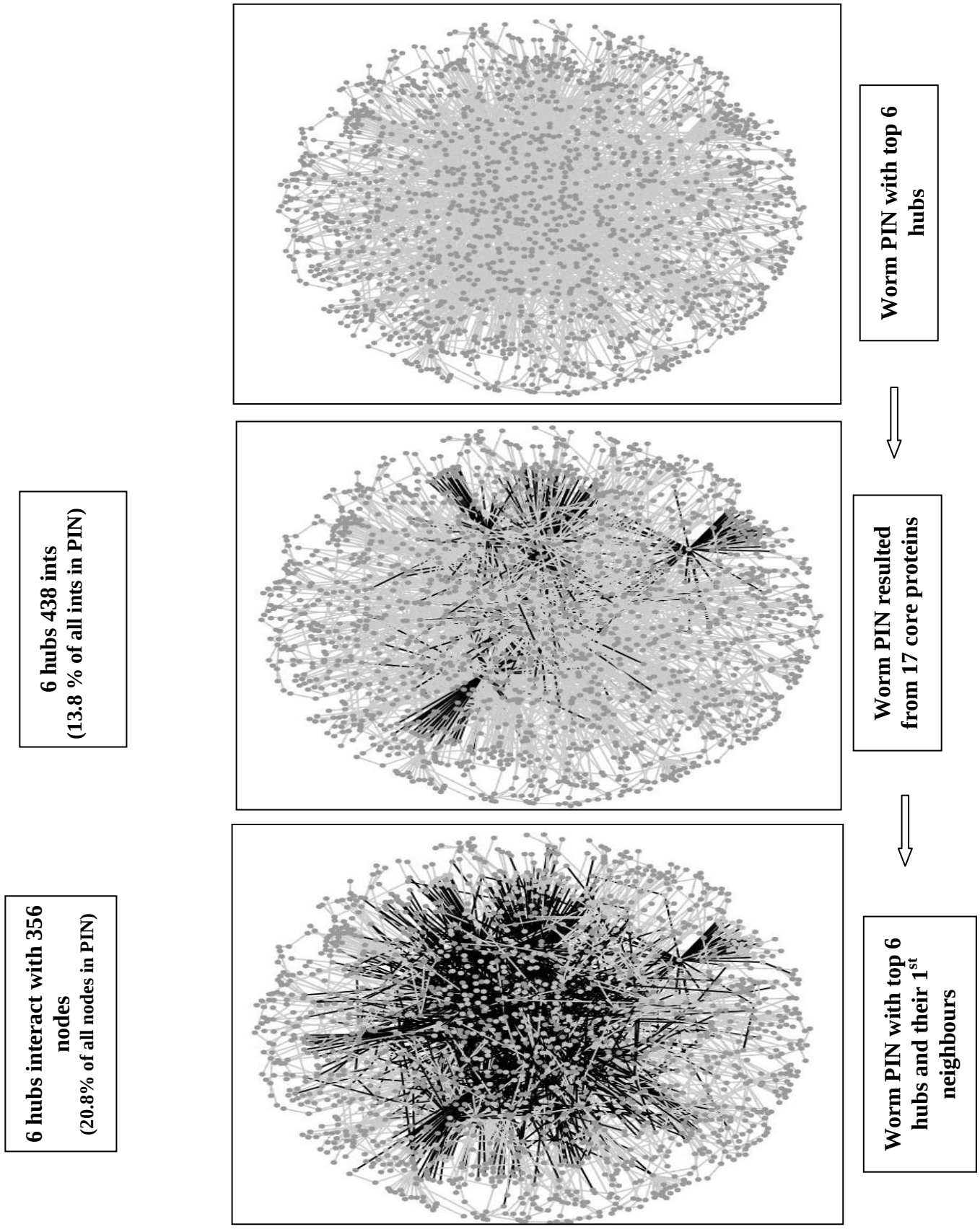


Figure 5.13. Hub proteins and their interaction in worm PIN

Finally, the hub proteins' dominant molecular function and biological process terms are also same with PIN terms; *nucleotide binding*, *ATP binding* taking place in the biological processes of *embryonic development ending in birth or egg hatching*, *reproduction*, *nematode larval development*, *positive regulation of growth rate*, *growth* and *locomotion* which possibly occur in *nucleus* or integral to *plasma membrane.*

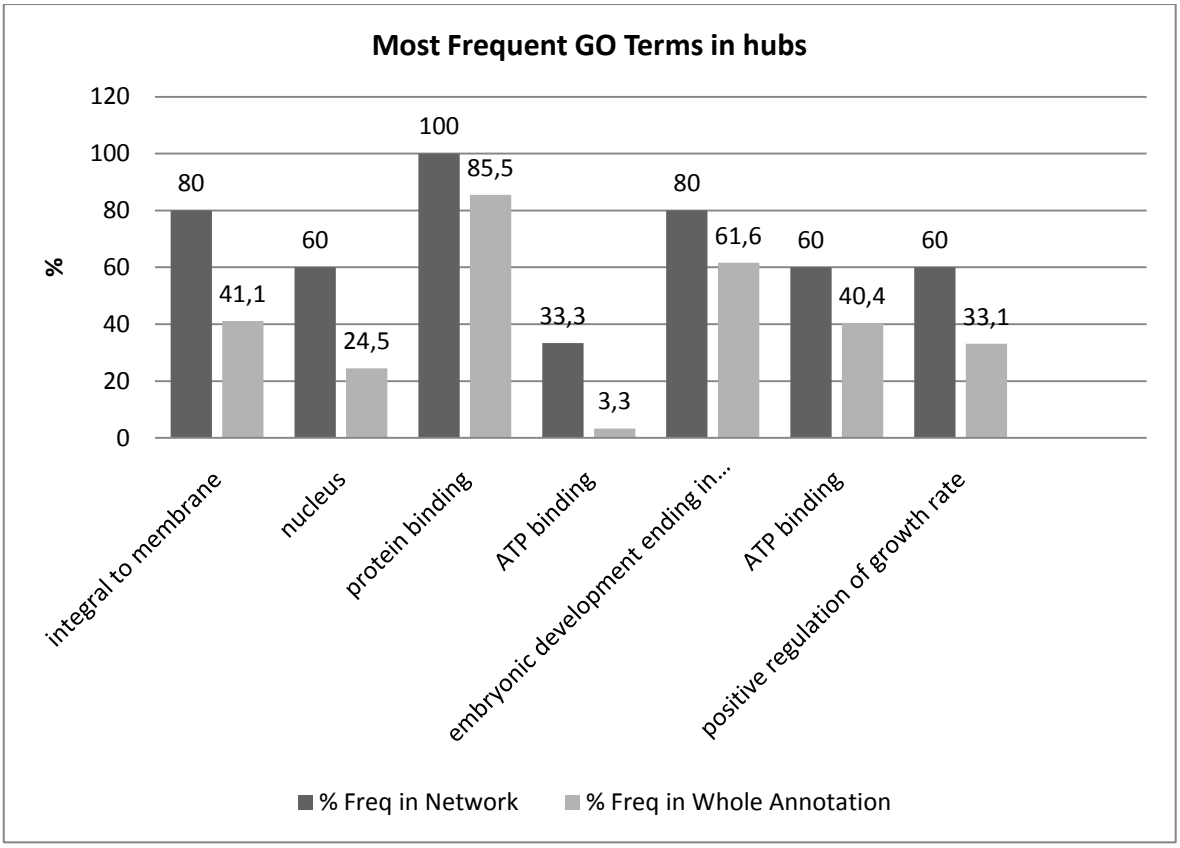


Figure 5.14. Predominant GO terms in worm hub proteins

5.3.3. Module Detection and Proteins with Unknown Function

In Worm PIN, there are 1026 proteins with no functional assignment at all; to which "molecular function unknown" term is given manually. Also there are 9 proteins with "molecular function unknown" term in whole genome given in Gene Ontology but not included in PIN.

Using MCODE, a Cytoscape plugin, 7 modules containing 200 functionally unknown proteins are determined (details and figures can be found in App F.4.) The

module selection criterion depends on the number of functionally known protein dominance in the mentioned module/subgroup interacting with the functionally unknown protein. However, out of these 7 modules, no modules can be chosen to be appropriate for function assignment to functionally unknown proteins since no dominance of known function proteins can be determined.

6. RESULTS & DISCUSSION FOR CALCIUM SIGNALING IN HUMAN

6.1. Core Proteins & Resultant Protein Interaction Network

6.1.1. Determination of Core Proteins for SPA

The Calcium signaling system in Figure 6.1 is taken from KEGG – Kyoto Encyclopedia of Genes and Genomes; under the title of 3.2 Signal Transduction – Calcium Signaling Pathway in Homo sapiens. Each box represents a group of proteins with the same function or structure.

As core proteins, the proteins which are directly related to calcium and expressed in brain regions are chosen (green boxes in Figure 6.1 and Table 6.1). The blue boxes, on the other hand, are the proteins that are not included in the core protein list but end up in the PIN obtained from SPA. Finally, the red boxes imply that those proteins are not present in the reconstructed PIN since they have no expression in any brain or brain region tissue but in other parts of the human body (PLCζ - phospholipase C, PLN - phospholamban, TnC - troponin C) or they are expressed in brain but are not able to be included in PIN (PHK – phosphorylase kinase, PLCδ) .

32 proteins under the groups:
NCX (Na/Ca exchangers),
PMCA (Plasma membrane Ca ATPases),
SOC (store operated channels),
VOCs (voltage operated channels),
NMDAR (NMDA receptors),
SERCA (Endoplasmic Reticulum Ca ATPases),
RYR (Ryanodine receptors),
IP3R (Inositol-3-phosphate receptors),
CALM (calmodulin),
VDAC (voltage dependent Ca channels in mitochondria),

ANT (mitochondrial Ca carriers) and

PPID (Ca related mitochondrial isomerases) are detected from the KEGG pathway as the core proteins.

The pumps (PMCA and SERCA groups) have ATPase and hydrolase domains. The channels (VOC and SOC groups except VDAC), on the other hand, contain ion transfer domains as expected. VDAC group has porin domains that form a voltage-dependent anion-selective channel (VDAC) that behaves as a general diffusion pore for small hydrophilic molecules.

NMDA receptors have three common domains; ANF receptor which is an extracellular ligand binding domain of a wide range of receptors, SBP bac 3 domain which is involved in active transport of solutes across the membrane and finally a domain specific to the glutamate receptors, Lig chan domain which includes the four transmembrane regions of the ionotropic glutamate receptors and NMDA receptors.

IP3R and RyR contain their specific domains like Ins145 P3 receptor domain, RyR domain and RIH associated domain. These receptors have an ion transfer domains as well.

Exchanger proteins function with their Na-Ca exchanger and Calx-beta domains. ANT proteins are substrate carrier proteins with mitochondrial carrier domains that are involved in energy transfer.

Finally, the calmodulin group has ef hand domains which are specific for calcium binding.

Table 6.1. Human Core Protein List

	HPRD Id	Uniprot Id	Name	A.a	Domain	Tissue expressed	Description
PMCA	159	P23634	PMCA4	1241	Cation ATPase N E1-E2 ATPase Hydrolase Cation ATPase C 2 ATP Ca Trans C	Brain, Cerebellum, Cerebral cortex, Hippocampus	ATPase, Ca++ transporting, plasma membrane 4
	160	Q01814	PMCA2	1243	Cation ATPase N E1-E2 ATPase Hydrolase Cation ATPase C ATP Ca Trans C	Brain	ATPase, Ca++ transporting, plasma membrane 2
NMDAR	339	P36544	CHRNA7	502	Neur chan LBD Neur chan memb	Brain, Hippocampus, Substantia nigra, Thalamus	Neuronal acetylcholine receptor subunit alpha-7
	698	Q12879	NMDAR2A	1464	ANF receptor SBP bac 3 Lig chan NMDAR2 C	Brain, Hippocampus	Glutamate [NMDA] receptor subunit epsilon-1
	699	Q14957	NMDAR2C	1236	ANF receptor SBP bac 3 Lig chan NMDAR2 C	Amygdala, Caudate nucleus, Corpus Callosum, Hippocampus, Hypothalamus, Substantia nigra, Subthalamic nucleus, Thalamus	Glutamate [NMDA] receptor subunit epsilon-3
	3977	Q99572	P2X7	595	P2X Receptor	Brain	P2X purinoceptor 7
	4095	O15399	NMDAR2D	1336	ANF receptor SBP bac 3 Lig chan	Brain, Cerebellar cortex, Cerebellum, Cerebral cortex, Purkinje Cells	Glutamate [NMDA] receptor subunit epsilon-4

Table 6.2. Human Core Protein List - cont.

	15926	Q05586	NMDAR1	938	ANF receptor SBP bac 3 Lig chan CaM bdg CO	Brain, Caudate nucleus, Cerebellum, Dentate gyrus, Frontal cortex, Hippocampus, Locus coeruleus, Nucleus accumbens, Pontine nucleus, Putamen, Temporal cortex, Temporal lobe	Glutamate [NMDA] receptor subunit zeta-1
	15984	Q9UBL9	P2X2	471	P2X Receptor	Brain, Amygdala	P2X purinoceptor 2
VOC	247	Q01668	Cav1.3	2161	4 Ion trans Ca chan IQ	Brain	Voltage-dependent L-type calcium channel subunit alpha-1D
	3004	O00555	Cav2.1	2505	4 Ion trans Ca chan IQ	Brain, Cerebellum	Voltage-dependent P/Q-type calcium channel subunit alpha-1A
	3005	Q00975	Cav2.2	2339	4 Ion trans Ca chan IQ	Dentate nucleus, Hippocampus, Parahippocampal gyrus, Purkinje cell, Pyramidal cell	Voltage-dependent N-type calcium channel subunit alpha-1B
	3006	Q15878	Cav2.3	2313	4 Ion trans Ca chan IQ	Brain, Cerebellum, Dentate gyrus, Entorhinal cortex, Hippocampus, Parahippocampal gyrus	Voltage-dependent R-type calcium channel subunit alpha-1E
	7442	O95180	Cav3.2	2353	4 Ion trans	Brain, Amygdala, Caudate nucleus, Cerebellum, Cerebral cortex, Corpus callosum, Frontal lobe, Hippocampus, Medulla oblongata, Occipital lobe, Putamen, Substantia nigra, Subthalamic Nucleus, Temporal lobe, Thalamus	Voltage-dependent T-type calcium channel subunit alpha-1H

Table 6.3. Human Core Protein List - cont.

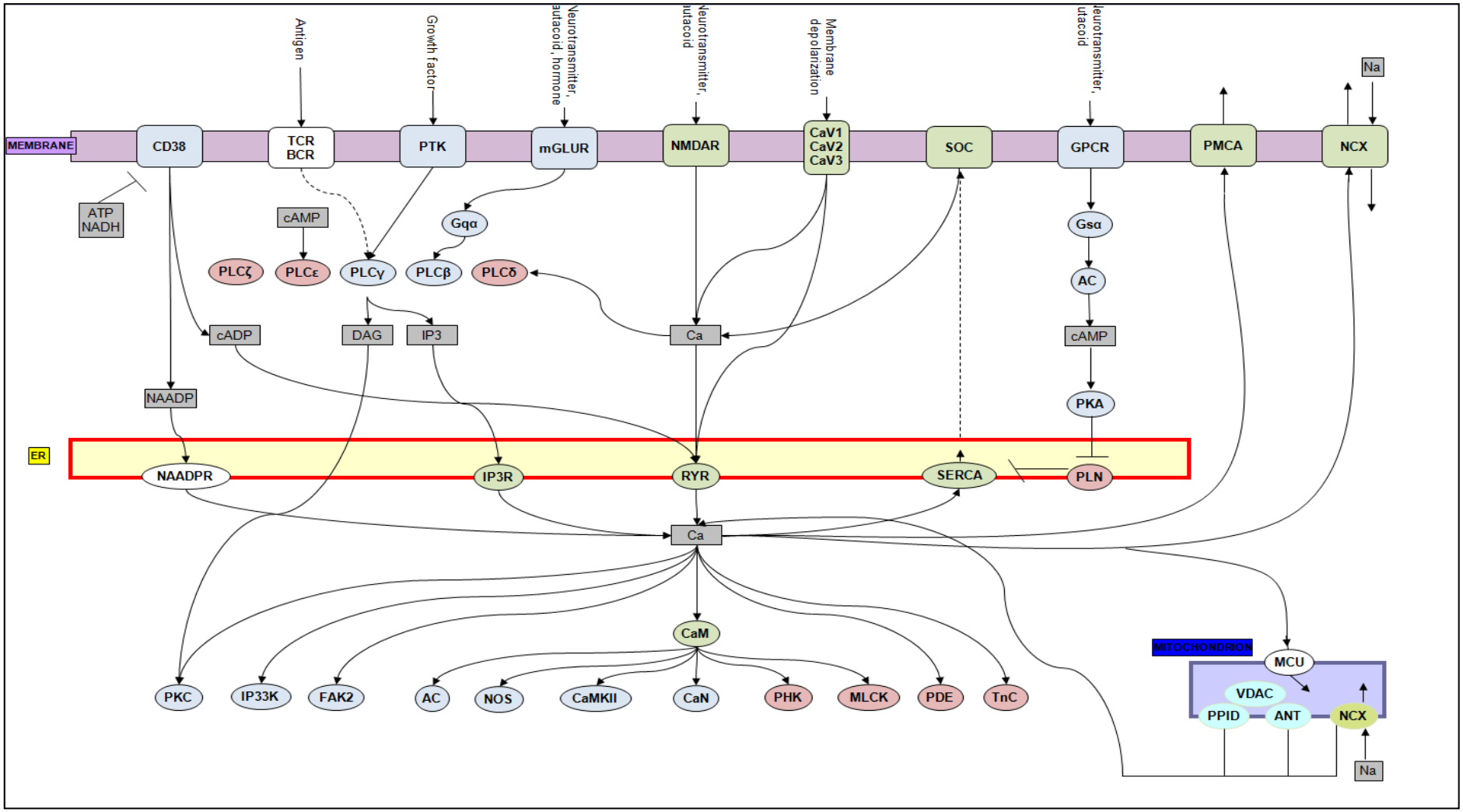
SOC	11894	P48995	TrpC1	793	TRP 2 Ion trans	Brain	Short transient receptor potential channel 1
RYR	1618	P21817	RYR-1	5038	Ins145 P3 rec MIR 2 RYDR ITPR 3 SPRY 4 RyR RIH assoc RR TM4-6 Ion trans	Cerebellum, Hippocampus	Ryanodine receptor 1
	1619	Q92736	RYR-2	4965	Ins145 P3 rec MIR 2 RYDR ITPR 3 SPRY 4 RyR RIH assoc RR TM4-6 Ion trans	Brain	Ryanodine receptor 2
	1620	Q15413	RYR-3	4870	Ins145 P3 rec MIR 2 RYDR ITPR 3 SPRY 4 RyR RIH assoc RR TM4-6 Ion trans	Brain, Caudate nucleus, Cerebellum, Hippocampus, Substantia nigra, Subthalamus, Thalamus	Ryanodine receptor 3

Table 6.4. Human Core Protein List - cont.

IP3R	925	Q14643	IP3R 1	2758	Ins145 P3 rec MIR 2 RYDR ITPR RIH assoc Ion trans	Caudate nucleus, Cerebellum, Hippocampus, Purkinje cell	Inositol 1,4,5-trisphosphate receptor type 1
SERCA	161	P16615	SERCA2	1042	Cation ATPase N E1-E2 ATPase Hydrolase Cation ATPase C	Brain	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2
	3568	Q93084	SERCA3	1043	Cation ATPase N E1-E2 ATPase Hydrolase Cation ATPase C	Brain, Amygdala, Caudate nucleus, Cerebellum, Hippocampus, Medulla oblongata, Occipital lobe, Putamen, Substantia nigra, Subthalamic Nucleus, Temporal lobe, Thalamus	Sarcoplasmic/endoplasmic reticulum calcium ATPase 3
NCX	1659	P32418	NCX1	973	2 Na Ca ex 2 Calx-beta	Hippocampus, Temporal lobe	Sodium/calcium exchanger 1
	9723	P57103	NCX3	927	2 Na Ca ex 2 Calx-beta	Brain	Sodium/calcium exchanger 3
	15990	Q9UPR5	NCX2	921	2 Na Ca ex 2 Calx-beta	Brain, Amygdala, Caudate nucleus, Cerebellum, Corpus callosum, Hippocampus, Substantia nigra, Subthalamic Nucleus, Thalamus	Sodium/calcium exchanger 2
ANT	58	P12235	ANT 1	298	3 Mito carr	Brain	ADP/ATP translocase 1, mitochondrial carrier
	2147	P05141	ANT 2	298	3 Mito carr	Brain	ADP/ATP translocase 2, mitochondrial carrier

Table 6.5. Human Core Protein List - cont.

PPID	11875	Q08752	PPID	370	Pro isomerase TPR 1	Brain	Peptidyl-prolyl cis-trans isomerase D
VDAC	1904	P45880	VDAC2	294	Porin 3	Cerebral cortex, Hippocampus	Voltage-dependent anion-selective channel protein 2
	5137	P21796	VDAC1	283	Porin 3	Brain, Cerebellum, Cerebral cortex	Voltage-dependent anion-selective channel protein 1
	6750	Q9Y277	VDAC3	283	Porin 3	Brain	Voltage-dependent anion-selective channel protein 3
CAM	242	P62158	CALM2	149	4 efhand	Brain	Calmodulin
	243	P62158	CALM3	149	4 efhand	Brain	Calmodulin 3



Core proteins Proteins in PIN Proteins not in PIN/core

Figure 6.1. Calcium signaling in human

6.1.2. Detection of Brain Specific Proteome/Interactome Data

The proteome data is downloaded from HPRD including tissue expression details. The Human Proteinpedia data (101629 tissue expression data) is combined with HPRD data (112158 tissue expression data). Among those proteome data (213737 tissue expressions for 19701 proteins), 8188 proteins are found to be expressed in brain regions.

For the Interactome data; HPRD, HUPA and Biogrid interaction data sets are combined eliminating genetic and in vitro interactions.

Table 6.6. Interactome Data Decomposition

	Biogrid	HPRD	HUPA
All	45726	39241	4607
Physical	34108	19256	4607
Brain related	18388	10507	2311

The self loops and duplicate interactions are deleted from the combined data yielding in *5222 brain related proteins* and *20710 interactions among themselves*.

Finally, the GO annotations of the human proteins required for SPA, are downloaded from Gene Ontology database.

The challenging part is to define each protein under a common identity. HPRD uses its own HPRD ids and corresponding protein abbreviations. Biogrid assigns Uniprot ids to the proteins. On the other hand, Gene Ontology uses protein abbreviations and sometimes Uniprot ids. An intractable matching process is carried out in order to assign a HPRD id to each protein.

In the Gene Ontology, some proteins have no associations (e.g. no cellular component term). To these proteins, corresponding “unknown” terms (Go id: 5575 for cellular component unknown, Go id: 3674 for molecular function unknown, Go id: 8150 for biological process unknown) are assigned since no term implies unknown annotation.

6.1.3. Resultant Protein Interaction Network

SPA with 32 core proteins resulted in a PIN consisting of ***2593*** proteins and ***9890*** undirected physical interactions after 7 runs. 1 core protein (SLC25A5 – HPRD no:2147) is lonely, so it is not included in the resultant PIN.

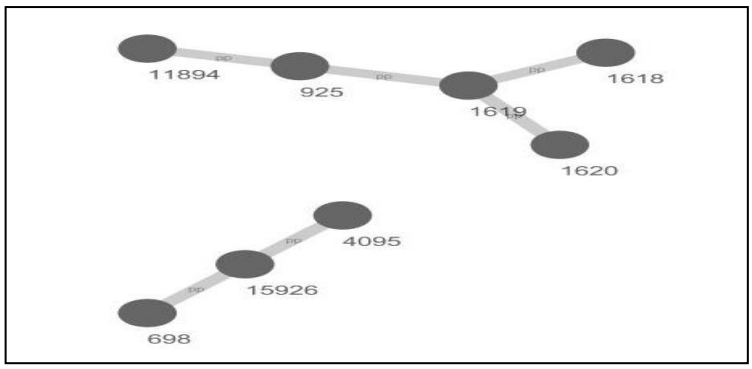


Figure 6.2. Human brain calcium signaling core protein network consisting of 8 nodes and 6 edges. 24 lonely proteins are not shown. Produced by Cytoscape 2.6.

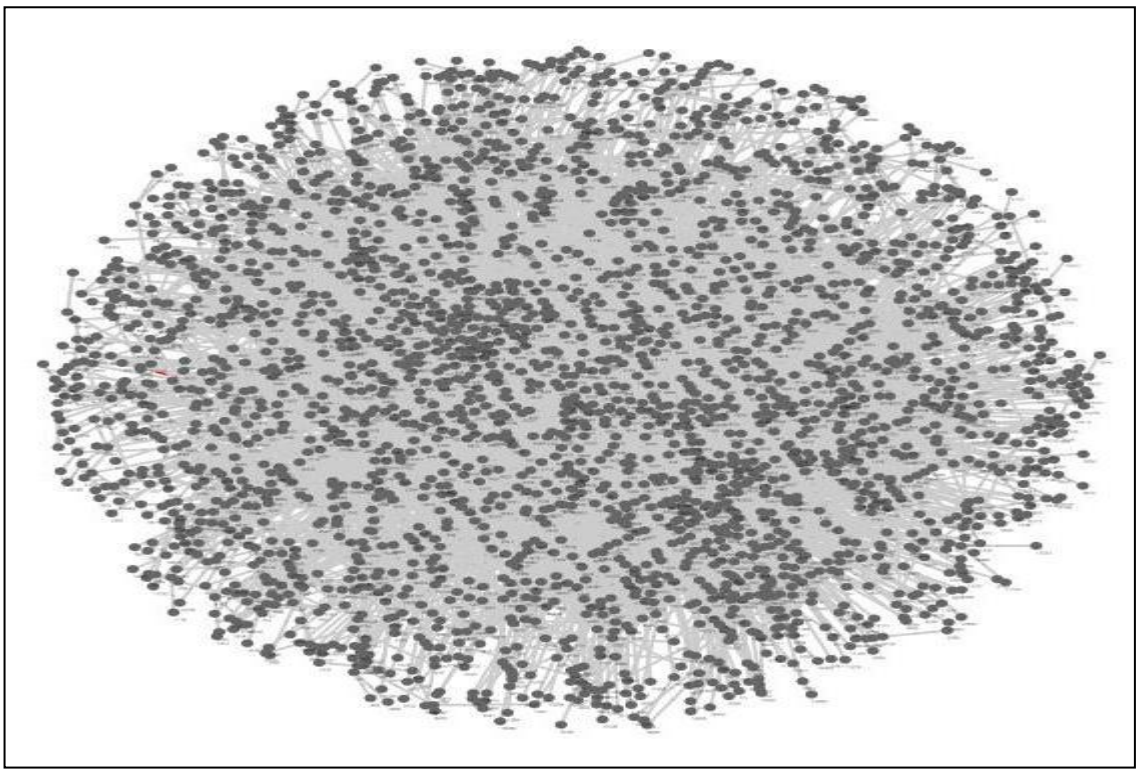


Figure 6.3. SPA resultant Human brain calcium signaling protein network consisting of 2593 nodes and 9890 edges. Produced by Cytoscape 2.6.

6.2. Analysis of Reconstructed Network

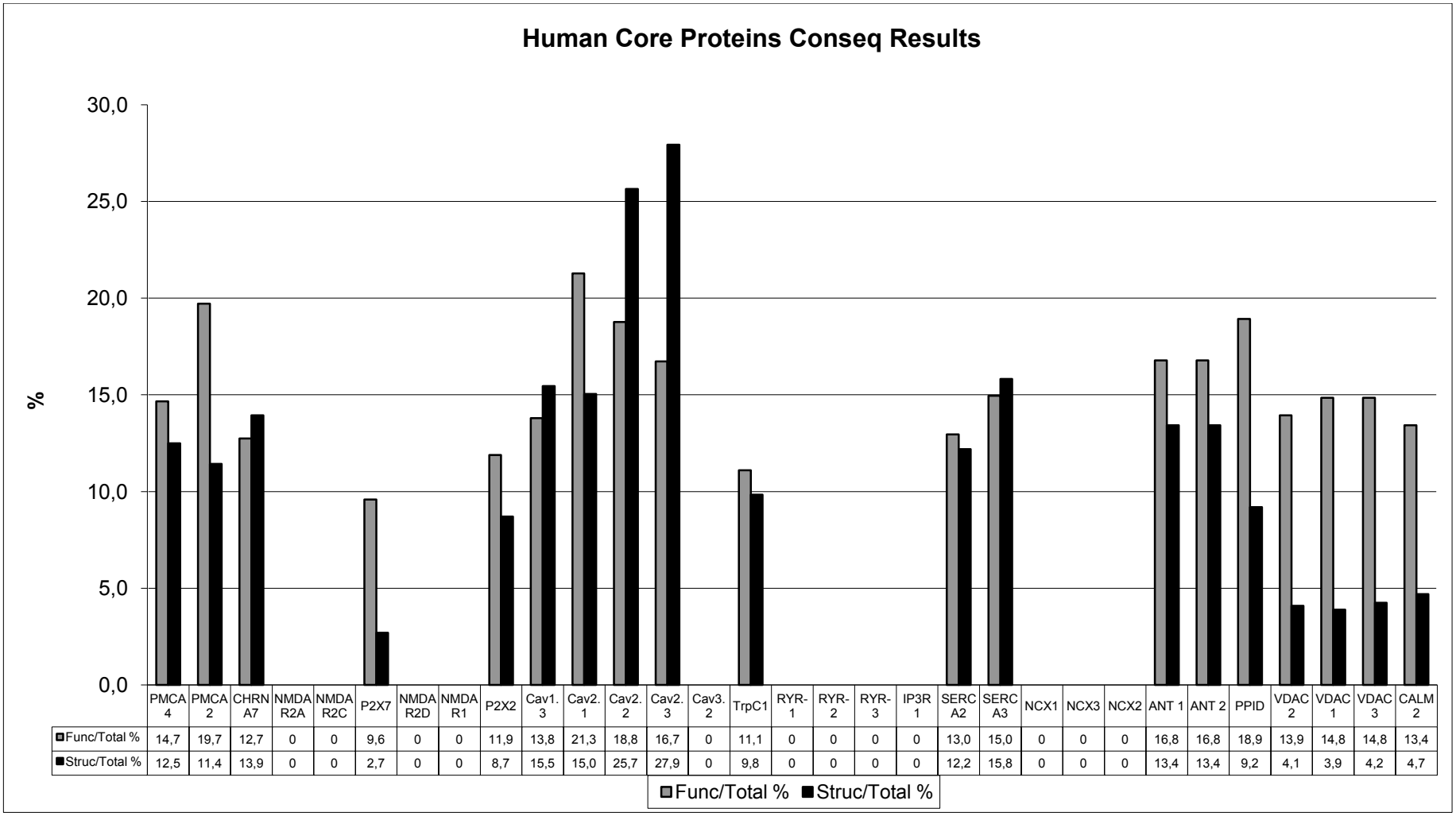
6.2.1. Evolutionary History of Core Proteins

The core proteins chosen among the KEGG proteins have an average of 33 % conservation score (Table 6.3), highly conserved ones being PMCA proteins and some VOCs and the low conservation scored one being the NMDAR protein P2X7.

Most of the NMDAR proteins as well as IP3R, RyR and Na-Ca exchangers do not have enough aligned sequence in ConSeq, so that no data can be extracted for them.

In the VDAC channel group and calmodulin proteins, the functionally conserved regions are much more than the structurally important ones (Figure 6.4). The PMCA pumps also have more functionally conserved regions than structurally conserved ones. On the other hand, SERCA pumps have an equal distribution in terms of functionally and structurally important sites.

The mitochondrial carriers also have more functionally conserved regions than structurally conserved ones. For voltage gated calcium channels, in contrary to VDAC, structurally important sites are dominant.



	PMCA4	PMCA2	CHRNA7	NMDAR2A	NMDAR2C	P2X7	NMDAR2D	NMDAR1	P2X2	Cav1.3	Cav2.1	Cav2.2	Cav2.3	Cav3.2	TrpC1	RYR-1	RYR-2	RYR-3	IP3R1	SERCA2	SERCA3	NCX1	NCX3	NCX2	ANT 1	ANT 2	PPID	VDAC2	VDAC1	VDAC3	CALM2
Func/Total %	14,7	19,7	12,7	0	0	9,6	0	0	11,9	13,8	21,3	18,8	16,7	0	11,1	0	0	0	0	13,0	15,0	0	0	0	16,8	16,8	18,9	13,9	14,8	14,8	13,4
Struc/Total %	12,5	11,4	13,9	0	0	2,7	0	0	8,7	15,5	15,0	25,7	27,9	0	9,8	0	0	0	0	12,2	15,8	0	0	0	13,4	13,4	9,2	4,1	3,9	4,2	4,7

Figure 6.4. ConSeq results for Human core proteins

Table 6.7. Conservation Scores of Human core proteins

HPRD Id	Name	A.a	Conservation		Residues	Exposed	Residues	Buried	Functionally	Structurally	Func/conserved	Struc/conserved	Func/total	Struc/total	Func/exposed	Struc/buried
			Total	% cons	exposed	%	buried	%	important res.	important res.	%	%	%	%	%	%
159	PMCA4	1241	435	35,1	672	54,1	569	45,9	182	155	41,8	35,6	14,7	12,5	27,1	27,2
160	PMCA2	1243	533	42,9	668	53,7	575	46,3	245	142	46,0	26,6	19,7	11,4	36,7	24,7
339	CHRNA7	502	191	38,0	270	53,8	232	46,2	64	70	33,5	36,6	12,7	13,9	23,7	30,2
698	NMDAR2A	1464	-	-	-	-	-	-	-	-	-	-	-	-	-	-
699	NMDAR2C	1236	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3977	P2X7	595	105	17,6	344	57,8	251	42,2	57	16	54,3	15,2	9,6	2,7	16,6	6,4
4095	NMDAR2D	1336	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15926	NMDAR1	938	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15984	P2X2	471	117	24,8	304	64,5	167	35,5	56	41	47,9	35,0	11,9	8,7	18,4	24,6
247	Cav1.3	2161	781	36,1	1160	53,7	1001	46,3	298	334	38,2	42,8	13,8	15,5	25,7	33,4
3004	Cav2.1	2505	721	28,8	1451	57,9	1054	42,1	533	377	73,9	52,3	21,3	15,0	36,7	35,8
3005	Cav2.2	2339	1072	45,8	1316	56,3	1023	43,7	439	600	41,0	56,0	18,8	25,7	33,4	58,7
3006	Cav2.3	2313	1055	45,6	1283	55,5	1030	44,5	387	646	36,7	61,2	16,7	27,9	30,2	62,7
7442	Cav3.2	2353	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11894	TrpC1	793	200	25,2	413	52,1	380	47,9	88	78	44,0	39,0	11,1	9,8	21,3	20,5
1618	RYR-1	5038	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1619	RYR-2	4965	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1620	RYR-3	4870	-	-	-	-	-	-	-	-	-	-	-	-	-	-
925	IP3R 1	2758	-	-	-	-	-	-	-	-	-	-	-	-	-	-
161	SERCA2	1042	354	34,0	539	51,7	503	48,3	135	127	38,1	35,9	13,0	12,2	25,0	25,2
3568	SERCA3	1043	399	38,3	544	52,2	499	47,8	156	165	39,1	41,4	15,0	15,8	28,7	33,1
1659	NCX1	973	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 6.8. Conservation Scores of Human core proteins - cont.

15990	NCX2	921	-	-	-	-	-	-	-	-	-	-	-	-	-	-
58	ANT 1	298	115	38,6	137	46,0	161	54,0	50	40	43,5	34,8	16,8	13,4	36,5	24,8
2147	ANT 2	298	114	38,3	137	46,0	161	54,0	50	40	43,9	35,1	16,8	13,4	36,5	24,8
11875	PPID	370	121	32,7	257	69,5	113	30,5	70	34	57,9	28,1	18,9	9,2	27,2	30,1
1904	VDAC2	294	74	25,2	176	59,9	118	40,1	41	12	55,4	16,2	13,9	4,1	23,3	10,2
5137	VDAC1	283	73	25,8	170	60,1	113	39,9	42	11	57,5	15,1	14,8	3,9	24,7	9,7
6750	VDAC3	283	73	25,8	170	60,1	113	39,9	42	12	57,5	16,4	14,8	4,2	24,7	10,6
242	CALM2	149	41	27,5	100	67,1	49	32,9	20	7	48,8	17,1	13,4	4,7	20,0	14,3

6.2.2. Annotation Table Enrichment in Human PIN

The annotation table, set as the main criteria for SPA method, is constructed from the GO terms of the 32 core proteins. With the intention of assigning new functions to the proteins not having any known function, the GO terms of "cellular component unknown" (GO id: 5575), "molecular function unknown" (GO id: 3674) and "biological process unknown" (GO id: 8150) are also included in the Annotation Table.

The Annotation Table of core proteins contains 74 cellular component terms, 78 molecular function terms and 213 biological process terms with a total number of 365 GO terms.

The dominant biological process terms in the core protein set are *calcium ion transport* and *transmembrane transport* in general as expected; since the core proteins chosen are mainly dealing with calcium signaling present in *plasma membrane, membrane and cytoplasm* in order to sustain calcium ion homeostasis in the cell. Hence, *plasma membrane, membrane* and *cytosol* being the dominant cellular component terms are consistent with the biological process dominance. Also the most frequent molecular function terms are *protein binding, receptor binding, calmodulin binding and ion channel activity* which are the main functions of core proteins (Figures 6.5-6.10).

On the other hand; *signal transduction* and *regulation of transcription* are the abundant biological process terms in the reconstructed PIN. Accordingly, the dominant molecular function terms are *protein binding* and *nucleotide binding* which take place in cytoplasm and *nucleus* (Figures 6.5-6.10).

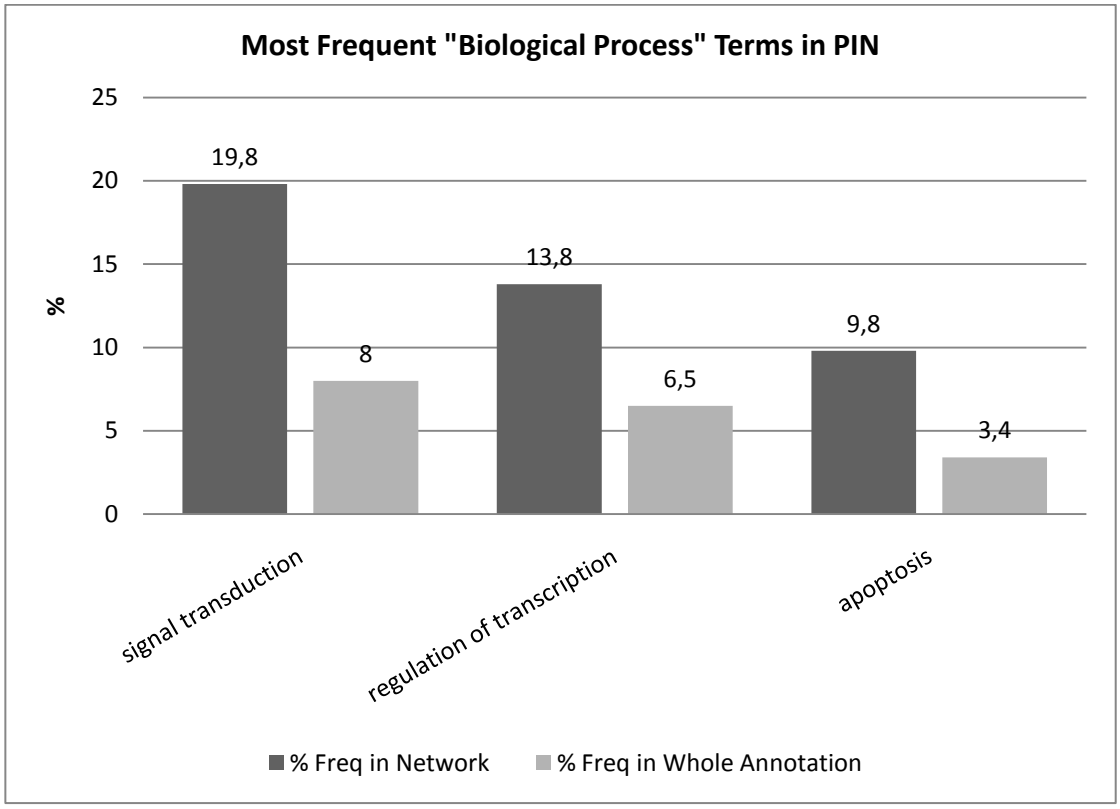


Figure 6.5. Predominant "Biological Process" terms in Human PIN

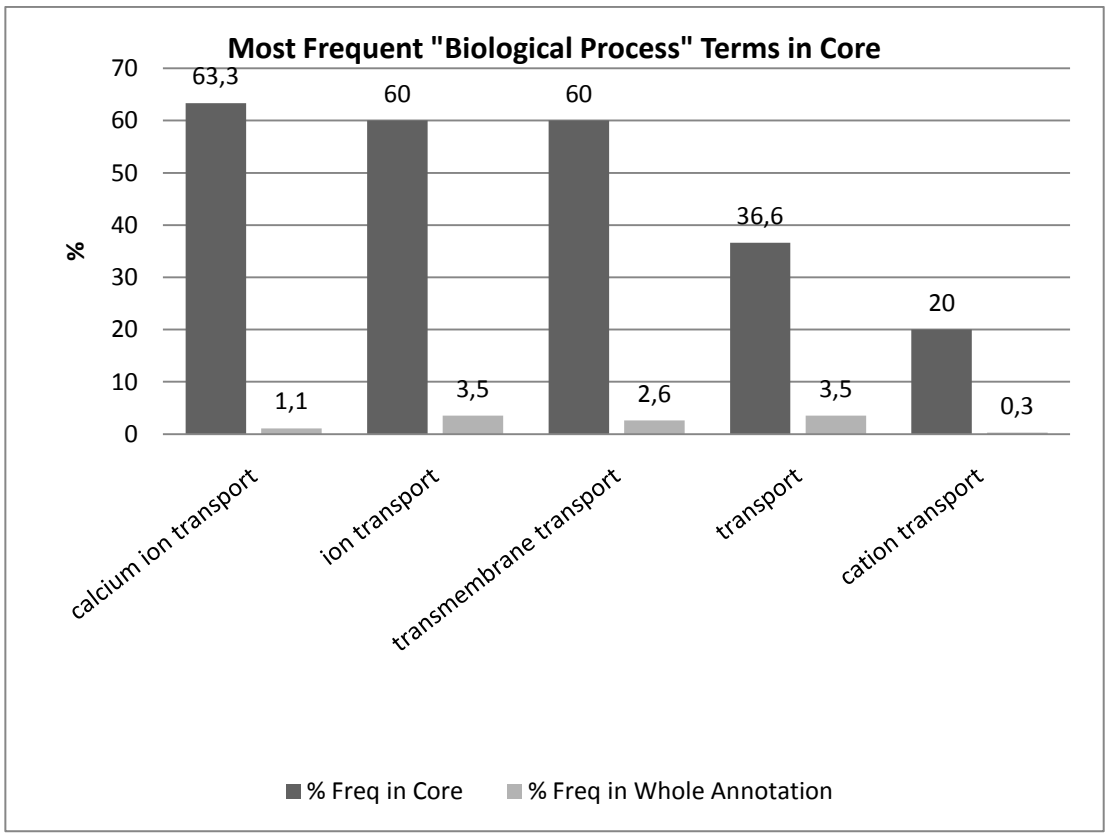


Figure 6.6. Predominant "Biological Process" terms in Human core

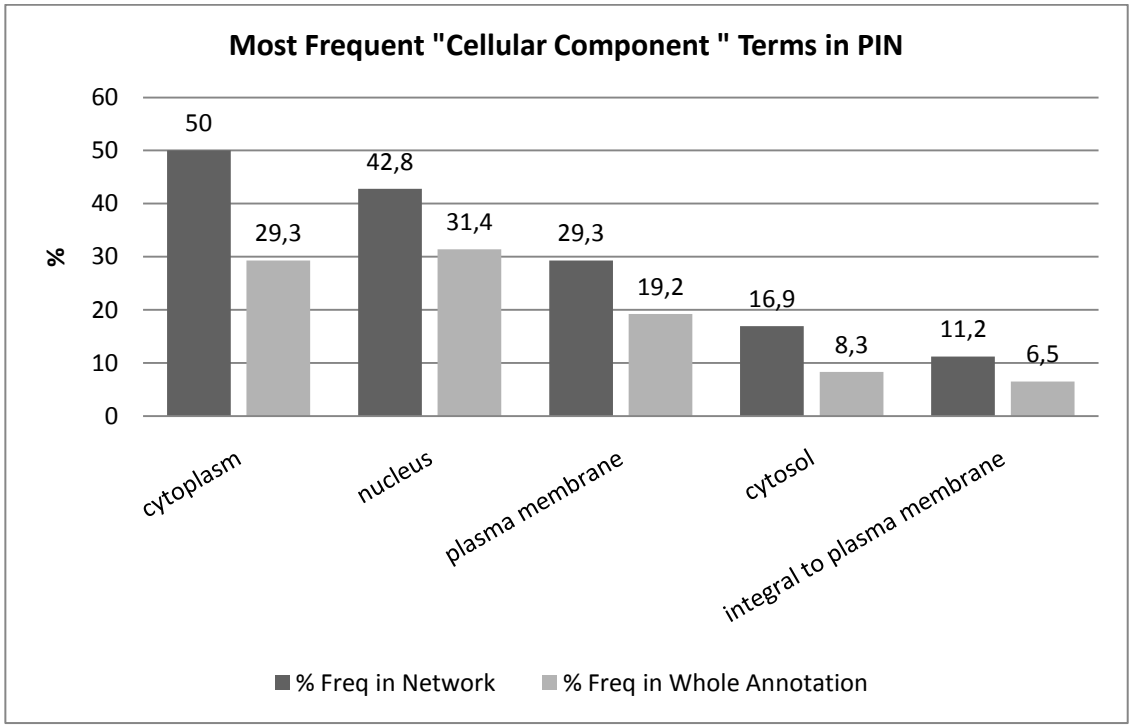


Figure 6.7. Predominant "Cellular Component" terms in Human PIN

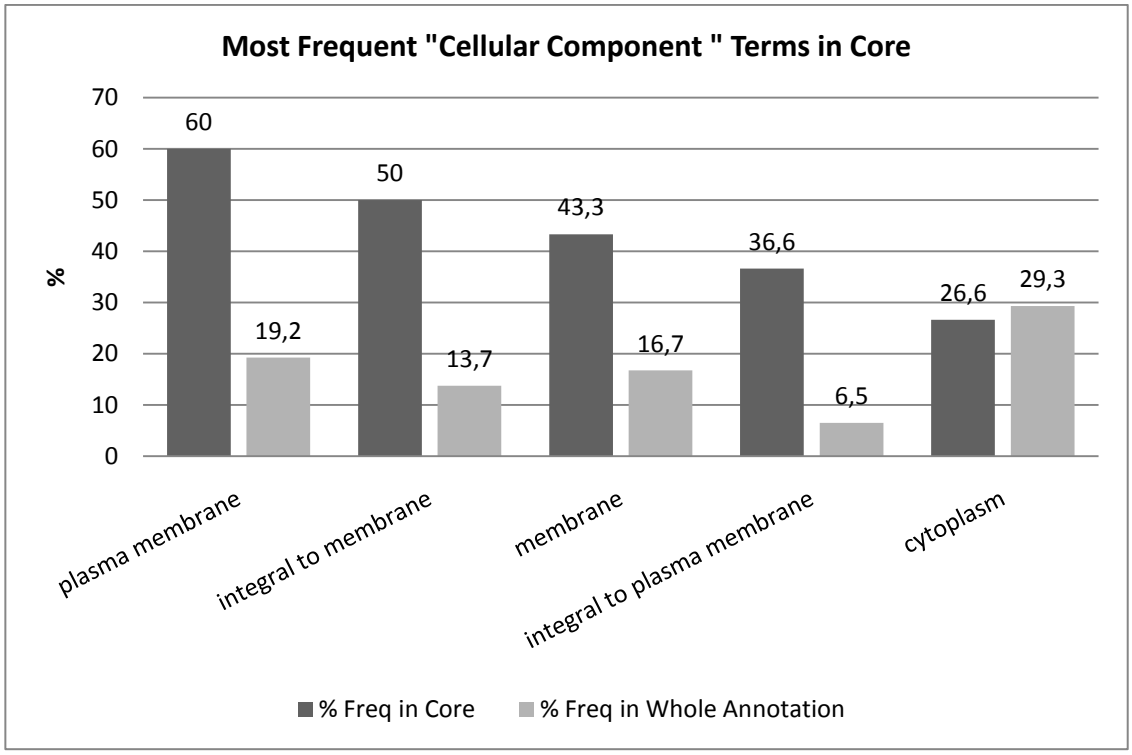


Figure 6.8. Predominant "Cellular Component" terms in Human core

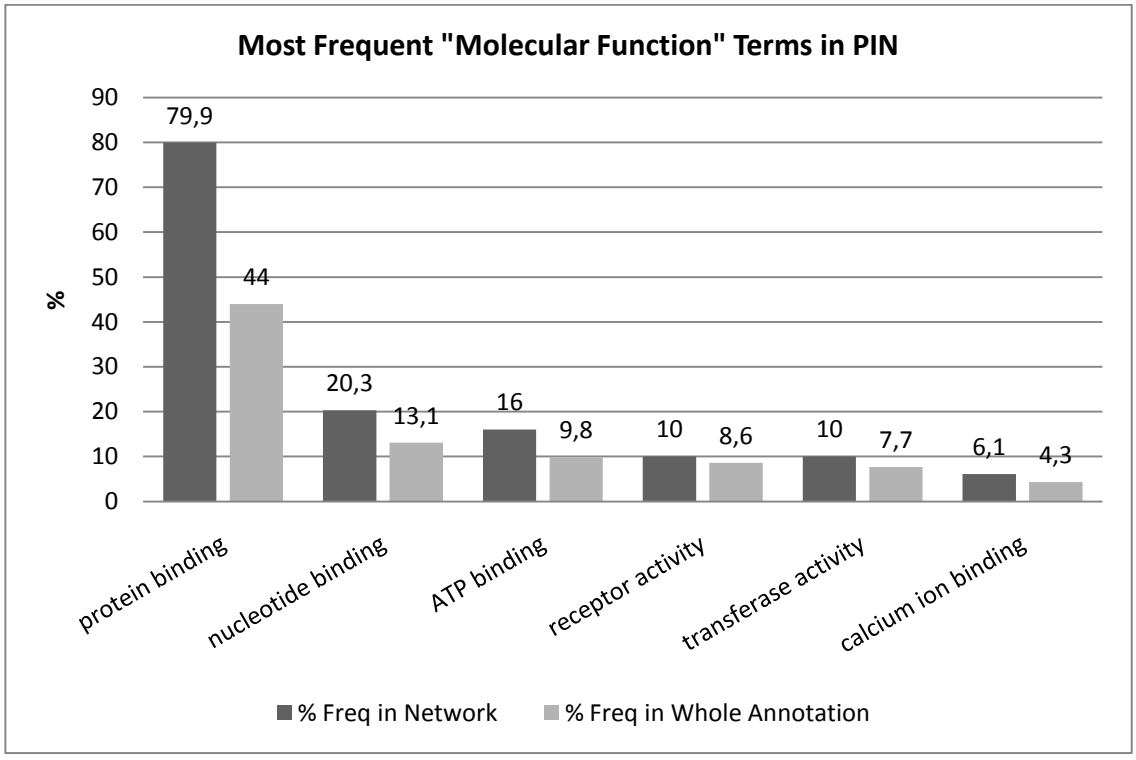


Figure 6.9. Predominant "Molecular Function" terms in Human PIN

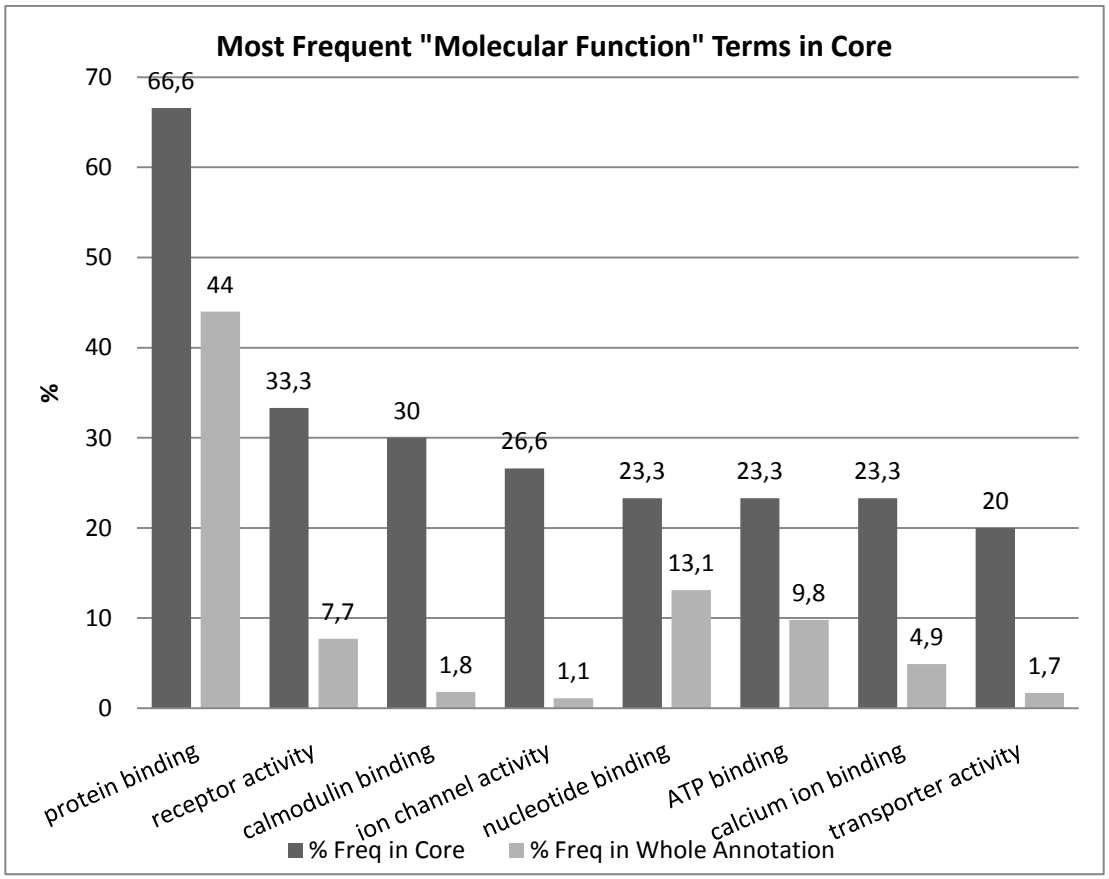


Figure 6.10. Predominant "Molecular Function" terms in Human core

6.3. Topology Based Analysis of Human PIN

6.3.1. Topological Analysis of Human Network

The undirected protein interaction graph of the human brain signaling network was analyzed using the Network Analysis algorithm, a Cytoscape plugin. The reconstructed signaling network includes 2593 proteins (represented as nodes) and 9890 interactions (represented as edges) among these proteins. (Table 6.4)

Table 6.9. Topological properties of the human signaling networks and whole brain interactome

HUMAN	*Core*	*PPI Network*	*Whole Brain Interactome*
# of nodes	8	2593	5124
# of edges	6	9890	20710
Mean connectivity (degree)	1,50	7,62	7,93
Mean path length	1,69	3,87	4,01
Network diameter	3	9	13
Network density	0,214	0,003	0,002
Network heterogenity	0,47	1,59	1,74
Clustering coefficient	0	0,14	0,12
Network centralization	0,29	0,06	0,05
Shortest paths	26	6721058	25265856
Power Law Constant (g)		1.64 (R^2:0.90)	1.74 (R^2:0.92)
# of isolated & lonely nodes	24	3	98
Total # of proteins	32	2596	5222

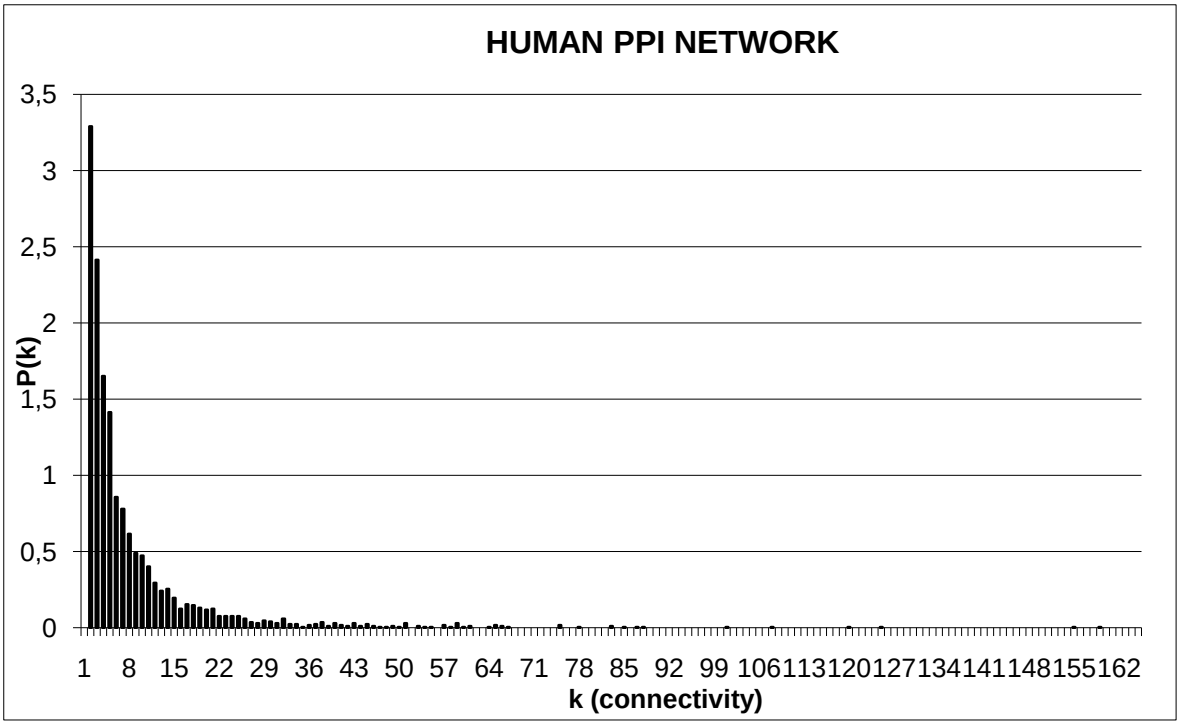


Figure 6.11. Degree distribution in the human signaling network

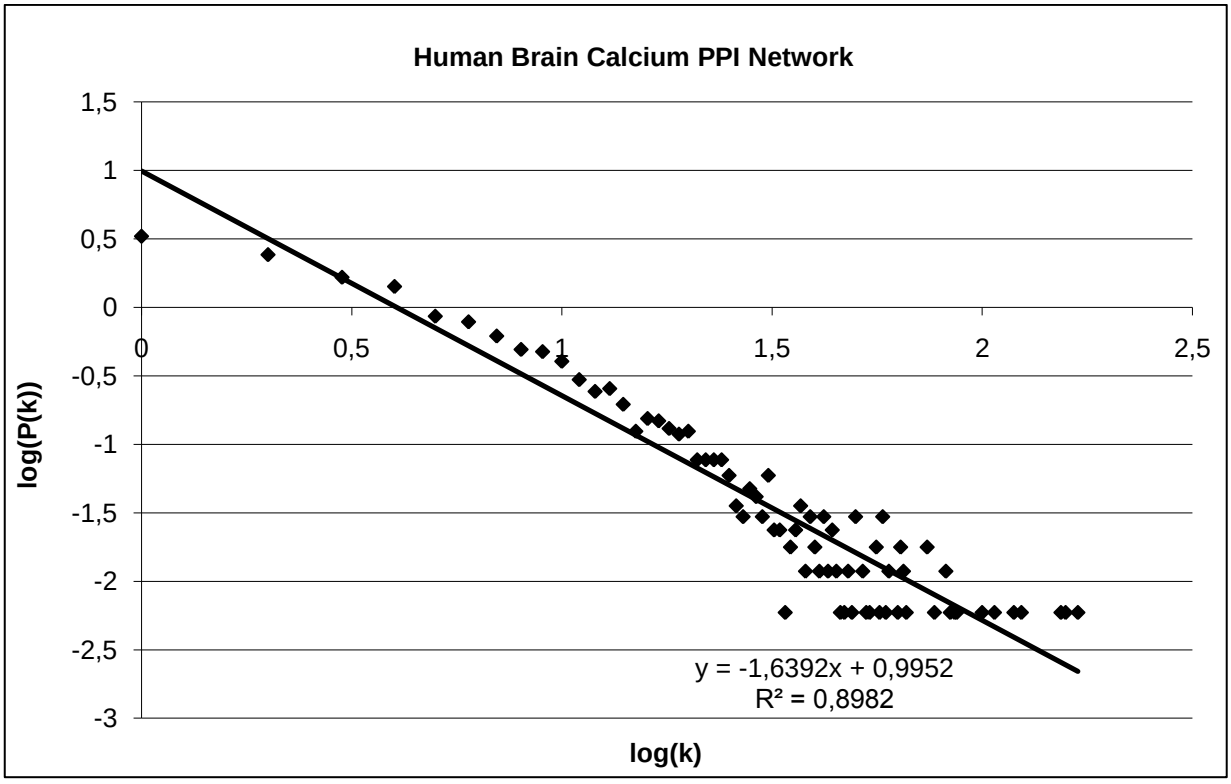


Figure 6.12. Log transform of degrees versus log transform of P(k) (probability of having connectivity, k) in human brain calcium signaling network

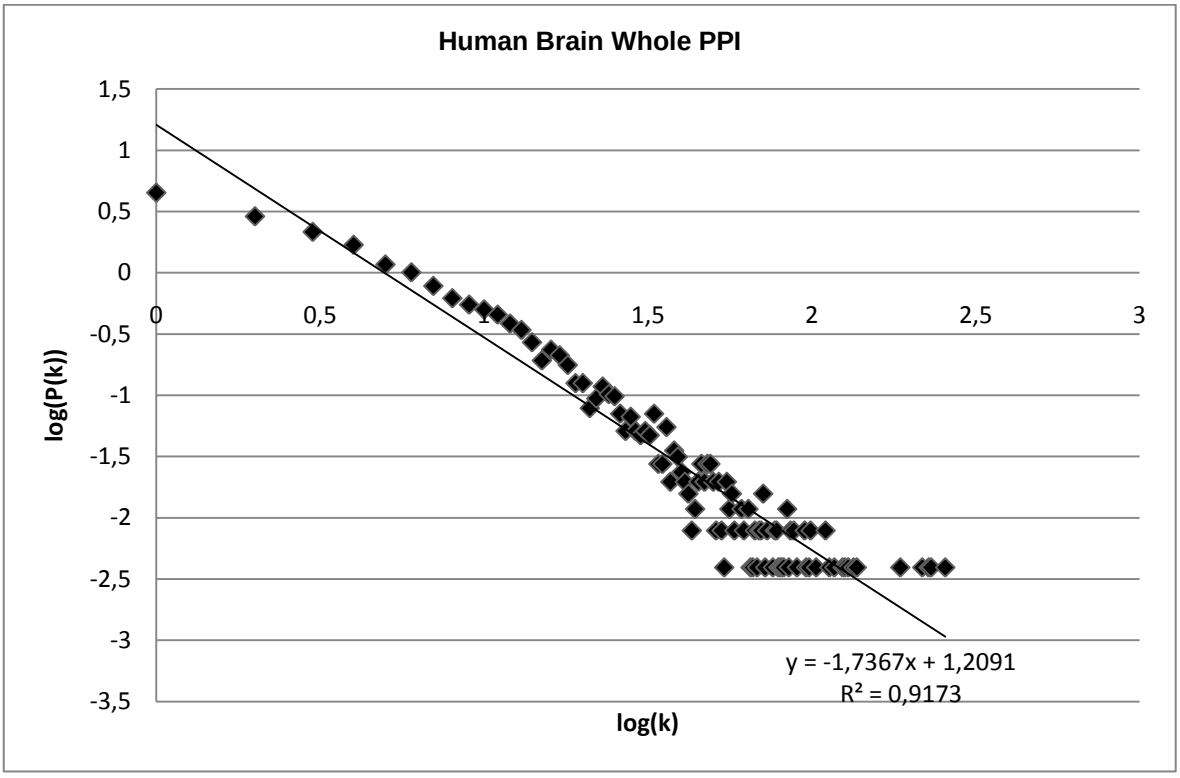


Figure 6.13. Log transform of degrees versus log transform of P(k) (probability of having connectivity, k) in human brain whole network

Scale-free property & power law: Both networks follow nearly a power law model ($P(k) \approx k^{-\gamma}$, γ=1.64 with R^2=0.90, γ=1.74 with R^2=0.92 for PINs resulted from 32 cores and whole brain interactome respectively), implying that networks can be characterized by a scale-free degree distribution with a general emergence of hubs, as many other biological networks. The exponent γ is almost the same between two networks indicating that the role of hubs does not differ considerably, 32 core calcium brain PIN being slightly more hub dominant than whole brain interactome PIN.

The network density is the normalized version of the degree and varies between 0 - 1. It shows how densely the network is populated with edges. Since, for both networks, it is close to 0, one can conclude that not every node has many edges, but some hubs. The network heterogeneity reflects the tendency of a network to contain hub nodes. Network heterogeneity of 1.59 and 1.74 supports the scale-free distribution, claiming that, the networks have tendency to contain hubs instead of a constant node/edge ratio.

When compared to other human networks for the power law exponent, the calcium signaling PIN in brain has one of the smallest γ values, implying that it is one of the PINs mostly dominated by hubs.

Small-world property, mean path length & diameter: The magnitude of log(nodes) for PINs of 32 cores and whole brain interactome respectively are 3,41 and 3,72 which are of the same order of magnitude as avg. path lengths (3,87 and 4.01) and diameters of the PINs (9 and 13) meeting the criteria for having small-world property.

Since the average path lengths of both networks are almost the same, it can be concluded that these networks have the same efficiency in the transfer of biological information. When compared to other networks found from literature, these networks are one of the most efficient networks in transfer of biological information, together with network of M.musculus and yeast calcium signaling PIN.

The diameter of a network essentially characterizes the network's interconnectivity. The diameter of calcium PIN is smaller than the whole brain PIN; so the calcium network has more interconnectivity than that of whole brain PIN. When compared to the diameters of networks found from literature, calcium PIN diameter is again the smallest being the most interconnected. The whole brain PIN, however, stays in between and also consistent with MDC-Y2H, CCSB-H1 and HOMOMINT networks.

Clustering coefficients & modularity: Both networks have the same values of clustering coefficients as well as HPRD human networks (Futschik et.al). When compared to other networks, they are on the average level implying the average level of redundancy and coherence.

It is seen that, most of the scale free networks listed in Table 6.5 and 6.6 are also highly modular like the networks studied.

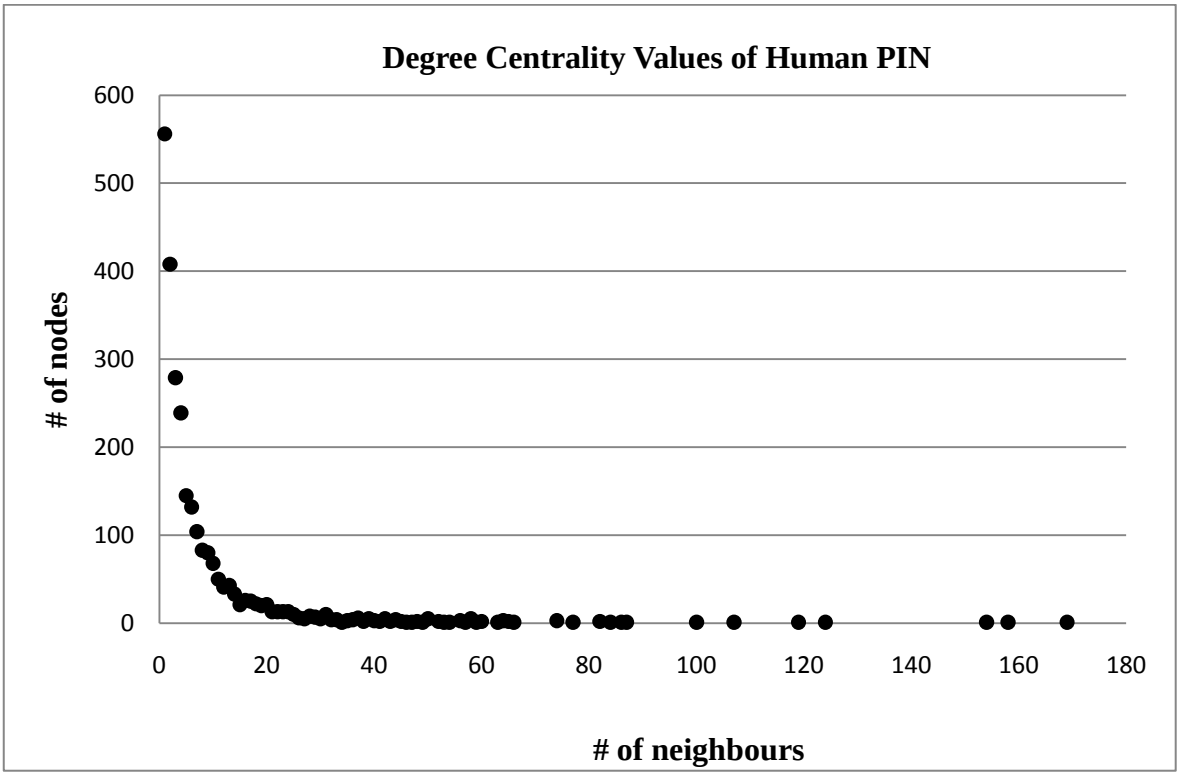


Figure 6.14. Degree Distribution of Human Brain Calcium Signaling PIN

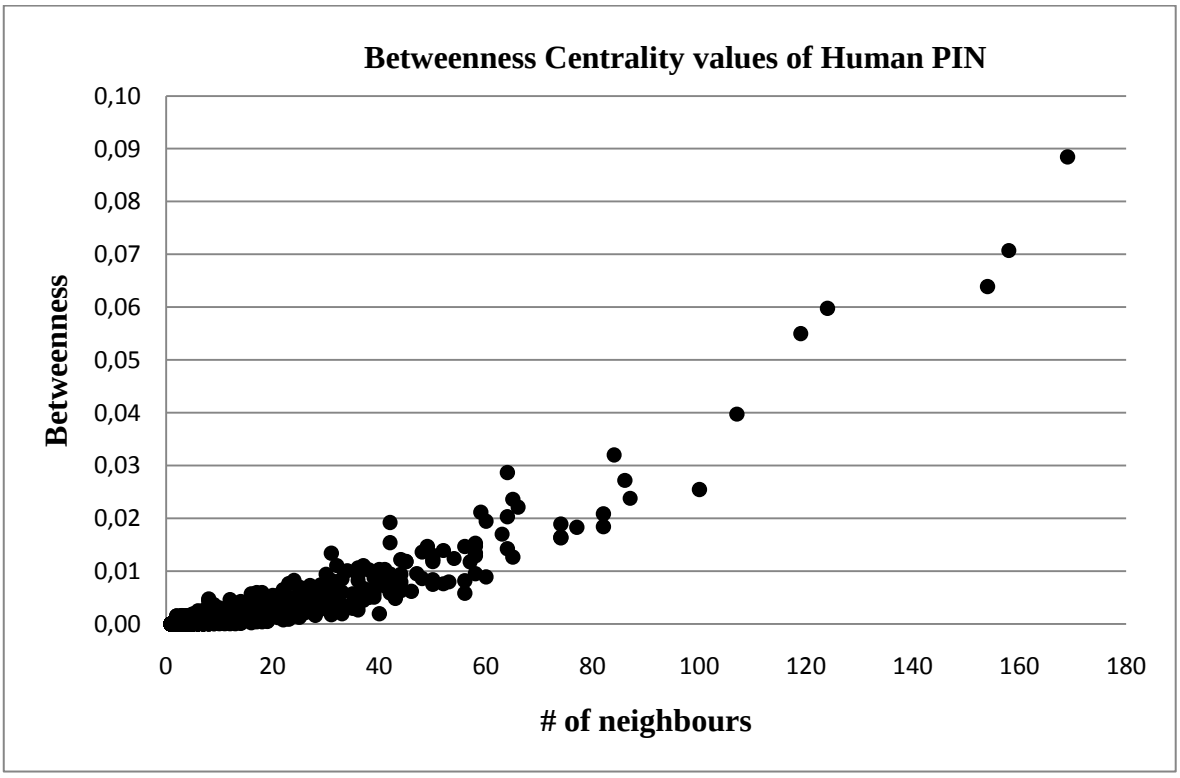


Figure 6.15. Betweenness Centrality Distribution of Human Brain Calcium Signaling PIN

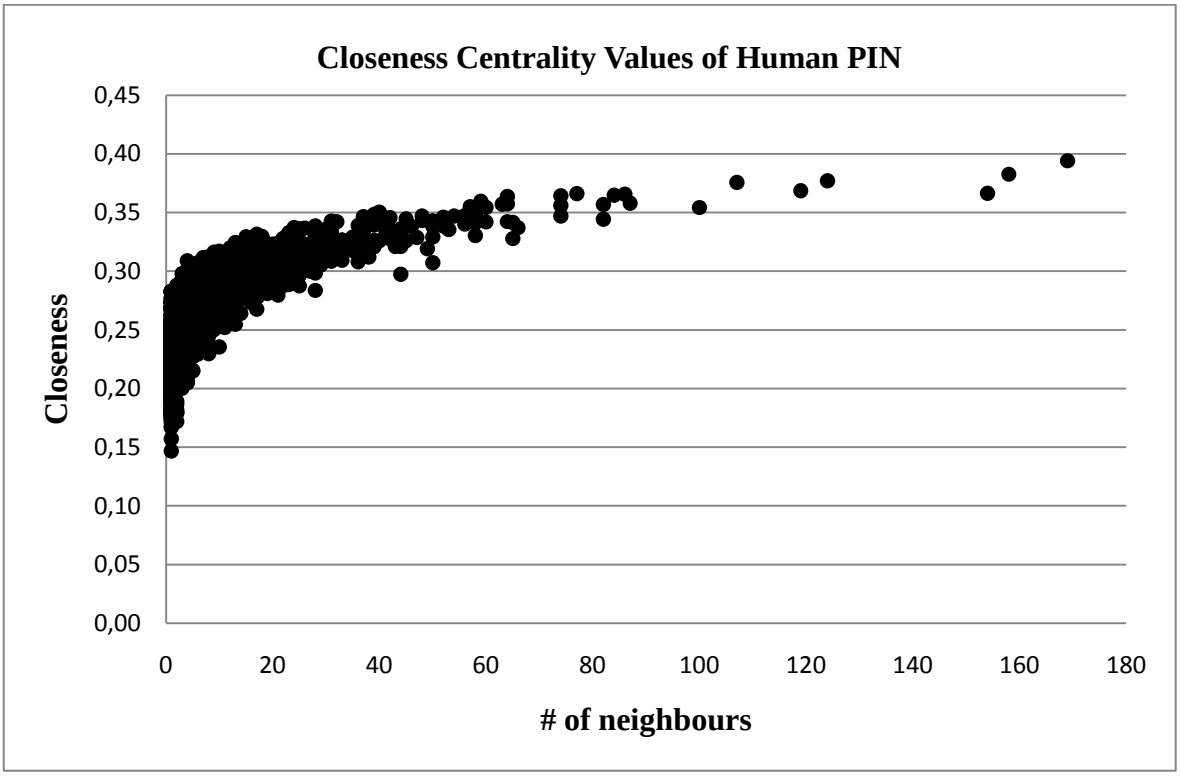


Figure 6.16. Closeness Centrality Distribution of Human Brain Calcium Signaling PIN

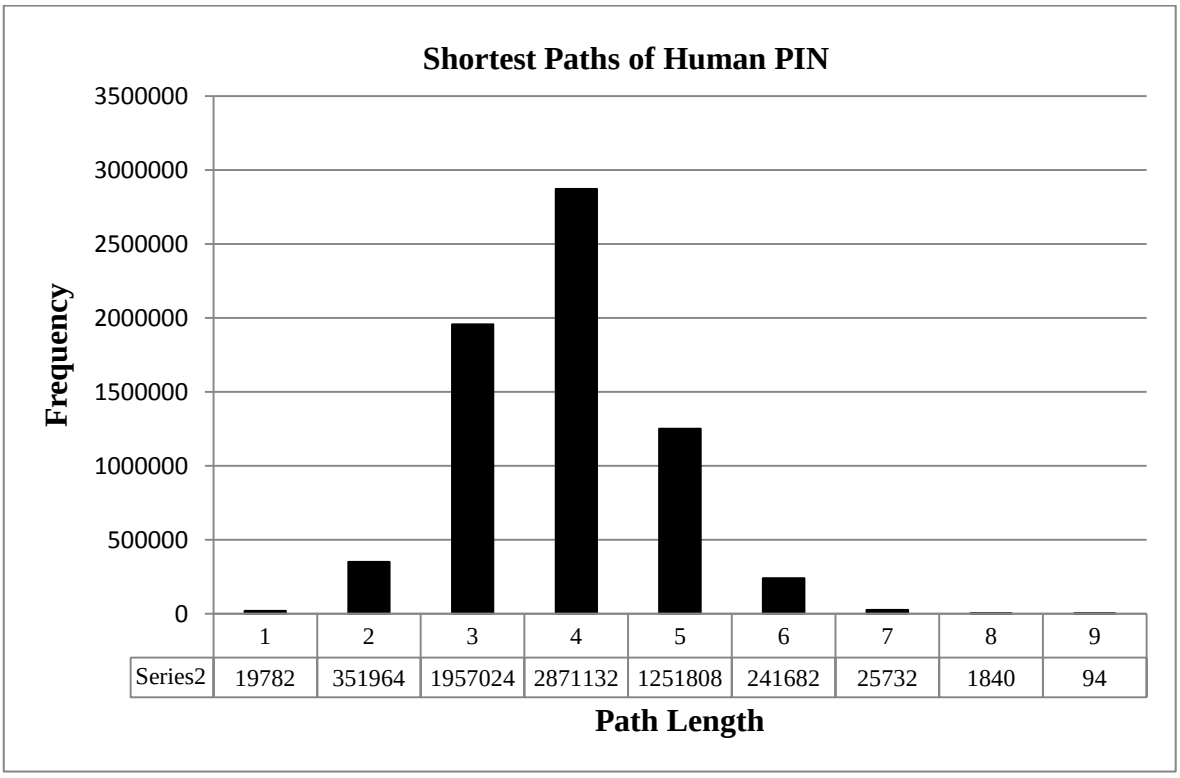


Figure 6.17. Shortest Path Distribution of Human Brain Calcium Signaling PIN

In human PIN, all the centrality parameters are consistent with the lethality-centrality rule and the scale-free property. In other words, there are many nodes with low connectivity and a few nodes with high connectivity, named hubs. Also the hub proteins tend to have the highest betweenness and closeness values within the network. There are no outliers observed as in the case of Joy *et al.* (2005).

In conclusion, both networks have almost the same topological properties although number of nodes and number of edges vary by a factor of 2,1 between networks.

Table 6.10. Topological properties of various networks found from literature

	# of nodes	# of edges	Mean connectivity (degree)	Mean path length	Network diameter	Clustering coefficient	Reference
S. cerevisiae (Ca signaling - 12 core)	1703	6299	7,4	3,82	10	0,16	Present study
S. cerevisiae (Ca signaling - 63 core)	3226	20167	12,5	3,43	9	0,15	Present study
S. cerevisiae - BIOGRID	5685	52141	18,3	2,78	6	0,24	Present study
C. elegans (Ca Signaling - 17 core)	1714	3174	3,7	4,78	12	0,03	Present study
C.elegans -BIOGRID & Intact	3741	7118	3,81	4,79	13	0,03	Present study
H. Sapiens (Ca Signaling in brain)	2593	9890	7,62	3,87	9	0,14	Present study
H. Sapiens - BIOGRID & HPRD & HUPA (Brain)	5222	20710	7,93	4,01	13	0,12	Present study
H. Sapiens (Ca Signaling in general) - KEGG	6309	27591	8,75	3,99	11	0,13	Present study
S.cerevisiae - core - DIP	2640	6600	5	5	13	0,33	Li *et al.*
S. cerevisiae	2115	2240	2,12	6,8	14		Jeong et.al.
S. cerevisiae	5253	65673	25	5,21	11		Arga et.al.
S. cerevisiae (signaling)	1363	3649	5,35	6,81	9		Arga et.al.
S.cerevisiae - DIP	4773	15444	6,5	5,01	13		Li *et al.*
S.cerevisiae - MIPS	2043	5434	5,32	7,71	16		Yook et.al
S.cerevisiae - DIP	5798	20098	6,93	4,9	12		Yook et.al
S.cerevisiae	3280	8868	5,41	4,9	13		Ito et.al.
D.melanogaster - DIP	7451	22819	6,2	4,4	11	0,02	Li *et al.*
C.elegans - DIP	2638	4030	3,1	4,8	14	0,06	Li *et al.*
H.pylori - DIP	710	1420	4	4,1	9	0,08	Li *et al.*
H.sapiens - DIP	1065	1369	2,6	6,8	21	0,21	Li *et al.*
E.coli - DIP	553	761	2,8	5,5	16	0,62	Li *et al.*
M.musculus - DIP	329	286	1,7	3,6	9	0,15	Li *et al.*

Table 6.11. Topological properties of various networks found from literature for H. Sapiens

H.sapiens networks	#of nodes	# of edges	Mean connectivity (degree)	Mean path length	Network diameter	Clustering coefficient	Power Law Constant (γ)	Method	
H. Sapiens (Ca Signaling in brain)	2593	9890	7,62	3,87	9	0,14	1,64		Present study
H. Sapiens - BIOGRID & HPRD & HUPA (Brain)	5222	20710	7,93	4,01	13	0,12	1,74		Present study
H. Sapiens (Ca Signaling in general) - KEGG	6309	27591	8,75	3,99	11	0,13	1,73		Present study
MDC-Y2H	1703	3186	3,7	4,9	13	0,01	1,63	Y2H-assay	Futschik *et al.*
CCSB-H1	1549	2754	3,5	4,4	12	0,05	1,46	Y2H-assay	Futschik *et al.*
HPRD	5908	15658	5,2	5,0	15	0,13	2,44	Literature	Futschik *et al.*
BIND	2677	4233	2,9	5,9	16	0,17	1,90	Literature	Futschik *et al.*
COCIT	3737	6580	3,5	5,9	20	0,43	2,18	Literature	Futschik *et al.*
OPHID	2284	8962	7,8	4,8	15	0,23	1,36	Orthology	Futschik *et al.*
ORTHO	3503	9641	5,4	6,5	17	0,19	2,14	Orthology	Futschik *et al.*
HOMOMINT	2556	5582	4,2	5,1	12	0,07	2,76	Orthology	Futschik *et al.*

6.3.2. Hub Proteins of Human Network

In human brain calcium signaling network constructed from 32 core proteins by SPA method, the top seven hubs (the nodes with the highest degree – Figure 6.18) are identified using the Cytoscape plugin cytohub. Those hubs have 923 interactions with 677 proteins in total covering 9,3 % of the interactions and 26,1 % of the proteins in PIN.

The hub proteins' dominant molecular function and biological process terms are; *protein binding, double stranded DNA binding* and *ATP binding* taking place in the biological processes of *cell proliferation and Ras protein signal transduction* which possibly occur in *cytosol or cytoplasm.* Since activation of Ras signalling causes cell proliferation and differentiation; dysregulated Ras signaling can ultimately lead to cancer.

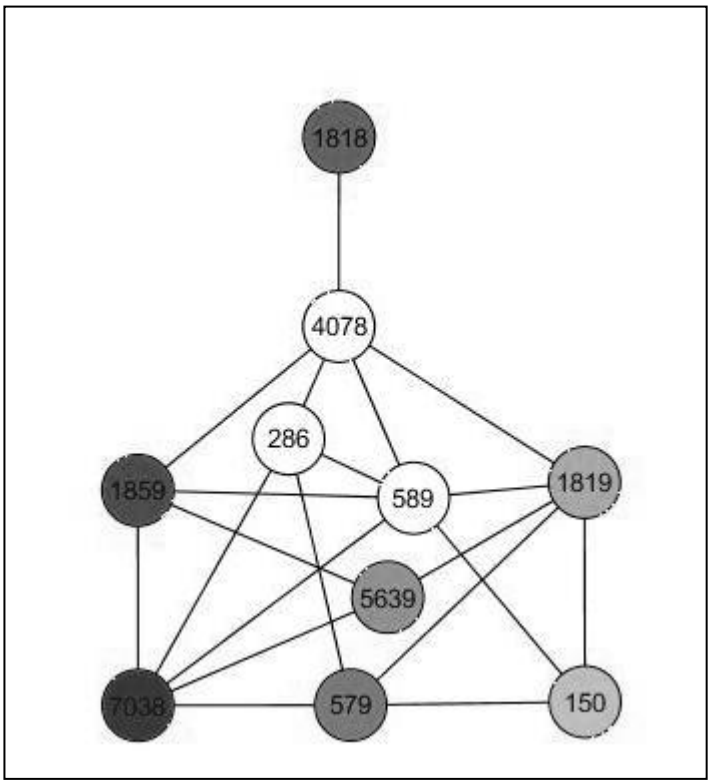


Figure 6.18. Top 10 Hub proteins of Human PIN

The white ones in Figure 6.18 are not considered as hub proteins since there was a gap between the number of interactions of the top 7 and that of the last 3 (white ones).

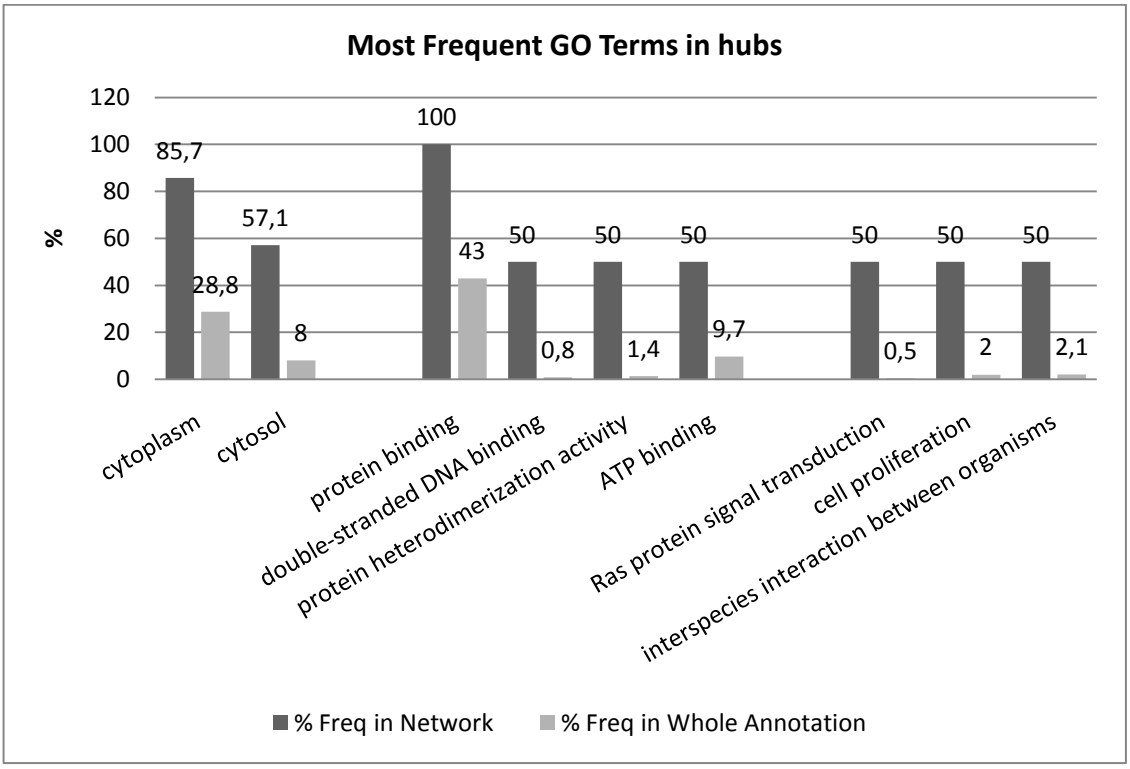


Figure 6.19. Predominant GO terms of hub proteins of Human PIN

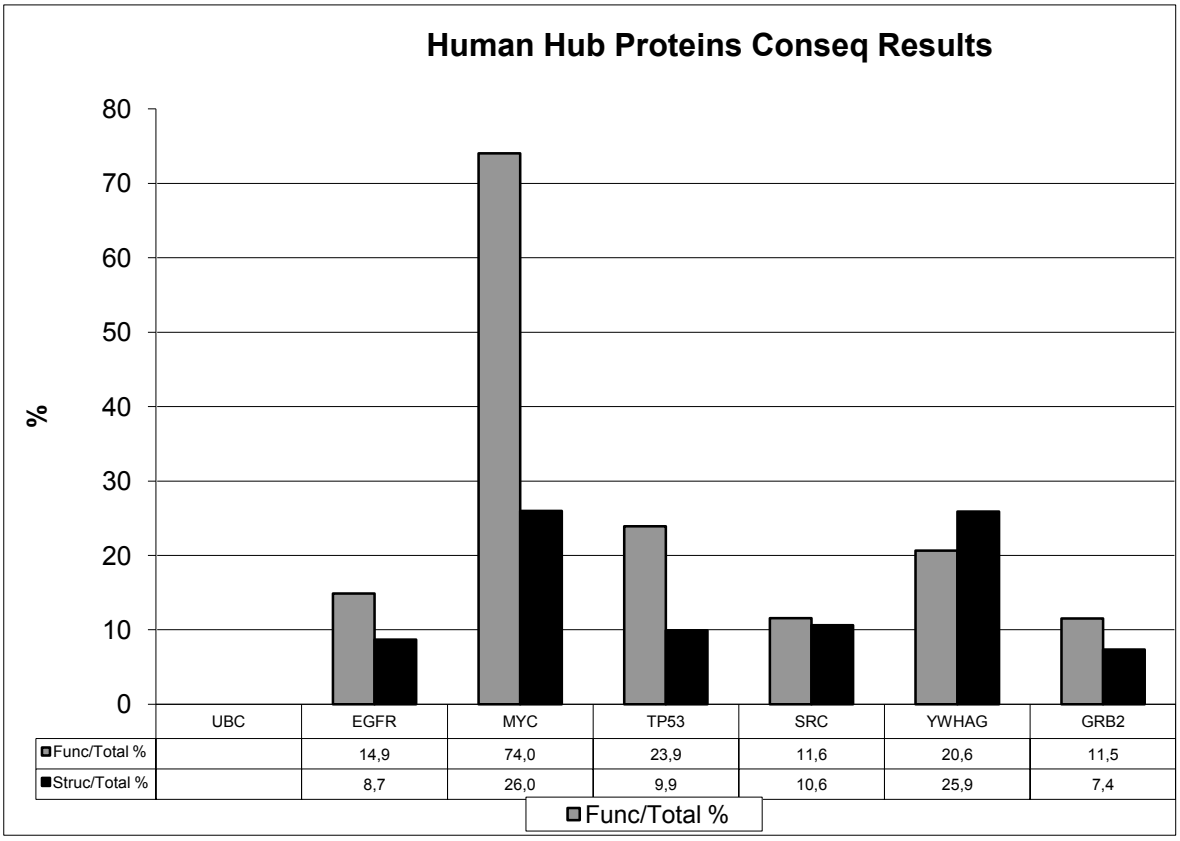


Figure 6.20. ConSeq Results of hub proteins of human PIN

In general, the number of functionally and structurally important residues is equally distributed in the hub proteins. MYC is deviated from this generalization, having functionally imported residues in vast majority. For UBC, no results can be drawn since there is not enough number of hits in multiple sequence alignment.

The EFGR protein domains are conserved with an average of 35 %. Hlh domain in MYC is the most conserved domain (75 %). Also, p53 domain in TP53 is the second most conserved domain (60.3 %) among hub protein domains. YWHAG's domain is 56.1 % conserved. GRB2 domains have the lowest conservation scores.

The conservation scores of proteins and their domains are consistent with each other.

Table 6.12. Hub proteins of the Human PIN

HPRD id	name	ints in PIN	ints in whole brain PIN	Uniprot	aa	PDB id	Domains	Description
7038	UBC	169	263	P62988	76		Ubiquitin	Ubiquitin C; protein modifier which can be covalently attached to target lysines either as a monomer or as a lysine-linked polymer
1859	TP53	158	295	P04637	393	1TSR 94-312	p53_TAD p53 p53_tetramer	Cellular tumor antigen p53; acts as a tumor suppressor in many tumor types by inducing growth arrest or apoptosis depending on the physiological circumstances and cell type
1818	MYC	154	279	P01106	439	1NKP 353-434	Myc_N HLH Myc_LZ	Myc proto-oncogene protein; participates in the regulation of gene transcription
579	EGFR	124	218	P00533	1210	1IVO 25-646 3GOP 669-1022	2 Recep_L_domain Furin_like Pkinase_Tyr	EGF Receptor; involved in the control of cell growth and differentiation
5639	YWHAG	119	290	P61981	247	2B05	14-3-3	Protein kinase C inhibitor protein 1; Adapter protein in the regulation of both general and specialized signaling pathway
1819	SRC	107	169	P12931	536	2H8H	SH3_1 SH2 Pkinase_Tyr	Proto-oncogene tyrosine-protein kinase Src
150	GRB2	100	184	P62993	217	1GRI	2 SH3_1 SH2	Growth factor receptor-bound protein 2; provides a critical link between cell surface growth factor receptors and the Ras signaling pathway

Table 6.13. Conservation Scores Hub proteins and their domains

HPRD id	Protein name	Tot. Res.	Cons %	Domain Cons. in Cons. Residues	Exposed %	Buried %	Func/conserved %	Domain F/Cons F %	Struc/conserved %	Domain S/Cons S %
7038	UBC	76	-	-	-	-	-	-	-	-
	Ubiquitin	68								
579	EGFR	1210	33,3		62,2	37,8	44,7		26,1	
	Recep_L_domain	111	43,2	11,9	55,0	45,9	27,1	7,2	39,6	18,1
	Furin_like	161	31,1	12,4	73,3	27,3	48,0	13,3	4,0	1,9
	Recep_L_domain	120	30,0	8,9	44,2	56,7	33,3	6,7	44,4	15,2
	Pkinase_Tyr	256	58,2	37,0	46,5	53,9	36,2	30,0	38,3	54,3
1818	MYC	439	35,3		74,0	26,0	65,2		25,8	
	Myc_N	344	30,5	67,7	75,0	25,3	64,8	67,3	25,7	67,5
	HLH	52	75,0	25,2	59,6	42,3	64,1	24,8	25,6	25,0
	Myc_LZ	31	25,8	5,2	87,1	16,1	62,5	5,0	37,5	7,5
1859	TP53	393	37,7		74,6	25,4	63,5		26,4	
	p53_TAD	24	37,5	6,1	70,8	33,3	44,4	4,3	44,4	10,3
	p53	194	60,3	79,1	69,1	36,1	60,7	75,5	29,1	87,2
	p53_tetramer	41	22,0	6,1	70,7	31,7	66,7	6,4	11,1	2,6
1819	SRC	536	30,2		67,5	32,5	38,3		35,2	
	SH3_1	47	23,4	6,8	72,3	29,8	45,5	8,1	18,2	3,5
	SH2	82	29,3	14,8	65,9	35,4	29,2	11,3	33,3	14,0
	Pkinase_Tyr	249	45,4	69,8	54,2	46,2	36,3	66,1	40,7	80,7
5639	YWHAG	247	54,3		57,9	42,1	38,1		47,8	
	14-3-3	237	56,1	99,3	56,5	43,9	37,6	98,0	48,1	100,0
150	GRB2	217	23,5		66,8	33,2	49,0		31,4	
	SH3_1	46	23,9	21,6	73,9	28,3	54,5	24,0	27,3	18,8
	SH2	75	32,0	47,1	64,0	37,3	29,2	28,0	54,2	81,3
	SH3_1	45	17,8	15,7	64,4	37,8	87,5	28,0	0,0	0,0

6.3.3. Module Detection and Proteins with Unknown Function

Among human proteins expressed in brain and brain regions, there are 38 proteins with no functional assignment at all; to which “molecular function unknown” term is given manually. Also there are 309 proteins with “molecular function unknown” term given in Gene Ontology for proteins expressed in brain. Out of these total 347 proteins, 60 of them are present in the reconstructed human brain calcium signaling PIN.

Using MCODE, a Cytoscape plugin, only one module is determined containing 1 functionally unknown protein (details and figures can be found in App.) The module selection criterion depends on the number of functionally known protein dominance in the mentioned module/subgroup interacting with the functionally unknown protein. However, the obtained module is not appropriate for function assignment to functionally unknown proteins since no dominance of known function proteins can be determined.

6.4. Neurodegenerative Disorders and Human PIN

The neurodegenerative disease pathways are conducted from KEGG. Since KEGG gives the group of proteins, and not the individual proteins, in the pathway maps; the proteins under the general group names are tabulated. Then, they are checked whether they are expressed in brain or not (Table 6.9). The final procedure is to search and identify the disease related proteins expressed in brain among the human PIN.

Table 6.14. Disease related proteins in KEGG and PIN

	Alzheimer	Parkinson	Huntington	ALS
# of proteins in KEGG disease network	159	104	158	53
# of proteins expressed in brain&brain regions in KEGG disease network	106	63	109	44
# of human PIN proteins in KEGG disease network	66	30	55	33
% of human PIN proteins in KEGG disease network	62,3	47,6	50,5	75,0

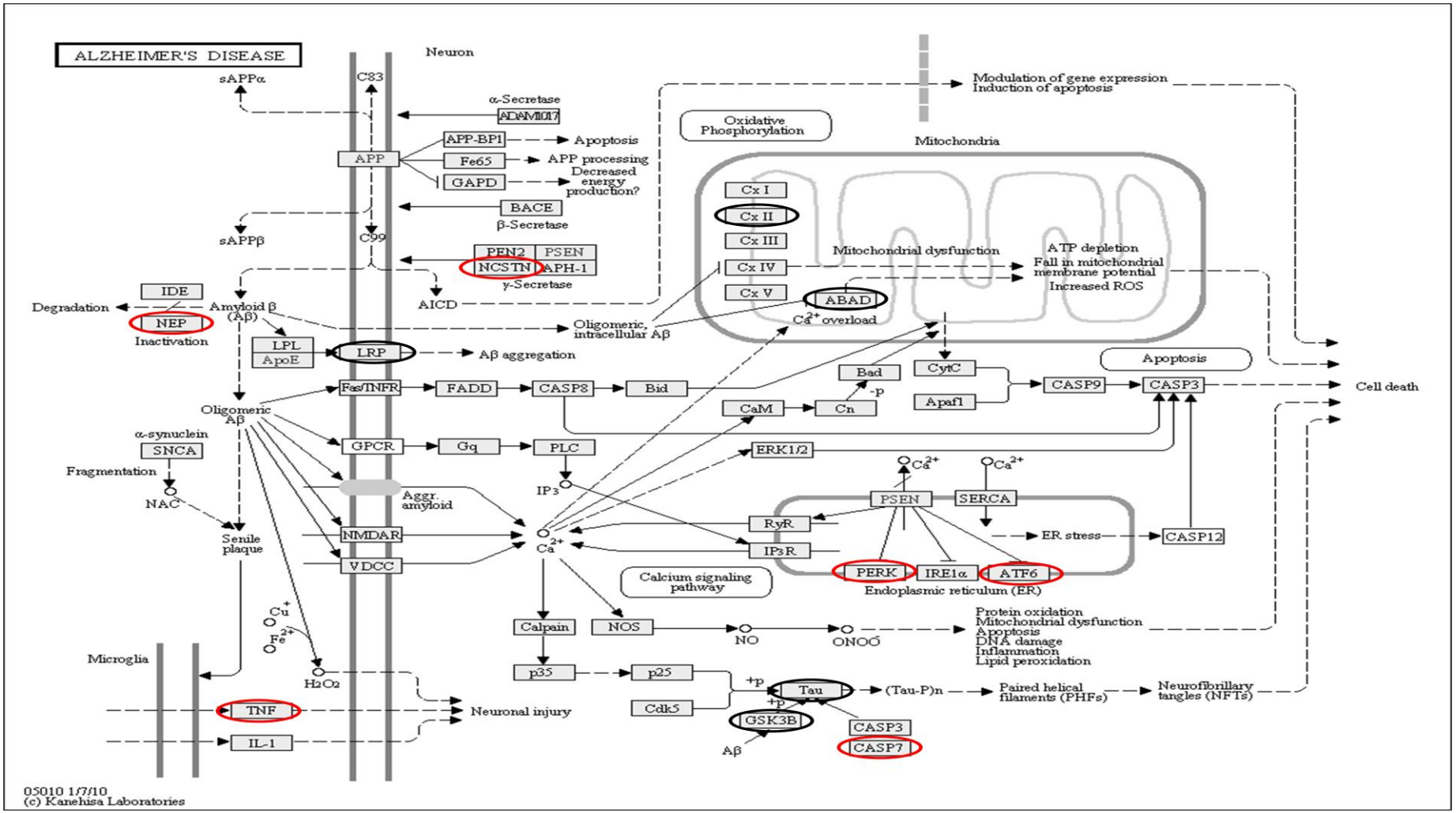


Figure 6.21. Alzheimer's Disease Pathway (taken from KEGG)

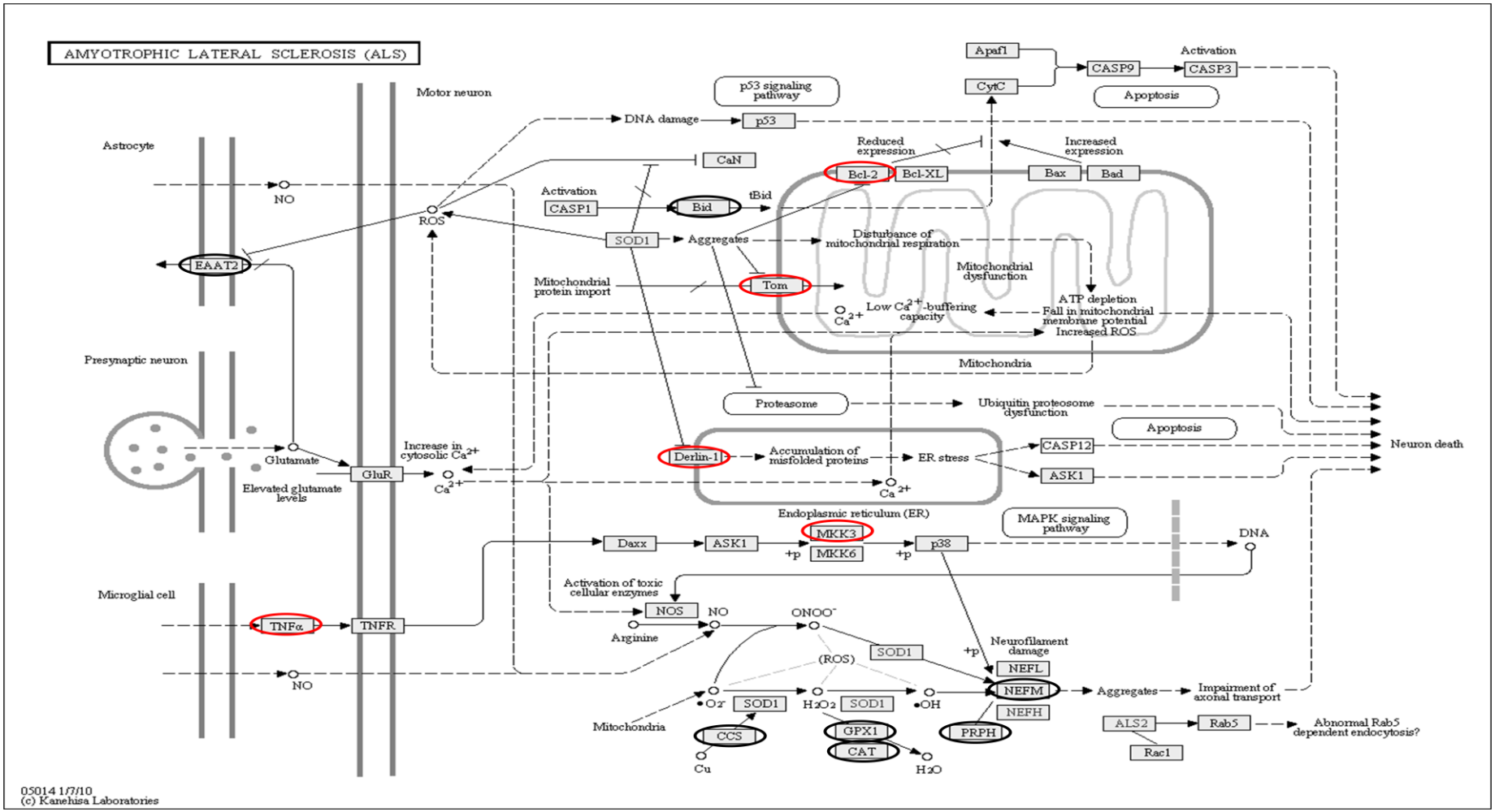


Figure 6.22. Amyotrophic Lateral Sclerosis Pathway (taken from KEGG)

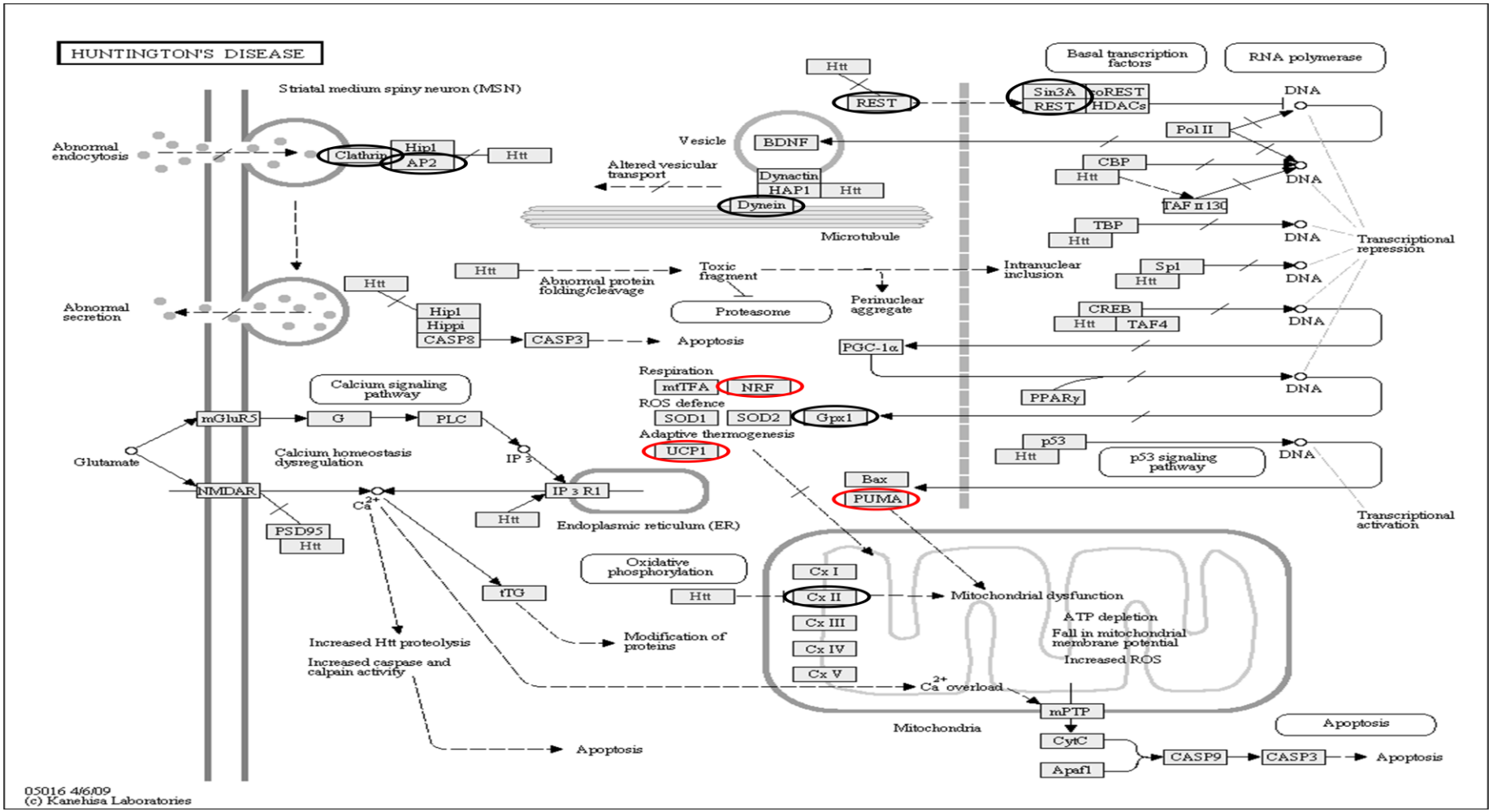


Figure 6.23. Huntington's Disease Pathway (taken from KEGG)

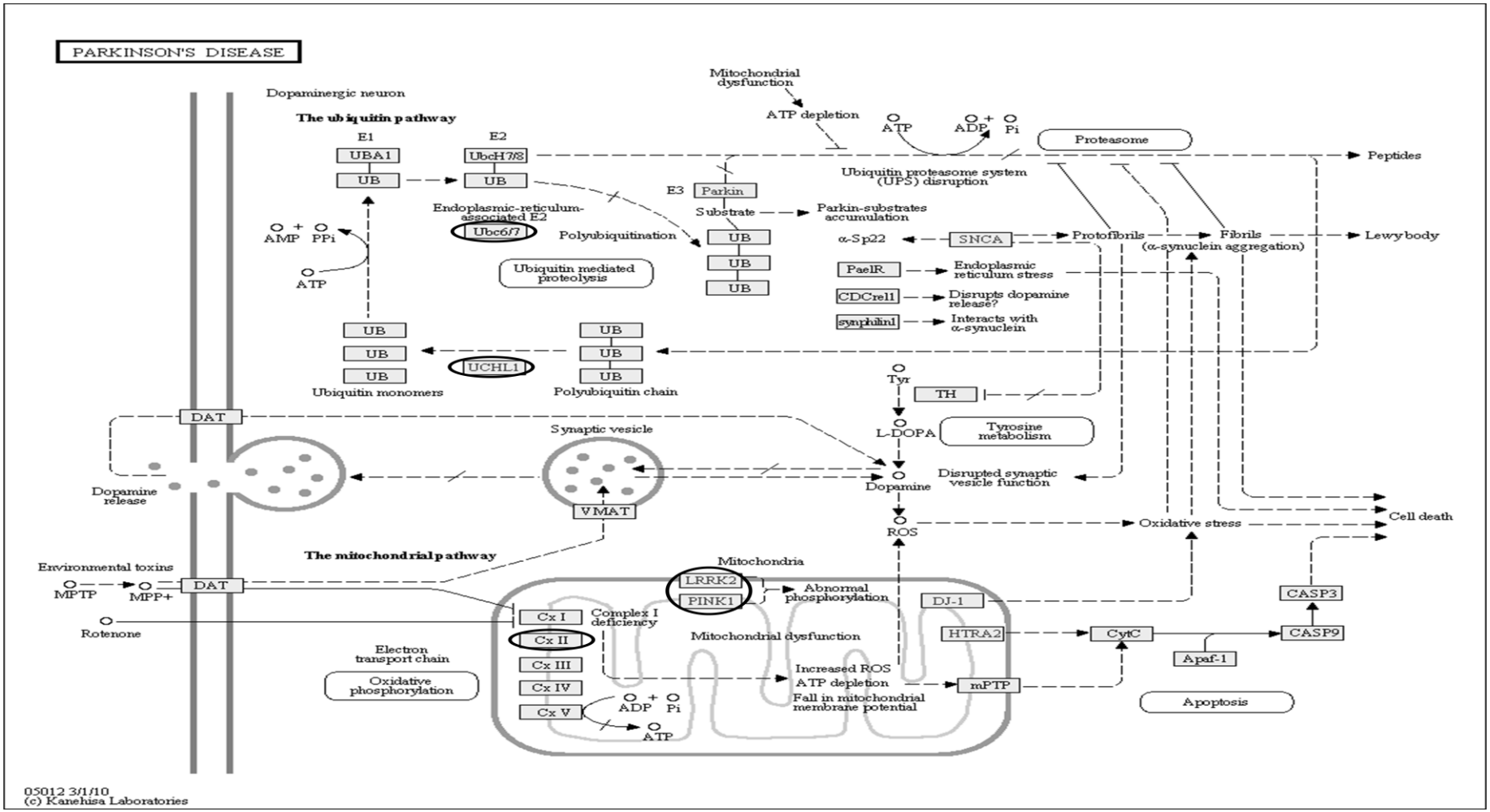


Figure 6.24. Parkinson's Disease Pathway (taken from KEGG)

The proteins with red circle are the ones that are not expressed in brain according to data obtained from HPRD and HuPa. The proteins with black circle are expressed in brain, but not found in reconstructed human PIN. In fact, it is expected that the proteins having a role in the diseases related to brain should be expressed in brain or brain regions. However, the presence of red circles in the figures (indicating that they are not expressed in brain according to the data obtained from HPRD and HuPa) can only be explained by the lack of expression data for these proteins.

In general, half of the proteins in disease maps are covered in reconstructed human PIN. ALS is the most and Parkinson Disease is the least covered diseases with 75 % and 47.6 % respectively.

The common proteins in disease pathways are mitochondrial cytochrome proteins under the groups Cx I, Cx II, Cx III, Cx IV and Cx V. The proteins in Cx II are not covered in human PIN.

In Alzheimer's disease, Amyloid-beta (Abeta), a major component of senile plaques, has various pathological effects on cell and organelle function. The key protein amyloid beta protein and APP precursor protein are found in human PIN. Also the three genes that may be linked to autosomal dominant or familial early onset AD (FAD), namely presenilin 1 (PS1), presenilin 2 (PS2) and apolipoprotein E (ApoE) are also contained in human PIN. However, tau protein is not among the human PIN proteins.

Mutations in alpha-synuclein (SNCA), UCHL1 (an ubiquitin carboxy-terminal hydrolase L1), parkin, DJ1 (a parkin-associated protein involved with oxidative stress), and PINK1 (a putative serine threonine kinase) are known to cause early-onset PD. Parkin protein, SNCA, DJ1 and DAT (dopamine transporter) are among the ones found in human PIN. However, UCHL1 and PINK1 are not covered.

HD is caused by a CAG repeat expansion in the IT15 gene, which results in a long stretch of polyglutamine close to the amino-terminus of the HD protein huntingtin (Htt) which is also in the human PIN. Mutant Htt (mHtt) has effects both in the cytoplasm and in

the nucleus. In the cytoplasm, full-length mHtt can interfere with BDNF (present in the PIN) vesicular transport on microtubules.

This mutant protein also may lead to abnormal endocytosis and secretion in neurons, because normal Htt form a complex with the proteins Hip1, clathrin and AP2 that are involved in endocytosis. Among these three proteins in endosyctosis, only Hip1 is found in human PIN. In addition, mHtt affects Ca signaling by sensitizing IP3R1 to activation by InsP3, stimulating NR2B/NR1 NMDAR activity, and destabilizing mitochondrial Ca handling. This cascade is fully covered in human PIN. As a result, stimulation of glutamate receptors leads to supranormal Ca responses in HD and mitochondrial Ca overload.

Mutant superoxide dismutase 1 (SOD1), as seen in some familial amyotrophic lateral sclerosis (ALS) cases, may be toxic because it is unstable, forming aggregates in the motor neuron cytoplasm, axoplasm and mitochondria. Within mitochondria, mutant SOD1 may interfere with the anti-apoptotic function of Bcl-2, affect mitochondrial import by interfering with the translocation machinery (TOM/TIM), and generate toxic free radicals (ROS) via aberrant superoxide chemistry. Although SOD1 is present in human PIN, the rest of the cascade is not covered in human PIN.

Reactive oxygen species (ROS), produced within mitochondria, inhibit the function of EAAT2, the main glial glutamate transporter protein, responsible for most of the reuptake of synaptically released glutamate. Glutamate excess causes neurotoxicity by increasing intracellular calcium, which enhances oxidative stress and mitochondrial damage. However, since the Bcl-2 and TOM/TIM machinery are absent in human PIN, EAAT2 is also absent.

Mutant SOD1 can also trigger oxidative reactions by various means including by increasing levels of peroxynitrite, which can then cause damage through the formation of hydroxyl radicals or via nitration of tyrosine residues on proteins. Nitration may target neurofilament proteins, disrupting their phosphorylation and affecting axonal transport. Collectively, these mechanisms (or a combination thereof) are predicted to disturb cellular homeostasis (within glial and/or motor neurons), ultimately triggering motor neuron death. This procedure is covered in human PIN as well.

6.5. Comparison of Calcium Signaling in Human Brain and in Human in General

6.5.1. Comparison of Core Proteins & Resultant Protein Interaction Network

In order to find out the differences as well as similarities between the calcium signaling in human brain and in general, the human interactome data is filtered with respect to the tissue expressions; i.e. the proteins that are expressed only in brain are not included in the interactome data since the aim is to compare the different aspects of these two PINs. a PIN is reconstructed starting with the 169 core proteins that are listed in KEGG under the pathway of "Calcium signaling in Human". In fact there are 172 core proteins in KEGG pathway; however 3 of them were only expressed in brain so they are excluded. The annotation table of 169 core proteins consists of 139 cellular component terms, 252 molecular function terms and 856 biological process terms.

The reconstructed network contains 6309 nodes with 27591 undirected physical edges after 7 runs. 2538 nodes are common with brain PIN, whereas 57 nodes that are present in brain PIN are not present in general calcium signaling PIN. Out of these 57 proteins, only 2 are expressed in only brain regions; others are expressed in various tissues apart from brain. One of them is the Dopamine receptor protein DRD5 whose defects are a cause of benign essential blepharospasm (BEB). BEB is a primary focal dystonia affecting the orbicularis oculi muscles leading to blindness in severe cases. The other one is HTR6, being Functions of -HT6 receptors include modulation of cholinergic and dopaminergic neurotransmission, and they have a role in spatial learning and memory.

When compared with the enriched terms of calcium signaling in brain, the most dominant terms are in agreement with each other (Figure 6.25-6.27). However, while "membrane" and "integral to membrane" terms are among the enriched cellular component terms of brain calcium signaling, these terms are rare in general calcium signaling in human despite of the fact that "plasma membrane" term is present in both. Also, it is observed that the same situation is present for the "metal ion binding" term in molecular function. The "regulation of transcription" term in general calcium signaling is more dominant than "apoptosis" term in general calcium signaling, whereas the opposite case is valid in calcium signaling in brain.

In addition, the enriched terms of 57 proteins reveal that they are the proteins mainly responsible for synaptic transmission, glutamate signaling pathway and transport; i.e. the neurotransmitter release mechanism.

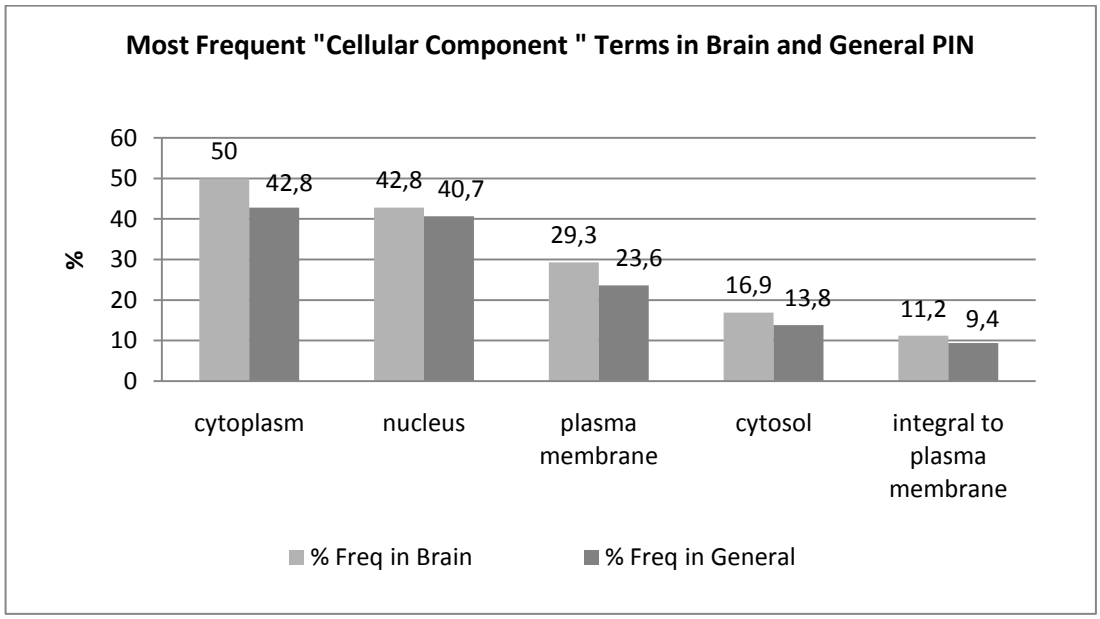


Figure 6.25. Comparison of Predominant "Cellular Component" terms in Human PIN in general and in brain

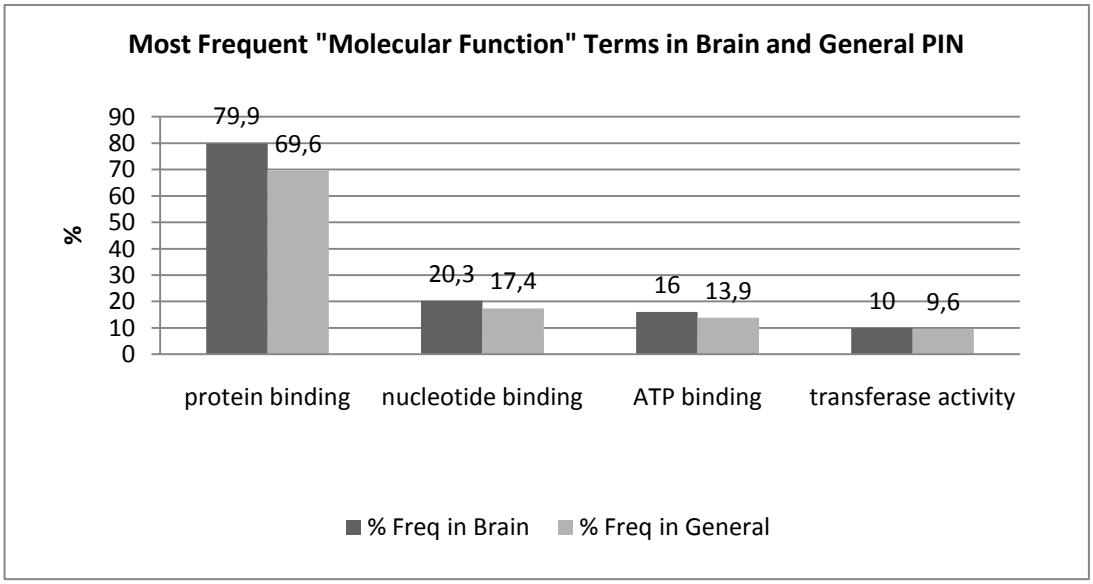


Figure 6.26. Comparison of Predominant "Molecular Function" terms in Human PIN in general and in brain

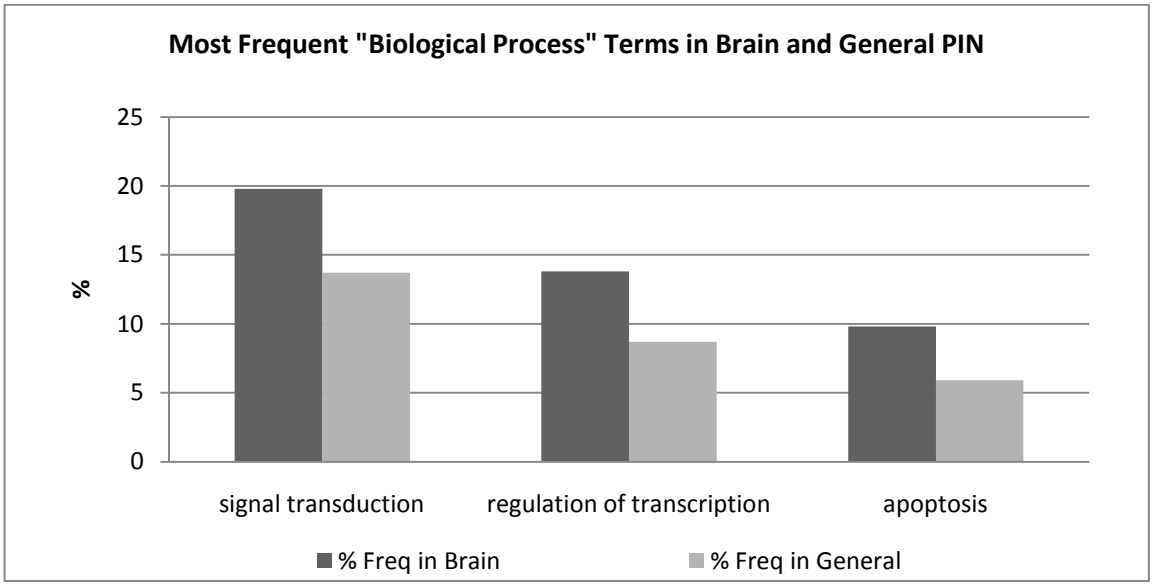


Figure 6.27. Comparison of Predominant "Biological Process" terms in Human PIN in general and in brain

6.5.2. Topology Based Comparison of General PIN and Brain PIN

In order to analyze and compare topological properties of general calcium signaling in human, the undirected protein interaction graph is analyzed using Network Analysis algorithm, a Cytoscape plugin. Also the hub proteins are determined and it is observed that general calcium signaling PIN and calcium signaling PIN in brain share the same hubs. Only the degrees of the hubs differ slightly.

The topological properties of the general calcium signaling PIN do not differ from the calcium signaling in brain (Table 6.9). They both have scale free properties (presence of hubs) as they follow power law well (R^2=0.92 and 0.90 respectively). But the PIN of calcium signaling in brain's γ value is the smallest (1.64 compared to 1.73 for calcium signaling in general PIN and 1.74 for whole brain interactome), implying that it is the most hub dominated PIN.

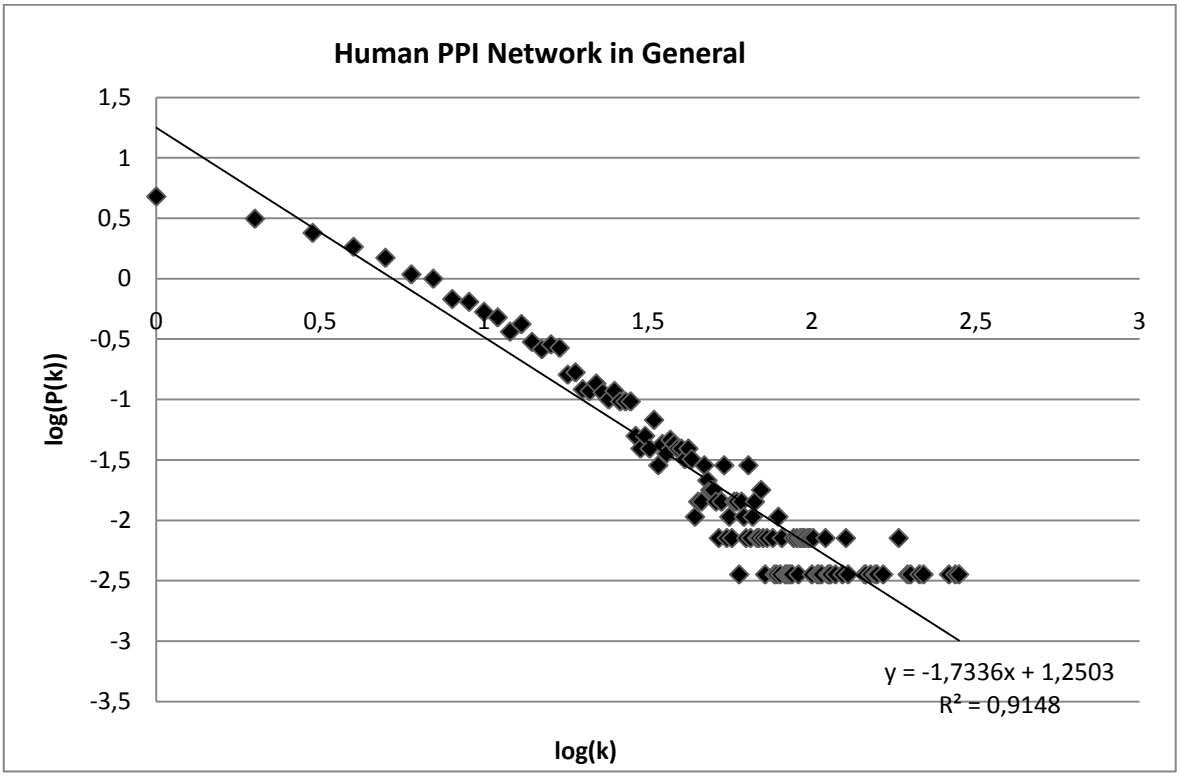


Figure 6.28. Log transform of degrees versus log transform of P(k) (probability of having connectivity, k) in general calcium signaling network in human

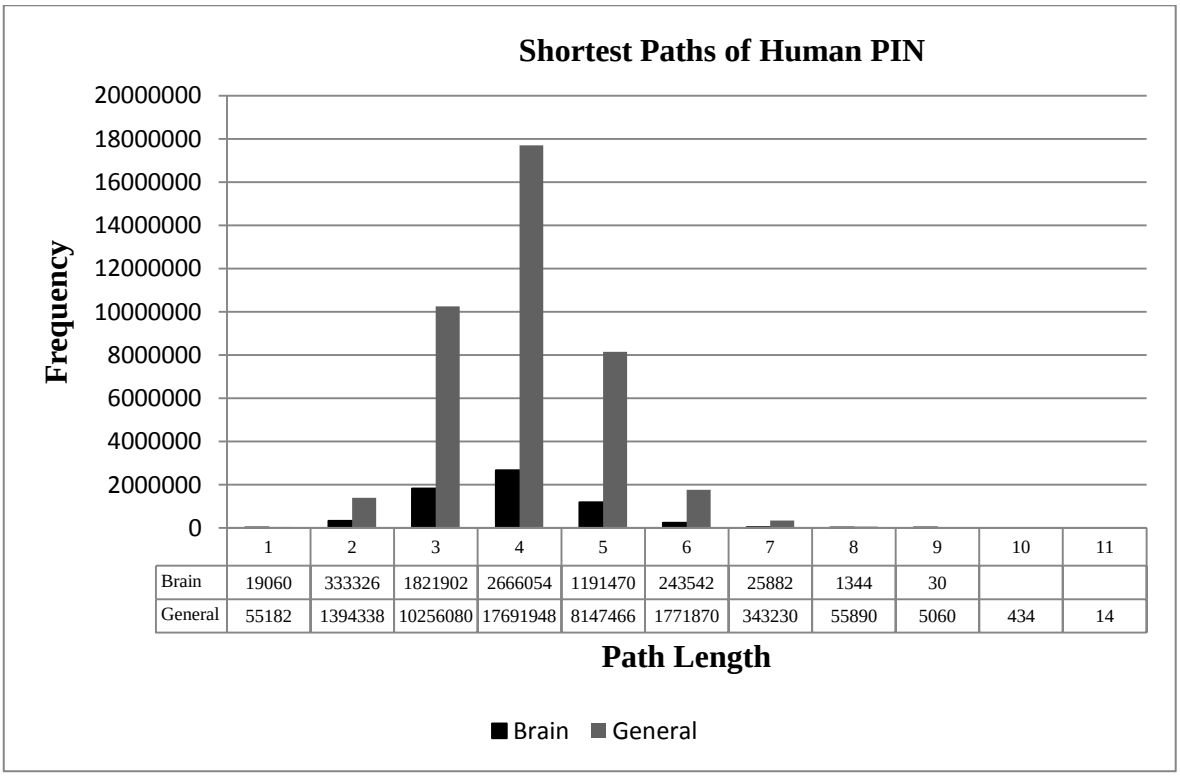
The general calcium signaling PIN is the least dense and most heterogeneous PIN when compared to the calcium signaling in brain as well as the whole brain interactome. Since its network size is bigger than the one for brain, the general calcium signaling pathway's hubs have higher degree than the brain although the hubs are common.

All three networks have small world property. But the mean path length comparison reveals that, the calcium signaling in brain is the fastest among them. This is an expected outcome since brain, apart from the other parts of the body, controls many processes throughout the body and also in itself. So the signal transmission should be as fast as possible to function properly and manipulate the body actions.

Table 6.15. Topological properties of calcium signaling in brain, in general and the whole brain interactome

HUMAN	*PIN Calcium Signaling in General*	*PIN Calcium Signaling in Brain*	*Whole Brain Interactome*
# of nodes	6309	2593	5124
# of edges	27591	9890	20710
Mean connectivity (degree)	8,75	7,62	7,93
Mean path length	3,99	3,87	4,01
Network diameter	11	9	13
Network density	0,001	0,003	0,002
Network heterogenity	1,81	1,59	1,74
Clustering coefficient	0,13	0,14	0,12
Network centralization	0,04	0,06	0,05
Shortest paths	39721512	6721058	25265856
Power Law Constant (g)	1,73 (R^2:0.92)	1,64 (R^2:0.90)	1,74 (R^2:0.92)
# of isolated & lonely nodes	28	3	98
Total # of proteins	6337	2096	5222

The shortest paths of general calcium signaling and that of brain are both accumulated at a step size of 4. Also, it has been observed that the shortest paths of general calcium signaling are much more than that of brain. This can be due to the fact that general calcium signaling covers all the shortest paths occurring in various regions of the body. The diversity of the regions and processes leads to a higher number of shortest paths.



	1	2	3	4	5	6	7	8	9	10	11
Brain	19060	333326	1821902	2666054	1191470	243542	25882	1344	30		
General	55182	1394338	10256080	17691948	8147466	1771870	343230	55890	5060	434	14

Figure 6.29. Comparison of Shortest path distributions

7. NETWORK COMPARISON

7.1. Topological Aspects

All three calcium signaling networks obtained by SPA algorithm, namely in yeast, worm and human brain, as well as the whole interactomes for these species are compared with respect to their topological characteristics (Table 7.1).

In general, yeast genome interactome is found to be the most robust since its corresponding average path length is the smallest one among all 7 PINs. Also having the highest clustering coefficient and lowest diameter, this PIN is more modular and higher interconnected than other 7 PINs. When the power law constant, γ, is considered; this network is most hub dominated one since it has the lowest γ value.

On the other hand, it was expected that the information transfer within network should speed up in parallel with evolution. In other words, human being the ultimate complex species should be more robust (having the lowest average path length) and interconnected (having the smallest diameter) than the other two. However, instead of following an increasing order of yeast-worm-human with respect to robustness and interconnectivity; the order is as worm-human-yeast. This outcome can be explained by the fact that yeast is a much simpler organism than human and worm. So the signals can be transferred faster within the organism. Also, it should be reminded that the concerning human interactome is not the whole human interactome, but only the one in brain. This regional discrimination may affect the topological values.

It has been observed that both yeast and human PINs and interactomes are in accordance with the data found in literature (Table 6.5). However, worm PIN does not seem to meet the expected outcomes at all. In this study, it has been observed that, worm genome is the least complete one containing so many hypothetical and undefined proteins when compared to that of human and yeast. This drawback may lead to these improbable outcomes.

Table 7.1. Topological parameters of Yeast, Worm and Human Networks

	HUMAN		WORM		YEAST		
	PPI Network	Whole Brain Interactome	PPI Network	Whole Genome	PIN 63 core	PIN 12 core	Whole Genome
# of nodes	2593	5124	1714	3741	3226	1703	5685
# of edges	9890	20710	3174	7118	20167	6299	52141
Mean connectivity (degree)	7,62	7,93	3,7	3,81	12,5	7,4	18,3
Mean path length	3,87	4,01	4,78	4,79	3,43	3,82	2,78
Network diameter	9	13	12	13	9	10	6
Network density	0,003	0,002	0,002	0,001	0,004	0,004	0,003
Network heterogenity	1,59	1,74	1,8	2,11	1,43	1,34	2,61
Clustering coefficient	0,14	0,12	0,03	0,03	0,15	0,16	0,24
Network centralization	0,06	0,05	0,08	0,05	0,08	0,08	0,45
Shortest paths	6721058	25265856	2936082	13991340	10403850	2898506	32313540
Power Law Constant (g)	1.70 (R^2:0.90)	1.74 (R^2:0.92)	1.71 (R^2:0.89)	1.73 (R^2:0.90)	1.67 (R^2:0.90)	1.62 (R^2:0.88)	1.41 (R^2:0.87)
# of isolated & lonely nodes	3	98	11	-	2	1	922
Total # of proteins	2596	5222	1725	3741	3228	1704	6607

In contrary to whole genome comparison, the topological properties of calcium signaling networks of yeast and human are similar (Table 7.1). Also these two calcium signaling networks seem to be faster and more modular than the worm signaling network.

The average degree values of yeast and human signaling PINs are almost the same (7.6 and 7.4 respectively) while worm's corresponding value is nearly half of those (3.7). On the contrary, the γ values of human and worm are almost the same (1.70 and 1.71) while γ value of yeast network (1.62) is the smallest among these three networks implying that it is the most hub dominated network like in the case of whole genome network comparison.

Also, average path lengths (3.82 for yeast and 3.87 for human) as well as diameters (10 for yeast, 9 for human) of these two networks agree well with each other. Again, worm is the outlier in these properties having an average path length of 4.78 and diameter of 12; hence being the least robust and interconnected network among these three networks.

The most striking difference of worm network is the clustering coefficient values it has (0.03), while the yeast and human networks' corresponding values are much higher than worm and close to each other (0.16 for yeast and 0.14 for human). This result verifies the fact that as the degree increases, the clustering coefficient behaves just the opposite. Hence, the worm network does not display strong neighborhood cohesiveness when compared to human and yeast networks.

In general, yeast and human calcium signaling networks alike being more hub-dominant, faster in information transfer, more modular and interconnected than worm network. In the light of these findings, it can be concluded that there is a topological conservation between yeast and human brain in calcium signaling. For this reason, yeast, being a simpler and easier to handle organism, may serve as an experimental model for studies concerning the much more complex organism human, particularly the human brain.

7.2. GO Enrichments of Yeast, Worm and Human

7.2.1. Comparison of GO Enrichments in Core Proteins

Yeast and human's core proteins partially agree in biological process term enrichment (Figure 7.1). In both core protein groups the enriched process is transport being the parent term for ion transport and in particular calcium ion transport. Human core proteins are also dominantly involved in transmembrane transport process whereas yeast core proteins are taking action in calcium ion homeostasis in general terms.

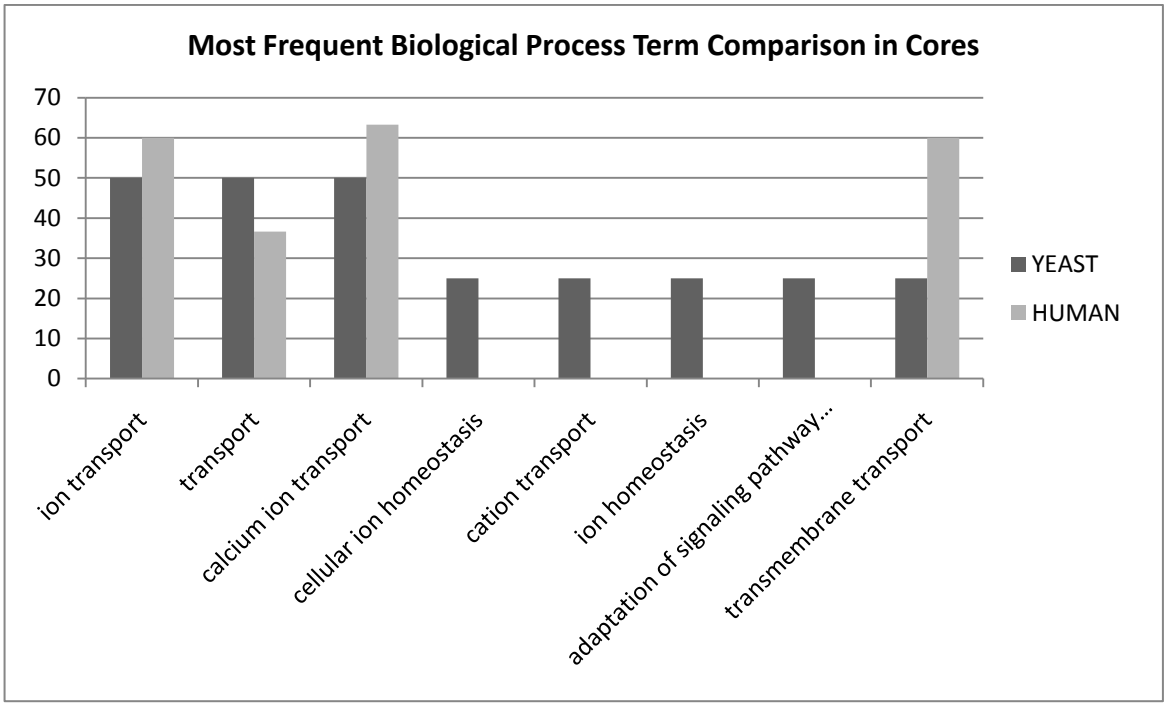


Figure 7.1. Predominant biological process terms of yeast and human core proteins

The process may be common but the functions of the proteins in these common process differ from yeast to human (Figure 7.2). Yeast core proteins have functions in calcium ion binding via activation of pumps (hydrolase activity and calcium channel activity as well as calcium in binding). However, human core proteins are mainly responsible for protein binding and receptor activity in transport. Also they try to help calcium ion transport by binding to calmodulin.

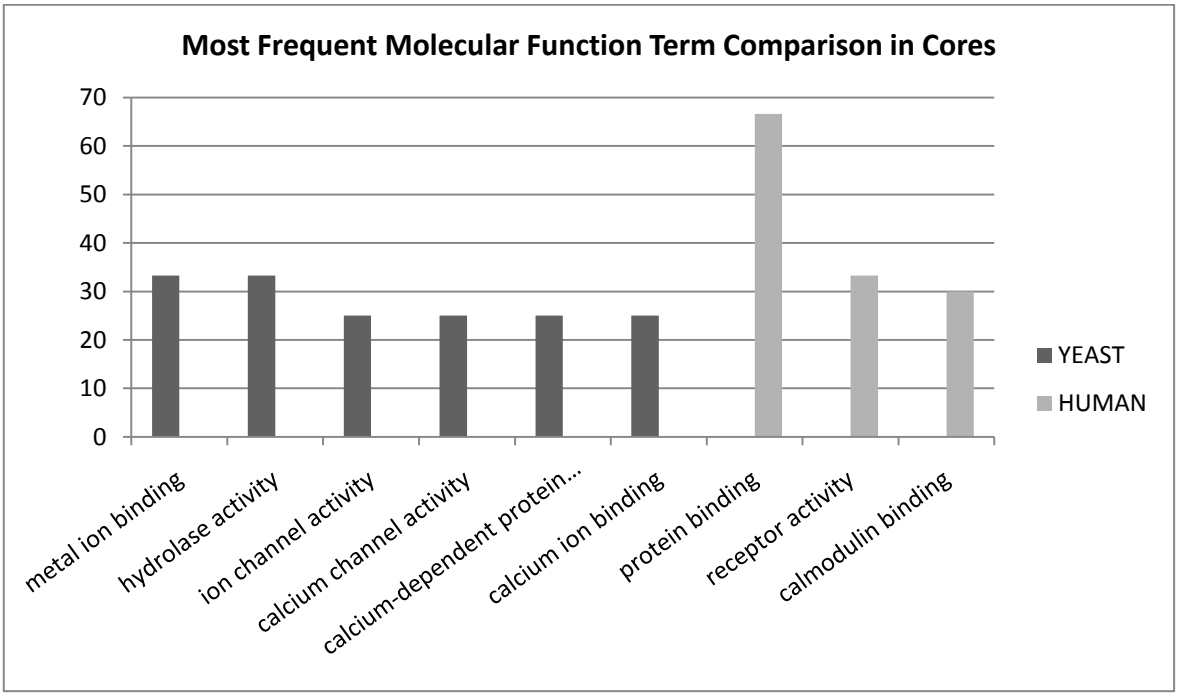


Figure 7.2. Predominant molecular function terms of yeast and human core proteins

7.2.2. Comparison of GO Enrichments in PINs

The agreement observed in core proteins' biological process terms can also be observed in PIN comparison. The enriched biological processes in yeast are transcription and transport and worm mainly deals with growth and reproduction processes. Moreover, signal transduction and apoptosis are the dominant process terms in human (Figure 7.3). One function of calcium signaling is to activate the immediate early genes responsible for inducing resting cells (G0) to re-enter the cell cycle. It may also promote the initiation of DNA synthesis at the G1/S transition. Finally, calcium contributes to the completion of the cell cycle by stimulating events at mitosis. All these processes eventually lead to transcription, development and cell proliferation as can be seen in all three PINs' enriched terms.

Although their enriched process terms do not agree, worm and human PIN proteins are both dominantly responsible for protein binding as in the case of their core proteins. Yeast PIN, on the other hand, now deals with nucleotide binding function mostly since the enriched process transcription requires this function (Figure 7.4).

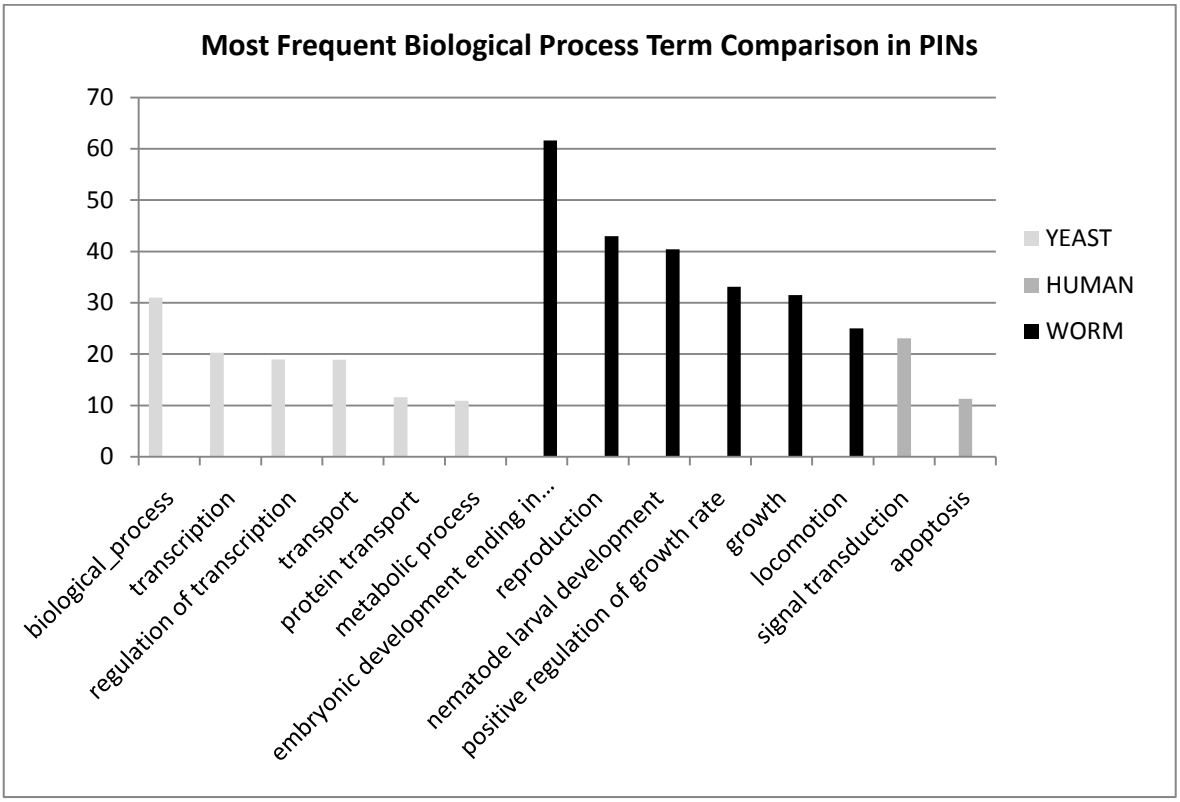


Figure 7.3. Predominant biological process terms of PINs

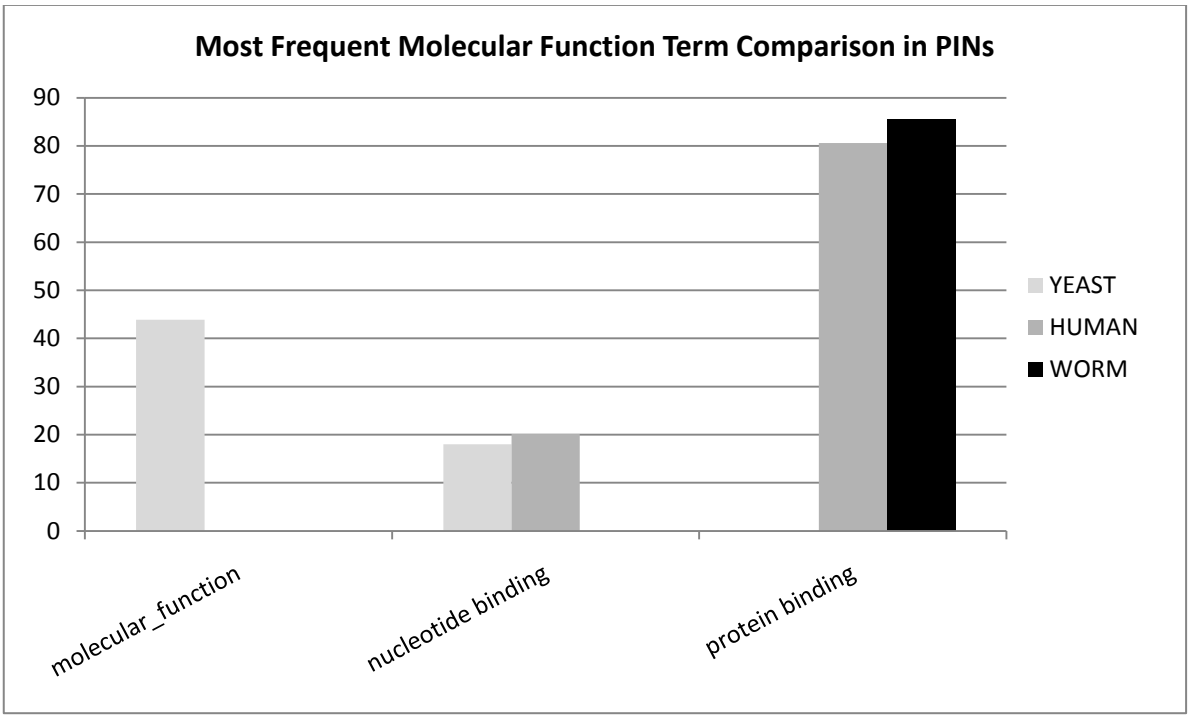


Figure 7.4. Predominant molecular function terms of PINs

7.2.3. Comparison of GO Enrichments in hubs

The comparison of hub proteins agrees with the results of the comparison of PINs (Figure 7.5). As yeast calcium signaling focuses on transcription and transport, worm specifically uses calcium signaling in growth and development. Additionally, the calcium signaling in human brain deals with cell proliferation and differentiation as well as cell cycle arrest.

The common function they carry out is ATP binding which is an important function for many processes, from transport to growth, taking place in the cell. Worm and human also shares the protein binding function (Figure 7.6).

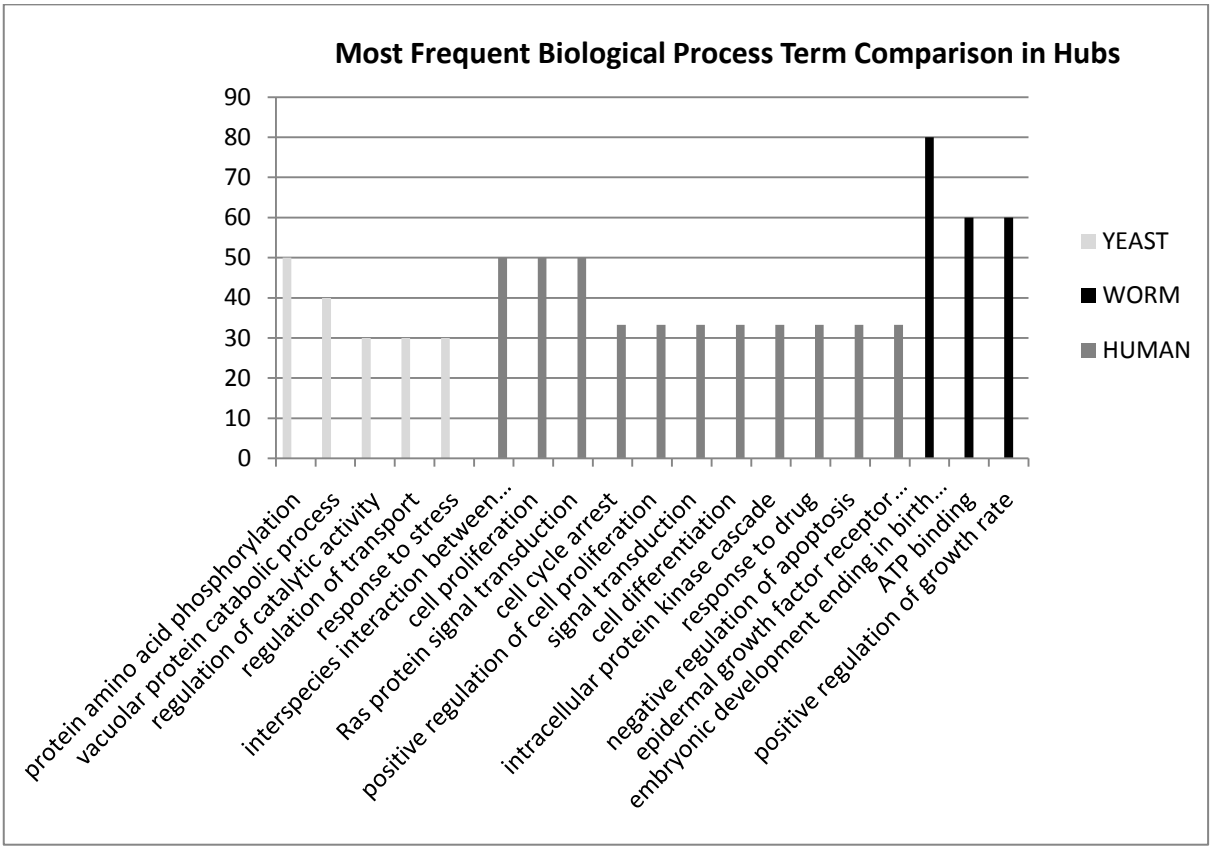


Figure 7.5. Predominant biological process terms of hubs

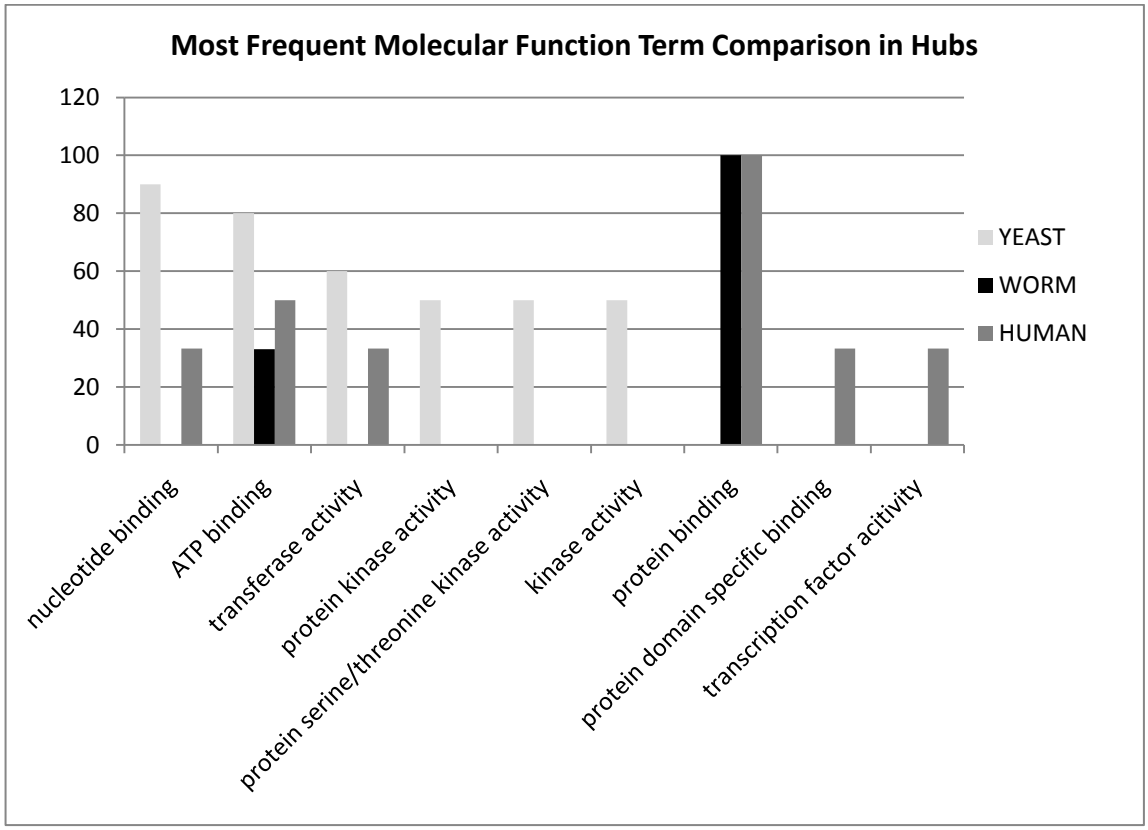


Figure 7.6. Predominant molecular function terms of hubs

7.3. Orthology Search for Core and Hub Proteins

The core proteins for each species namely yeast, worm and human are fed into blastp algorithm and the corresponding orthologs in between these species are revealed. In general, core proteins are conserved throughout the evolution with an approximate average above 35 % (Table 7.2).

Human core proteins' orthologs in worm are 50% conserved in average, 6 of the orthologs being the core proteins of worm. Maximum conservation is for cmd-1, which is not included in the worm PIN, with 98% identity. On the other hand, the mean identity score is 35 % for yeast orthologs. No orthologs in yeast could be found for calcium channels (VOC, SOC and VDAC groups) and receptors (NMDAR, RYR and IP3R groups) as well as exchangers (NCX group) of human. The most conserved protein in yeast is the calmodulin which has also crucial roles in calcium signaling in yeast.

Human hub proteins, MYC and TP53 are the least conserved ones both in worm and yeast (avg 5% in worm, no conservation in yeast). When these outliers are excluded, hub proteins of human in worm are 57% conserved on average. UBC has the maximum conservation score (98.7%) both in worm and yeast. Although worm PIN contains this protein, yeast does not (Table 7.3).

YWHAG and GRB2 are also well conserved in worm (over 55%) both being included in the worm PIN with 31 interactions. In yeast network, only GRB2 ortholog is present with 15 interactions. Although YWHAG is also well conserved in yeast, the yeast PIN does not contain YWHAG orthologs.

EGFR and SRC are 20% conserved in worm on average with 11 interactions in the worm PIN. However, there are no orthologs in yeast for these hubs.

Yeast hub proteins are conserved in worm and human with averages of 45% and 48% respectively. 8 of worm and 9 of human orthologs take place in their PINs. PTK2 and ATG1 are the least conserved ones. If their identity % are excluded, both human and worm conservation rises above 50 %. The orthologs in worm PIN has an average of 4 interactions with a maximum of 11, none of them being hubs.

In human, the average number of interactions is higher (19) than that of worm with a maximum of 53 interactions. They are not hubs either. However, they are the proteins that connect the hub proteins in the human brain network. For example, HSP90-A, being the human ortholog of the worm protein HSP82, interacts with MYC, TP53 and SRC, three hubs of human PIN. Also, worm protein CDC28's human ortholog CDK1 interacts with both MYC and TP53.

Table 7.2. Orthologs of core proteins of yeast, worm and human

Grey areas indicate the query proteins for orthology. Bold ones are the ortholog core proteins

YEAST						WORM						HUMAN					
% identity	e-val	prot name	ORF	# of ints.	ID	% identity	e-val	prot name	WB id	# of ints.	ID	% identity	e-val	prot name	HPRD id	# of ints.	ID
		CMD1	YBR109C	46		59,90	4,58E-50	CalModulin family member (cmd-1)	WBGene00000552	-	NP_503386	59,90	7,07E-50	calmodulin	241/**242/243**	37/10/11	NP_001734
		CNB1	YKL190W	5		52,60	5,24E-49	CalciNeurin B family member (cnb-1)	**WBGene00000554**	-	NP_505885	52,60	9,95E-49	calcineurin subunit B type 1	3195	-	NP_000936
		CNA1	YLR433C	7		47,90	8,44E-155	Serine/threonine-protein phosphatase (tax-6)	**WBGene00006527**	2	NP_001021292	47,00	9,90E-149	serine/threonine-protein phosphatase 2B	236	-	NP_005596
		CMP2	YML057W	7		47,00	1,20E-153	Serine/threonine-protein phosphatase (tax-6)	**WBGene00006527**	2	NP_001021293	45,20	1,47E-154	serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform isoform 1	234	7	NP_000935
		CRZ1	YNL027W	13		6,00	7,15E-13	Zinc finger Transcription Factor family member (ztf-23)	WBGene00021846	-	NP_491096	8,00	1,48E-13	zinc finger protein 397OS	19434	-	NP_001106205
		VCX1	YDL128W	5								8,00	1,25E+00	sodium/potassium/calcium exchanger 3 precursor	15352	-	NP_065740
		PMR1	YGL167C	7		45,30	0,00E+00	PMR-type Golgi ATPase family member (pmr-1)	**WBGene00004063**	-	NP_001021860	47,80	0,00E+00	calcium-transporting ATPase type 2C member 1 isoform	5089	-	NP_055197
		MID1	YNL291C	3								3,10	1,48E+00	TPA_inf: olfactory receptor OR14-7	14983	-	DAA04757
		PMC1	YGL006W	16		35,50	2,56E-146	calcium ATPase	WBGene00003153	-	CAA09303	34,90	1,89E-154	plasma membrane calcium-transporting ATPase 2 isoform 2	**160**	2	NP_001674
		RCN1	YKL159C	5		16,60	1,61E+00	Regulator of CalciNeurin family member (rcn-1)	**WBGene00004321**	4	NP_498223	12,30	7,35E-01	regulator of calcineurin 3 isoform 2,3,4b,5	5792	-	AAF01685

Table 7.3. Orthologs of core proteins of yeast, worm and human - cont.

		YVC1	YOR087W	-		9,30	4,15E-01	TRP (transient receptor potential) channel family member (trp-2)	WBGene00006615	7	NP_001022703	17,50	1,49E-01	short transient receptor potential channel 7 isoform 1	18228	1	NP_065122
		CCH1	YGR217W	1		15,20	3,95E-57	four domain-type voltage-gated ion channel alpha-1 subunit	-		ABI94563	10,84	3,09E-39	L-type calcium channel (HFCC)	246	-	AAA17030
27,55	5,30E-32	YPL141C	YPL141C	25	NP_015184			ckk-1	WBGene00000518	-		48,15	1,76E-105	calcium/calmodulin-dependent protein kinase kinase 2 isoform 2	6467	-	NP_705719
24,54	6,52E-38	TOS3	YGL179C	21	NP_011336												
37,64	3,80E-59	CMK2	YOL016C	-	NP_014626			cmk-1	WBGene00000553	1		61,49	5,50E-123	calcium/calmodulin-dependent protein kinase type 1	5412	2	NP_003647
37,07	4,37E-62	CMK1	YFR014C	57	NP_116669												
53,80	1,86E-49	CNB1	**YKL190W**	5	NP_012731			cnb-1	WBGene00000554	-		78,36	1,06E-74	calcineurin subunit B type 1	3195	-	NP_000936
28,10	3,50E-29	CNE1	YAL058W	-	NP_009343			crt-1	WBGene00000802	4		57,72	2,78E-135	calreticulin precursor	169	17	NP_004334
								csq-1	WBGene00000822	-		29,02	4,72E-44	calsequestrin-1	15915	-	NP_001222
8,10	2,74E-41	CCH1	**YGR217W**	1	NP_011733			egl-19	WBGene00001187	-		57,49	0,00E+00	voltage-dependent L-type calcium channel subunit alpha-1D isoform c	**247**	2	NP_001122311
								itr-1	WBGene00002173	-		37,76	0,00E+00	inositol 1,4,5-trisphosphate receptor type 3	926	-	NP_002215
22,31	5,83E-80	PMR1	**YGL167C**	7	NP_011348			mca-1	WBGene00003151	-		50,48	0,00E+00	plasma membrane calcium-transporting ATPase 3 isoform 3b	2057	-	NP_001001344
18,97	7,45E-98	PMC1	**YGL006W**	16	NP_011509												

Table 7.4. Orthologs of core proteins of yeast, worm and human - cont.

23,66	4,30E-93	PMR1	**YGL167C**	7	NP_011348			mca-3	WBGene00003153	-		61,47	0,00E+00	plasma membrane calcium-transporting ATPase 3 isoform 3b	2057	-	NP_001001344
47,72	0,00E+00	PMR1	**YGL167C**	7	NP_011348			pmr-1	WBGene00004063	-		58,27	0,00E+00	calcium-transporting ATPase type 2C member 1 isoform 1a	5089	-	NP_055197
30,52	2,49E-105	ENA2	YDR039C	4	NP_010324												
30,52	3,17E-105	ENA5	YDR038C	3	NP_010323												
30,52	1,14E-104	ENA1	YDR040C	2	NP_010325												
								rcn-1	WBGene00004321	4		34,30	2,92E-32	calcipressin-3	5792	-	NP_038469
								sca-1	WBGene00004736	3		66,95	0,00E+00	sarcoplasmic/endoplasmic reticulum calcium ATPase 2 isoform 1	**161**	3	NP_733765
								ser-1	WBGene00004776	1		19,62	2,07E-44	D(2) dopamine receptor isoform long	539	9	NP_000786
								slo-1	WBGene00004830	-		56,14	0,00E+00	calcium-activated potassium channel subunit alpha-1 isoform c	15967	2	NP_001154824
								spe-41	WBGene00004976	-		17,80	9,08E-33	short transient receptor potential channel 3 isoform a	15999	-	NP_001124170
52,34	5,07E-155	CMP2	**YML057W**	7	NP_013655			tax-6	WBGene00006527	2		68,60	0,00E+00	serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform isoform 2	234	7	NP_001124163

Table 7.5. Orthologs of core proteins of yeast, worm and human - cont.

20,23	8,71E-66	ENA2	YDR039C	4	NP_010324	55,36	0,00E+00	Membrane Calcium ATPase family member (mca-3)	**WBGene00003153**	-	NP_500294			pmca4	159	7	
20,23	1,19E-65	ENA5	YDR038C	3	NP_010323	52,22	0,00E+00	hypothetical protein R05C11.3	WBGene00019875	-	NP_500161						
20,15	7,94E-66	ENA1	YDR040C	2	NP_010325	49,80	0,00E+00	Membrane Calcium ATPase family member (mca-1)	**WBGene00003151**	-	NP_501709						
17,86	4,04E-93	PMC1	**YGL006W**	16	NP_011509	56,88	0,00E+00	Membrane Calcium ATPase family member (mca-3)	**WBGene00003153**	-	NP_500294			pmca2	160	2	
15,85	5,78E-71	PMR1	**YGL167C**	7	NP_011348	53,82	0,00E+00	hypothetical protein R05C11.3	WBGene00019875	-	NP_500161						
						44,42	6,31E-113	AcetylCholine Receptor family member (acr-19)	WBGene00000058	-	NP_492429			chrna7	339	2	
						44,02	5,11E-119	AcetylCholine Receptor family member (acr-16)	WBGene00000055	-	NP_505207						
						20,06	2,08E-104	NMDA class glutamate receptor family member (nmr-2)	WBGene00003775	-	NP_506694			nmdar2c	699	2	
						18,86	7,60E-102	NMDA class glutamate receptor family member (nmr-2)	WBGene00003775	-	NP_506694			nmdar2d	4095	16	
						30,60	3,17E-129	NMDA class glutamate receptor family member (nmr-1)	WBGene00003774	-	NP_495033			nmdar1	15926	25	
7,03	2,37E-38	CCH1	**YGR217W**	1	NP_011733	46,65	0,00E+00	EGg Laying defective family member (egl-19)	**WBGene00001187**	-	NP_741442			cav 1.3	247	2	
						34,01	0,00E+00	UNCoordinated family member (unc-2)	WBGene00006742	-	NP_741732						
						18,88	0,00E+00	UNCoordinated family member (unc-2)	WBGene00006742	-	NP_741732			cav 2.1	3004	3	

Table 7.6. Orthologs of core proteins of yeast, worm and human - cont.

						14,77	0,00E+00	EGg Laying defective family member (egl-19)	**WBGene00001187**	-	NP_741442						
7,44	2,27E-39	CCH1	**YGR217W**	1	NP_011733	21,25	0,00E+00	UNCoordinated family member (unc-2)	WBGene00006742	-	NP_741732			cav 2.2	3005	2	
						16,25	0,00E+00	EGg Laying defective family member (egl-19)	**WBGene00001187**	-	NP_741442						
						21,92	0,00E+00	UNCoordinated family member (unc-2)	WBGene00006742	-	NP_741733			cav 2.3	3006	-	
						16,30	0,00E+00	EGg Laying defective family member (egl-19)	**WBGene00001187**	-	NP_741442						
						25,29	0,00E+00	Calcium Channel, Alpha subunit family member (cca-1)	WBGene00000367	-	NP_741848			cav 3.2	7442	-	
						38,21	1,97E-147	TRP (transient receptor potential) channel family member (trp-2)	WBGene00006615	7	NP_497311			trpc1	11894	10	
						35,18	3,89E-136	TRP (transient receptor potential) channel family member (trp-1)	WBGene00006614	-	NP_001040889						
						15,30	0,00E+00	UNCoordinated family member (unc-68)	WBGene00006801	-	NP_504753			ryr1	1618	4	
						21,65	0,00E+00	UNCoordinated family member (unc-68)	WBGene00006801	-	NP_504753			ryr2	1619	11	
						21,19	0,00E+00	UNCoordinated family member (unc-68)	WBGene00006801	-	NP_504753			ryr3	1620	2	
						24,91	0,00E+00	Inositol Triphosphate Receptor family member (itr-1)	**WBGene00002173**	-	NP_001023170			ip3r1	925	17	
27,06	1,40E-116	PMR1	**YGL167C**	7	NP_011348	68,04	0,00E+00	SERCA (Sarco-Endoplasmic Reticulum Calcium ATPase) family member (sca-1)	**WBGene00004736**	3	NP_499386			serca2	161	3	

Table 7.7. Orthologs of core proteins of yeast, worm and human - cont.

30,30	1,08E-123	PMR1	**YGL167C**	7	NP_011348	65,68	0,00E+00	SERCA (Sarco-Endoplasmic Reticulum Calcium ATPase) family member (sca-1)	**WBGene00004736**	3	NP_499385			serca3	3568	1	
						30,87	2,73E-126	PMR-type Golgi ATPase family member (pmr-1)	**WBGene00004063**	-	NP_001021860						
						45,63	0,00E+00	Na/Ca eXchangers family member (ncx-2)	WBGene00003567	-	NP_504414			ncx1	1659	1	
						43,37	0,00E+00	Na/Ca eXchangers family member (ncx-1)	WBGene00003566	-	NP_001041182						
						46,58	0,00E+00	Na/Ca eXchangers family member (ncx-2)	WBGene00003567	-	NP_504415			ncx2	15990	2	
						45,17	0,00E+00	Na/Ca eXchangers family member (ncx-1)	WBGene00003566	-	NP_001041182						
						47,57	0,00E+00	Na/Ca eXchangers family member (ncx-2)	WBGene00003567	-	NP_504415			ncx3	9723	1	
						45,31	0,00E+00	Na/Ca eXchangers family member (ncx-1)	WBGene00003566	-	NP_001041182						
53,02	1,42E-78	PET9	YBL030C	-	NP_009523	68,79	1,40E-114	Adenine Nucleotide Translocator family member (ant-1.3)	WBGene00010485	-	NP_501727			ant1	58	5	
53,02	3,55E-77	AAC1	YMR056C	-	NP_013772	68,79	2,50E-114	Adenine Nucleotide Translocator family member (ant-1.44)	WBGene00020140	-	NP_501440						
51,34	1,50E-78	AAC3	YBR085W	-	NP_009642	68,12	7,77E-114	Adenine Nucleotide Translocator family member (ant-1.1)	WBGene00006439	9	NP_001022799						
						65,10	2,46E-110	Adenine Nucleotide Translocator family member (ant-1.2)	WBGene00007057	-	NP_491927						
						50,00	6,52E-83	hypothetical protein C47E12.2	WBGene00008147	-	NP_501803						
						46,98	1,25E-80	hypothetical protein F25B4.7	WBGene00017770	-	NP_504498						
						44,63	3,04E-71	hypothetical protein R07E3.4	WBGene00001105	-	NP_509733						

Table 7.8. Orthologs of core proteins of yeast, worm and human - cont.

53,69	6,69E-81	AAC3	YBR085W	-	NP_009642	69,46	2,88E-113	Adenine Nucleotide Translocator family member (ant-1.3)	WBGene00010485	-	NP_501727			ant2	2147	-	
53,69	7,71E-80	PET9	YBL030C	-	NP_009523	69,46	3,58E-113	Adenine Nucleotide Translocator family member (ant-1.44)	WBGene00020140	-	NP_501440						
52,35	3,08E-77	AAC1	YMR056C	-	NP_013772	68,46	7,39E-114	Adenine Nucleotide Translocator family member (ant-1.1)	WBGene00006439	9	NP_001022799						
						65,77	2,30E-108	Adenine Nucleotide Translocator family member (ant-1.2)	WBGene00007057	-	NP_491927						
						50,67	4,79E-82	hypothetical protein C47E12.2	WBGene00008147	-	NP_501803						
						47,32	6,42E-82	hypothetical protein F25B4.7	WBGene00017770	-	NP_504498						
						44,63	2,87E-71	hypothetical protein R07E3.4	WBGene00001105	-	NP_509733						
45,95	1,78E-78	CPR6	YLR216C	-	NP_013317	30,81	1,10E-59	CYclophyliN family member (cyn-3)	WBGene00000879	-	NP_506751			ppid	11875	6	
						30,81	1,62E-59	CYclophyliN family member (cyn-1)	WBGene00000877	-	NP_506561						
						30,54	2,56E-59	CYclophyliN family member (cyn-7)	WBGene00000883	-	NP_506749						
						30,00	9,99E-53	CYclophyliN family member (cyn-2)	WBGene00000878	-	NP_499828						
						39,58	4,56E-55	hypothetical protein R05G6.7	WBGene00019900	-	NP_501211			vdac3	6750	2	
						39,93	5,43E-56	hypothetical protein R05G6.7	WBGene00019900	-	NP_501211			vdac1	5137	14	
						39,46	6,25E-59	hypothetical protein R05G6.7	WBGene00019900	-	NP_501211			vdac2	1904	2	
59,06	5,53E-49	CMD1	**YBR109C**	46	NP_009667	97,99	2,65E-78	CalModulin family member (cmd-1)	WBGene00000552	-	NP_503386			cam	242/243	10/11	
						65,77	8,74E-52	CALmodulin related genes family member (cal-1)	WBGene00000285	-	NP_506138						
						63,09	3,74E-48	CALmodulin related genes family member (cal-2)	WBGene00000286	-	NP_495906						

Table 7.9. Orthologs of core proteins of yeast, worm and human - cont.

						59,73	1,51E-41	hypothetical protein Y73B3A.12	WBGene00022214	-	NP_508066						
						57,72	4,81E-42	CALmodulin related genes family member (cal-4)	WBGene00000288	-	NP_501943						
						57,05	1,14E-40	CALmodulin related genes family member (cal-3)	WBGene00000287	-	NP_500421						

Table 7.10. Orthologs of hub proteins

S.CEREVISIAE (yeast)						C.ELEGANS (worm)						H.SAPIENS (Human)					
% identity	e-val	prot name	ORF	Ints	ID	% identity	e-val	prot name	WB Id	Ints	ID	% identity	e-val	prot name	HPRD Id	Ints	ID
98,68	3,31E-37	Ribosomal protein L40A	YIL148W	-	NP_012118	98,68	4,58E-38	UBiQuitin family member (ubq-2)	WBGene00006728	8	NP_499695			Ubiquitin C	7038	138	P62988
						98,68	1,21E-37	UBiQuitin family member (ubq-1)	WBGene00006727	10	NP_741157						
						98,68	2,14E-11	UBiquitin-Like family member (ubl-1)	WBGene00006725	3	NP_741102						
		-				15,95	1,07E-73	LEThal family member (let-23)	WBGene00002299	-	NP_001129831			EGF Receptor	579	117	P00533
						15,95	1,26E-46	related to oncogene ABL family member (abl-1)	WBGene00000018	4	NP_509777						
5,47	2,79E+00	Retrograde regulation protein 3	YBL103C	4	NP_009447	5,47	2,42E-05	MaX-Like family member (mxl-3)	WBGene00003511	-	NP_510223			Myc proto-oncogene protein	1818	112	P01106
		-				4,07	5,77E+00	defective PHArynx development family member (pha-2)	WBGene00004011	-	NP_508131			Cellular tumor antigen p53	1859	111	P04637
		-				27,43	1,04E-143	SRC oncogene related family member (src-2)	WBGene00005078	3	NP_493502			Proto-oncogene tyrosine-protein kinase Src	1819	95	P12931
						24,81	9,20E-96	related to oncogene ABL family member (abl-1)	WBGene00000018	4	NP_509777						
61,94	2,53E-84	14-3-3 protein	YDR099W	-	NP_010384	56,68	1,09E-98	abnormal embryonic PARtitioning of cytoplasm family member (par-5)	WBGene00003920	9	NP_502235			Protein kinase C inhibitor protein 1	5639	90	P61981
39,27	3,59E-83	14-3-3 protein, major isoform	YER177W	-	NP_011104	72,47	1,63E-101	Fourteen-Three-Three family member (ftt-2)	WBGene00001502	11	NP_509939						
22,58	1,93E-14	[PSI+] inducibility protein 3	YPR154W	15	NP_015480	58,06	2,26E-70	SEx Muscle abnormal family member (sem-5)	WBGene00004774	11	NP_509342			Growth factor receptor-bound protein 2	150	89	P62993
		SSB1	YDL229W	147	P11484	59,05	0,00E+00	Heat Shock Protein family member (hsp-1)	WBGene00002005	4	NP_503068	60,03	0,00E+00	heat shock-related 70 kDa protein 2	7174	-	P54652
						55,79	0,00E+00	Heat Shock Protein family member (hsp-70)	WBGene00002026	1	NP_492485	59,05	0,00E+00	heat shock cognate 71 kDa protein	7205	26	P11142
						54,16	0,00E+00	Heat Shock Protein family member (hsp-3)	WBGene00002007	-	NP_509019	59,05	0,00E+00	heat shock 70 kDa protein 1A/1B	774/6784	0/2	P08107
						53,02	0,00E+00	Heat Shock Protein family member (hsp-4)	WBGene00002008	3	NP_495536	58,56	0,00E+00	heat shock 70 kDa protein 6	775	-	P17066
		HSP82	YPL240C	124	P02829	61,07	0,00E+00	abnormal DAuer Formation family member (daf-21)	WBGene00000915	11	NP_506626	61,78	0,00E+00	heat shock protein HSP 90-beta	6763	10	P08238

Table 7.11. Orthologs of hub proteins - cont.

												61,07	0,00E+00	heat shock protein HSP 90-alpha	777	53	P07900
		CDC28	YBR160W	107	P00546	54,70	1,03E-94	Cyclin-Dependent Kinase family member (cdk-1)	WBGene00000405	2	NP_001022747	62,75	1,28E-108	cell division protein kinase 3	446	-	Q00526
												58,72	8,30E-104	cell division protein kinase 1	302	39	P06493
												55,37	3,65E-90	cell division protein kinase 5	449	20	Q00535
		TPK1	YJL164C	86	P06244	41,56	1,05E-95	protein KINase family member (kin-1)	WBGene00002189	-	NP_740958	40,81	1,56E-97	cAMP-dependent protein kinase catalytic subunit beta	1482	1	P22694
		PTK2	YJR059W	76	P47116	9,29	1,30E-14	hypothetical protein Y43D4A.6	WBGene00012792	-	NP_502989	13,94	1,69E-14	serine/threonine-protein kinase MARK2	2753	-	Q7KZI7
		SSA1	YAL005C	76	P10591	74,61	0,00E+00	Heat Shock Protein family member (hsp-1)	WBGene00002005	4	NP_503068	73,36	0,00E+00	heat shock 70 kDa protein 1-like	776	1	P34931
						64,95	0,00E+00	Heat Shock Protein family member (hsp-70)	WBGene00002026	1	NP_492485	70,40	0,00E+00	heat shock 70 kDa protein 6	775	-	P17066
						61,84	0,00E+00	Heat Shock Protein family member (hsp-3)	WBGene00002007	-	NP_509019	70,09	0,00E+00	heat shock-related 70 kDa protein 2	7174	-	P54652
						59,97	0,00E+00	Heat Shock Protein family member (hsp-4)	WBGene00002008	3	NP_495536						
		YRA1	YDR381W	72	Q12159	21,24	3,32E-07	Ref/ALY RNA export adaptor family member (aly-1)	WBGene00000120	-	NP_501588	17,70	2,86E-03	eukaryotic translation initiation factor 3 subunit G	4886	-	O75821
		ATG1	YGL180W	63	P53104	13,04	1,78E-43	UNCoordinated family member (unc-51)	WBGene00006786	2	NP_507869	13,15	1,21E-50	serine/threonine-protein kinase ULK3	15609	-	Q6PHR2
		CMK1	YFR014C	57	P27466	28,92	2,81E-61	CaM Kinase family member (cmk-1)	WBGene00000553	1	NP_500139	28,70	2,83E-62	calcium/calmodulin-dependent protein kinase type 1D	6409	-	Q8IU85
						27,58	1,41E-49	UNCoordinated family member (unc-43)	WBGene00006779	6	NP_001023300	28,03	8,82E-61	calcium/calmodulin-dependent protein kinase type 1	5412	2	Q14012
		RPO21	YDL140C	56	P04050	46,39	0,00E+00	AMAnitin resistant family member (ama-1)	WBGene00000123	-	NP_500523	47,55	0,00E+00	DNA-directed RNA polymerase II subunit RPB1	8916	-	P24928
						28,51	0,00E+00	RNA Polymerase, Class III (C) family member (rpc-1)	WBGene00004411	-	NP_501127	28,56	0,00E+00	DNA-directed RNA polymerase III subunit RPC1	17879	-	O14802

8. CONCLUSION

8.1. Methods & Databases

For the SPA method used in this work, the core proteins affect the resultant PIN directly. This outcome can be observed in the yeast PIN starting with 63 core proteins and 12 proteins. As the core protein number increases, enlarging the annotation table, the resultant PIN becomes larger and it may skip the focus from a significant cellular event to a general cascade occurring in the cell. Starting with large number core proteins to SPA may eventually result in a PIN that covers most of the whole interactome and makes it harder to focus on the main process in question. One has to find the most relevant proteins to the process in question and start using this set of proteins as core proteins.

In conjunction with this issue, the completeness of GO terms comes into prominence since the main criteria for the SPA method is strictly dependent on GO terms of the candidate proteins. In yeast, the GO term data is almost complete posing no problem. However, this situation is not valid for worm and human GO data. There are some proteins with no GO term assigned at all leaving these proteins out in SPA method since they are not recognized by the program written in MATLAB. In order to overcome the absence of some GO terms, "unknown" GO terms are assigned to those proteins.

For calcium signaling in human brain, the tissue expression data is also critical since this network is reconstructed using the interactions taking place only between the proteins that are expressed in brain / brain regions. The tissue expression data is taken from HPRD and HuPa in this work. However, the completeness of this data is susceptible jeopardizing the reliability of the resultant PINs. For example, some of the proteins listed in KEGG's Neurodegenerative Disorder Pathways are found not be expressed in brain according to HuPa and HPRD; although being present in neurodegenerative disorder pathway implies expression in brain.

8.2. Calcium Signaling Networks

This work represents the first attempt to explore protein interaction network for calcium signaling in yeast, worm and human by utilizing computational methods. Up-to-date, in most of the published experimental and computational works, the focus is on a small portion of the calcium signaling; e.g. the effects of calcium related proteins in a definite process (Spang *et al.*, 1993, Cunningham and Fink, 1994, Batiza *et al.*, 1996, Dennis and Cyert, 2002, Kafadar and Cyert, 2004, Kane P., 2006, Rispail *et al.*, 2009). However, calcium signaling affects many different processes that seem to be distinct but interrelated in the cell. In order to unravel the insights of calcium signaling and its role in the cell more clearly and more completely in yeast and in higher organisms afterwards, a network reconstruction process integrating interactome data with Gene Ontology is accomplished, depicting a global picture of calcium signaling and its ultimate effects in yeast. Moreover, for the proteins that seem to be in the calcium signaling network but have no known/reported function, a functional annotation strategy is used which enabled us to predict the role of these proteins in certain biological processes.

8.2.1. Core Proteins and Reconstructed Networks

Since the core proteins for reconstructed networks are chosen to be directly related to calcium (pumps, channels, calcium binding proteins), they show a reasonable amount of conservation between themselves indicating that the key proteins to calcium signaling are conserved throughout the evolutionary time span.

The largest network is found to be the calcium signaling network in human body having also the highest edge-to-node ratio. Yeast and human brain networks have almost equal edge-to-node ratios and worm, having a more incomplete interactome compared to other two, has the smallest one.

8.2.2. Enriched Terms for Reconstructed Networks

Calcium signaling occupies a preeminent position in the signal transduction system of the cell by virtue of its participation in a wide range of physiological functions together

with the biological events associated with genetic expression, cell proliferation and apoptosis, as well as cell differentiation and morphogenesis. Yeast calcium network mainly deals with transcription; worm network on the other hand focuses on development and growth. Human calcium networks in brain and in general are mainly responsible for signal transduction, regulation of transcription and apoptosis. Basically all four reconstructed networks, in agreement with the literature knowledge (Lombardia *et al.*, 2002, Tombes and Borisy, 1995, Kahl and Means, 2003), are found to have a main role in transcription and eventually cell proliferation and growth processes.

The enriched term for molecular function in yeast is nucleotide binding and in the rest of three networks is protein binding. In fact, protein binding is a common function for proteins in general since biological processes are carried by the interactions among them. However since calcium-binding proteins (CBPs) appear to function heavily in signal transduction, the protein binding term becomes more significant. Another outcome is that the hub proteins of calcium signaling in human brain are mainly proteins related to cancer. It is known that several CBPs have recently been demonstrated to be actively engaged in the regulation of the cell cycle in normal as well as in aberrant cell proliferation and in apoptotic cell death (Sherbet, 2001), making the outcome meaningful.

8.2.3. Topological Properties

All reconstructed networks show scale-free and small world properties. Networks of calcium signaling in brain and in yeast have almost the same degree and average path length values being more robust and modular than worm network.

The comparison of whole interactome vs calcium signaling PIN reveals that as the network becomes larger and complex the average path length shortens. This outcome can be observed for yeast and worm networks. However calcium signaling in brain has smaller mean path length than the whole brain interactome and calcium signaling in general. Brain, apart from the other parts of the body, controlling many processes throughout the body and also in itself and using calcium as the starting point to activate other responsive mechanisms; the signal transmission should be as fast as possible to function properly and manipulate the body actions.

8.3. Future Prospects and Recommendations

In this work, network reconstructions are based on the SPA method using Gene Ontology. The modularity of the networks, as well as annotational enrichments in direct interactions is utilized for functional annotations. More accurate and false-free analyses of biological systems will be possible as the interactome data becomes more complete.

Construction of networks containing physical interactions of proteins based on Gene Ontology; and investigating the topological/modular properties of such networks has become popular in functional annotation studies for almost a decade. On the other hand, the future trend is the addition of dynamics into the interactions for a better understanding of the roles of proteins in biological systems.

8.4. Dynamic Networks

Since the dynamics of protein interactions play a central role in regulating the functioning of cells and organisms, and since high-throughput experimental techniques are increasingly helpful in elucidating interaction dynamics, it is expected that computational approaches for inferring and analyzing network dynamics will be a cornerstone in future efforts to understand biological systems.

As increasingly diverse high-throughput experimental data sets are gathered, it is likely that novel computational models, or new applications of existing models, will be needed to incorporate the data into predictive quantitative models of the dynamic interactome. While many probabilistic network models have already been applied to incorporate disparate data sets in a single framework, new methods that can intelligently relate different types of interaction data with different sorts of temporal and contextual biases may be helpful. For example, hypergraph models might be used for explicit representations of different types of interaction data. Important questions arising from efforts to merge disparate data types include how to propagate different types of dynamic information among an integrated data set, how to represent explicit bias and how to identify and resolve discrepancies.

An important consideration is that temporal variations affecting network dynamics can range over several orders of magnitude. This wide range not only raises questions about increasing and expanding the temporal resolution of experimental methods but also demands the development of appropriate multi-timescale computational methods. Another simplification made by many of the approaches is that they typically model biological networks with a single node for each protein; in reality, each node represents a population of proteins that may be in different phosphorylation, activation or interaction states. Dynamical systems approaches for analysis and simulation provide a framework for handling these complexities, and integrating these approaches with more high-throughput graph-theoretic methods may be particularly fruitful.

Temporally, spatially and contextually dependent molecular interactions play a key role in regulating the functioning of cells and organisms. Incorporating and analyzing these dynamics are the next major challenges for computational interactomics.

While most large-scale technologies to determine protein interactions do not provide spatial, temporal or contextual information about detected interactions, dynamic views of the interactome can be obtained by integrating static molecular interaction data sets with other types of dynamic data, such as gene expression data, phenotypic responses to perturbations such as gene knock-outs or knock-downs, and information about expression QuantitativeTrait Loci (eQTLs) (Przytycka, 2010)

It is essential that computational biologists studying molecular networks increasingly shift their focus to develop methods that incorporate information about the dynamic nature of cellular systems.

8.5. Domain Based Interaction Prediction

There are several machine learning approaches to predict domain interactions. The five classifiers, SVM (Support Vector Machine), Nearest Neighbor, Naïve Bayes, J48 and Random Forest which have been commonly used in protein-protein interaction prediction problems. In SVM algorithm, feature vectors are non-linearly mapped on a high dimensional feature space and a set of hyperplanes are constructed to be used for

classification or regression. The simplest one among used classifiers is Nearest Neighbor which classifies instances according to their closeness to the training examples. The basic idea behind Naïve Bayes is to predict the class of an instance by learning conditional probability of each attribute. J48, also known as C4.5 grows an initial tree by using divide-and-conquer algorithm and then rank test instances. Random Forest developed by Breiman generates many classification trees simultaneously where each node uses a random subset of the features and outputs the classification based on majority voting over all trees in the forest. Random Forest algorithm is found to outperform other classifiers which were previously shown to be the best classification algorithm (e.g. SVM) (Kalyoncu *et al.*, 2010).

A new method is developed based on Random Forest algorithm, additionally converting primary sequence information into numerical feature input and calculating trigram and bigram amino acid frequencies. The discovery of features within primary sequences of known protein interaction pairs could be subsequently developed by using other features (binding affinities, secondary/tertiary structure, etc.) in the learning model. Further improvements on these lines may generate a powerful computational virtual screening technique that significantly reduces the search space for putative candidate target proteins of PDZ domains as well as other domains (Kalyoncu *et al.*, 2010).

8.6. Gene Ontology

The accumulation of data produced by genome-scale research requires explicitly defined vocabularies to describe the biological attributes of genes in order to allow integration, retrieval and computation of the data. Arguably, the most successful example of systematic description of biology is the Gene Ontology (GO) project. GO is widely used in biological databases, annotation projects and computational analyses for annotating newly sequenced genomes, text mining, network modeling and clinical applications, among others.

However, GO is still an incomplete source, having its own drawbacks. In this work, it has been observed that there are some proteins with no annotation at all causing

problems in computational analyses. It will be logical to assign the appropriate “unknown” terms to these proteins.

Also, a consensus should be reached regarding the identification of the genes and proteins. It has been seen that, databases may use different styles in identifying a gene/protein. Since the systems biology studies require the combination and evaluation of various sources, the collaboration of the data will be much easier if there is a consensus on a common identity.

APPENDIX A

A.1 Interaction Classification with respect to Experiment Type

Table A.1. Experimental methods in protein-protein interaction determination

Experiment Type	Interaction Type
Affinity capture MS	Physical
Affinity capture RNA	Physical
Affinity capture Western	Physical
Biochemical Activity	Physical
Co-crystal Structure	Physical
Co-fractination	Physical
Co-localization	Physical
Co-purification	Physical
Dosage Growth Effect	Genetic
Dosage Lethality	Genetic
Dosage Rescue	Genetic
Far Western	Physical
FRET	Physical
Phenotypic Enhancement	Genetic
Phenotypic Supression	Genetic
Protein - Peptide	Physical
Protein - RNA	Physical
Reconstituted Complex	Physical
Synthetic Growth Defect	Genetic
Synthetic Lethality	Genetic
Synthetic Rescue	Genetic
Two-Hybrid	Physical

A.2 Gene Ontology Evidence Code Sources

Table A.2. Evidence Code Sources

Experimental Evidence Codes	
EXP	Inferred from Experiment
IDA	Inferred from Direct Assay
IPI	Inferred from Physical Interaction
IMP	Inferred from Mutant Phenotype
IGI	Inferred from Genetic Interaction
IEP	Inferred from Expression Pattern
Computational Analysis Evidence Codes	
ISS	Inferred from Sequence or Structural Similarity
ISO	Inferred from Sequence Orthology
ISA	Inferred from Sequence Alignment
ISM	Inferred from Sequence Model
IGC	Inferred from Genomic Context
RCA	inferred from Reviewed Computational Analysis
Author Statement Evidence Codes	
TAS	Traceable Author Statement
NAS	Non-traceable Author Statement
Curator Statement Evidence Codes	
IC	Inferred by Curator
ND	No biological Data available
Automatically-assigned Evidence Codes	
IEA	Inferred from Electronic Annotation
Obsolete Evidence Codes	
NR	Not Recorded

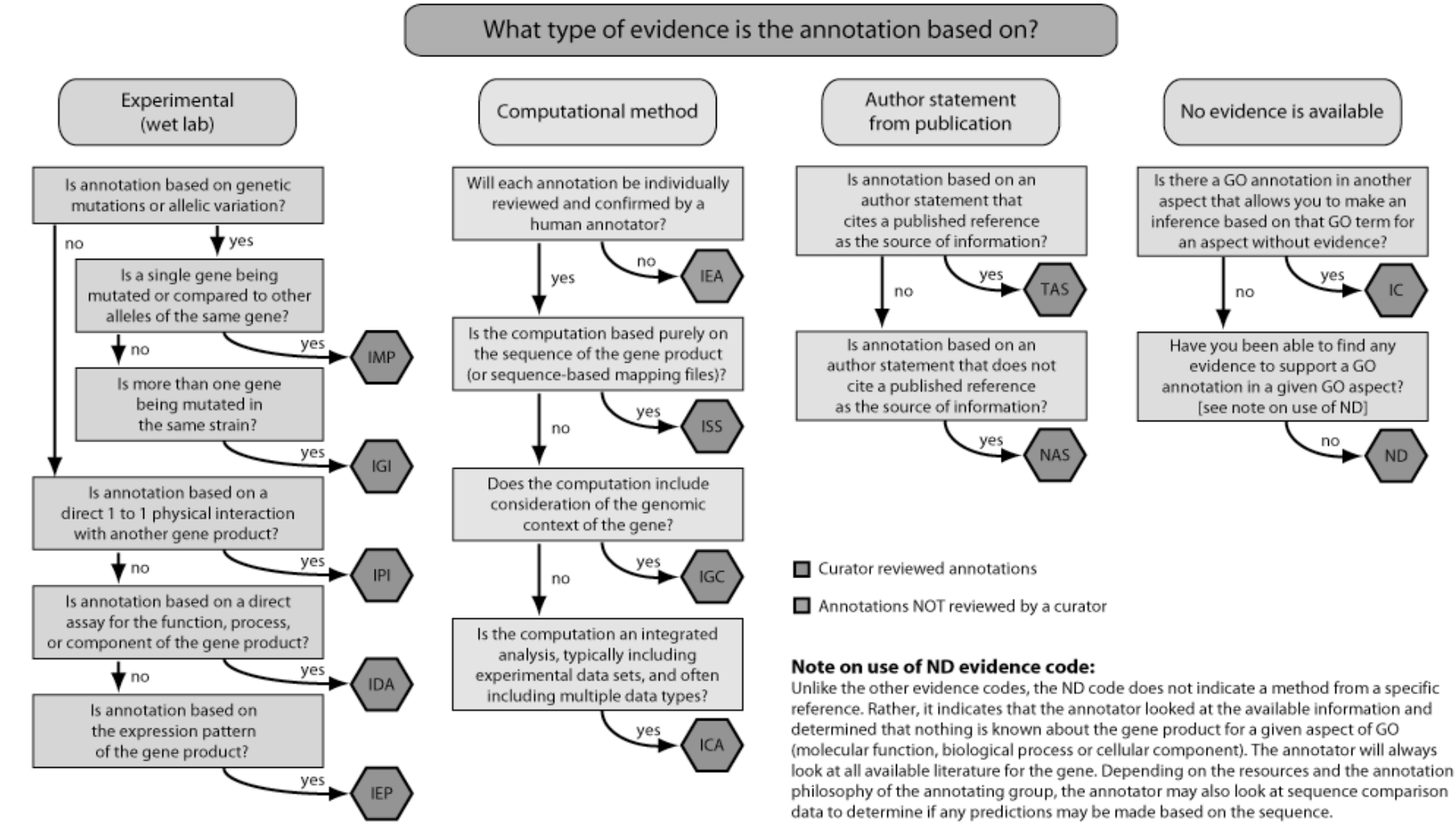


Figure A. 2 Go Evidence Code Decision Tree (Taken from Gene Ontology website)

A.3. Annotation Table Construction Code Algorithm

For the Selective Permissibility Algorithm implementation in MATLAB (Mathworks), the first step is to obtain the Annotation Table of core proteins. The Annotation Table is prepared as seperate three lists annottablecomp (AC), annottablefunc (AF) and annottableproc (AP); each column representing the reference GO annotations (cellular component, molecular function, cellular process).

- Textread “core proteins” list.
- Textread Gene Ontology data obtained from the corresponding web server in three groups (cellular component, molecular function, cellular process).
- Search each core protein in GO data lists
- List the found GO terms in three groups
- Delete the repeating data

A.4. SPA Code Algorithm

In order to make the algorithm robust, the time consuming step (finding the interaction partners of core/accepted proteins) is carried out by Excel, using the "match" command.

The MATLAB code steps are as follows:

- Textread "core/accepted proteins" and "candidate proteins" (interaction partners)
- Textread annotation table in three lists
- Search candidate proteins for repetition and delete them.
- Find the GO terms of the candidate proteins
- Match the GO terms of candidates in Annotation Table lists
- Find the proteins that have at least one element matching with each Annotation Table groups. These are accepted proteins.
- Search if the newly accepted protein list contains any formerly obtained accepted protein or core protein. Exclude them.
- Repeat the procedure until no new accepted protein is found.

The code is given as:

```
clear all;
[proteinlist] = textread('coreprot.txt','%s');
[candidateprots] = textread('candidate8run.txt','%s');
[acceptedprots] = textread('accepted1run.txt','%s');
[accepted2prots] = textread('accepted2run.txt','%s');
[accepted3prots] = textread('accepted3run.txt','%s');
[accepted4prots] = textread('accepted4run.txt','%s');
[accepted5prots] = textread('accepted5run.txt','%s');
[accepted6prots] = textread('accepted6run.txt','%s');
[accepted7prots] = textread('accepted7run.txt','%s');

% Read Annotation Table
AC = textread('annottablecomp.txt','%s');
AF = textread('annottablefunc.txt','%s');
AP = textread('annottableproc.txt','%s');
[GOCL,         PROTC]         =         textread('gocomplistun.txt','%s
%s','delimiter',':\t','commentstyle','shell');
[GOFL,         PROTF]         =         textread('gofunclistun.txt','%s
%s','delimiter',':\t','commentstyle','shell');
[GOPL,         PROTP]         =         textread('goproclistun.txt','%s
%s','delimiter',':\t','commentstyle','shell');

%-------------------------------------------------------
```

```
% Deletion of repeated proteins
for i = 1:length(candidateprots)
    Y2 = candidateprots(1:(i-1));
    Z2 = candidateprots((i+1):length(candidateprots));

    y1 = strmatch(candidateprots(i),Y2,'exact');
    if length(y1) ~= 0
        for j = 1:length(y1)
            Y2(y1(j)) = {'deleted'};
        end
    end

    y2 = strmatch(candidateprots(i),Z2,'exact');
    if length(y2) ~= 0
        for j = 1:length(y2)
            Z2(y2(j)) = {'deleted'};
        end
    end
    candidateprots = [Y2;candidateprots(i);Z2];
end

%Deletion of core proteins if exists in the candidate protein list

for i = 1:length(proteinlist)
    y = strmatch(proteinlist(i),candidateprots,'exact');
    if length(y) ~= 0
        for j = 1:length(y)
            candidateprots(y(j)) = {'deleted'};
        end
    end
end

%-------------------------------------------------------------
% Deletion of the label 'deleted' from the matrix
binaryX = zeros(length(candidateprots),1);
for i = 1:length(candidateprots)
    a = strmatch('deleted',candidateprots,'exact');

    for j = 1:length(a)
        %'deleted'ler 1 normal gen yazanlar 0
        binaryX(a(j)) = [1];
    end
end
k = 1;
for i = 1:length(binaryX)
    if binaryX(i) == 0
        coreprot(k,1) = candidateprots(i);
        k = k + 1;
    end
end

%Now we have the filtered candidate list 'coreprot'

%-----------------------------------------------------------------------
% Search the GO terms of proteins and create lists
ACe = {'start', 'start'};
AFe = {'start', 'start'};
```

```
APe = {'start', 'start'};
for i = 1:length(coreprot)
   acs = strmatch(coreprot(i),PROTC(:,:),'exact');
   if length(acs) ~= 0
        for j = 1:length(acs)
            ACe = [ACe; {PROTC{acs(j)}, GOCL{acs(j)}}];
        end
   end
end
for i = 1:length(coreprot)
   afs = strmatch(coreprot(i),PROTF(:,:),'exact');
   if length(afs) ~= 0
        for j = 1:length(afs)
            AFe = [AFe; {PROTF{afs(j)}, GOFL{afs(j)}}];
        end
   end
end
for i = 1:length(coreprot)
   aps = strmatch(coreprot(i),PROTP(:,:),'exact');
   if length(aps) ~= 0
        for j = 1:length(aps)
            APe = [APe; {PROTP{aps(j)}, GOPL{aps(j)}}];
        end
   end
end

%-------------------------------------------------------------------------
% Search Go terms of proteins in Annotation Table
ACC = {'start', 'start'};
AFF = {'start', 'start'};
APP = {'start', 'start'};
for i = 1:length(ACe(:,2))
   cc = strmatch(ACe(i,2),AC,'exact');
   %Cxde olanlar
    if length(cc) ~= 0
        for j = 1:length(cc)
            ACC = [ACC; {ACe{i,1}, ACe{i,2}}];
        end
    end
end
for i = 1:length(AFe(:,2))
   fc = strmatch(AFe(i,2),AF,'exact');
   %Fxde olanlar
    if length(fc) ~= 0
        for j = 1:length(fc)
            AFF = [AFF; {AFe{i,1}, AFe{i,2}}];
        end
    end
end
for i = 1:length(APe(:,2))
    %Pxde olanlar
    pc = strmatch(APe(i,2),AP,'exact');
    if length(pc) ~= 0
        for j = 1:length(pc)
            APP = [APP; {APe{i,1}, APe{i,2}}];
        end
    end
```

```
end
%-------------------------------------------------------------------------

%Deletion of the repeated terms

accproc = ACC(:,1);
accprof = AFF(:,1);
accprop = APP(:,1);

for i = 1:length(accproc)
    Y2 = accproc(1:(i-1));
    Z2 = accproc((i+1):length(accproc));

    y1 = strmatch(accproc(i),Y2,'exact');
    if length(y1) ~= 0
        for j = 1:length(y1)
            Y2(y1(j)) = {'deleted'};
        end
    end

    y2 = strmatch(accproc(i),Z2,'exact');
    if length(y2) ~= 0
        for j = 1:length(y2)
            Z2(y2(j)) = {'deleted'};
        end
    end
    accproc = [Y2;accproc(i);Z2];
end
for i = 1:length(accprof)
    Y2 = accprof(1:(i-1));
    Z2 = accprof((i+1):length(accprof));

    y1 = strmatch(accprof(i),Y2,'exact');
    if length(y1) ~= 0
        for j = 1:length(y1)
            Y2(y1(j)) = {'deleted'};
        end
    end

    y2 = strmatch(accprof(i),Z2,'exact');
    if length(y2) ~= 0
        for j = 1:length(y2)
            Z2(y2(j)) = {'deleted'};
        end
    end
    accprof = [Y2;accprof(i);Z2];
end
for i = 1:length(accprop)
    Y2 = accprop(1:(i-1));
    Z2 = accprop((i+1):length(accprop));

    y1 = strmatch(accprop(i),Y2,'exact');
    if length(y1) ~= 0
        for j = 1:length(y1)
            Y2(y1(j)) = {'deleted'};
        end
    end
```

```
    y2 = strmatch(accprop(i),Z2,'exact');
    if length(y2) ~= 0
        for j = 1:length(y2)
            Z2(y2(j)) = {'deleted'};
        end
    end
    accprop = [Y2;accprop(i);Z2];
end
%-------------------------------------------------------------------------

%Deletion of the label 'deleted' from the matrix
binaryX = zeros(length(accproc),1);
for i = 1:length(accproc)
    a = strmatch('deleted',accproc,'exact');

    for j = 1:length(a)
        %'deleted'ler 1 normal gen yazanlar 0
        binaryX(a(j)) = [1];
    end
end
k = 1;
for i = 1:length(binaryX)
    if binaryX(i) == 0
        %tekrarsýz candidate listesi
        accprocn(k,1) = accproc(i);
        k = k + 1;
    end
end
binaryX = zeros(length(accprof),1);
for i = 1:length(accprof)
    a = strmatch('deleted',accprof,'exact');

    for j = 1:length(a)
        %'deleted'ler 1 normal gen yazanlar 0
        binaryX(a(j)) = [1];
    end
end
k = 1;
for i = 1:length(binaryX)
    if binaryX(i) == 0
        %tekrarsýz candidate listesi
        accprofn(k,1) = accprof(i);
        k = k + 1;
    end
end
binaryX = zeros(length(accprop),1);
for i = 1:length(accprop)
    a = strmatch('deleted',accprop,'exact');

    for j = 1:length(a)
        %'deleted'ler 1 normal gen yazanlar 0
        binaryX(a(j)) = [1];
    end
end
k = 1;
for i = 1:length(binaryX)
```

```
    if binaryX(i) == 0
        %tekrarsýz candidate listesi
        accpropn(k,1) = accprop(i);
        k = k + 1;
    end
end
% Repeated terms are cleared from the data
%-------------------------------------------------------------------

% Finding the proteins that have at least a term in each Annotation
Table % column


accproff = {'start'};
accprocc = {'start'};
for i = 1:length(accpropn)
    fcc = strmatch(accpropn(i),accprofn,'exact');
   %Fxde olanlar
    if length(fcc) ~= 0
        for j = 1:length(fcc)
            accproff = [accproff; {accprofn{fcc(j)}}];
        end
    end
end

% The intersection of Function and Process terms are found. Now
component
for i = 1:length(accproff)
   pcc = strmatch(accproff(i),accprocn,'exact');
   %Fxde olanlar
    if length(pcc) ~= 0
        for j = 1:length(pcc)
            accprocc = [accprocc; {accprocn{pcc(j)}}];
        end
    end
end

%
% Deletion of the core proteins or any repeated accepted proteins
from %the previously found ones if any
%
for i = 1:length(proteinlist)
    y = strmatch(proteinlist(i),accprocc,'exact');
    if length(y) ~= 0
        for j = 1:length(y)
           accprocc(y(j)) = {'deleted'};
        end
    end
end
%
for i = 1:length(acceptedprots)
    y = strmatch(acceptedprots(i),accprocc,'exact');
    if length(y) ~= 0
        for j = 1:length(y)
           accprocc(y(j)) = {'deleted'};
        end
    end
```

```
end
% % % % % %
for i = 1:length(accepted2prots)
    y = strmatch(accepted2prots(i),accprocc,'exact');
    if length(y) ~= 0
        for j = 1:length(y)
           accprocc(y(j)) = {'deleted'};
        end
    end
end
% % %
for i = 1:length(accepted3prots)
    y = strmatch(accepted3prots(i),accprocc,'exact');
    if length(y) ~= 0
        for j = 1:length(y)
           accprocc(y(j)) = {'deleted'};
        end
    end
end
% % % % % %
for i = 1:length(accepted4prots)
    y = strmatch(accepted4prots(i),accprocc,'exact');
    if length(y) ~= 0
        for j = 1:length(y)
           accprocc(y(j)) = {'deleted'};
        end
    end
end
% % % % %
for i = 1:length(accepted5prots)
    y = strmatch(accepted5prots(i),accprocc,'exact');
    if length(y) ~= 0
        for j = 1:length(y)
           accprocc(y(j)) = {'deleted'};
        end
    end
end
for i = 1:length(accepted6prots)
    y = strmatch(accepted6prots(i),accprocc,'exact');
    if length(y) ~= 0
        for j = 1:length(y)
           accprocc(y(j)) = {'deleted'};
        end
    end
end
for i = 1:length(accepted7prots)
    y = strmatch(accepted7prots(i),accprocc,'exact');
    if length(y) ~= 0
        for j = 1:length(y)
           accprocc(y(j)) = {'deleted'};
        end
    end
end
%
% ‘accprocc’ is the accepted protein list
```

APPENDIX B

B.1. Calcium Related Gene Ontology Terms

Bold ones are excluded since they contain “calcium-independent” term.

Table B.1. GO terms with the keyword “calcium” in Gene Ontology for Yeast

Go ID	Go Term
0004698	calcium-dependent protein kinase C activity
0004699	**calcium-independent protein kinase C activity**
0004723	calcium-dependent protein serine/threonine phosphatase activity
0005218	intracellular ligand-gated calcium channel activity
0005219	ryanodine-sensitive calcium-release channel activity
0005220	inositol 1,4,5-triphosphate-sensitive calcium-release channel activity
0005227	calcium activated cation channel activity
0005229	intracellular calcium activated chloride channel activity
0005245	voltage-gated calcium channel activity
0005246	calcium channel regulator activity
0005262	calcium channel activity
0005388	calcium-transporting ATPase activity
0005432	calcium sodium antiporter activity
0005509	calcium ion binding
0005513	detection of calcium ion
0005544	calcium-dependent phospholipid binding
0005891	voltage-gated calcium channel complex
0005954	calcium- and calmodulin-dependent protein kinase complex
0006816	calcium ion transport
0006851	mitochondrial calcium ion transport
0006874	cellular calcium ion homeostasis
0007036	vacuolar calcium ion homeostasis

Table B.1. GO terms with the keyword “calcium” in Gene Ontology for Yeast - cont.

0007161	**calcium-independent cell-matrix adhesion**
0007204	elevation of cytosolic calcium ion concentration
0007223	Wnt receptor signaling pathway, calcium modulating pathway
0008048	calcium sensitive guanylate cyclase activator activity
0008086	light-activated voltage-gated calcium channel activity
0008087	light-activated voltage-gated calcium channel complex
0008273	calcium, potassium sodium antiporter activity
0008294	calcium- and calmodulin-responsive adenylate cyclase activity
0008331	high voltage-gated calcium channel activity
0008332	low voltage-gated calcium channel activity
0008377	light-induced release of internally sequestered calcium ion
0008427	calcium-dependent protein kinase inhibitor activity
0008597	calcium-dependent protein serine/threonine phosphatase regulator activity
0009931	calcium-dependent protein serine/threonine kinase activity
0015085	calcium ion transmembrane transporter activity
0015269	calcium-activated potassium channel activity
0015270	dihydropyridine-sensitive calcium channel activity
0015275	stretch-activated, cation-selective, calcium channel activity
0015278	calcium-release channel activity
0015279	store-operated calcium channel activity
0015368	calcium cation antiporter activity
0015369	calcium hydrogen antiporter activity
0016286	small conductance calcium-activated potassium channel activity
0016338	**calcium-independent cell-cell adhesion**
0016339	calcium-dependent cell-cell adhesion
0016340	calcium-dependent cell-matrix adhesion
0016931	vasopressin activated calcium mobilizing receptor activity
0017156	calcium ion-dependent exocytosis
0017158	regulation of calcium ion-dependent exocytosis
0019722	calcium-mediated signaling
0019855	calcium channel inhibitor activity

Table B.1. GO terms with the keyword “calcium” in Gene Ontology for Yeast - cont.

0021808	cytosolic calcium signaling involved in the initiation of cell movement in glial-mediated radial cell migration
0021945	calcium-mediated positive regulation of granule cell migration
0022849	glutamate-gated calcium ion channel activity
0022894	Intermediate conductance calcium-activated potassium channel activity
0030899	calcium-dependent ATPase activity
0030925	calcium incorporation into metallo-oxygen cluster
0030926	calcium incorporation into metallo-oxygen cluster via bis-L-aspartato tris-L-glutamato L-histidino calcium tetramanganese tetroxide
0030927	manganese incorporation into metallo-oxygen cluster via bis-L-aspartato tris-L-glutamato L-histidino calcium tetramanganese tetroxide
0032234	regulation of calcium ion transport via store-operated calcium channel
0032235	negative regulation of calcium ion transport via store-operated calcium channel
0032236	positive regulation of calcium ion transport via store-operated calcium channel
0032237	activation of store-operated calcium channel activity
0032468	Golgi calcium ion homeostasis
0032469	endoplasmic reticulum calcium ion homeostasis
0032470	elevation of endoplasmic reticulum calcium ion concentration
0032471	reduction of endoplasmic reticulum calcium ion concentration
0032472	Golgi calcium ion transport
0043006	calcium-dependent phospholipase A2 activation
0045955	negative regulation of calcium ion-dependent exocytosis
0045956	positive regulation of calcium ion-dependent exocytosis
0046586	regulation of calcium-dependent cell-cell adhesion
0046587	positive regulation of calcium-dependent cell-cell adhesion
0046588	negative regulation of calcium-dependent cell-cell adhesion
0046904	calcium oxalate binding
0047498	calcium-dependent phospholipase A2 activity
0047499	**calcium-independent phospholipase A2 activity**

Table B.1. GO terms with the keyword “calcium” in Gene Ontology for Yeast - cont.

0048101	calcium- and calmodulin-regulated 3',5'-cyclic-GMP phosphodiesterase activity
0048306	calcium-dependent protein binding
0048763	calcium-induced calcium release activity
0048791	calcium ion-dependent exocytosis of neurotransmitter
0048792	**calcium ion-independent exocytosis of neurotransmitter**
0050429	calcium-dependent phospholipase C activity
0050848	regulation of calcium-mediated signaling
0050849	negative regulation of calcium-mediated signaling
0050850	positive regulation of calcium-mediated signaling
0051040	**regulation of calcium-independent cell-cell adhesion**
0051041	**positive regulation of calcium-independent cell-cell adhesion**
0051042	**negative regulation of calcium-independent cell-cell adhesion**
0051208	sequestering of calcium ion
0051209	release of sequestered calcium ion into cytosol
0051279	regulation of release of sequestered calcium ion into cytoplasm
0051280	negative regulation of release of sequestered calcium ion into cytosol
0051281	positive regulation of release of sequestered calcium ion into cytosol
0051282	regulation of sequestering of calcium ion
0051283	negative regulation of sequestering of calcium ion
0051284	positive regulation of sequestering of calcium ion
0051480	cytosolic calcium ion homeostasis
0051481	reduction of cytosolic calcium ion concentration
0051482	elevation of cytosolic calcium ion concentration during G-protein signaling, coupled to IP3 second messenger (phospholipase C activating)
0051560	mitochondrial calcium ion homeostasis
0051561	elevation of mitochondrial calcium ion concentration
0051562	reduction of mitochondrial calcium ion concentration
0051563	smooth endoplasmic reticulum calcium ion homeostasis
0051564	elevation of smooth endoplasmic reticulum calcium ion concentration
0051565	reduction of smooth endoplasmic reticulum calcium ion concentration

Table B.1. GO terms with the keyword “calcium” in Gene Ontology for Yeast - cont.

0051592	response to calcium ion
0051924	regulation of calcium ion transport
0051925	regulation of calcium ion transport via voltage-gated calcium channel
0051926	negative regulation of calcium ion transport
0051927	negative regulation of calcium ion transport via voltage gated calcium channel
0051928	positive regulation of calcium ion transport
0051929	positive regulation of calcium ion transport via voltage gated calcium channel
0052065	positive regulation by organism of defense-related calcium ion flux in other organism during symbiotic interaction
0052102	positive regulation by symbiont of defense-related host calcium-dependent protein kinase pathway
0052162	modulation by symbiont of defense-related host calcium ion flux
0052168	modulation by symbiont of defense-related host calcium-dependent protein kinase pathway
0052265	induction by organism of defense-related calcium ion flux in other organism during symbiotic interaction
0052287	positive regulation by organism of defense-related calcium-dependent protein kinase pathway in other organism during symbiotic interaction
0052301	modulation by organism of defense-related calcium ion flux in other organism during symbiotic interaction
0052307	modulation by organism of defense-related calcium-dependent protein kinase pathway in other organism during symbiotic interaction
0052389	positive regulation by symbiont of defense-related host calcium ion flux
0052391	induction by symbiont of defense-related host calcium ion flux
0055074	calcium ion homeostasis
0060072	large conductance calcium-activated potassium channel activity

B.2. Annotation Tables for Yeast

The GO terms in grey shaded boxes are the terms of 12 core protein group.

Table B.2.1. Cellular Component Terms for Yeast

GO id	GO name
111	nucleotide-excision repair factor 2 complex
131	**incipient cellular bud site**
139	**Golgi membrane**
142	cellular bud neck contractile ring
221	vacuolar proton-transporting V-type ATPase, V1 domain
307	cyclin-dependent protein kinase holoenzyme complex
324	**vacuole, cell cycle-correlated morphology**
329	**membrane of vacuole with cell cycle-correlated morphology**
407	pre-autophagosomal structure
775	chromosome, pericentric region
778	condensed nuclear chromosome kinetochore
1950	PME fraction
5575	**cellular_component**
5622	**intracellular**
5624	membrane fraction
5625	soluble fraction
5634	**nucleus**
5643	nuclear pore
5737	**cytoplasm**
5739	mitochondrion
5741	mitochondrial outer membrane
5743	mitochondrial inner membrane
5768	endosome
5773	**vacuole**
5774	**vacuolar membrane**
5783	**endoplasmic reticulum**

Table B.2.1. Cellular Component Terms for Yeast - cont.

5788	endoplasmic reticulum lumen
5789	endoplasmic reticulum membrane
5794	**Golgi apparatus**
5802	trans-Golgi network
5816	spindle pole body
5823	**central plaque of spindle pole body**
5825	half bridge of spindle pole body
5829	cytosol
5853	eukaryotic translation elongation factor 1 complex
5856	cytoskeleton
5886	**plasma membrane**
5887	integral to plasma membrane
5933	cellular bud
5934	**cellular bud tip**
5935	**cellular bud neck**
5940	septin ring
5953	CAAX-protein geranylgeranyltransferase complex
5955	calcineurin complex
10008	endosome membrane
16020	**membrane**
16021	**integral to membrane**
16459	myosin complex
16469	proton-transporting two-sector ATPase complex
30176	integral to endoplasmic reticulum membrane
30427	site of polarized growth
30479	actin cortical patch
31902	late endosome membrane
31982	vesicle
32432	actin filament bundle
32541	cortical endoplasmic reticulum
33101	cellular bud membrane

Table B.2.1. Cellular Component Terms for Yeast - cont.

33178	proton-transporting two-sector ATPase complex, catalytic domain
33180	proton-transporting V-type ATPase, V1 domain
43332	**mating projection tip**
43529	GET complex
70390	transcription export complex 2

Table B.2.2. Molecular Function Terms for Yeast

GO id	GO name
166	**nucleotide binding**
287	magnesium ion binding
3674	**molecular_function**
3676	**nucleic acid binding**
3677	**DNA binding**
3700	**transcription factor activity**
3702	RNA polymerase II transcription factor activity
3712	transcription cofactor activity
3713	transcription coactivator activity
3746	translation elongation factor activity
3774	motor activity
3779	actin binding
3824	**catalytic activity**
3924	GTPase activity
4143	diacylglycerol kinase activity
4252	serine-type endopeptidase activity
4435	phosphoinositide phospholipase C activity
4437	inositol or phosphatidylinositol phosphatase activity
4555	alpha,alpha-trehalase activity
4571	mannosyl-oligosaccharide 1,2-alpha-mannosidase activity
4614	phosphoglucomutase activity
4629	phospholipase C activity
4659	prenyltransferase activity
4661	protein geranylgeranyltransferase activity
4662	CAAX-protein geranylgeranyltransferase activity
4672	protein kinase activity
4674	protein serine/threonine kinase activity
4683	calmodulin-dependent protein kinase activity
4693	cyclin-dependent protein kinase activity
4721	**phosphoprotein phosphatase activity**

Table B.2.2. Molecular Function Terms for Yeast - cont.

4723	**calcium-dependent protein serine/threonine phosphatase activity**
4871	signal transducer activity
5085	guanyl-nucleotide exchange factor activity
5089	Rho guanyl-nucleotide exchange factor activity
5096	GTPase activator activity
5097	Rab GTPase activator activity
5200	structural constituent of cytoskeleton
5216	**ion channel activity**
5227	**calcium activated cation channel activity**
5244	**voltage-gated ion channel activity**
5245	**voltage-gated calcium channel activity**
5262	**calcium channel activity**
5267	**potassium channel activity**
5272	**sodium channel activity**
5384	manganese ion transmembrane transporter activity
5388	**calcium-transporting ATPase activity**
5471	ATP ADP antiporter activity
5488	binding
5509	**calcium ion binding**
5515	**protein binding**
5516	**calmodulin binding**
5524	**ATP binding**
5525	GTP binding
5529	sugar binding
5543	phospholipid binding
5544	calcium-dependent phospholipid binding
8047	enzyme activator activity
8081	phosphoric diester hydrolase activity
8117	sphinganine-1-phosphate aldolase activity
8233	peptidase activity
8236	serine-type peptidase activity

Table B.2.2. Molecular Function Terms for Yeast - cont.

8270	**zinc ion binding**
8289	lipid binding
8324	**cation transmembrane transporter activity**
8481	sphinganine kinase activity
8597	**calcium-dependent protein serine/threonine phosphatase regulator activity**
8934	inositol-1(or 4)-monophosphatase activity
15085	**calcium ion transmembrane transporter activity**
15093	ferrous iron transmembrane transporter activity
15275	**stretch-activated, cation-selective, calcium channel activity**
15410	**manganese-transporting ATPase activity**
15662	**ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism**
16301	kinase activity
16538	cyclin-dependent protein kinase regulator activity
16740	transferase activity
16787	**hydrolase activity**
16798	hydrolase activity, acting on glycosyl bonds
16820	**hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances**
16829	lyase activity
16831	carboxy-lyase activity
16853	isomerase activity
16868	intramolecular transferase activity, phosphotransferases
16887	ATPase activity
17050	D-erythro-sphingosine kinase activity
30170	pyridoxal phosphate binding
30234	enzyme regulator activity
30246	carbohydrate binding
30346	**protein phosphatase 2B binding**
30674	protein binding, bridging

Table B.2.2. Molecular Function Terms for Yeast - cont.

31489	myosin V binding
32038	myosin II heavy chain binding
42392	sphingosine-1-phosphate phosphatase activity
43130	ubiquitin binding
43495	protein anchor
43565	**sequence-specific DNA binding**
45159	myosin II binding
46872	**metal ion binding**
46933	hydrogen ion transporting ATP synthase activity, rotational mechanism
46961	hydrogen ion transporting ATPase activity, rotational mechanism
48306	**calcium-dependent protein binding**
51015	actin filament binding
51082	unfolded protein binding

Table B.2.3. Biological Process Terms for Yeast

GO id	GO name
1	mitochondrion inheritance
3	**reproduction**
79	regulation of cyclin-dependent protein kinase activity
83	G1/S-specific transcription in mitotic cell cycle
122	negative regulation of transcription from RNA polymerase II promoter
147	actin cortical patch assembly
183	chromatin silencing at rDNA
282	cellular bud site selection
742	**karyogamy during conjugation with cellular fusion**
746	conjugation
747	conjugation with cellular fusion
750	pheromone-dependent signal transduction during conjugation with cellular fusion
753	cellular morphogenesis during conjugation with cellular fusion
754	**adaptation to pheromone during conjugation with cellular fusion**
910	cytokinesis
915	cytokinesis, contractile ring formation
916	cytokinesis, contractile ring contraction
1402	signal transduction during filamentous growth
1403	invasive growth (sensu Saccharomyces)
5975	carbohydrate metabolic process
5978	glycogen biosynthetic process
5991	trehalose metabolic process
5992	trehalose biosynthetic process
5993	trehalose catabolic process
6006	glucose metabolic process
6011	UDP-glucose metabolic process
6033	chitin localization
6066	alcohol metabolic process
6109	regulation of carbohydrate metabolic process

Table B.2.3. Biological Process Terms for Yeast - cont.

6112	energy reserve metabolic process
6350	**transcription**
6355	**regulation of transcription, DNA-dependent**
6366	transcription from RNA polymerase II promoter
6390	transcription from mitochondrial promoter
6412	translation
6414	translational elongation
6457	protein folding
6468	protein amino acid phosphorylation
6486	**protein amino acid glycosylation**
6487	protein amino acid N-linked glycosylation
6508	proteolysis
6525	arginine metabolic process
6607	**NLS-bearing substrate import into nucleus**
6629	lipid metabolic process
6661	**phosphatidylinositol biosynthetic process**
6665	sphingolipid metabolic process
6688	glycosphingolipid biosynthetic process
6754	**ATP biosynthetic process**
6810	**transport**
6811	**ion transport**
6812	**cation transport**
6816	**calcium ion transport**
6828	**manganese ion transport**
6839	mitochondrial transport
6873	**cellular ion homeostasis**
6874	**cellular calcium ion homeostasis**
6875	cellular metal ion homeostasis
6879	cellular iron ion homeostasis
6886	intracellular protein transport
6887	**exocytosis**

Table B.2.3. Biological Process Terms for Yeast - cont.

6890	retrograde vesicle-mediated transport, Golgi to ER
6897	endocytosis
6898	**receptor-mediated endocytosis**
6903	vesicle targeting
6914	**autophagy**
6950	**response to stress**
6970	response to osmotic stress
6974	response to DNA damage stimulus
6986	response to unfolded protein
7005	mitochondrion organization and biogenesis
7010	**cytoskeleton organization and biogenesis**
7015	actin filament organization
7020	microtubule nucleation
7029	endoplasmic reticulum organization and biogenesis
7035	vacuolar acidification
7039	**vacuolar protein catabolic process**
7047	**cell wall organization and biogenesis**
7049	cell cycle
7067	mitosis
7096	regulation of exit from mitosis
7103	spindle pole body duplication in nuclear envelope
7114	**cell budding**
7118	budding cell apical bud growth
7119	budding cell isotropic bud growth
7120	axial cellular bud site selection
7121	bipolar cellular bud site selection
7124	pseudohyphal growth
7165	signal transduction
7205	protein kinase C activation
7264	small GTPase mediated signal transduction
7323	peptide pheromone maturation

Table B.2.3. Biological Process Terms for Yeast - cont.

7329	positive regulation of transcription from RNA polymerase II promoter by pheromones
8104	protein localization
8150	**biological_process**
8152	**metabolic process**
8298	intracellular mRNA localization
9060	aerobic respiration
9117	**nucleotide metabolic process**
9266	response to temperature stimulus
9267	cellular response to starvation
9395	phospholipid catabolic process
9408	response to heat
9851	auxin biosynthetic process
9966	**regulation of signal transduction**
10675	protein transport
15031	ferrous iron transport
15684	organic acid transport
15849	ADP transport
15866	ATP transport
15867	ATP synthesis coupled proton transport
15986	proton transport
15992	lipid catabolic process
16042	vesicle-mediated transport
16192	microautophagy
16237	**positive regulation of macroautophagy**
16239	negative regulation of macroautophagy
16242	protein amino acid geranylgeranylation
18348	response to pheromone
19236	glucose 1-phosphate metabolic process
19255	galactose catabolic process
19388	calcium-mediated signaling

Table B.2.3. Biological Process Terms for Yeast - cont.

19722	**carboxylic acid metabolic process**
19752	sexual reproduction
19953	**cellular cation homeostasis**
23034	establishment of cell polarity
30003	**cellular manganese ion homeostasis**
30010	sphingolipid biosynthetic process
30026	ER-associated protein catabolic process
30148	sporulation
30433	filamentous growth
30435	spore wall assembly (sensu Fungi)
30447	unfolded protein response
30476	microtubule organizing center organization and biogenesis
30968	chitin- and beta-glucan-containing cell wall organization and biogenesis
31023	regulation of protein stability
31505	**CVT pathway**
31647	regulation of Rab GTPase activity
32258	regulation of establishment and/or maintenance of cell polarity
32313	regulation of protein localization
32878	inositol phosphate biosynthetic process
32880	regulation of Rho protein signal transduction
32958	vacuole fusion, non-autophagic
34501	ribosome biogenesis and assembly
34613	response to drug
35023	transcription factor import into nucleus
42144	**proteasomal ubiquitin-dependent protein catabolic process**
42254	regulation of carbohydrate biosynthetic process
42493	negative regulation of transcription factor activity
42991	**regulation of generation of precursor metabolites and energy**
43161	cellular lipid metabolic process
43255	cellular carbohydrate metabolic process
43433	protein insertion into ER membrane

Table B.2.3. Biological Process Terms for Yeast - cont.

43467	establishment of protein localization
44255	regulation of transcription
44262	negative regulation of glycogen biosynthetic process
45048	negative regulation of phosphate metabolic process
45184	positive regulation of transcription from RNA polymerase II promoter
45449	**negative regulation of calcium ion-dependent exocytosis**
45719	ATP metabolic process
45936	inositol phosphate dephosphorylation
45944	mitochondrion transport along microtubule
45955	inositol lipid-mediated signaling
46034	organelle inheritance
46855	regulation of catalytic activity
47497	ion homeostasis
48017	negative regulation of calcium-mediated signaling
48308	mRNA transport
50790	regulation of transport
50801	**glucose 6-phosphate metabolic process**
50849	cofactor transport
51028	spindle pole body organization and biogenesis
51049	**cell division**
51156	glucose 6-phosphate metabolic process
51181	**cofactor transport**
51300	**spindle pole body organization and biogenesis**
51301	**cell division**
55085	**transmembrane transport**

B.3. Homology & Conservation Search of 12 Core Proteins for Yeast

Table B.3.1. Homologs of 12 Yeast Core Proteins through Species

	C.ELEGANS (worm)			D.MELANOGASTER (Fruit Fly)			D.RERIO (Zebra Fish)			X.LAEVIS (Frog)			M.MUSCULUS (Mouse)			H.SAPIENS (Human)		
Yeast	*% id.*	*e-val*	*prot name*	*% id.*	*e-val*	*prot name*	*% id.*	*e-val*	*prot name*	*% id.*	*e-val*	*prot name*	*% id.*	*e-val*	*prot name*	*% id.*	*e-val*	*prot name*
CMD1	59,9	4,58E-50	CalModulin family member (cmd-1)	59,9	4,82E-50	calmodulin	59,2	2,07E-49	calmodulin 2-like	46,9	1,84E-37	Cam protein	59,9	1,11E-49	calmodulin	59,9	7,07E-50	calmodulin
CNB1	52,6	5,24E-49	CalciNeurin B family member (cnb-1)	52	6,09E-47	Calcineurin subunit B type 1	52,6	2,80E-48	protein phosphatase 3	52	4,43E-48	CanB protein	52,6	1,20E-48	Calcineurin subunit B type 1	52,6	9,95E-49	calcineurin subunit B type 1
CNA1	47,9	8,44E-155	Serine/threonine-protein phosphatase (tax-6)	49,5	2,06E-157	calcineurin A1	46,5	3,03E-145	serine/threonine-protein phosphatase 2B	47,7	1,57E-148	Serine/threonine-protein phosphatase	47,4	4,75E-149	Serine/threonine-protein phosphatase 2B	47	9,90E-149	serine/threonine-protein phosphatase 2B
CMP2	47	1,20E-153	Serine/threonine-protein phosphatase (tax-6)	47,8	1,28E-157	calcineurin A2	44,7	2,74E-147	Serine/threonine-protein phosphatase	44,9	6,19E-153	Serine/threonine-protein phosphatase	44,5	1,62E-150	Serine/threonine-protein phosphatase 2B catalytic subunit beta isoform	45,2	1,47E-154	serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform isoform 1
CRZ1	6	7,15E-13	Zinc finger Transcription Factor family member (ztf-23)	7,2	4,48E-13	CG9797	7,8	1,29E-12	Zinc finger protein 569-like	8,4	7,46E-14	Fez family zinc finger protein 2	7,7	1,15E-13	zinc finger protein 90	8	1,48E-13	zinc finger protein 397OS
VCX1				7,5	2,50E-01	Nckx30C, isoform A	23,8	9,65E-29	type II CAX cation/proton exchanger	10,5	1,96E+00	solute carrier family 24 (sodium/potassium/calcium exchanger), member 6	9,5	1,54E+00	solute carrier family 24 (sodium/potassium/calcium exchanger), member 3	8	1,25E+00	sodium/potassium/calcium exchanger 3 precursor
PMR1	45,3	0,00E+00	PMR-type Golgi ATPase family member (pmr-1)	48,3	0,00E+00	secretory pathway calcium atpase, isoform D	33,8	1,01E-111	ATPase, Na+/K+ transporting, alpha 1a.3 polypeptide	47,4	0,00E+00	ATPase, Ca++ transporting, type 2C, member 2	47,9	0,00E+00	calcium-transporting ATPase type 2C member 1	47,8	0,00E+00	calcium-transporting ATPase type 2C member 1 isoform
MID1				3,8	2,92E+00	middleman of seventy-eight signaling							5,8	8,94E-01	mCG3819	3,1	1,48E+00	TPA_inf: olfactory receptor OR14-7
PMC1	35,5	2,56E-146	calcium ATPase	22,8	5,21E-80	sarco/endoplasmic reticulum-type Ca^{-2+}-ATPase	34,8	1,15E-156	plasma membrane calcium ATPase 4	24,5	7,85E-79	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2	35	1,34E-154	plasma membrane calcium-transporting ATPase 2 isoform 1	34,9	1,89E-154	plasma membrane calcium-transporting ATPase 2 isoform 2
RCN1	16,6	1,61E+00	Regulator of CalciNeurin family member (rcn-1)	10	2,61E-01	CG3209, isoform B	13,3	2,59E-02	regulator of calcineurin family member 3	9	3,74E+00	enhancer of split related epidermal protein-6	13,3	3,01E+00	calcipressin-2 isoform 2	12,3	7,35E-01	regulator of calcineurin 3 isoform 2,3,4b,5

Table B.3.1. Homologs of 12 Yeast Core Proteins through Species - cont.

YVC1	9,3	4,15E-01	TRP (transient receptor potential) channel family member (trp-2)	11,4	1,68E-04	TRPgamma cation channel short form	11,7	8,50E-03	transient receptor potential cation channel, subfamily A, member 1a	6,7	1,70E-05	calcium transporter 2	17,3	1,12E-01	short transient receptor potential channel 7	17,5	1,49E-01	short transient receptor potential channel 7 isoform 1
CCH1	15,20	3,95E-57	four domain-type voltage-gated ion channel alpha-1 subunit	8,53	1,41E-49	alpha1U protein	10,79	8,92E-34	calcium channel, voltage-dependent, L type, alpha 1S subunit, b				10,74	1,26E-43	voltage-dependent L-type calcium channel subunit alpha-1C isoform 2	10,84	3,09E-39	L-type calcium channel (HFCC)

Table B.3.2. Conserved Residues of 12 Yeast Core Proteins and their Domains obtained by ConSeq

ORF name	Protein name	Tot. Res.	Conservation		Residues	Exposed	Residues	Buried	Functionally important res.	Structurally important res.	Func/conserved	Struc/conserved	Func/total	Struc/total	Func/exposed	Struc/buried
			Total	% cons	exposed	%	buried	%			%	%	%	%	%	%
YBR109C	CMD1	147	37	25,2	101	68,7	46	31,3	22	7	59,5	18,9	15,0	4,8	21,8	15,2
	EF hand	29	11	37,9	18	62,1	11	37,9	7	3	63,6	27,3	24,1	10,3	38,9	27,3
	EF hand	29	6	20,7	17	58,6	12	41,4	2	2	33,3	33,3	6,9	6,9	11,8	16,7
YDL128W	VCX1	411	109	26,5	152	37,0	259	63,0	22	51	20,2	46,8	5,4	12,4	14,5	19,7
	Na Ca ex	150	53	35,3	50	33,3	100	66,7	13	23	24,5	43,4	8,7	15,3	26,0	23,0
	Na Ca ex	134	49	36,6	26	19,4	108	80,6	9	25	18,4	51,0	6,7	18,7	34,6	23,1
YGL006W	PMC1	1173	464	39,6	652	55,6	521	44,4	173	207	37,3	44,6	14,7	17,6	26,5	39,7
	Cation_ATPase_N	71	23	32,4	45	63,4	26	36,6	9	8	39,1	34,8	12,7	11,3	20,0	30,8
	E1-E2_ATPase	277	133	48,0	136	49,1	141	50,9	49	56	36,8	42,1	17,7	20,2	36,0	39,7
	Hydrolase	135	71	52,6	74	54,8	61	45,2	32	36	45,1	50,7	23,7	26,7	43,2	59,0
	Cation_ATPase_C	196	84	42,9	70	35,7	126	64,3	22	43	26,2	51,2	11,2	21,9	31,4	34,1
YGL167C	PMR1	950	277	29,2	488	51,4	462	48,6	117	110	42,2	39,7	12,3	11,6	24,0	23,8
	Cation_ATPase_N	71	10	14,1	42	59,2	29	40,8	5	5	50,0	50,0	7,0	7,0	11,9	17,2
	E1-E2_ATPase	243	91	37,4	117	48,1	126	51,9	36	37	39,6	40,7	14,8	15,2	30,8	29,4
	Hydrolase	197	67	34,0	112	56,9	85	43,1	25	29	37,3	43,3	12,7	14,7	22,3	34,1
	Cation_ATPase_C	174	37	21,3	66	37,9	108	62,1	16	11	43,2	29,7	9,2	6,3	24,2	10,2
YGR217W	CCH1	2039	590	28,9	1087	53,3	952	46,7	203	225	34,4	38,1	10,0	11,0	18,7	23,6
	Ion trans	245	120	49,0	104	42,4	141	57,6	39	51	32,5	42,5	15,9	20,8	37,5	36,2
	Ion trans	194	69	35,6	48	24,7	146	75,3	12	30	17,4	43,5	6,2	15,5	25,0	20,5
	Ion trans	229	112	48,9	78	34,1	151	65,9	29	50	25,9	44,6	12,7	21,8	37,2	33,1
	Ion trans	205	92	44,9	66	32,2	139	67,8	32	43	34,8	46,7	15,6	21,0	48,5	30,9

Table B.3.2. Conserved Residues of 12 Yeast Core Proteins and their Domains obtained by ConSeq - cont.

YKL159C	RCN1	211	43	20,4	146	69,2	65	30,8	27	12	62,8	27,9	12,8	5,7	18,5	18,5
	Calcipressin	168	32	19,0	113	67,3	55	32,7	22	7	68,8	21,9	13,1	4,2	19,5	12,7
YKL190W	CNB1	175	43	24,6	118	67,4	57	32,6	25	7	58,1	16,3	14,3	4,0	21,2	12,3
YLR433C	CNA1	553	221	40,0	326	59,0	227	41,0	103	98	46,6	44,3	18,6	17,7	31,6	43,2
	Metallophos	210	110	52,4	98	46,7	112	53,3	48	55	43,6	50,0	22,9	26,2	49,0	49,1
YML057W	CMP2/CNA2	604	219	36,3	367	60,8	237	39,2	100	101	45,7	46,1	16,6	16,7	27,2	42,6
	Metallophos	227	112	49,3	110	48,5	117	51,5	49	56	43,8	50,0	21,6	24,7	44,5	47,9
YNL027W	CRZ1	678														
	Zf_C2H2															

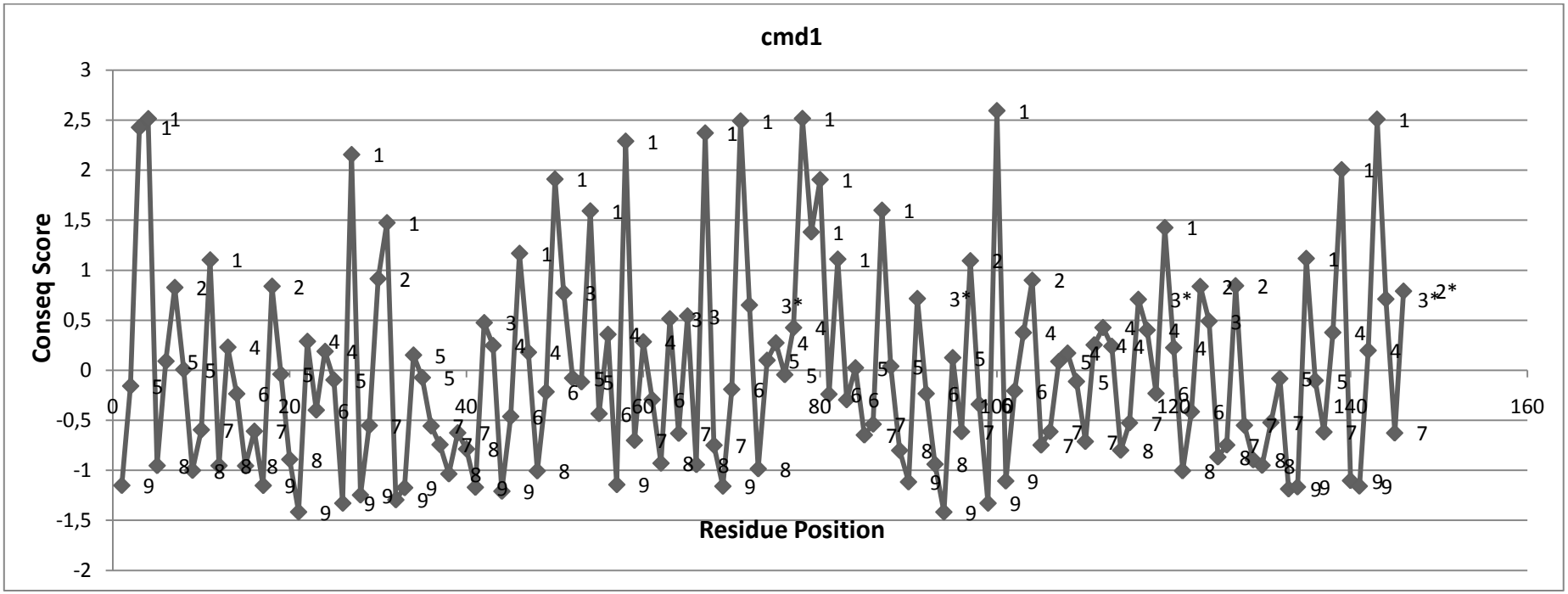


Figure B.3.1. Conseq scores vs Residue position for cmd1. The more negative score a residue has, the more it is conserved.

APPENDIX C

C.1. Hub Proteins of Yeast PIN with 12 core

SSB1/YDL229W:

Cytoplasmic ATPase that is a ribosome-associated molecular chaperone, functions with J-protein partner Zuo1p; may be involved in folding of newly-made polypeptide chains; member of the HSP70 family; interacts with phosphatase subunit Reg1p.

Heat shock protein may aid in the passage of the nascent polypeptide chain through the ribosome channel into the cytosol. Such an interaction could be crucial for continuous transport of the polypeptide; could serve to prevent the nascent polypeptide from interfering with translation by clogging the ribosome channel. (Uniprot)

Many chaperones are heat shock proteins, that is, proteins expressed in response to elevated temperatures or other cellular stresses. The reason for this behaviour is that protein folding is severely affected by heat and, therefore, some chaperones act to repair the potential damage caused by misfolding. Other chaperones are involved in folding newly made proteins as they are extruded from the ribosome. Although most newly synthesized proteins can fold in absence of chaperones, a minority strictly requires them. (Wikipedia)

The most abundant Hsp70s in the budding yeast Saccharomyces cerevisiae are the cytosolic subfamilies of Ssa and Ssb.

Cytosolic Hsp70 is encoded by two close homologs, SSB1 and SSB2, which are functionally redundant. It is commonly accepted that Ssb, which can interact with nascent polypeptides in the context of the ribosome also assists the folding of proteins. The current model suggests that ribosome-bound Ssb functions early during protein biogenesis and Ssa takes over, after synthesis is complete and newly synthesized proteins are released from the

ribosome. Recent results revealed that Ssb does not enhance the productive folding of structurally destabilized proteins in vivo. Moreover, the Dssb1Dssb2 strain displays a strong growth defect at low temperature and growth approaches that of wild type when the temperature is raised. The transcription of SSB1–SSB2 is not induced but rather diminished at elevated temperatures. Instead, Ssb levels increase when late log-phase cells are transferred to fresh glucose-containing medium and decrease during late stages of the stationary phase. There is evidence that Ssb might function in processes other than protein folding. For instance, interaction of Ssb with the heat-shock transcription factor HSF1 was shown to down-regulate the activity of the transcription factor. It is suggested that Ssb binding sterically interferes with the activation domains of HSF1.

The kinase SNF1 is one of the major components of the yeast glucose-sensing system. SNF1 is activated in response to glucose starvation and other stresses, however, not in response to heat shock. Ssb, most likely in concert with a bona fide Glc7 targeting subunit, affects the dephosphorylation of Snf1. (Plehve et al., 2009)

HSP82/YPL240C:

Hsp90 chaperone required for pheromone signaling and negative regulation of Hsf1p; docks with Tom70p for mitochondrial preprotein delivery; promotes telomerase DNA binding and nucleotide addition; interacts with Cns1p, Cpr6p, Cpr7p, Sti1p.

In S. cerevisiae, the molecular chaperone HSP90 exists as two isoforms, encoded by the genes HSC82 and HSP82. HSP90 family member activity is required for folding a specific set of difficult-to-fold proteins from nascent polypeptides into biologically active structures as well as for the refolding of denatured proteins back into native conformations. Although most cellular proteins do not require Hsp82p/Hsc82p chaperone activity for correct folding under normal conditions, Hsp82p and Hsc82p are required for the activation of many key cellular regulatory and signaling proteins, like kinases and transcription factors, such as Swe1p, Gcn2p, and Hap1p.

HSP82 and HSC82 share ~97% sequence identity and together, the encoded proteins comprise 1-2% of all the protein in the cytosol. While HSC82 is expressed constitutively at

high levels and only slightly induced by heat shock, HSP82 transcription is strongly induced by both heat and stress. Increased HSP82 expression is mediated by the transcription factor Hsf1p, which binds cooperatively to three heat shock elements (HSE) in the HSP82 promoter. It has been shown that Hsp1p-HSE interaction is also required for stimulating HSP82 basal transcription.

Hsp82p and Hsc82p associate with many co-chaperones which both positively and negatively regulate Hsp82p/Hsc82p function. Hsp82p/Hsc82p co-chaperones include Sti1p, Cdc37p, Cns1p, Sba1p, Cpr6p, Cpr7p, Sse1p, Hch1p, and Aha1p. Cdc37p is required for stress-induced high-osmolarity glycerol and protein kinase C mitogen-activated protein kinase pathway functionality by interaction with Hog1p and Slt2p (Mpk1p) (Hawle P, et al., 2007). One common method by which these partner proteins regulate Hsp82p/Hsc82p activity is through the inhibition/enhancement of ATP hydrolysis. Mechanistically, this can occur through direct interference of ATP binding or by alteration of Hsp82p/Hsc82p protein conformation such that ATP binding is affected.

HSP82 and HSC82 are highly conserved among eukaryotes and the presence of at least one of the HSP90 gene product family members is essential for viability in yeast, Drosophila, and humans. Overexpression of HSP82 in S. cerevisiae has been shown to increase the virulence of the yeast in mice. Human HSP90 (OMIM and OMIM) is being explored as a target for cancer therapeutics due to its role in folding oncogenic protein kinases and because inhibition of chaperone activity by ATP analogs has been shown to promote substrate protein degradation via the ubiquitin-dependent proteasomal pathway.

CDC28/YBR160W:

Catalytic subunit of the main cell cycle cyclin-dependent kinase (CDK); alternately associates with G1 cyclins (CLNs) and G2/M cyclins (CLBs) which direct the CDK to specific substrates.

Cdc28 is the catalytic subunit of the main cell cycle cyclin-dependent kinase (CDK). Homologs include CDK1 in animals and cdc2 in S. pombe. Waves of CDK activity drive events of the cell cycle through phosphorylation of key substrates. To accomplish these

waves of activity, Cdc28p associates with different regulators throughout the cell cycle. The expression of several of these regulators is periodic which serves to limit their window of action to the proper time in the cell cycle.

The first wave of CDK activity occurs when Cdc28p associates with G1 cyclins and Cks1p. The G1 cyclins are Cln1p, Cln2p, and Cln3p. Cln3p/Cdc28p activity is required for setting the size threshold at which cells pass through START (commitment to duplication and division). Once committed Cln3p/Cdc28p inactivates a repressor of G1 transcription, Whi5p, which in turn leads to active SBF (Swi4p-Swi6p) and MBF (Mbp1p-Swi6p), transcription factors that promote transcription of CLN1, CLN2, and genes required for S-phase. Cln1p and Cln2p are important for initiating polarized growth at the site of bud emergence, promoting spindle pole body (SPB) duplication and inhibiting Sic1p and Cdh1p, two CDK inhibitors. During G1, Sic1p binds and inhibits the growing pool of Cdc28p/B-type cyclin complexes. Towards the end of G1, Cdc28p/Cln1p and Cdc28p/Cln2p complexes phosphorylate Sic1p and target it for degradation. The absence of Sic1p allows a wave of CDK/B-cyclin activity that drives DNA replication and entry into mitosis.

B-type cyclins Clb1p, Clb2p, Clb3p, Clb4p, Clb5p, and Clb6p regulate Cdc28p during S, G2, and M phases. Cdc28p association with Clb5p and Clb6p drives DNA replication. Association with Clb3p, Clb4p, and Clb5p promotes maturation and separation of spindle pole bodies, and proper spindle segregation while Cdc28p association with Clb2p (and to some extent Clb1p, Clb3p, and Clb4p) promotes entry into mitosis and triggers a switch in bud growth from polarized to isotropic. The metaphase to anaphase transition occurs when securin (Pds1p), an inhibitor of DNA segregation is destroyed by the proteosome. Mitotic CDK activity is required to target Pds1p for degradation by directly phosphorylating Pds1p and activating the Anaphase Promoting Complex/Cyclosome (APC/C).

Once DNA is segregated, exit from mitosis (spindle disassembly, cytokinesis and transition to the next G1) requires that mitotic CDK activity be turned off. This is accomplished by degradation of mitotic cyclins and inhibition of remaining mitotic activity by Sic1p. In the absence of mitotic CDK activity, G1 cyclins can once again accumulate.

In addition to being regulated by binding partners, Cdc28p is regulated by post-translational modifications. Cak1p phosphorylation of Cdc28p on threonine 169 is essential for CDK activity and is thought to precede cyclin binding. Phosphorylation of Cdc28p on tyrosine 19 (Y19) by Swe1p kinase (wee1 in S. pombe) inhibits mitotic CDK activity and hence, entry into mitosis. This phosphorylation is removed by the phosphatase Mih1p (homolog of cdc25 in S. pombe). Phosphorylation of Y19 is critical for enforcing the morphogenesis checkpoint. When cells experience an environmental perturbation that disrupts bud formation, the morphogenesis checkpoint delays entry into mitosis until a bud is formed. The checkpoint impinges on Swe1p and Mih1p to inactivate the mitotic CDK, insuring that mitosis does not occur in an unbudded cell. Unlike in S. pombe and metazoans, Swe1p and/or Mih1p are not direct targets of the spindle checkpoints or DNA checkpoints.

Developmental programs such as mating, meiosis and sporulation, and pseudohyphal growth require alterations in cell cycle control. For example, during mating, pheromone-dependent inhibition of Cln/Cdc28p complexes by Far1p arrests cells in G1 so cell-cell fusion can occur. When sporulation is induced, cells enter meiosis from G1, but CLN1 and CLN2 are repressed by a mechanism that makes meiosis and mitosis incompatible.

TPK1/YJL164C:

cAMP-dependent protein kinase catalytic subunit; promotes vegetative growth in response to nutrients via the Ras-cAMP signaling pathway; inhibited by regulatory subunit Bcy1p in the absence of cAMP; partially redundant with Tpk2p and Tpk3p.

TPK1, TPK2, and TPK3 encode isoforms of the catalytic subunit of cAMP-dependent protein kinase (PKA), the effector kinase of the Ras-cAMP signaling pathway. Through phosphorylation of various targets, PKA activity regulates processes involved in cell growth and response to nutrients and stress, such as nutrient sensing, energy metabolism, carbohydrate utilization, cell cycle progression, thermotolerance, osmotic shock tolerance, sporulation, bud site selection, pseudohyphal growth, aging, and

autophagy. PKA substrates include transcription factors (e.g., Rap1p, Hsf1p, Adr1p, Msn2p / Msn4p, Ssn2p), metabolic enzymes (e.g., Cho1p, Pfk2p, Nth1p), and other regulatory kinases (e.g., Rim15p, Atg1p). The PKA target site is R-[KR]-x-S, where S is the site of substrate phosphorylation. PKA is conserved from yeast to man and TPK homologs have been identified in fission yeast, flies, worms, mice, pigs, cows, and humans.

In the absence of cAMP, the catalytic subunits (Tpk1p, 2p, or 3p) form an inactive heterotetrameric complex with the PKA regulatory subunit Bcy1p, with two catalytic subunits bound to two regulatory subunits. In the presence of cAMP, Bcy1p binds to cAMP and dissociates from the complex as a homodimer, releasing the two catalytic subunits as active monomers. PKA is also regulated through a cAMP feedback inhibition loop: PKA activity leads to a reduction in cAMP levels by directly phosphorylating enzymes which regulate cAMP production, (e.g. the GTPase Ras2p and the cAMP phosphodiesterase Pde1p), and the resulting decrease in cAMP levels promotes formation of the inactive heterotetrameric complex. All three Tpk isozymes are phosphorylated and the phosphorylation state of Tpk1p has been shown to affect its substrate specificity constant and to be correlated with the availability of a fermentable carbon source.

The three PKA catalytic subunits are highly similar only in their C-terminal regions. Although the C-terminal 320 amino acids of Tpk1p and Tpk2p share 77% identity and are 88% and 75% identical to Tpk3p respectively, the N-terminal regions of these proteins share no sequence similarity and are heterogeneous in length. While no single TPK gene is essential, the presence of at least one of the isomers is required for normal growth. Mutations affecting levels of PKA activity lead to pleiotropic phenotypes because of the wide range of PKA substrates and their involvement in many cellular processes. Phenotypes due to hyperactive PKA signaling include rapamycin resistance, accelerated pseudohyphal growth, transient cell cycle arrest in G1, and increased sensitivity to stresses such as heat shock and nutrient starvation while some of the phenotypes resulting from reduced PKA activity are hyperaccumulation of cAMP, defective glucose repression, and decreased levels of ribosomal protein mRNAs.

In rapidly proliferating cells, Bcy1p and Tpk1p are predominantly localized in the nucleus. In these cells, cAMP activation of PKA results in Tpk1p translocation to the cytoplasm while Bcy1p remains in the nucleus. In nonproliferating cells or cells growing on a non-fermentable carbon source, Bcy1p and Tpk1p are present in both the nucleus and the cytoplasm (33). Nuclear versus cytoplasmic localization of Tpk1p is also partially regulated by the TOR pathway.

Although the Tpk isomers are functionally redundant for cell viability, they appear to regulate different processes. One high-throughput study showed that different numbers of distinct proteins were targeted by Tpk1p, Tpk2p, and Tpk3p and greater than 87% of all PKA substrates were uniquely phosphorylated by just one of the Tpks. In addition to differing protein targets, the three isomers also recognize and affect the transcription of different sets of gene targets. Tpk1p was shown to bind the coding regions of most actively transcribed genes and in particular is required for the expression of genes involved in branched chain amino acid biosynthesis.

PTK2/YJR059W:

Putative serine/threonine protein kinase involved in regulation of ion transport across plasma membrane; enhances spermine uptake.

Ptk2 is an intrinsic plasma membrane protein although the exact nature of its membrane association is not clear. Ptk2 is the kinase responsible for the in vivo phosphorylation of Ser899 during glucose activation of Pma1.

The exact molecular mechanisms and events leading to Pma1 activation by glucose are not yet completely understood. It has been suggested that glucose activation of ATPase is modulated by a combination of a down-regulating CK1- dependent phosphorylation of the MgATP-binding domain and another up-regulating phosphorylation(s) of the C-terminal domain

SSA1/YAL005C:

ATPase involved in protein folding and nuclear localization signal (NLS)-directed nuclear transport; member of heat shock protein 70 (HSP70) family; forms a chaperone complex with Ydj1p; localized to the nucleus, cytoplasm, and cell wall.

SSA1, SSA2, SSA3, and SSA4 encode chaperone proteins that comprise the S. cerevisiae SSA subfamily of cytosolic HSP70 proteins. HSP70 is a large family of proteins that has been evolutionarily conserved from bacteria (DnaK) to humans (HSP72/73). HSP70 proteins were originally classified based upon their induction by heat shock and their size of ~70kDa. The main function of these proteins is to serve as molecular chaperones, binding newly-translated proteins to assist in proper folding and prevent aggregation/misfolding. In yeast, HSP70s are also involved in disassembling aggregates of misfolded proteins, translocating select proteins into the mitochondria and ER, degrading aberrant proteins, and regulating the expression of other heat shock proteins. S. cerevisiae has at least 9 cytosolic forms of HSP70 (SSA1, SSA2, SSA3, SSA4, SSB1, SSB2, SSE1, SSE2, SSZ1), 2 HSP70s which are found in the endoplasmic reticulum (KAR2, LHS1), and 3 mitochondrial HSP70s (SSC1, SSQ1, ECM10).

The 4 genes of the SSA subfamily are closely related, with Ssa1p sharing 99%, 84%, and 85% amino acid identity with Ssa2p, Ssa3p, and Ssa4p, respectively. Although SSA1 basal expression is significant under normal growth conditions, transcription is also upregulated before diauxic shift or upon heat shock. Increased expression after heat shock is mediated by the transcriptional activator Hsf1p, which recognizes and binds two heat shock elements (HSE) in the SSA1 promoter. Of the two HSEs, only one contributes to basal expression. Although the majority of Ssa protein is found in the cytosol, Ssa1p and Ssa2p can also be detected in the cell wall. Additionally, Ssa1p and Ssa2p have been implicated in DNA-damage as they have been identified as members of Rad9 DNA-checkpoint complexes.

Most of the structural knowledge of the S. cerevisiae HSP70 proteins is based on experimental evidence from bacterial DnaK, mammalian HSP70, and Ssa1p. All Hsp70s contain an N-terminal ATPase domain and a C-terminal peptide binding domain. ATPase

activity of HSP70s is intrinsically weak but can be enhanced by interaction with DnaJ/HSP40 proteins. It has been shown for Ssa1p, and based on similarity is implicated for the remaining Ssa subfamily, that activity is stimulated by interaction with the DnaJ/HSP40 co-chaperones Ydj1p, Sis1p, Sti1p, and Cns1p. Substrate binding is regulated by ATP turnover; in the presence of ATP, peptide exchange is rapid and the binding constant is low while when ADP is bound, peptide exchange is slower and the substrate affinity higher. The rate of Ssa protein ATP/ADP exchange is stimulated by the nucleotide exchange factors Fes1p and Snl1p.

The effect of SSA1 expression has also been studied in yeast models of human prion disease such as Creutzfeldt-Jakob disease (OMIM). For the [PSI+] prion (an isoform of Sup35p), overexpression of any one of the Ssa proteins promotes prion formation and suppresses the ability of Hsp104p to cure prion propagation. In contrast, overexpression of SSA1 is able to cure cells infected with the yeast prion [URE3] (an isoform of Ure2p). In cells carrying the [PIN+] prion (an isoform of Rnq1p), ssa1ssa2 double null mutations result in the loss of polyglutamine aggregate expansion and toxicity, which are two hallmarks of Huntington disease (OMIM).

YRA1/YDR381W:

RNA binding protein required for export of poly(A)+ mRNA from the nucleus; proposed to couple mRNA export with 3' end processing via its interactions with Mex67p and Pcf11p; functionally redundant with Yra2p, another REF family member. (SGD)

Yra1p is a nuclear RNA-binding protein, which is essential for nuclear mRNA export. Strikingly, this RNA-binding protein physically contacts the mRNA export factor Mex67p. This points to a direct link between the intranuclear pre-mRNA assembly and nuclear export machinery.

The mechanism of Mex67p-mediated mRNA export from the nucleus to the cytoplasm is still unknown. To date, it is uncertain whether in vivo Mex67p binds to mRNA with the help of co-factors.

Because an interaction between Mex67p and the major yeast hnRNP-like protein Npl3p could not be found, which is also implicated in nuclear mRNA export, putative 'upstream' factors of Mex67p are investigated. Yrap1 can directly associate with Mex67p observed in in vitro reconstitution experiments, which revealed that Yra1p is a Mex67p-binding protein.

ATG1/YGL180W:

Protein ser/thr kinase required for vesicle formation in autophagy and the cytoplasm-to-vacuole targeting (Cvt) pathway; structurally required for phagophore assembly site formation; during autophagy forms a complex with Atg13p and Atg17p (SGD).

ATG1 encodes a cytosolic protein kinase required for vesicle formation during autophagy and the Cvt pathway. Atg1p kinase activity is required for initiation of the Cvt pathway as well as for proper localization and cycling of autophagy proteins such as Atg23p. Atg1p also has a role in autophagy unrelated to its kinase activity; it is thought that Atg1p is a structural protein required for efficient PAS organization and assembly.

Atg1p kinase activity is stimulated by interaction with Atg13p and Atg17p, and formation of this complex is specific for the role of Atg1p in autophagy initiation. Atg1p also interacts with the Cvt pathway-specific protein Atg11p. Complementation experiments and the presence of PKA phosphorylation sites on Atg1p suggest that regulation of autophagy by the kinases Snf1p, Pho85p, and PKA may occur via regulation of Atg1p.

atg1 mutants are defective in autophagy, Cvt transport, sporulation, and survival under starvation condition. Overexpression of ATG1 inhibits filamentous growth. ATG1 is highly conserved, and homologs have been identified in organisms such as soil amoeba (atg1), worms (UNC-51), Drosophila (ATG1), and human (ULK1).

CMK1/YFR014C:

Calmodulin-dependent protein kinase; may play a role in stress response, many CA++/calmodulan dependent phosphorylation substrates demonstrated in vitro, amino acid sequence similar to Cmk2p and mammalian Cam Kinase II.

Yeast cells display an adaptive stress response when exposed to weak organic acids at low pH. This adaptation is important in the spoilage of preserved foods, as it allows growth in the presence of weak acid food preservatives. In Saccharomyces cerevisiae, this stress response leads to strong induction of the Pdr12 ATP-binding cassette (ABC) transporter, which catalyses the active efflux of weak acid anions from the cytosol of adapted cells. S. cerevisiae cells lacking the Cmk1 isoform of Ca^{2+}-calmodulin-dependent protein kinase are intrinsically resistant to weak acid stress, in that they do not need to spend a long adaptive period in lag phase before resuming growth after exposure to this stress. This resistance of the cmk1 mutant is Pdr12 dependent and, unlike with wild-type S. cerevisiae, cmk1 cells are capable of performing Pdr12-specific functions such as energy-dependent cellular extrusion of fluorescein and benzoate. However, they have neither higher PDR12 gene promoter activity nor higher Pdr12 protein levels. The increased Pdr12 activity in cmk1 cells is therefore caused by Cmk1 exerting a negative post-transcriptional influence over the activity of the Pdr12 ABC transporter, a transporter protein that is constitutively expressed in low-pH yeast cultures. This is the first preliminary evidence that shows a protein kinase, either directly or indirectly, regulating the activity of a yeast ABC transporter.

RPO21/YDL140C:

RNA polymerase II largest subunit B220, part of central core; phosphorylation of C-terminal heptapeptide repeat domain regulates association with transcription and splicing factors; similar to bacterial beta-prime.

Nuclear transcription in S. cerevisiae is performed by three multisubunit nuclear RNA polymerases (RNAPs) that are conserved in all eukaryotes. The roles of these three RNA polymerases are generally conserved across eukaryotes, particularly with respect to

production of rRNAs, mRNAs, and tRNAs, though production of other small RNAs is somewhat variable between RNAP II and RNAP III in different species. In S. cerevisiae, RNA polymerase I transcribes rDNA to produce the 35S primary rRNA transcript that is processed to produce three of the four mature ribosomal rRNAs: 25S, 18S, and 5.8S. RNA polymerase II produces all nuclear mRNAs, all of the snoRNAs except snR52, four of the five snRNAs (U1, U2, U4, and U5), the RNase MRP RNA encoded by NME1, and the telomerase RNA encoded by TLC1. RNA polymerase III produces the 5S rRNA, all nuclear tRNAs, the U6 snRNA, the snR52 snoRNA, the RNase P RNA encoded by RPR1, and the 7SL RNA component of the signal recognition particle encoded by SCR1.

Coordinate regulation of these three RNA polymerases is essential, since in rapidly growing yeast cells, much of the transcriptional output of the cell is devoted to the production of ribosomes. About 60% of total cellular transcription is devoted to transcription by RNAP I of the rRNA genes, which comprise about 10% of the entire genome. While mRNAs generally only comprise 5% of total cellular RNA and the 137 ribosomal protein (RP) genes represent only 2% of the genome, it is estimated that 50% of RNAP II transcription occurs on RP genes. RNAP II is also responsible for production of the majority of the snoRNAs, which are collectively involved in maturation of the ribosome. RNAP III plays a similarly important role in production of ribosomes and the process of translation, producing both the 5S rRNA and all nuclear tRNAs, which constitute about 15% of total cellular RNA. The TOR pathway is a major factor in this coordinate regulation as it regulates the activity of all three nuclear RNAPs in response to nutrient availability and growth.

In addition to producing the majority of cellular RNA, RNAP I and RNAP III may also play roles in nuclear architecture and genome organization. RNAP I activity may be involved in organizing the rDNA repeats into the nucleolus. Active tRNA genes transcribed by RNAP III appear to act as chromatin boundary elements that affect both transcription and DNA replication.

Additionally, recombination between dispersed tRNA genes may be a source of genetic instability and evolutionary change.

Five genes (RPB5, RPO26, RPB8, RPB10, and RPC10) encode subunits common to all three of the nuclear RNA polymerases. Two genes (RPC40 and RPC19) encode subunits present in both RNAP I and RNAP III; RPB3 and RPB11 encode the corresponding RNAP II subunits. Five more subunits are encoded by a separate gene for each polymerase, but are considered functional equivalents of each other. Thus there are twelve subunits that are conserved in all three of the nuclear RNA polymerases, eleven of which correspond to subunits of Archaeal RNAPs, and five of which also correspond to the subunits of E. coli RNAP. In each, ten of these comprise the enzyme cores, while Rpb4/7 (RNAP II), Rpa14/43 (RNAP I), and Rpc17/25 (RNAP III) form heterodimers which associate with this core and have roles in initiation. RNAPs I and III also have two subunits which are homologous to the subunits of the TFIIF general initiation factor for RNAP II, and RNAP III has three additional unique subunits.

Table C.1.1. Conserved Residues of Hub Proteins and their Domains in 12 core PIN obtained by Conseq

ORF Name	Protein name	Tot. Res.	Conservation		Residues	Exposed	Residues	Buried	Functionally	Structurally	Func/conserved	Struc/conserved	Func/total	Struc/total
			Total	% cons	exposed	%	buried	%	important res.	important res.	%	%	%	%
YDL229W	SSB1	613	296	48,3	356	58,1	257	41,9	148	126	50,0	42,6	24,1	20,6
	HSP70	605	292	48,3	351	58,0	254	42,0	145	125	49,7	42,8	24,0	20,7
YPL240C	HSP82	709	324	45,7	413	58,3	296	41,7	152	126	46,9	38,9	21,4	17,8
	HATPase_c	155	88	56,8	79	51,0	76	49,0	33	41	37,5	46,6	21,3	26,5
	HSP90	527	217	41,2	321	60,9	206	39,1	112	74	51,6	34,1	21,3	14,0
YBR160W	CDC28	298	134	45,0	163	54,7	135	45,3	63	49	47,0	36,6	21,1	16,4
	Pkinase	288	133	46,2	153	53,1	135	46,9	62	49	46,6	36,8	21,5	17,0
YJL164C	TPK1	397	108	27,2	269	67,8	128	32,2	52	29	48,1	26,9	13,1	7,3
	Pkinase	255	98	38,4	142	55,7	113	44,3	45	28	45,9	28,6	17,6	11,0
YJR059W	PTK2	818	101	12,3	599	73,2	219	26,8	57	14	56,4	13,9	7,0	1,7
	Pkinase	337	70	20,8	230	68,2	107	31,8	33	14	47,1	20,0	9,8	4,2
YAL005C	SSA1	642	366	57,0	377	58,7	265	41,3	184	145	50,3	39,6	28,7	22,6
	HSP70	604	359	59,4	340	56,3	264	43,7	178	145	49,6	40,4	29,5	24,0
YDR381W	YRA1	226	61	27,0	181	80,1	45	19,9	34	15	55,7	24,6	15,0	6,6
	RRM_1	71	28	39,4	45	63,4	26	36,6	11	10	39,3	35,7	15,5	14,1
YGL180W	ATG1	897	247	27,5	631	70,3	266	29,7	111	62	44,9	25,1	12,4	6,9
	Pkinase	302	114	37,7	183	60,6	119	39,4	50	27	43,9	23,7	16,6	8,9
	DUF3543	312	93	29,8	211	67,6	101	32,4	38	25	40,9	26,9	12,2	8,0
YFR014C	CMK1	446	103	23,1	311	69,7	135	30,3	41	30	39,8	29,1	9,2	6,7
	Pkinase	263	102	38,8	156	59,3	107	40,7	40	30	39,2	29,4	15,2	11,4
YDL140C	RPO21	1733	439	25,3	1232	71,1	501	28,9	245	87	55,8	19,8	14,1	5,0
	RNA_pol_Rpb1_1	330	75	22,7	236	71,5	94	28,5	40	15	53,3	20,0	12,1	4,5
	RNA_pol_Rpb1_2	166	82	49,4	102	61,4	64	38,6	47	20	57,3	24,4	28,3	12,0
	RNA_pol_Rpb1_3	160	31	19,4	96	60,0	64	40,0	17	5	54,8	16,1	10,6	3,1
	RNA_pol_Rpb1_4	108	36	33,3	73	67,6	35	32,4	18	6	50,0	16,7	16,7	5,6
	RNA_pol_Rpb1_5	592	150	25,3	390	65,9	202	34,1	73	34	48,7	22,7	12,3	5,7

Table C.1.1. Conserved Residues of Hub Proteins and their Domains in 12 core PIN obtained by ConSeq - cont.

	RNA_pol_Rpb1_6	184	2	1,1	140	76,1	44	23,9	1	0	50,0	0,0	0,5	0,0
	RNA_pol_Rpb1_7	134	15	11,2	92	68,7	42	31,3	5	1	33,3	6,7	3,7	0,7
	RNA_pol_Rpb1_R	14	0	0,0	14	100,0	0	0,0	0	0	-	-	0,0	0,0
	RNA_pol_Rpb1_R	14	1	7,1	14	100,0	0	0,0	1	0	100,0	0,0	7,1	0,0
	RNA_pol_Rpb1_R	14	2	14,3	14	100,0	0	0,0	2	0	100,0	0,0	14,3	0,0
	RNA_pol_Rpb1_R	14	2	14,3	13	92,9	1	7,1	2	0	100,0	0,0	14,3	0,0
	RNA_pol_Rpb1_R	14	2	14,3	13	92,9	1	7,1	2	0	100,0	0,0	14,3	0,0
	RNA_pol_Rpb1_R	14	3	21,4	14	100,0	0	0,0	3	0	100,0	0,0	21,4	0,0
	RNA_pol_Rpb1_R	14	7	50,0	14	100,0	0	0,0	7	0	100,0	0,0	50,0	0,0
	RNA_pol_Rpb1_R	14	2	14,3	14	100,0	0	0,0	2	0	100,0	0,0	14,3	0,0
	RNA_pol_Rpb1_R	14	4	28,6	14	100,0	0	0,0	4	0	100,0	0,0	28,6	0,0
	RNA_pol_Rpb1_R	14	1	7,1	14	100,0	0	0,0	1	0	100,0	0,0	7,1	0,0
	RNA_pol_Rpb1_R	14	5	35,7	14	100,0	0	0,0	5	0	100,0	0,0	35,7	0,0
	RNA_pol_Rpb1_R	14	3	21,4	14	100,0	0	0,0	3	0	100,0	0,0	21,4	0,0
	RNA_pol_Rpb1_R	14	6	42,9	14	100,0	0	0,0	6	0	100,0	0,0	42,9	0,0
	RNA_pol_Rpb1_R	14	1	7,1	14	100,0	0	0,0	1	0	100,0	0,0	7,1	0,0
	RNA_pol_Rpb1_R	14	0	0,0	14	100,0	0	0,0	0	0	-	-	0,0	0,0

APPENDIX D

D.1 Module List Obtained from 12 yeast core PIN

Bold ones are the proteins with unknown function.

Parameters:

Network Scoring: *Include Loops:* false

Cluster Finding:

Node Score Cutoff: 0.2 *Haircut:* true *Fluff:* false *K-Core:* 2 *Max. Depth from Seed:* 100

Table D.1. Modules of yeast network

Cluster	Score (Density*#Nodes)	Nodes	Edges	Node IDs
1	4,816	38	183	YGR005C, YIL021W, ***YDL115C***, YJL140W, YER139C, YDL140C, YGR186W, YOL005C, YOR151C, ***YBR231C***, YGR002C, YOL012C, YNL107W, YDR334W, YDR190C, YDR485C, ***YLR385C***, ***YAL011W***, YPL235W, YDR156W, YOR341W, YJR063W, YJL148W, YPR010C, YNR003C, YKL144C, YJL011C, YNL151C, YDL150W, YDR045C, YOR207C, YPR190C, YPR110C, YNL248C, YNL113W, YOR116C, YGL070C, YGR063C
2	4,038	26	105	YFR036W, ***YIR025W***, YOR249C, YDR448W, YGL112C, YMR223W, YDR392W, YDR145W, YMR236W, YNL172W, ***YDR260C***, YDR113C, YDR118W, YDR254W, ***YBR211C***, ***YBR107C***, YPR046W, ***YLR315W***, ***YDR383C***, ***YIR010W***, YGR179C, YHR166C, YLR127C, YBL084C, YLR102C, YKL022C

Table D.1. Modules of yeast network (cont.)

3	3,222	27	87	YOR210W, YFR037C, YLR176C, YGR274C, YBR089C-A, YBR200W, YDR166C, YBR102C, YPR055W, YGL233W, YJL085W, YIL068C, YLR166C, YFL039C, YBR245C, YER164W, ***YCL010C***, YPL047W, YPL082C, YGL133W, YOR304W, YKR001C, YIL126W, YEL009C, YHR143W-A, YPR187W, YBR154C
4	3	14	42	YHR039C-A, YKL080W, YBR127C, YIL017C, ***YGL227W***, ***YBL049W***, ***YMR135C***, ***YBR105C***, ***YIL097W***, ***YCL039W***, YDL185W, YJR033C, ***YDR202C***, YEL051W
5	2,805	41	115	***YPL101W***, ***YLR384C***, YPL086C, ***YHR187W***, ***YMR312W***, ***YNL078W***, YJR076C, ***YNL166C***, YDR507C, YCL024W, ***YAL034W-A***, ***YLR052W***, ***YNL215W***, ***YFL013C***, ***YER092W***, YGL150C, ***YEL044W***, YLR314C, YHR107C, YMR001C, YDL017W, YEL032W, YLR274W, YPR019W, YBR202W, YBL023C, YGL201C, YBR160W, YKL185W, ***YDL076C***, YPL181W, YNL330C, YNL097C, YDR207C, YDL225W, YCR002C, YGL137W, YFR051C, YDR238C, YGL051W, ***YGR200C***
6	2,769	26	72	***YLR033W***, ***YGR275W***, YPL010W, YIL076W, ***YML067C***, YDL145C, YDR372C, YNL287W, YGR120C, ***YGL223C***, ***YGL005C***, ***YPR105C***, ***YML071C***, ***YNL051W***, YDR381W, YDR138W, YML062C, YDL084W, YNL139C, ***YNL253W***, YHR167W, ***YNL041C***, YDR303C, YGR056W, YHR056C, YLR357W
7	2,15	20	43	YPL007C, YOR110W, YBR123C, YGR047C, YAL001C, YPL106C, YNL064C, YBR169C, YLL024C, ***YMR237W***, YJL099W, ***YOR299W***,

Table D.1. Modules of yeast network - cont.

				YLR330W, ***YKR027W***, YMR186W, YAL029C
8	2,1	10	21	YJR102C, YKL041W, YPR173C, YKL002W, ***YKR035W-A***, YLR025W, YPL002C, YLR119W, YLR417W, YPL065W
9	2,1	10	21	***YMR025W***, ***YGR134W***, YNL288W, YER068W, YAL021C, YNR052C, ***YDR179C***, ***YOL117W***, **YJR084W**, YDL216C
10	2	5	10	YDR325W, YFR031C, YLR086W, ***YBL097W***, YLR272C
11	2	6	12	***YGL124C***, YML001W, YMR231W, YLR396C, YPL195W, YLR148W
12	1,833	18	33	YGR172C, ***YNL263C***, YNL093W, YKR014C, YOR089C, YBL016W, YDR103W, YDL159W, YLR362W, YPL218W, YIL109C, YPR181C, YNL298W, YMR109W, YBR234C, YKL129C, YIR006C, YDL029W
13	1,8	5	9	YJR135C, YDR318W, ***YPL233W***, YPL018W, YLR381W
14	1,778	9	16	YNL216W, YLR442C, YDL042C, YCR033W, YOL068C, YGL194C, ***YBR103W***, YDR227W, YKR101W
15	1,625	16	26	YGL023C, YDR011W, ***YDL222C***, YCR021C, ***YKR007W***, YDR264C, ***YHL042W***, ***YEL017C-A***, ***YDR414C***, ***YBL102W***, ***YCR043C***, ***YJL151C***, YPR036W, ***YDL123W***, YGR163W, YML121W
16	1,614	44	71	***YGL213C***, ***YKL023W***, YHR170W, YIR026C, YGL099W, YPL169C, YOR098C, YBL021C, YKL109W, YOR358W, YOL086C, YOR357C, YPL246C, YHR098C, ***YDR517W***, YGL237C, YIL053W, YIL095W, YPR161C, YCL014W, YJL052W, YJL034W, YER166W, YGR229C, ***YOR104W***, YAL026C, ***YCR077C***, YNL118C, YBL037W, YJR058C, ***YOL062C***, YJR005W,

Table D.1. Modules of yeast network - cont.

				YLR187W, YAL005C, YBR126C, YDR074W, YMR261C, YML100W, YHL034C, YGL173C
17	1,5	4	6	YOR115C, ***YGR166W***, YDR407C, ***YEL048C***
18	1,5	4	6	YNL099C, YHL029C, YNL032W, ***YCR095C***
19	1,4	5	7	***YKR077W***, YLR182W, YDL056W, ***YOR066W***, YER111C
20	1,25	4	5	YBR091C, ***YKL195W***, YEL020W-A, YHR005C-A
21	1	3	3	***YEL018W***, ***YNL136W***, YDR359C
22	1	3	3	***YPL264C***, YER005W, ***YJL004C***
23	1	3	3	***YJL058C***, YKL203C, YNL006W
24	1	3	3	YKL138C-A, ***YFR042W***, YDR320C-A
25	1	3	3	***YLR360W***, YLR240W, YBR097W
26	1	3	3	YNL116W, YHR115C, ***YLR215C***
27	1	3	3	***YGR130C***, ***YMR031C***, ***YMR086W***
28	1	3	3	***YGR041W***, ***YLR353W***, YOR301W
29	1	3	3	YDR091C, ***YNL260C***, ***YJR067C***
30	1	3	3	YLR106C, ***YGR245C***, YER006W

D.2. Module Figures for Yeast

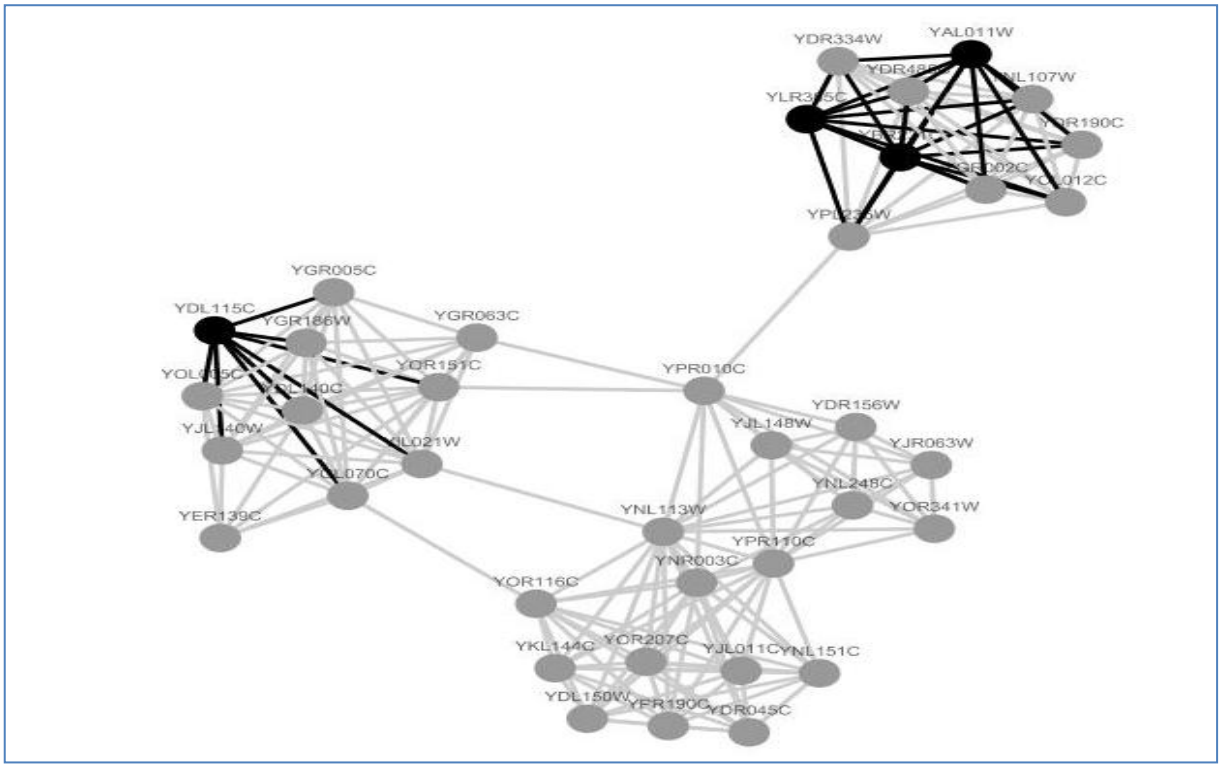


Figure D.2.1. Module 1 for yeast

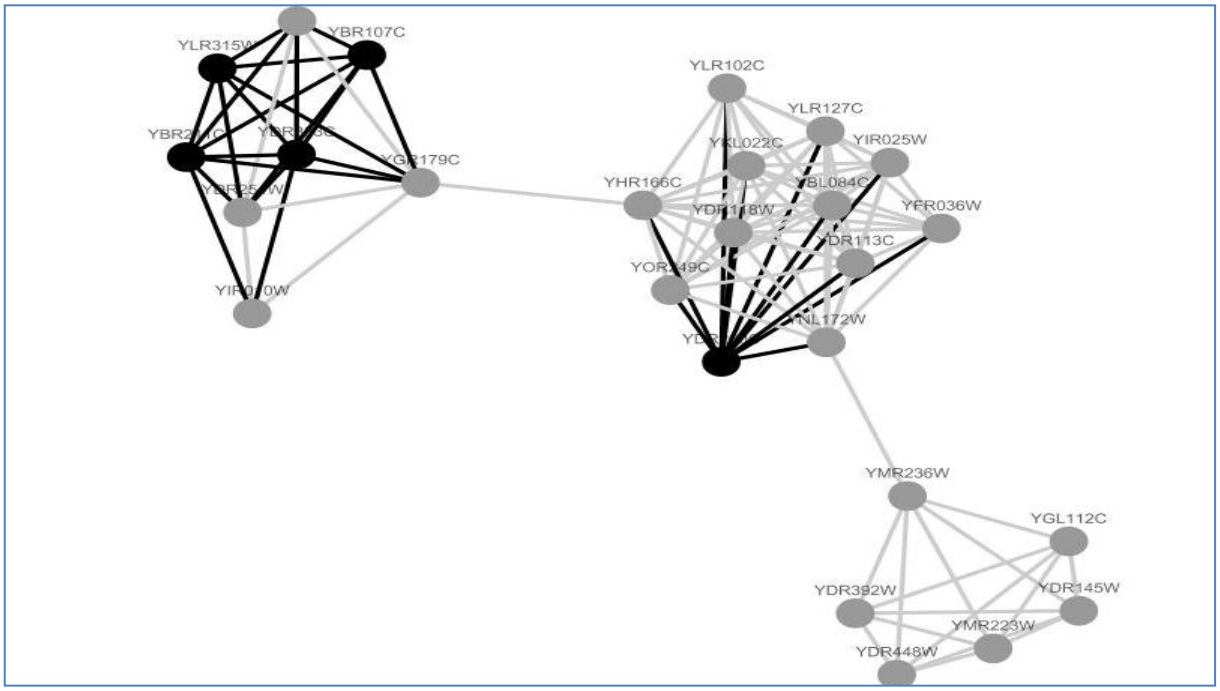


Figure D.2.2. Module 2 for yeast

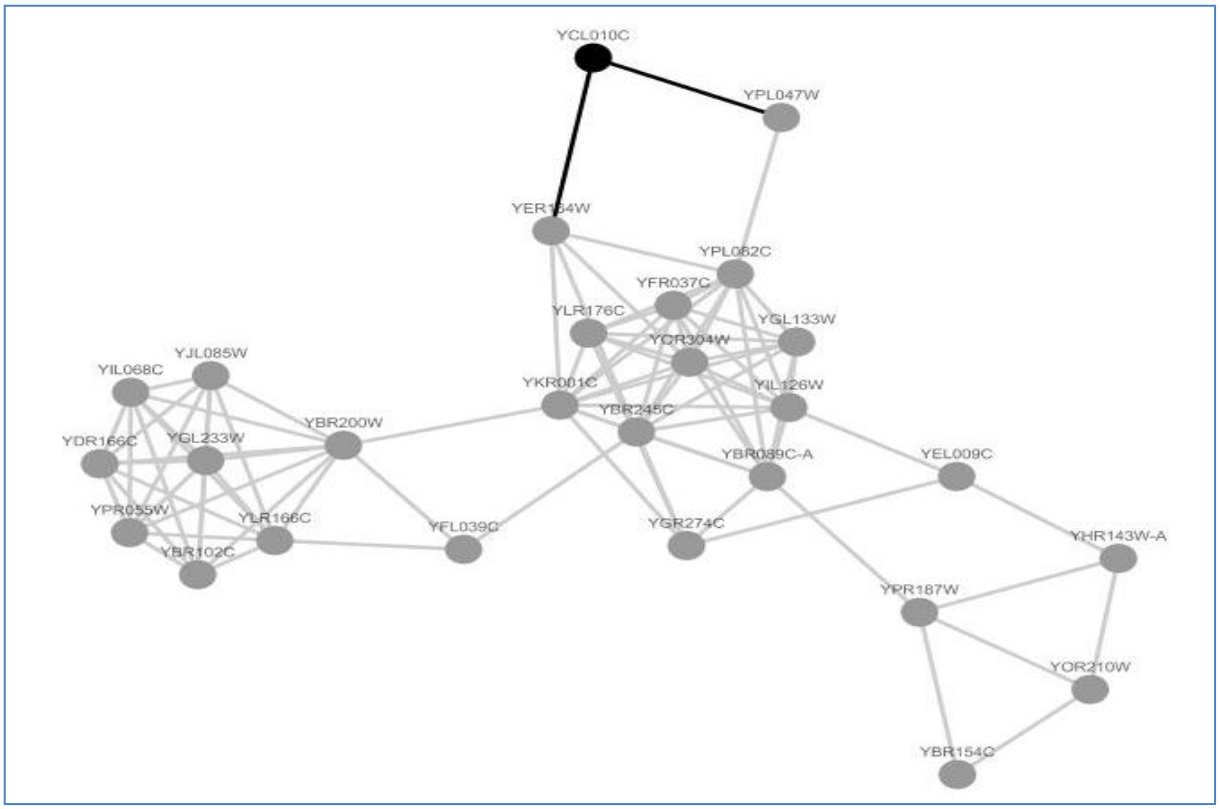


Figure D.2.3. Module 3 for yeast

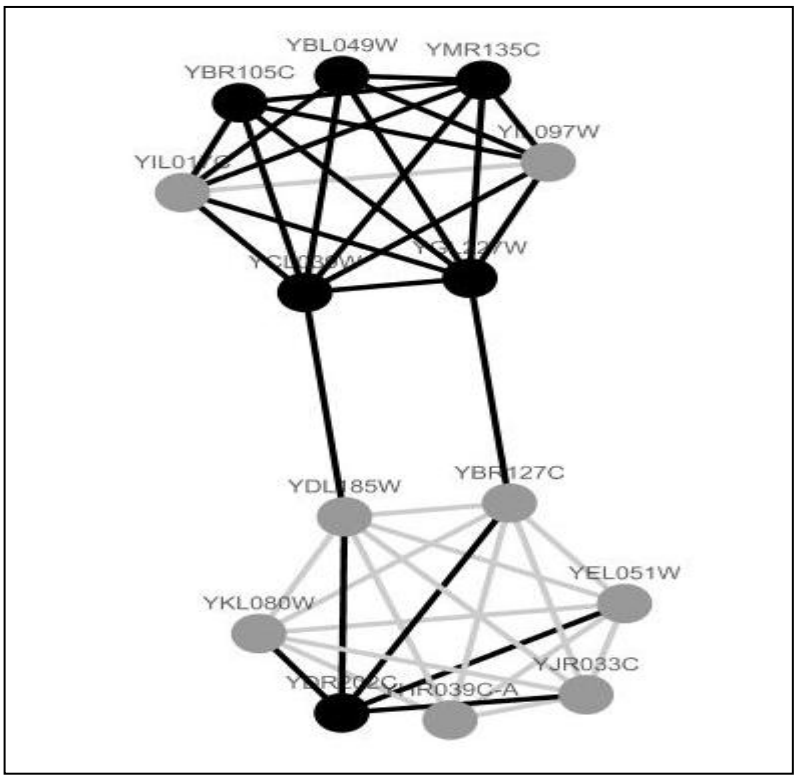


Figure D.2.4. Module 4 for yeast

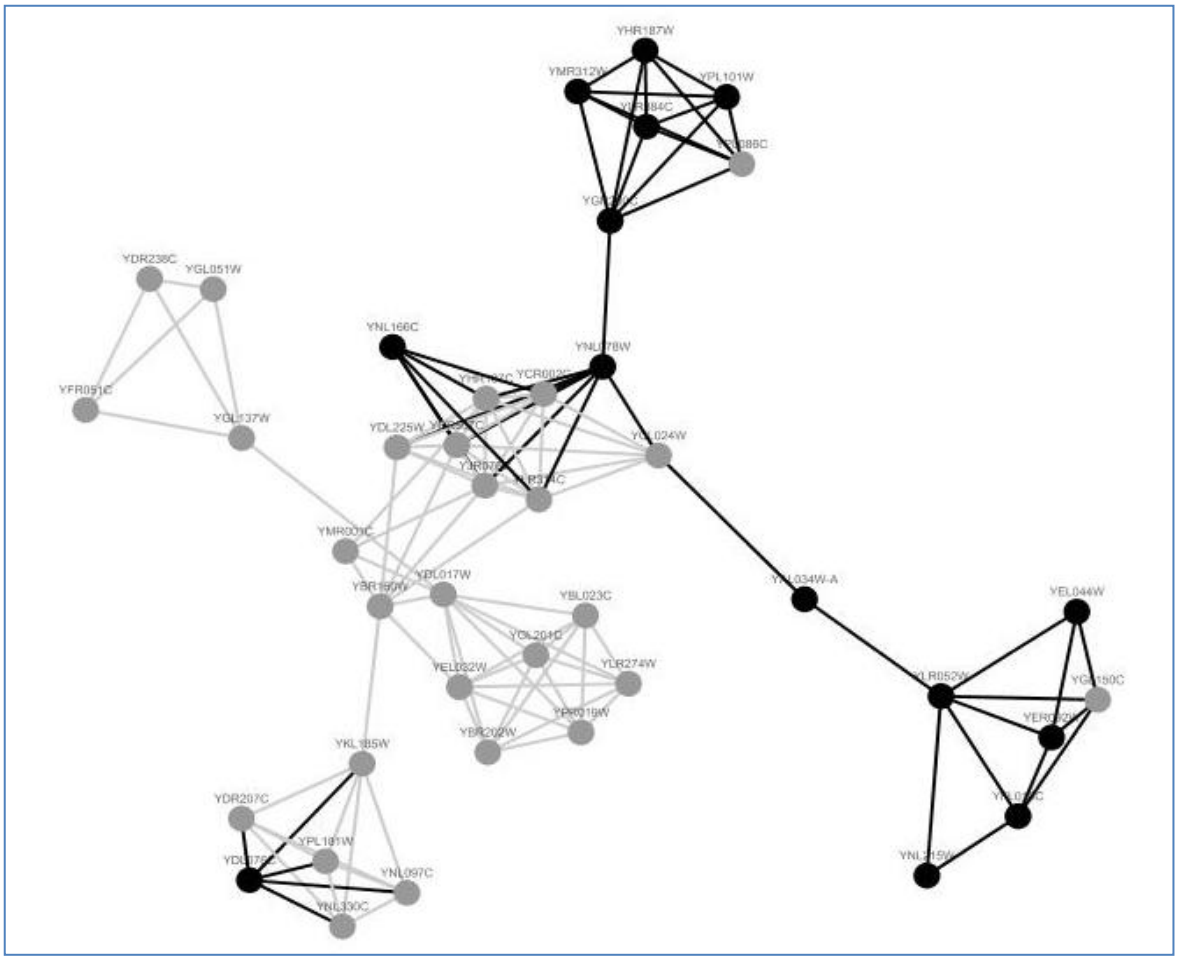


Figure D.2.5. Module 5 for yeast

Figure D.2.6. Module 6 for yeast

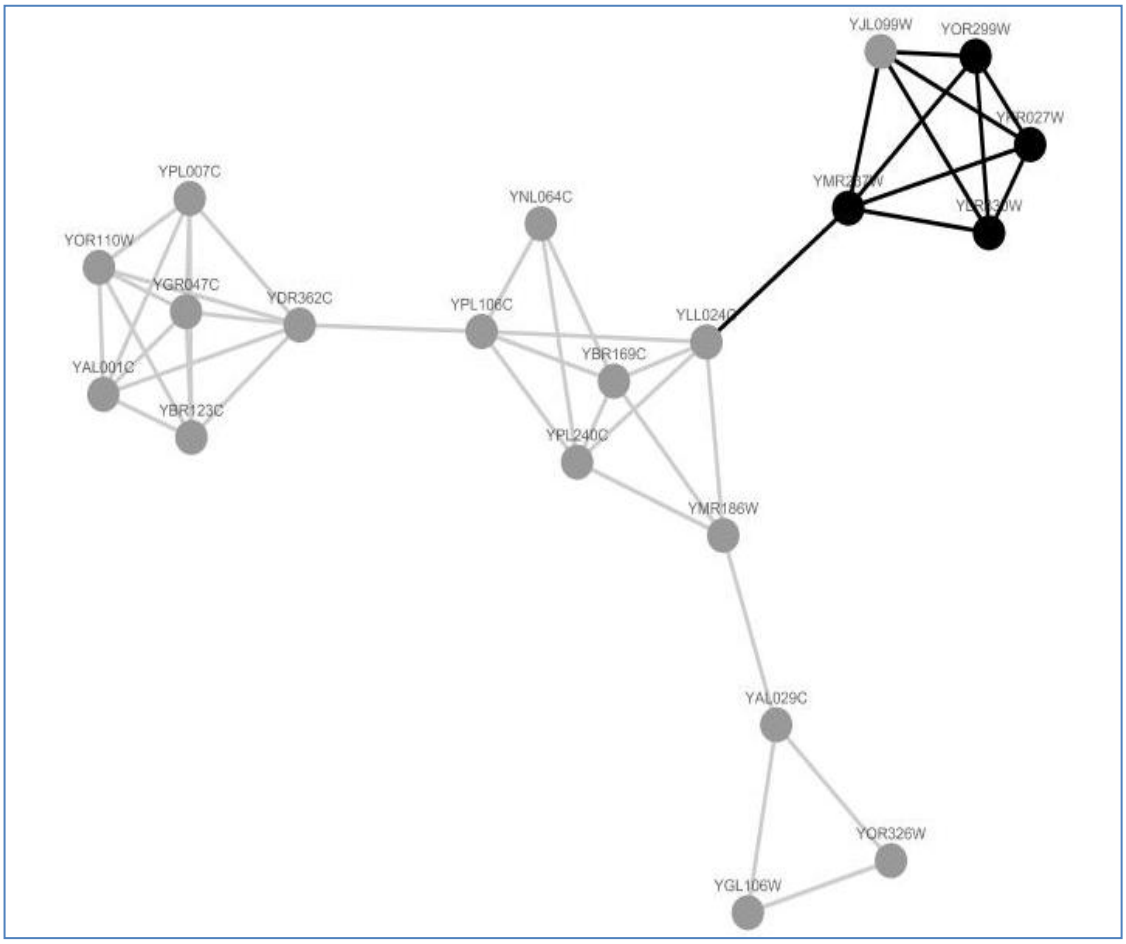


Figure D.2.7. Module 7 for yeast

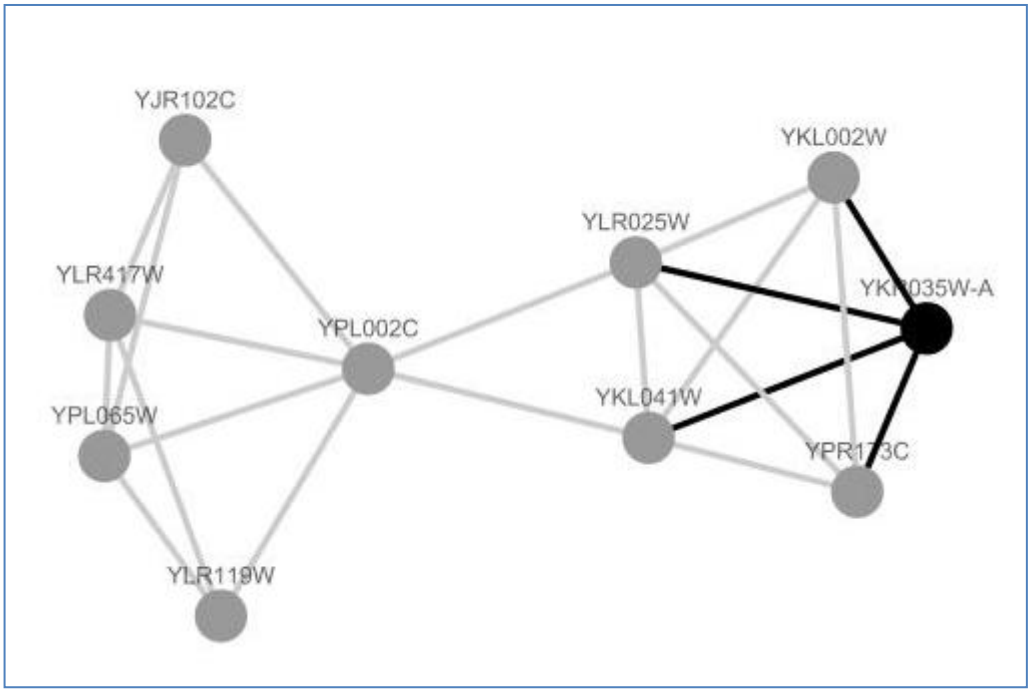


Figure D.2.8. Module 8 for yeast

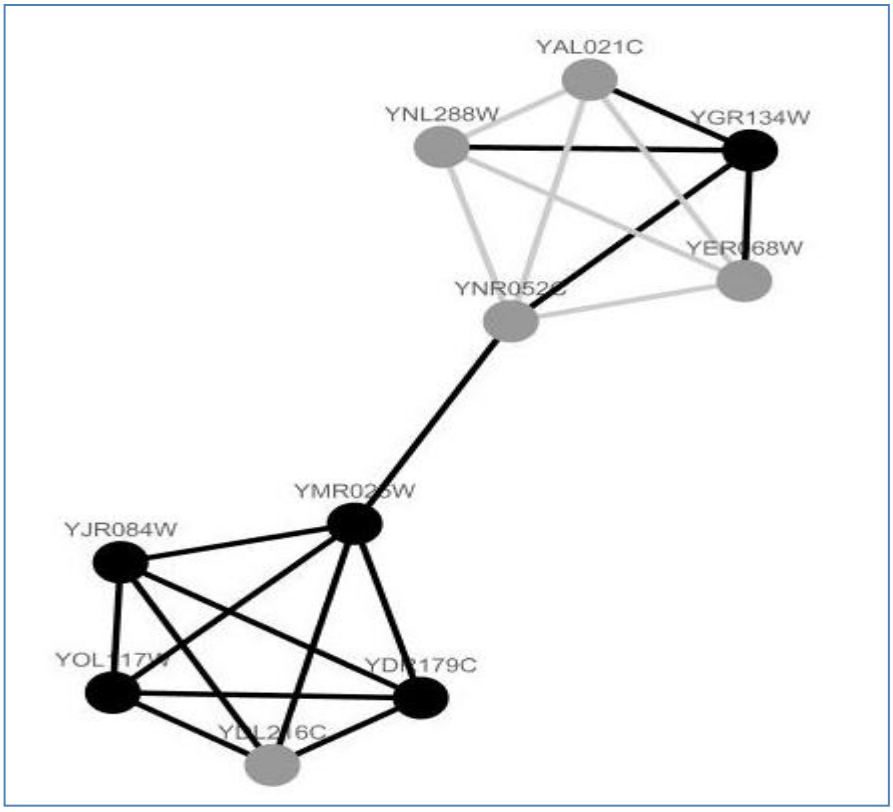


Figure D.2.9. Module 9 for yeast

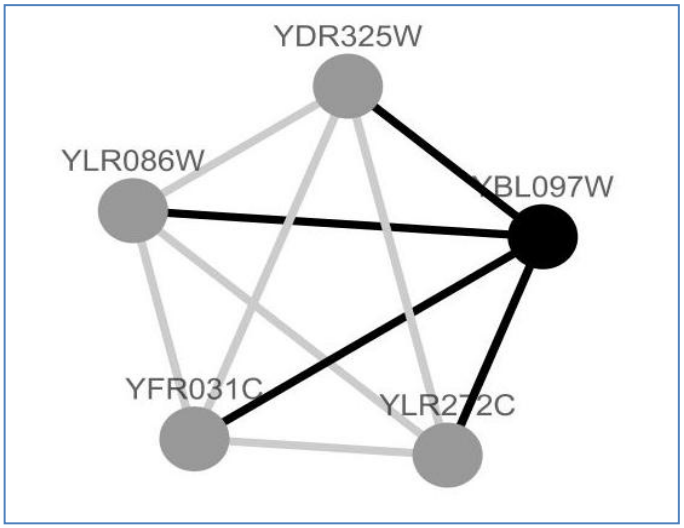


Figure D.2.10. Module 10 for yeast

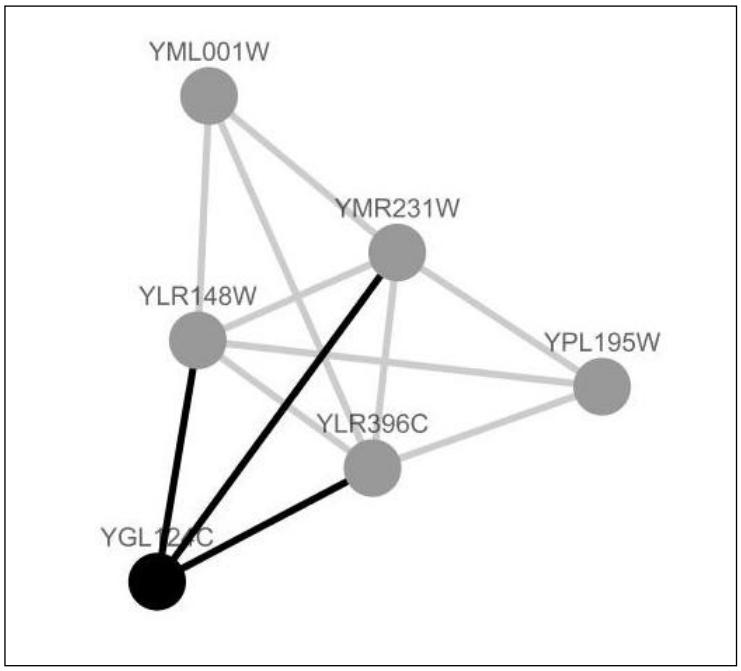


Figure D.2.11. Module 11 for yeast

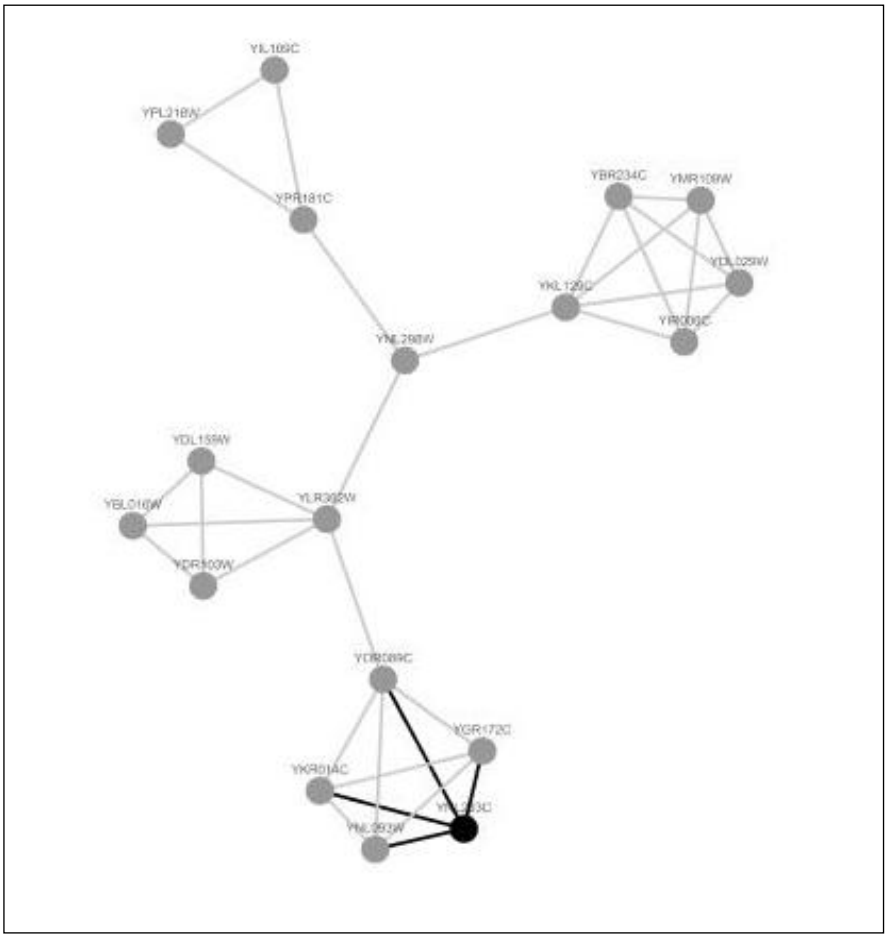


Figure D.2.12. Module 12 for yeast

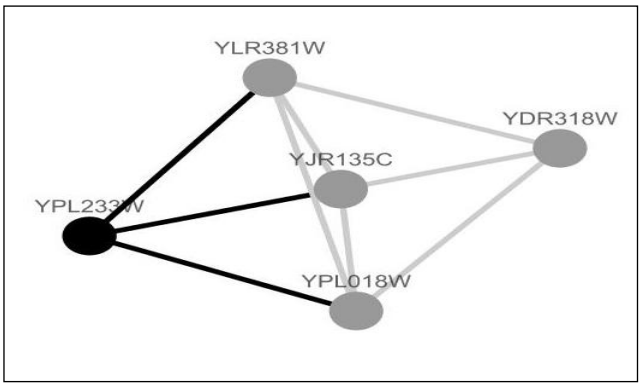


Figure D.2.13. Module 13 for yeast

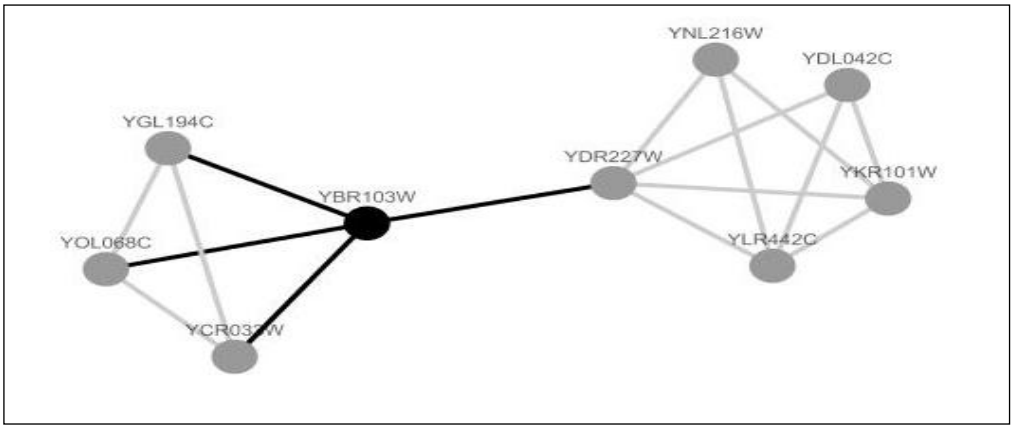


Figure D.2.14. Module 14 for yeast

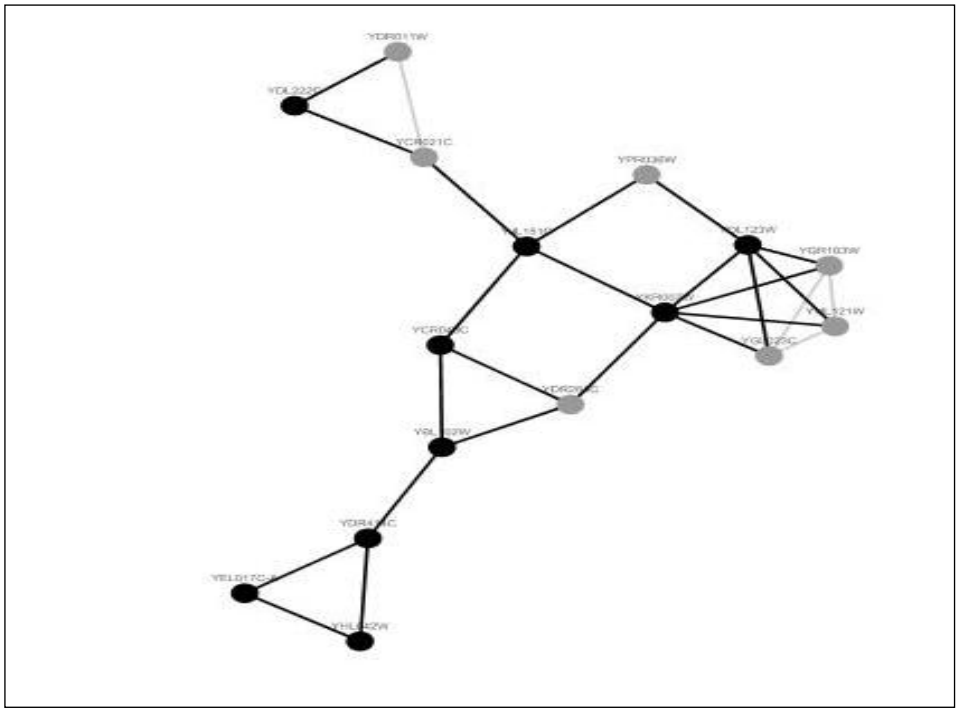


Figure D.2.15. Module 15 for yeast

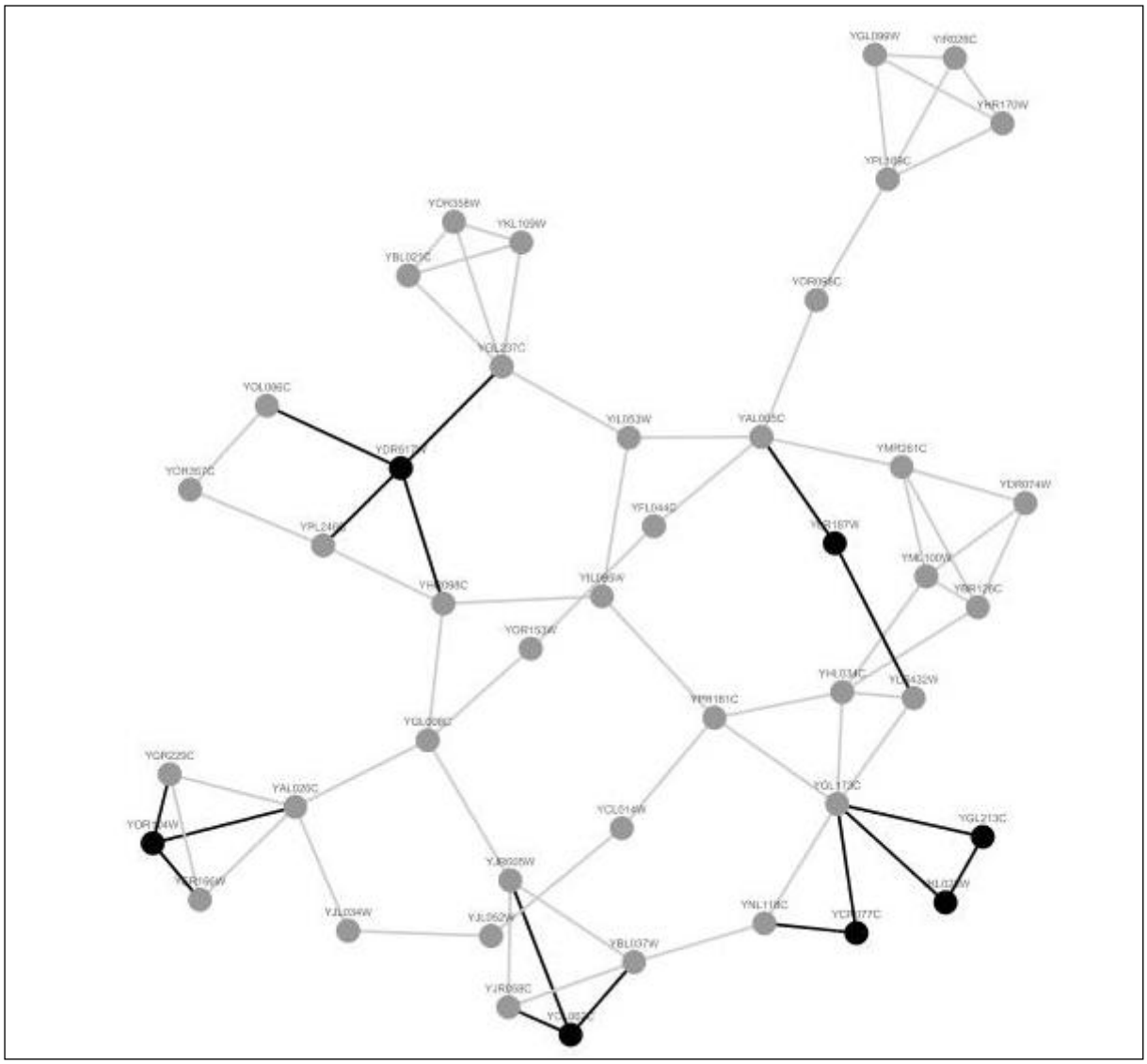


Figure D.2.16. Module 16 for yeast

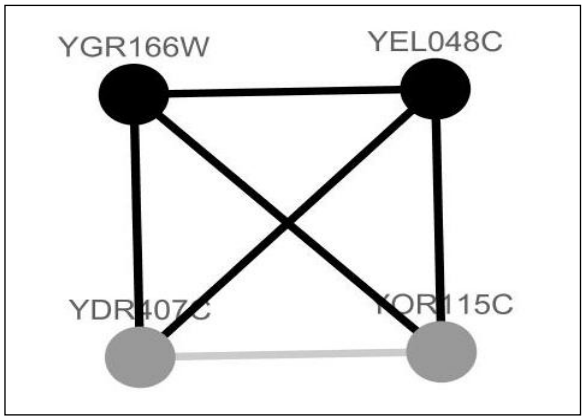


Figure D.2.17. Module 17 for yeast

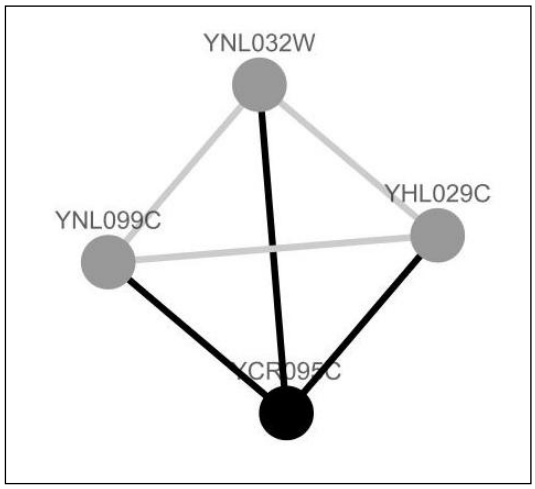


Figure D.2.18. Module 18 for yeast

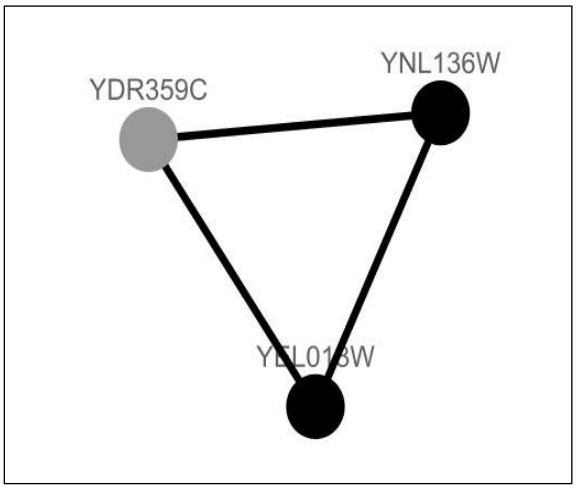


Figure D.2.21. Module 21 for yeast

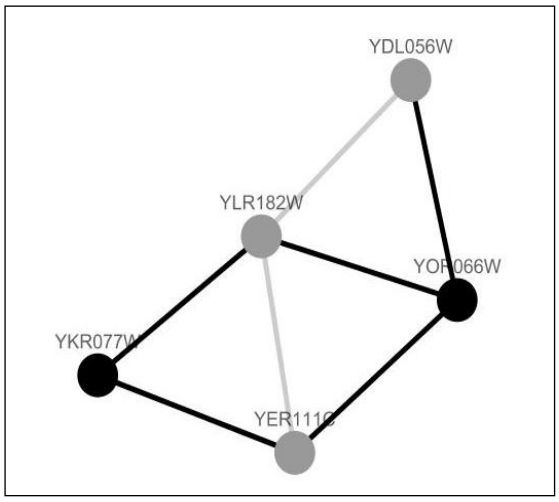


Figure D.2.19. Module 19 for yeast

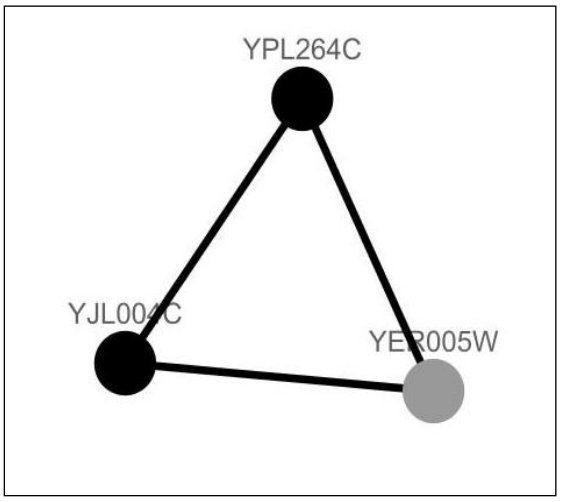


Figure D.2.22. Module 22 for yeast

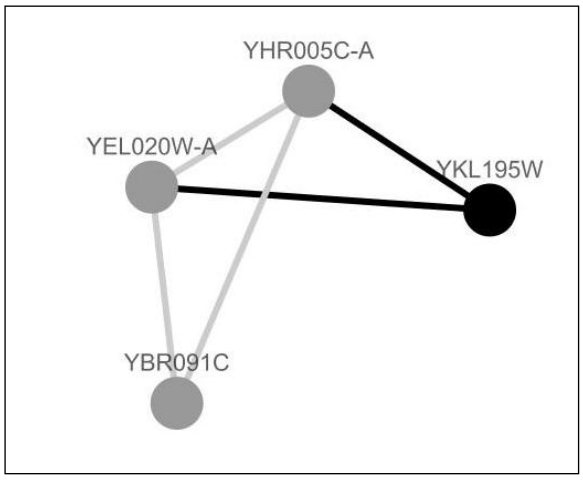


Figure D.2.20. Module 20 for yeast

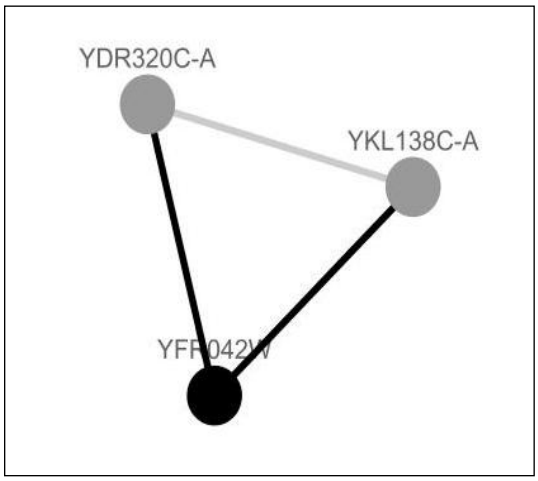


Figure D.2.23. Module 23 for yeast

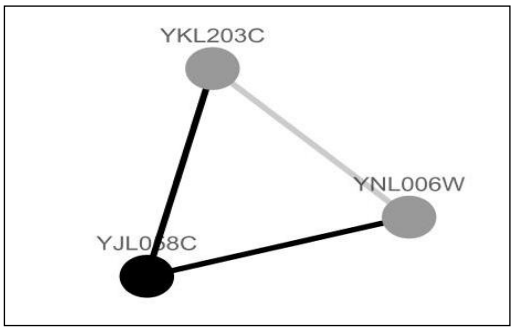


Figure D.2.24. Module 24 for yeast

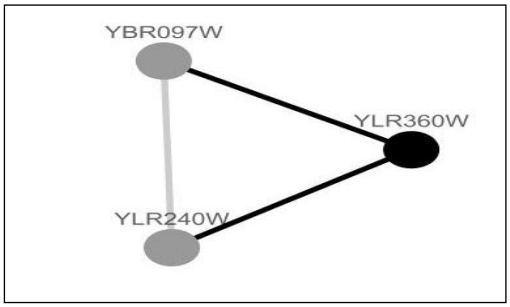


Figure D.2.25. Module 25 for yeast

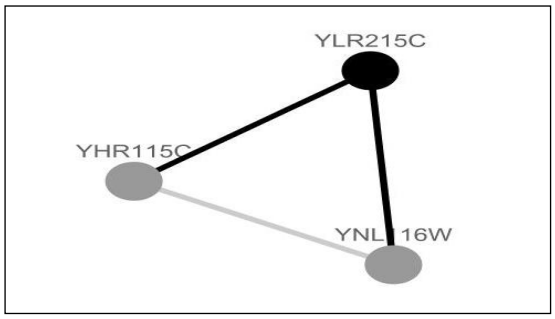


Figure D.2.26. Module 26 for yeast

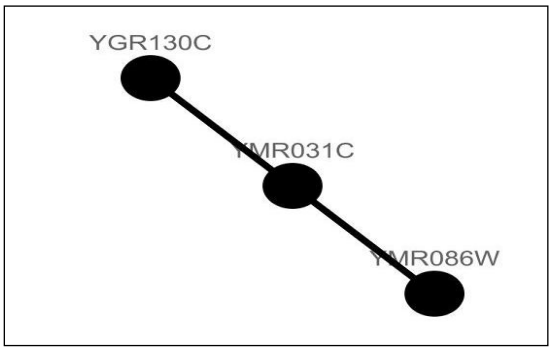


Figure D.2.27. Module 27 for yeast

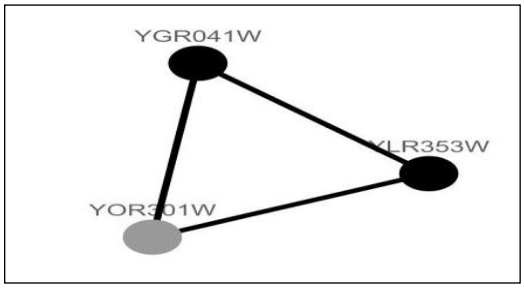


Figure D.2.28. Module 28 for yeast

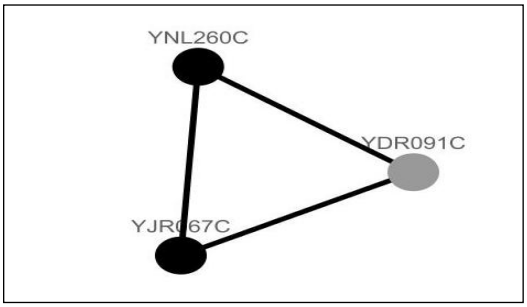


Figure D.2.29. Module 29 for yeast

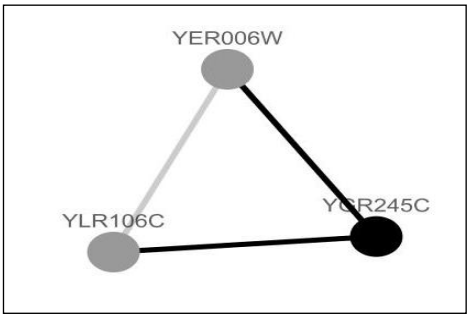


Figure D.2.30. Module 30 for yeast

D.3. Annotations of Functionally Unknown Yeast Proteins with Evidence Scores

Table D.3.1. GO's of Functionally Unknown Proteins

**APS: Assigned in Present Study

GO id	Go Term	Ref		Name	ORF
6338	chromatin remodeling	IMP	P	**IWR1**	**YDL115C**
6350	transcription	IEA	P		
6914	autophagy	RCA	P		
32568	general transcription from RNA polymerase II promoter	IMP	P		
32569	gene-specific transcription from RNA polymerase II promoter	IMP	P		
45449	regulation of transcription	IEA	P		
5634	nucleus	IDA	C		
5634	nucleus	IEA	C		
5737	cytoplasm	IDA	C		
5737	cytoplasm	IEA	C		
16591	DNA-directed RNA polymerase II, holoenzyme	IPI	C		
3677	DNA binding	APS	F		
46872	metal ion binding	APS	F		
3899	DNA-directed RNA polymerase activity	APS	F		
7049	cell cycle	IEA	P	**NKP1**	**YDR383C**
7067	mitosis	IEA	P		
7126	meiosis	IEA	P		
8150	biological_process	ND	P		
51301	cell division	IEA	P		
776	kinetochore	IEA	C		
776	kinetochore	IDA	C		
777	condensed chromosome kinetochore	IEA	C		
777	condensed chromosome kinetochore	RCA	C		
779	condensed chromosome, centromeric region	RCA	C		
5634	nucleus	IEA	C		
5515	protein binding	APS	F		
4842	ubiquitin-protein ligase activity	APS	F		
70	mitotic sister chromatid segregation	IMP	P	**MND2**	**YIR025W**
7049	cell cycle	IEA	P		
7067	mitosis	IEA	P		

Table D.3.1. GO's of Functionally Unknown Proteins – cont.

7131	reciprocal meiotic recombination	IMP	P		
16567	protein ubiquitination	IPI	P		
30071	regulation of mitotic metaphase/anaphase transition	IMP	P		
30435	sporulation resulting in formation of a cellular spore	RCA	P		
31145	anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	IPI	P		
51301	cell division	IEA	P		
5634	nucleus	IDA	C		
5680	anaphase-promoting complex	IPI	C		
5680	anaphase-promoting complex	IDA	C		
5515	protein binding	APS	F		
4842	ubiquitin-protein ligase activity	APS	F		
6810	transport	IEA	P	**RAV2**	**YDR202C**
7035	vacuolar acidification	IGI	P		
7035	vacuolar acidification	IMP	P		
9117	nucleotide metabolic process	RCA	P		
15031	protein transport	IEA	P		
43254	regulation of protein complex assembly	IMP	P		
43254	regulation of protein complex assembly	IGI	P		
43254	regulation of protein complex assembly	IPI	P		
50801	ion homeostasis	RCA	P		
5737	cytoplasm	IEA	C		
5768	endosome	IEA	C		
16020	membrane	IEA	C		
31901	early endosome membrane	IEA	C		
43291	RAVE complex	IPI	C		
46961	proton-transporting ATPase activity, rotational mechanism	APS	F		
16820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	APS	F		
16787	hydrolase activity	APS	F		
6350	transcription	IEA	P	**RXT3**	**YDL076C**
16568	chromatin modification	IEA	P		
16575	histone deacetylation	IDA	P		
45449	regulation of transcription	IEA	P		
5634	nucleus	IEA	C		
5634	nucleus	IDA	C		

Table D.3.1. GO's of Functionally Unknown Proteins – cont.

33698	Rpd3L complex	IDA	C		
70210	Rpd3L-Expanded complex	IDA	C		
4407	histone deacetylase activity	APS	F		
8270	zinc ion binding	APS	F		
46872	metal ion binding	APS	F		
43565	sequence-specific DNA binding	APS	F		
3	reproduction	RCA	P	**BNI5**	**YNL166C**
910	cytokinesis	IGI	P		
921	septin ring assembly	IPI	P		
921	septin ring assembly	IMP	P		
921	septin ring assembly	IGI	P		
7047	cellular cell wall organization	RCA	P		
7049	cell cycle	IEA	P		
51301	cell division	RCA	P		
144	cellular bud neck septin ring	IDA	C		
144	cellular bud neck septin ring	IPI	C		
5737	cytoplasm	IEA	C		
5933	cellular bud	IEA	C		
5935	cellular bud neck	IDA	C		
166	nucleotide binding	APS	F		
5525	GTP binding	APS	F		
5545	phosphatidylinositol binding	APS	F		
5200	structural constituent of cytoskeleton	APS	F		
3	reproduction	RCA	P	**NIS1**	**YNL078W**
6414	translational elongation	RCA	P		
7088	regulation of mitosis	IPI	P		
51181	cofactor transport	RCA	P		
51301	cell division	RCA	P		
5634	nucleus	IDA	C		
5737	cytoplasm	IEA	C		
5935	cellular bud neck	IDA	C		
5935	cellular bud neck	IEA	C		
5938	cell cortex	IEA	C		
166	nucleotide binding	APS	F		
5525	GTP binding	APS	F		
5545	phosphatidylinositol binding	APS	F		

Table D.3.1. GO's of Functionally Unknown Proteins – cont.

5200	structural constituent of cytoskeleton	APS	F		
6350	transcription	IEA	P	**TEX1**	**YNL253W**
6406	mRNA export from nucleus	IC	P		
45449	regulation of transcription	IEA	P		
346	transcription export complex	IPI	C		
347	THO complex	RCA	C		
445	THO complex part of transcription export complex	RCA	C		
446	nucleoplasmic THO complex	RCA	C		
5634	nucleus	IEA	C		
3676	nucleic acid binding	APS	F		
6337	nucleosome disassembly	IDA	P	**RTT102**	**YGR275W**
6338	chromatin remodeling	IPI	P		
6350	transcription	IEA	P		
6368	RNA elongation from RNA polymerase II promoter	IDA	P		
7059	chromosome segregation	IGI	P		
45449	regulation of transcription	IEA	P		
5634	nucleus	IDA	C		
5634	nucleus	IEA	C		
16514	SWI/SNF complex	IDA	C		
16514	SWI/SNF complex	IPI	C		
16586	RSC complex	IPI	C		
3677	DNA binding	APS	F		
8094	DNA-dependent ATPase activity	APS	F		
82	G1/S transition of mitotic cell cycle	IMP	P	**RSC58**	**YLR033W**
6337	nucleosome disassembly	IDA	P		
6338	chromatin remodeling	IDA	P		
6350	transcription	IEA	P		
6368	RNA elongation from RNA polymerase II promoter	IDA	P		
16568	chromatin modification	IEA	P		
45449	regulation of transcription	IEA	P		
5634	nucleus	IEA	C		
16586	RSC complex	IDA	C		
3677	DNA binding	APS	F		
8094	DNA-dependent ATPase activity	APS	F		

Table D.3.1. GO's of Functionally Unknown Proteins – cont.

					YKR035W-
6623	protein targeting to vacuole	IMP	P	**DID2**	**A**
6810	transport	IEA	P		
15031	protein transport	IEA	P		
32511	late endosome to vacuole transport via multivesicular body sorting pathway	IMP	P		
45184	establishment of protein localization	RCA	P		
45324	late endosome to vacuole transport	IMP	P		
5737	cytoplasm	IDA	C		
5768	endosome	IEA	C		
5768	endosome	IDA	C		
5770	late endosome	IDA	C		
10008	endosome membrane	IEA	C		
12505	endomembrane system	IEA	C		
16020	membrane	IEA	C		
44440	endosomal part	RCA	C		
5515	protein binding	APS	F		
6350	transcription	IEA	P	**CAF130**	**YGR134W**
6357	regulation of transcription from RNA polymerase II promoter	IPI	P		
45449	regulation of transcription	IEA	P		
5634	nucleus	IEA	C		
5737	cytoplasm	IEA	C		
30015	CCR4-NOT core complex	IPI	C		
46872	metal ion binding	APS	F		
3723	RNA binding	APS	F		
6623	protein targeting to vacuole	IMP	P	**MON1**	**YGL124C**
6810	transport	IEA	P		
6914	autophagy	RCA	P		
6914	autophagy	IMP	P		
6914	autophagy	IEA	P		
15031	protein transport	IEA	P		
48278	vesicle docking	IMP	P		
329	fungal-type vacuole membrane	IDA	C		
5768	endosome	IEA	C		
5773	vacuole	IEA	C		
5774	vacuolar membrane	IEA	C		

Table D.3.1. GO's of Functionally Unknown Proteins – cont.

5829	cytosol	IDA	C		
16020	membrane	IEA	C		
31902	late endosome membrane	IEA	C		
32585	multivesicular body membrane	IEA	C		
35091	phosphoinositide binding	APS	F		
6810	transport	IEA	P	**YIF1**	**YNL263C**
6888	ER to Golgi vesicle-mediated transport	IPI	P		
15031	protein transport	IEA	P		
16192	vesicle-mediated transport	IEA	P		
139	Golgi membrane	IEA	C		
5783	endoplasmic reticulum	IEA	C		
5789	endoplasmic reticulum membrane	IPI	C		
5794	Golgi apparatus	IEA	C		
16020	membrane	IEA	C		
16021	integral to membrane	IEA	C		
30134	ER to Golgi transport vesicle	IEA	C		
30134	ER to Golgi transport vesicle	IDA	C		
30173	integral to Golgi membrane	IDA	C		
31410	cytoplasmic vesicle	IEA	C		
5525	GTP binding	APS	F		
166	nucleotide binding	APS	F		
3924	GTPase activity	APS	F		
70	mitotic sister chromatid segregation	IMP	P	**BRN1**	**YBL097W**
278	mitotic cell cycle	IEA	P		
7049	cell cycle	IEA	P		
7067	mitosis	IEA	P		
7076	mitotic chromosome condensation	IMP	P		
7076	mitotic chromosome condensation	IGI	P		
30261	chromosome condensation	IEA	P		
30261	chromosome condensation	RCA	P		
51301	cell division	IEA	P		
70058	tRNA gene clustering	IMP	P		
799	nuclear condensin complex	IPI	C		
799	nuclear condensin complex	ISS	C		
799	nuclear condensin complex	TAS	C		
5634	nucleus	IDA	C		

Table D.3.1. GO's of Functionally Unknown Proteins – cont.

5634	nucleus	IEA	C		
5737	cytoplasm	IEA	C		
32116	cohesin loading complex	IDA	C		
3677	DNA binding	APS	F		
278	mitotic cell cycle	RCA	P	**NSL1**	**YPL233W**
279	M phase	RCA	P		
7049	cell cycle	IEA	P		
7059	chromosome segregation	IDA	P		
7067	mitosis	IEA	P		
51301	cell division	IEA	P		
776	kinetochore	IDA	C		
776	kinetochore	IEA	C		
777	condensed chromosome kinetochore	IEA	C		
818	nuclear MIS12/MIND type complex	IDA	C		
922	spindle pole	IDA	C		
5634	nucleus	IEA	C		
5515	protein binding	APS	F		
16568	chromatin modification	IEA	P	**SIF2**	**YBR103W**
16575	histone deacetylation	IDA	P		
30435	sporulation resulting in formation of a cellular spore	RCA	P		
31939	negative regulation of chromatin silencing at telomere	IPI	P		
31939	negative regulation of chromatin silencing at telomere	IMP	P		
32874	positive regulation of stress-activated MAPK cascade	IMP	P		
45449	regulation of transcription	IMP	P		
45835	negative regulation of meiosis	IPI	P		
5634	nucleus	IEA	C		
5634	nucleus	IDA	C		
34967	Set3 complex	IDA	C		
70210	Rpd3L-Expanded complex	IDA	C		
3677	DNA binding	APS	F		

Table D.3.2. The Neighborhood Enrichment of Functionally Unknown Proteins

iwr1	3677	DNA binding	21/70 30.0	487/6310 7.7
	166	nucleotide binding	16/70 22.8	802/6310 12.7
	5524	ATP binding	13/70 18.5	657/6310 10.4
	16740	transferase activity	12/70 17.1	648/6310 10.2
	3899	DNA-directed RNA polymerase activity	11/70 15.7	36/6310 0.5
	46872	metal ion binding	9/70 12.8	679/6310 10.7
	16787	hydrolase activity	8/70 11.4	637/6310 10.0
	16301	kinase activity	7/70 10.0	206/6310 3.2
	4672	protein kinase activity	7/70 10.0	126/6310 1.9
	4674	protein serine/threonine kinase activity	7/70 10.0	122/6310 1.9
swm1	5515	protein binding	13/37 35.1	242/6310 3.8
	4842	ubiquitin-protein ligase activity	12/37 32.4	82/6310 1.2
	166	nucleotide binding	8/37 21.6	802/6310 12.7
	5524	ATP binding	8/37 21.6	657/6310 10.4
	46872	metal ion binding	6/37 16.2	679/6310 10.7
	3824	catalytic activity	5/37 13.5	482/6310 7.6
	16787	hydrolase activity	5/37 13.5	637/6310 10.0
	3677	DNA binding	5/37 13.5	487/6310 7.7
	5488	binding	4/37 10.8	274/6310 4.3
	16874	ligase activity	4/37 10.8	132/6310 2.0
mnd2	4842	ubiquitin-protein ligase activity	12/26 46.1	82/6310 1.2
	5515	protein binding	12/26 46.1	242/6310 3.8
	5488	binding	5/26 19.2	274/6310 4.3
	46872	metal ion binding	4/26 15.3	679/6310 10.7
	3677	DNA binding	4/26 15.3	487/6310 7.7
rav2	46961	proton-transporting ATPase activity, rotational mechanism	5/11 45.4	20/6310 0.3
	166	nucleotide binding	4/11 36.3	802/6310 12.7
	16820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	3/11 27.2	26/6310 0.4
	16787	hydrolase activity	3/11 27.2	637/6310 10.0
	3677	DNA binding	2/11 18.1	487/6310 7.7
	46933	hydrogen ion transporting ATP synthase activity, rotational mechanism	2/11 18.1	18/6310 0.2
	5525	GTP binding	2/11 18.1	102/6310 1.6
	5524	ATP binding	2/11 18.1	657/6310 10.4
	3924	GTPase activity	2/11 18.1	72/6310 1.1
rxt3	3677	DNA binding	14/52 26.9	487/6310 7.7
	4407	histone deacetylase activity	10/52 19.2	20/6310 0.3
	46872	metal ion binding	9/52 17.3	679/6310 10.7
	43565	sequence-specific DNA binding	8/52 15.3	191/6310 3.0
	166	nucleotide binding	7/52 13.4	802/6310 12.7
	8270	zinc ion binding	6/52 11.5	323/6310 5.1

Table D.3.2. The Neighborhood Enrichment of Functionally Unknown Proteins - cont.

bni5	166	nucleotide binding	7/15 46.6	802/6310 12.7
	5545	phosphatidylinositol binding	5/15 33.3	12/6310 0.1
	5525	GTP binding	5/15 33.3	102/6310 1.6
	5200	structural constituent of cytoskeleton	5/15 33.3	51/6310 0.8
	15631	tubulin binding	3/15 20.0	8/6310 0.1
	51082	unfolded protein binding	3/15 20.0	84/6310 1.3
	16740	transferase activity	3/15 20.0	648/6310 10.2
	4672	protein kinase activity	2/15 13.3	126/6310 1.9
	4674	protein serine/threonine kinase activity	2/15 13.3	122/6310 1.9
	5524	ATP binding	2/15 13.3	657/6310 10.4
	16301	kinase activity	2/15 13.3	206/6310 3.2
	3924	GTPase activity	2/15 13.3	72/6310 1.1
nis1	166	nucleotide binding	13/31 41.9	802/6310 12.7
	5200	structural constituent of cytoskeleton	8/31 25.8	51/6310 0.8
	5524	ATP binding	8/31 25.8	657/6310 10.4
	16740	transferase activity	6/31 19.3	648/6310 10.2
	16301	kinase activity	5/31 16.1	206/6310 3.2
	5545	phosphatidylinositol binding	5/31 16.1	12/6310 0.1
	5525	GTP binding	5/31 16.1	102/6310 1.6
	3677	DNA binding	4/31 12.9	487/6310 7.7
	4672	protein kinase activity	4/31 12.9	126/6310 1.9
	4674	protein serine/threonine kinase activity	4/31 12.9	122/6310 1.9
tex1	3676	nucleic acid binding	9/47 19.1	309/6310 4.8
	166	nucleotide binding	9/47 19.1	802/6310 12.7
	3723	RNA binding	9/47 19.1	446/6310 7.0
	16787	hydrolase activity	8/47 17.0	637/6310 10.0
	3677	DNA binding	8/47 17.0	487/6310 7.7
	5524	ATP binding	6/47 12.7	657/6310 10.4
	16740	transferase activity	5/47 10.6	648/6310 10.2
rtt102	3677	DNA binding	14/51 27.4	487/6310 7.7
	8094	DNA-dependent ATPase activity	14/51 27.4	37/6310 0.5
	16251	general RNA polymerase II transcription factor activity	11/51 21.5	51/6310 0.8
	46872	metal ion binding	10/51 19.6	679/6310 10.7
	166	nucleotide binding	8/51 15.6	802/6310 12.7
	3735	structural constituent of ribosome	8/51 15.6	227/6310 3.5
	5524	ATP binding	7/51 13.7	657/6310 10.4
	16740	transferase activity	6/51 11.7	648/6310 10.2
rsc58	3677	DNA binding	23/58 39.6	487/6310 7.7
	8094	DNA-dependent ATPase activity	16/58 27.5	37/6310 0.5
	46872	metal ion binding	11/58 18.9	679/6310 10.7
	166	nucleotide binding	11/58 18.9	802/6310 12.7
	5524	ATP binding	10/58 17.2	657/6310 10.4
	16787	hydrolase activity	9/58 15.5	637/6310 10.0

Table D.3.2. The Neighborhood Enrichment of Functionally Unknown Proteins - cont.

	16887	ATPase activity	7/58 12.0	139/6310 2.2
	3682	chromatin binding	7/58 12.0	60/6310 0.9
	16740	transferase activity	7/58 12.0	648/6310 10.2
did2	5515	protein binding	3/8 37.5	242/6310 3.8
	5524	ATP binding	2/8 25.0	657/6310 10.4
	166	nucleotide binding	2/8 25.0	802/6310 12.7
	4843	ubiquitin-specific protease activity	1/8 12.5	19/6310 0.3
	1671	ATPase activator activity	1/8 12.5	12/6310 0.1
	3674	molecular_function	1/8 12.5	1994/6310 31.6
	8233	peptidase activity	1/8 12.5	160/6310 2.5
	4175	endopeptidase activity	1/8 12.5	27/6310 0.4
	8234	cysteine-type peptidase activity	1/8 12.5	30/6310 0.4
	16787	hydrolase activity	1/8 12.5	637/6310 10.0
	16887	ATPase activity	1/8 12.5	139/6310 2.2
	51082	unfolded protein binding	1/8 12.5	84/6310 1.3
	16741	transferase activity, transferring one-carbon groups	1/8 12.5	3/6310 0.0
	4221	ubiquitin thiolesterase activity	1/8 12.5	19/6310 0.3
	17111	nucleoside-triphosphatase activity	1/8 12.5	92/6310 1.4
	42623	ATPase activity, coupled	1/8 12.5	29/6310 0.4
caf130	51082	unfolded protein binding	5/25 20.0	84/6310 1.3
	3723	RNA binding	5/25 20.0	446/6310 7.0
	46872	metal ion binding	4/25 16.0	679/6310 10.7
	16787	hydrolase activity	4/25 16.0	637/6310 10.0
	3676	nucleic acid binding	4/25 16.0	309/6310 4.8
	175	3'-5'-exoribonuclease activity	3/25 12.0	16/6310 0.2
	15631	tubulin binding	3/25 12.0	8/6310 0.1
	5515	protein binding	3/25 12.0	242/6310 3.8
mon1	166	nucleotide binding	15/46 32.6	802/6310 12.7
	5524	ATP binding	11/46 23.9	657/6310 10.4
	46872	metal ion binding	8/46 17.3	679/6310 10.7
	3677	DNA binding	7/46 15.2	487/6310 7.7
	16740	transferase activity	7/46 15.2	648/6310 10.2
yif1	166	nucleotide binding	14/33 42.4	802/6310 12.7
	5525	GTP binding	11/33 33.3	102/6310 1.6
	3924	GTPase activity	11/33 33.3	72/6310 1.1
	16740	transferase activity	7/33 21.2	648/6310 10.2
	32266	phosphatidylinositol-3-phosphate binding	6/33 18.1	24/6310 0.3
	35091	phosphoinositide binding	6/33 18.1	32/6310 0.5
	4672	protein kinase activity	4/33 12.1	126/6310 1.9
	5524	ATP binding	4/33 12.1	657/6310 10.4
	16301	kinase activity	4/33 12.1	206/6310 3.2
brn1	5524	ATP binding	4/14 28.5	657/6310 10.4

Table D.3.2. The Neighborhood Enrichment of Functionally Unknown Proteins - cont.

	166	nucleotide binding	4/14 28.5	802/6310 12.7
	3682	chromatin binding	3/14 21.4	60/6310 0.9
	16887	ATPase activity	3/14 21.4	139/6310 2.2
	16787	hydrolase activity	2/14 14.2	637/6310 10.0
	217	DNA secondary structure binding	2/14 14.2	6/6310 0.0
	3680	AT DNA binding	2/14 14.2	5/6310 0.0
	3677	DNA binding	2/14 14.2	487/6310 7.7
nsl1	5488	binding	5/22 22.7	274/6310 4.3
	51082	unfolded protein binding	5/22 22.7	84/6310 1.3
	3677	DNA binding	4/22 18.1	487/6310 7.7
	5524	ATP binding	4/22 18.1	657/6310 10.4
	166	nucleotide binding	4/22 18.1	802/6310 12.7
	5515	protein binding	4/22 18.1	242/6310 3.8
	16887	ATPase activity	3/22 13.6	139/6310 2.2
sif2	166	nucleotide binding	16/82 19.5	802/6310 12.7
	3677	DNA binding	14/82 17.0	487/6310 7.7
	5524	ATP binding	12/82 14.6	657/6310 10.4
	46872	metal ion binding	10/82 12.1	679/6310 10.7
	4407	histone deacetylase activity	9/82 10.9	20/6310 0.3
	16787	hydrolase activity	9/82 10.9	637/6310 10.0

D.4 Domains of Function Assigned Yeast Proteins

IWR2: Protein of unknown function (DUF1762)

InterPro entry IPR013883

This is a group of proteins of unknown function. is known to interact with RNA polymerase II and deletion of this protein results in hypersensitivity to the K1 killer toxin.

NKP1:

No domains

MND2:

Apc15p protein (PF05841)

InterPro entry IPR008402

The anaphase-promoting complex (APC) is a multi-subunit E3 protein ubiquitin ligase that is responsible for the metaphase to anaphase transition and the exit from mitosis. Anaphase is initiated when the APC triggers the destruction of securin, thereby

allowing the protease, separase, to disrupt sister-chromatid cohesion. Securin ubiquitination by the APC is inhibited by cyclin-dependent kinase 1 (Cdk1)-dependent phosphorylation.

Forkhead Box M1 (FoxM1), which is a transcription factor that is over-expressed in many cancers, is degraded in late mitosis and early G1 phase by the APC/cyclosome (APC/C) E3 ubiquitin ligase. The APC/C targets mitotic cyclins for destruction in mitosis and G1 phase and is then inactivated at S phase. It thereby generates alternating states of high and low cyclin-Cdk activity, which is required for the alternation of mitosis and DNA replication.

RAV2:

Rogdi leucine zipper containing protein (PF10259)

InterPro entry IPR019386

This is a family of conserved proteins which, it has been suggested, contain leucine-zipper domains. A leucine zipper domain is a region of 30 amino acids with leucines repeating every seven or eight residues; a pattern which these proteins match. The protein in Drosophila comes from the gene ROGDI.

RXT3:

Histone deacetylation protein Rxt3 (PF08642)

InterPro entry IPR013951

Rxt3 has been shown in yeast to be required for histone deacetylation.

BNI5:

No domain

NIS1:

No domain

TEX1:

WD domain, G-beta repeat (PF00400)

InterPro entry IPR019781

WD-40 repeats (also known as WD or beta-transducin repeats) are short ~40 amino acid motifs, often terminating in a Trp-Asp (W-D) dipeptide. WD40 repeats usually assume a 7-8 bladed beta-propeller fold, but proteins have been found with 4 to 16 repeated units, which also form a circularised beta-propeller structure. WD-repeat proteins are a large family found in all eukaryotes and are implicated in a variety of functions ranging from signal transduction and transcription regulation to cell cycle control and apoptosis. Repeated WD40 motifs act as a site for protein-protein interaction, and proteins containing WD40 repeats are known to serve as platforms for the assembly of protein complexes or mediators of transient interplay among other proteins. The specificity of the proteins is determined by the sequences outside the repeats themselves. Examples of such complexes are G proteins (beta subunit is a beta-propeller), TAFII transcription factor, and E3 ubiquitin ligase.

RTT102:

Rtt102p-like transcription regulator protein (PF09510)

InterPro entry IPR018304

Rtt102p (Regulator of Ty1 Transposition Protein 102) is a transcription regulator protein found in fungi that appears to be integrally associated with both the Swi-Snf and the RSC chromatin remodelling complexes. RSC is involved in transcription regulation and nucleosome positioning, and is responsible for the transfer of a histone octamer from a nucleosome core particle to naked DNA. The reaction requires ATP and involves an activated RSC-nucleosome intermediate. Remodelling reaction also involves DNA translocation, DNA twist and conformational change. As a reconfigurer of centromeric and flanking nucleosomes, RSC complex is required both for proper kinetochore function in chromosome segregation and, via a PKC1-dependent signalling pathway, for organisation of the cellular cytoskeleton. It is a probable component of the SWI/SNF complex, an ATP-dependent chromatin-remodelling complex, is required for the positive and negative regulation of gene expression of a large number of genes. It changes chromatin structure by altering DNA-histone contacts within a nucleosome, leading eventually to a change in nucleosome position, thus facilitating or repressing binding of gene-specific transcription factors.

RSC58:

Bromodomain (PF00439)

InterPro entry IPR001487

Bromodomains are found in a variety of mammalian, invertebrate and yeast DNA-binding proteins. Bromodomains can interact with acetylated lysine. In some proteins, the classical bromodomain has diverged to such an extent that parts of the region are either missing or contain an insertion (e.g., mammalian protein HRX, Caenorhabditis elegans hypothetical protein ZK783.4, yeast protein YTA7). The bromodomain may occur as a single copy, or in duplicate.

The precise function of the domain is unclear, but it may be involved in protein-protein interactions and may play a role in assembly or activity of multi-component complexes involved in transcriptional activation.

DID2:

Snf7

InterPro entry IPR005024

This is a family of eukaryotic proteins which are variously described as either hypothetical protein, developmental protein or related to yeast SNF7. The family contains human CHMP1. CHMP1 (CHromatin Modifying Protein; CHarged Multivesicular body Protein), is encoded by an alternative open reading frame in the PRSM1 gene PUBMED:8863740 and is conserved in both complex and simple eukaryotes. CHMP1 contains a predicted bipartite nuclear localisation signal and distributes as distinct forms to the cytoplasm and the nuclear matrix in all cell lines tested.

Human CHMP1 is strongly implicated in multivesicular body formation. A multivesicular body is a vesicle-filled endosome that targets proteins to the interior of lysosomes. Immunocytochemistry and biochemical fractionation localise CHMP1 to early endosomes and CHMP1 physically interacts with SKD1/VPS4, a highly conserved protein directly linked to multivesicular body sorting in yeast. Similar to the action of a mutant SKD1 protein, over expression of a fusion derivative of human CHMP1 dilates endosomal compartments and disrupts the normal distribution of several endosomal markers. Genetic studies in Saccharomyces cerevisiae (Baker's yeast) further support a conserved role of CHMP1 in vesicle trafficking. Deletion of CHM1, the budding yeast homologue of

CHMP1, results in defective sorting of carboxypeptidases S and Y and produces abnormal, multi-lamellar prevacuolar compartments. This phenotype classifies CHM1 as a member of the class E vacuolar protein sorting gene.

CAF130:
No domain.

MON1:

Trafficking protein Mon1 (PF03164)
InterPro entry IPR004353

Members of this family have been called SAND proteins although these proteins do not contain a SAND domain. In Saccharomyces cerevisiae a protein complex of Mon1 and Ccz1 functions with the small GTPase Ypt7 to mediate vesicle trafficking to the vacuole. The Mon1/Ccz1 complex is conserved in eukaryotic evolution and members of this family (previously known as DUF254) are distant homologues to domains of known structure that assemble into cargo vesicle adapter (AP).

YIF1:
YIF1 (PF03878)
InterPro entry IPR005578

This family includes a number of eukaryotic proteins. It is an integral membrane protein, conserved in at least 1 copy in all sequenced eukaryotes. The gene name in Schizosaccharomyces pombe (Fission yeast) is hrf1+ for Heavy metal Resistance Factor 1.

BRN1:
Barren protein (PF05786)
InterPro entry IPR008418

This family consists of several Barren protein homologues from several eukaryotic organisms. In Drosophila Barren (barr) is required for sister-chromatid segregation in mitosis. barr encodes a novel protein that is present in proliferating cells and has homologues in Saccharomyces cerevisiae and Homo sapiens. Mitotic defects in barr embryos become apparent during cycle 16, resulting in a loss of PNS and CNS neurons.

Centromeres move apart at the metaphase-anaphase transition and Cyclin B is degraded, but sister chromatids remain connected, resulting in chromatin bridging. Barren protein localises to chromatin throughout mitosis. Colocalisation and biochemical experiments indicate that Barren associates with Topoisomerase II throughout mitosis and alters the activity of Topoisomerase II. It has been suggested that this association is required for proper chromosomal segregation by facilitating the decatenation of chromatids at anaphase.

Q9Y7R3 condensin, Cnd2 [Schizosaccharomyces pombe] NP_587811 742 29,17771883 5,261E-72 protein binding

NSL1:

Kinetochore protein Mis14 like (PF08641)

InterPro entry IPR013950

Mis14 is a kinetochore protein which is known to be recruited to kinetochores independently of CENP-A.

Q9P6M3 Kinetochore protein mis14 s.pombe protein binding IPI

APPENDIX E

E.1.1 Details of Worm Core Proteins

Table E.1.1. Worm core protein list

Grey ones are the 17 core proteins selected.

adp-1	WBGene00000078		Uncloned locus that affects calcium-dependent adaptation to AWC-sensed volatile odorants and adaptation to salts.
aex-1	WBGene00000084	C2	AEX-1 is a novel C2 domain protein; C2 is a proposed calcium binding domain; AEX-1 is weakly similar to the rat protein Munc13, which regulates vesicle exocytosis; AEX-1 regulates UNC-13 localization; mutation of dgk-1, or a gain-of-function mutation of egl-30, suppresses the aex-1 mutant phenotype.
aex-5	WBGene00000088		an ortholog of calcium-dependent serine endoproteinases, the kex2/subtilisin-like proprotein convertase that affects defecation cycle and anterior body contractions; and is expressed in all muscle cells except pharyngeal muscles.
C24H10.1	WBGene00016067		C24H10.1 encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; C24H10.1 is worm-specific, with obvious homologs only in C. elegans; C24H10.1 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.
C56C10.9	WBGene00016965	EF Hand	Calcium binding protein (IEA)
cal-1	WBGene00000285	EF Hand	cal-1 encodes one of five predicted C. elegans calcium-binding calmodulin homologs (the others being CAL-2, CAL-3, CAL-4, and CMD-1); as loss of cal-1 function via RNA-mediated interference (RNAi) does not result in any abnormalities, the precise role of CAL-1 in C. elegans development and/or behavior is not yet known.
cal-4	WBGene00000288	EF Hand	cal-4 encodes one of five predicted C. elegans calcium-binding calmodulin homologs (the others being CAL-1, CAL-2, CAL-3 and CMD-1); as loss of cal-4 function via RNA-mediated interference (RNAi) does not result in any abnormalities, the precise role of CAL-4 in C. elegans development and/or behavior is not yet known.
cal-5	WBGene00006861	EF Hand	uvt-2 encodes a protein that contains four EF-hand calcium binding motifs with similarity to human calmodulin; mRNA weakly expressed in L1 through L4 larval stages and in the adult hermaphrodite.
calu-1	WBGene00019760	EF Hand	
cbn-1	WBGene00019285	EF Hand	cbn-1 encodes a protein that contains a predicted signal peptide and calcium-binding EF-hand motifs.

Table E.1.1. Worm core protein list - cont.

cca-1	WBGene00000367		cca-1 encodes a calcium channel alpha subunit that is homologous to vertebrate T-type calcium channel alpha 1 subunits; CCA-1 is required for regulation of pharyngeal pumping, specifically for the efficient initiation of action potentials in the pharynx in response to excitatory inputs; a CCA-1::GFP fusion protein is expressed strongly in pharyngeal muscle as well as in many neurons, including particular subsets of neurons in the head, pharynx, ventral nerve cord, and anal ganglia
ccb-1	WBGene00000368		ccb-1 encodes a homolog of the beta subunits of dihydropyridine sensitive L-type calcium channels that, in RNAi screens, is dispensable for viability or grossly normal morphology; a mutant allele of ccb-1 exists, but its phenotype has not yet been described.
ccb-2	WBGene00021125		L-type voltage-dependent Ca^{2+} channel, beta subunit
ccz-1	WBGene00010283		ccz-1 encodes an ortholog of human C7orf28A and C7orf28B, with more distant similarity to human HPS4 (OMIM:606682) and S. cerevisiae Ccz1p in its N-terminal CHiPS domain; the CHiPS domain of CCZ-1 is predicted to have a longin-like fold; by analogy with Ccz1p and C. elegans SAND-1, CCZ-1 might be required for normal (RAB-7-mediated) traffic between early and late endosomes, or for autophagy; CCZ-1 is bound by SAND-1.
ckk-1	WBGene00000518		ckk-1 encodes a serine/threonine protein kinase that is the C. elegans calcium and calmodulin-dependent protein kinase kinase (CaM-KK) ortholog; ckk-1 activity is required for CRE-dependent transcription in neurons; when expressed in vitro, CKK-1 is capable of stimulating the kinase activity of a downstream kinase, either C. elegans CMK-1 or recombinant mouse CaM-KIV, in a Ca^{2+}/calmodulin-dependent manner; when expressed in COS-7 cells, CKK-1 and CMK-1 positively regulate the transcriptional activity of endogenous CREB.
clc-1	WBGene00000522		clc-1 encodes a claudin homolog, closely similar to CLC-2, that is required for normal cohesion of apical junctions in epithelia; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water; CLC-1 maintains the impermeability ('barrier function') of epithelia, since clc-1(RNAi) animals have abnormal permeability of the pharynx to dyes; clc-1 is expressed in spermatheca, pharynx, intestine, hypodermis, the excretory-secretory system, and the cell-cell junctions of the vulva; in pharyngeal cells, CLC-1 colocalizes with AJM-1 in long thin lines, parallel to the pharyngeal axis and lining the lumenal surface, that appear to correspond with apical intercellular junctions.
clc-2	WBGene00000523		clc-2 encodes a claudin homolog, closely similar to CLC-1, that is required for normal cohesion of apical junctions in epithelia; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water; CLC-2 maintains the impermeability ('barrier function') of epithelia, since clc-1(RNAi) animals have abnormal permeability of the hypodermis to dyes; clc-2 is expressed in hypodermal seam cells, with two diffuse lines of CLC-2 protein.
clc-3	WBGene00000524		clc-3 encodes a claudin homolog that may regulate ion channels; CLC-3 is worm-specific, with a highly divergent sequence weakly similar to mammalian voltage-dependent calcium channel gamma subunits that are known or suspected to prevent epilepsy in vivo (e.g., stargazin; MGI:1316660); claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.
clc-4	WBGene00000525		clc-4 encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; CLC-4 is worm-specific, with obvious homologs only in C. elegans; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.

Table E.1.1. Worm core protein list - cont.

clc-5	WBGene00000526		clc-5 encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; CLC-5 is worm-specific, with obvious homologs only in C. elegans; CLC-5 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.
clhm-1	WBGene00016626		
clp-2	WBGene00000543		clp-2 encodes a large calpain subunit that contains two predicted MIT (microtubule interacting and trafficking) domains and is homologous to the mammalian Calpain 7 proteins and the atypical, nuclear-localized Aspergillus calpain, PalB; by homology, CLP-2 is predicted to function as a calcium-dependent, cysteine protease that is involved in intracellular proteolysis and peptidolysis; however, as loss of clp-2 activity via RNA-mediated interference (RNAi) does not result in any obvious abnormalities, the precise role of CLP-2 in C. elegans development and/or behavior is not yet known.
clp-4	WBGene00000545		clp-4 encodes a large calpain subunit that is homologous to Drosophila CalpainB and the mammalian muscle-specific Calpain 3 proteins (OMIM:114240, mutations are associated with limb-girdle muscular dystrophy); by homology, CLP-4 is predicted to function as a nonlysosomal, calcium-dependent, cysteine protease that is involved in intracellular proteolysis and peptidolysis; however, as loss of clp-4 activity via RNA-mediated interference (RNAi) does not result in any obvious abnormalities, the precise role of CLP-4 in C. elegans development and/or behavior is not yet known.
cnb-1	WBGene00000554	EF Hand	cnb-1 encodes an ortholog of calcineurin B, the regulatory subunit of the protein phosphatase 2B with four EF-hand motifs for calcium binding, that binds TAX-6 (a calcineurin A ortholog) in a calcium-dependent manner; CNB-1 binds calcium, enhances the phosphatase activity of TAX-6 in vitro, promotes transcription of rcn-1, and is required for normal cuticle formation, sperm morphology, and brood size; cnb-1 null mutants exhibit an extended lifespan which is characterized by enhanced autophagy and dependent upon the presence of functional bec-1 and atg-7.
cnx-1	WBGene00000567		cnx-1 encodes the C. elegans ortholog of calnexin, a type I Ca^{2+}-binding integral membrane protein of the endoplasmic reticulum (ER); CNX-1 binds calcium and is predicted to function as a molecular chaperone required for glycoprotein folding and maturation as well as regulation of intracellular calcium homeostasis; in C. elegans, cnx-1 activity is required at 25 degrees C for wild-type levels of fertility and normal embryonic and larval development; cnx-1 activity is also required for maintaining viability in response to ER stress; in addition, cnx-1(RNAi) can suppress necrotic-like cell death induced by hyperactivated MEC-4 and DEG-1 ion channels, suggesting that CNX-1 also plays a role in regulating necrotic cell death; CNX-1 expression is first detected ubiquitously in the early embryo, with expression then becoming restricted to embryonic head and tail regions; post-embryonic expression is seen in the excretory cell, head and tail neurons, spermatheca, intestine, germ cells, and spicules in the male tail; antibody staining of early embryos indicates that CNX-1 localizes to the endoplasmic reticulum; in regulating reproduction at 25 degrees C, cnx-1 functions redundantly with crt-1, which encodes calreticulin, an additional Ca^{2+}-binding ER chaperone protein.
cpna-3	WBGene00007221		cpna-3 encodes a putative calcium-dependent phospholipid binding protein of the copine family; cpna-3 is required for sperm to normally migrate towards a PUFA-based signal exuded by oocytes; cpna-3(RNAi) hermaphrodites are infertile, both through aberrant loss of their own sperm and through failure of males to effectively inseminate them.

Table E.1.1. Worm core protein list - cont.

crt-1	WBGene00000802		crt-1 encodes an ortholog of calreticulin (a calcium-binding molecular chaperone of the endoplasmic reticulum); crt-1 is dispensable for viability, but required for normal sperm development, male mating efficiency, some forms of necrotic cell death, and hermaphrodite fertility at high temperatures (which may reflect a stress response); crt-1 expression is induced by stress; CRT-1 protein binds $Ca^{(2+)}$ and can suppress heat-induced protein aggregation in vitro; crt-1(bz29) and crt-1(bz30) mutants suppress necrotic cell death induced either by mec-4(d) or by a constitutively activated Gas subunit; crt-1(bz29); crt-1's suppression of necrosis is itself partly reversed by thapsigargin, which enhances calcium release from the endoplasmic reticulum; crt-1(bz30), and crt-1(jh101) mutants grow more slowly, and have have reduced broods at 25 deg. C.; crt-1(jh101) mutants are slightly shorter than normal and have defective sperm; CRT-1 and ITR-1 serve partly redundant functions in vivo, since crt-1;itr-1(sa73) are highly infertile and slow-growing with greatly slowed defecation; a number of proteins, notably HSP-3/4 and PDI-2/3, are over-expressed in crt-1(jh101) mutants and crt-1(jh101);cnx-1(nr2009) double mutants.
cul-5	WBGene00000840		cul-5 encodes a cullin, orthologous to vasopressin-activated calcium-mobilizing receptor-1 in mammals, that is thought to regulate the cell cycle by virtue of its paralogy to cul-1 and cul-2; however, CUL-5 is dispensable for viability and gross morphology in mass RNAi screens, and does not appear to interact with any of 17 Skp1-related proteins (SKR-1 through SKR-21).
efk-1	WBGene00001160		efk-1 encodes a calcium/calmodulin-dependent protein kinase with homology to human elongation factor-2 kinase; expressed in only a few cells in embryos.
egl-19	WBGene00001187		egl-19 encodes an ortholog of the alpha subunit of mammalian L-type calcium ion channels that affects muscle contraction in late embryonic morphogenesis, movement, egg-laying, mating and feeding; egl-19 is expressed in muscle cells and some neurons.
F10A3.1	WBGene00008631		F10A3.1 encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; F10A3.1 is worm-specific, with obvious homologs only in C. elegans; F10A3.1 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.
F44G3.10	WBGene00009710		F44G3.10 encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; F44G3.10 is worm-specific, with obvious homologs only in C. elegans; F44G3.10 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.
F53B3.5	WBGene00018743		F59C6.11 encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; F59C6.11 is worm-specific, with obvious homologs only in C. elegans; F59C6.11 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.
F59C6.11	WBGene00010328		F53B3.5 encodes a claudin homolog that may regulate ion channels; F53B3.5 is distantly similar to mammalian voltage-dependent calcium channel gamma subunits that are known or suspected to prevent epilepsy in vivo (e.g., stargazin; MGI:1316660); F53B3.5 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.

Table E.1.1. Worm core protein list - cont.

gon-2	WBGene00001651		The gon-2 gene encodes a predicted transmembrane cation channel homologous to the long transient receptor potential channel (LTRPC) subfamily of TRP channels including Drosophila DLTRP and mammalian melastatin and LTRPC7; GON-2 regulates the onset and continuation of post-embryonic mitotic cell divisions in the somatic gonad precursor cells perhaps by mediating an increase in intracellular calcium in response to activating signals.
grk-2	WBGene00001709		grk-2 encodes a serine/threonine protein kinase that is most closely related to G protein-coupled receptor (GPCR) kinases; in C. elegans, grk-2 activity is required in chemosensory neurons for normal calcium influx during chemosensation, but is not required for GPCR downregulation; in regulating chemosensation, GRK-2 likely acts upstream of the ODR-3/G alpha protein, as overexpression of ODR-3 can rescue abnormal avoidance behavior in grk-2 mutants; GRK-2 is broadly expressed in the adult nervous system, with expression beginning as early as the 20-30-cell stage of embryonic development and continuing into adulthood.
gsnl-1	WBGene00010593		gsnl-1 encodes one of three C. elegans gelsolin-related proteins; GSNL-1 contains four gelsolin-like (G) domains; GSNL-1 is predicted to function in vivo as an actin regulatory protein and in vitro, demonstrates actin filament severing activity; GSNL-1 also exhibits calcium-dependent capping of the barbed ends of actin filaments and stays bound to the side of actin filaments; GSNL-1 additionally binds to G-actin.
gtl-2	WBGene00001796		gtl-2 encodes a predicted receptor-activated calcium channel and represents a family of calcium channels, the other two family members being gtl-1 and gon-2.
hmr-1	WBGene00001980		hmr-1 encodes two isoforms of a classical cadherin that contain extracellular cadherin, EGF-like, and laminin G domains as well as a highly conserved intracellular domain that binds beta-catenin; by homology, the HMR-1 proteins are predicted to function as calcium-dependent, homophilic cell-cell adhesion receptors; in vivo, HMR-1A activity is required for mediating the cell migrations, cell shape changes, and presumably actin cytoskeleton rearrangements that occur during embryonic morphogenesis; HMR-1B is required for fasciculation and outgrowth of a subset of motor neuron processes; HMR-1 antibodies detect expression in all embryonic blastomeres at early stages of development but as hypodermal adherens junctions form and morphogenesis begins, HMR-1 becomes highly expressed in the apical junctions of all hypodermal cells; later, HMR-1 is expressed primarily at the apical margins of hypodermal, pharyngeal, and intestinal cells in a pattern consistent with localization to adherens junctions; an HMR-1B reporter fusion is expressed in neurons, including the DD, VD, and AS class of motor neurons; HMR-1 activity is required for proper localization of other junctional components, such as HMP-1/alpha-catenin and HMP-2/beta-catenin, but not for localization of AJM-1, the junctional component recognized by the MH27 antibody.
K10D6.2	WBGene00010742		K10D6.2 encodes three isoforms of a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; K10D6.2A-C are worm-specific, with obvious homologs only in C. elegans; K10D6.2A-C are required for embryonic viability in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.
kcnl-1	WBGene00007176		Ca^{2+}-activated K+ channel proteins (intermediate/small conductance classes)
kcnl-2	WBGene00008570		Ca^{2+}-activated K+ channel proteins (intermediate/small conductance classes)
kpc-1	WBGene00002232		kpc-1 encodes two isoforms of a Kex2/subtilisin-like proprotein convertase that is a member of the subtilase family of calcium-dependent serine endoproteinases that includes mammalian furin (OMIM:136950) and Saccharomyces cerevisiae Kex2p; KPC-1 is required for normal growth and locomotion and is likely responsible for the processing of transforming growth factor (TGF) beta proproteins encoded by daf-7, dbl-1, and unc-129.
lin-2	WBGene00002991		lin-2 encodes a protein belonging to the membrane associated guanylate kinase (MAGUK) family, with several domains (L27, PDZ, SH3, and guanylate kinase) thought to assemble specific multiprotein complexes in particular regions of the cell; in vivo, LIN-2 is required for the correct localization of LET-23 (and, presumably, other membrane proteins) to specific regions of the plasma membrane.

Table E.1.1. Worm core protein list - cont.

mca-1	WBGene00003151		mca-1 encodes one of three C. elegans plasma membrane Ca^{2+} ATPases (PMCAs); by homology, MCA-1 is predicted to function as a molecular pump that couples ATP hydrolysis to extrusion of cytosolic Ca^{2+}; when expressed in insect cells, MCA-1 exhibits calcium-pumping ATPase activity and binding to human calmodulin; large-scale RNAi screens indicate that mca-1 activity is required for normal coordinated locomotion and rates of growth; an mca-1::gfp reporter fusion is expressed solely in the excretory canal.
mca-2	WBGene00019875		mca-2 encodes a calcium-transporting ATPase; large-scale RNAi screens indicate that mca-2 activity is required for locomotion and normal rates of postembryonic growth. [
mca-3	WBGene00003153		mca-3 encodes one of three C. elegans plasma membrane Ca^{2+} ATPases (PMCAs); by homology, MCA-3 is predicted to function as a molecular pump that couples ATP hydrolysis to extrusion of cytosolic Ca^{2+}; mca-3 activity is required for coelomocyte endocytosis, normal coordinated locomotion, and embryonic and larval development; in regulating endocytosis, MCA-3 is required specifically in the early stages for recruitment of the endocytic machinery to the plasma membrane; an mca-3::gfp reporter is expressed from embryogenesis through adulthood and seen in body wall muscle, the nervous system, intestine, and coelomocytes; in coelomocytes, an MCA-3::GFP fusion protein localizes to the plasma membrane.
nca-2	WBGene00003558		nca-2 encodes a novel, four-domain alpha1U Ca^{2+} channel subunit; nca-2 functions redundantly with a paralog, nca-1, to regulate locomotion and the response to specific volatile anesthetics; in regulating these processes, nca-1 and nca-2 function in the same pathway as unc-79, which encodes a conserved novel protein that positively regulates levels of NCA-1 and NCA-2 protein; nca-2::gfp reporters are expressed in the nervous system in a few pharyngeal neurons, many nerve ring interneurons, most ventral nerve motor cord neurons, and in neurons and muscles in the tail; diffuse expression is also seen in vulval muscles and the intestine.
ncs-1	WBGene00003563	EF Hand	The ncs-1 gene encodes a neuronal calcium sensor protein that, along with calcium, is essential for thermotaxis along isothermal paths in thermal gradients.
ncs-2	WBGene00003564	EF Hand	Ca^{2+} sensor
ncs-3	WBGene00003565	EF Hand	ncs-3 encodes a neuronal calcium sensor (NCS) protein with no obvious null phenotype.
nex-1	WBGene00003588		nex-1 encodes an annexin, a member of a family of calcium-dependent phospholipid binding proteins; NEX-1 is required for efficient engulfment of apoptotic cell corpses in the pharynx, and may also function in other membrane fusion events, such as exocytosis; NEX-1 in vitro can bind phosphatidylserine, phosphatidylinositol, heparin, heparan sulfate, and chondroitin sulfate; NEX-1 is expressed in the pharynx, hypodermal cells, the vulva, the uterus, the spermathecal valve, and yolk granules of maturing oocytes.
nex-2	WBGene00003589		nex-2 encodes an annexin, a member of a family of calcium-dependent phospholipid binding proteins; by homology, NEX-2 could function in a number of processes, such as membrane fusion, cytoskeletal interactions, and intracellular signaling; however, as loss of nex-2 function via RNA-mediated interference (RNAi) does not result in any abnormalities, the precise role of NEX-2 in C. elegans development and/or behavior is not yet known
nex-3	WBGene00003590		nex-3 encodes an annexin, a member of a family of calcium-dependent phospholipid binding proteins; by homology, NEX-3 could function in a number of processes, such as membrane fusion, cytoskeletal interactions, and intracellular signaling; however, as loss of nex-3 function via RNA-mediated interference (RNAi) does not result in any abnormalities, the precise role of NEX-3 in C. elegans development and/or behavior is not yet known.

Table E.1.1. Worm core protein list - cont.

nra-1	WBGene00012128	C2	nra-1 encodes a putative calcium-dependent phospholipid binding protein of the copine family; nra-1 is required for sperm to normally migrate towards a PUFA-based signal exuded by oocytes; nra-1(RNAi) hermaphrodites are infertile, both through aberrant loss of their own sperm and through failure of males to effectively inseminate them.
nsy-4	WBGene00021415		Y38F2AL.1 encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; Y38F2AL.1 is worm-specific, with obvious homologs only in C. elegans; Y38F2AL.1 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.
nucb-1	WBGene00009674	EF Hand	F44A6.1 encodes two isoforms of a nucleobindin homolog that each contain a calcium-binding EF-hand domain and a glutamine/asparagine-rich domain; F44A6.1 is expressed in muscle (pharyngeal, body wall, and vulval), with weak expression in head and tail neurons and in (mainly posterior) intestinal cells; F44A6.1 has no obvious function in vivo, since it has no obvious phenotype when deleted or inactivated by RNAi (whether in a normal genetic background, or in goa-1 and egl-30 mutant backgrounds). [
orai-1	WBGene00015648		CAP-binding protein complex interacting protein 2
otpl-1	WBGene00015035		otpl-1 encodes an ortholog of otopetrins such as human OTOP1 (OMIM:607806), OTOP2 (OMIM:607827), and OTOP3 (OMIM:607828), along with other homologs in fish and insects (such as crustacean OTOPLb); otopetrins are predicted to have 12 transmembrane helices; by homology with OTOP1 and OTOPLb, OTPL-1 might regulate intracellular calcium fluxes during extracellular stimuli, promote extracellular calcium carbonate biomineralization, or both; OTPL-1's paralogs include OTPL-2 through OTPL-8; OTPL-1 has no function in mass RNAi assays.
otpl-2	WBGene00017034		otpl-2 encodes an ortholog of otopetrins such as human OTOP1 (OMIM:607806), OTOP2 (OMIM:607827), and OTOP3 (OMIM:607828), along with other homologs in fish and insects (such as crustacean OTOPLb); otopetrins are predicted to have 12 transmembrane helices; by homology with OTOP1 and OTOPLb, OTPL-2 might regulate intracellular calcium fluxes during extracellular stimuli, promote extracellular calcium carbonate biomineralization, or both; OTPL-2's paralogs include OTPL-1 through OTPL-8; OTPL-2 has no function in mass RNAi assays.
otpl-3	WBGene00017449		otpl-3 encodes an ortholog of otopetrins such as human OTOP1 (OMIM:607806), OTOP2 (OMIM:607827), and OTOP3 (OMIM:607828), along with other homologs in fish and insects (such as crustacean OTOPLb); otopetrins are predicted to have 12 transmembrane helices; by homology with OTOP1 and OTOPLb, OTPL-3 might regulate intracellular calcium fluxes during extracellular stimuli, promote extracellular calcium carbonate biomineralization, or both; OTPL-3's paralogs include OTPL-1 through OTPL-8; OTPL-3 has no function in mass RNAi assays.
otpl-4	WBGene00018477		otpl-4 encodes an ortholog of otopetrins such as human OTOP1 (OMIM:607806), OTOP2 (OMIM:607827), and OTOP3 (OMIM:607828), along with other homologs in fish and insects (such as crustacean OTOPLb); otopetrins are predicted to have 12 transmembrane helices; by homology with OTOP1 and OTOPLb, OTPL-4 might regulate intracellular calcium fluxes during extracellular stimuli, promote extracellular calcium carbonate biomineralization, or both; OTPL-4's paralogs include OTPL-1 through OTPL-8; OTPL-4 has no function in mass RNAi assays
otpl-5	WBGene00018478		otpl-5 encodes an ortholog of otopetrins such as human OTOP1 (OMIM:607806), OTOP2 (OMIM:607827), and OTOP3 (OMIM:607828), along with other homologs in fish and insects (such as crustacean OTOPLb); otopetrins are predicted to have 12 transmembrane helices; by homology with OTOP1 and OTOPLb, OTPL-5 might regulate intracellular calcium fluxes during extracellular stimuli, promote extracellular calcium carbonate biomineralization, or both; OTPL-5's paralogs include OTPL-1 through OTPL-8; OTPL-5 has no function in mass RNAi assays.

Table E.1.1. Worm core protein list - cont.

otpl-6	WBGene00018479		otpl-6 encodes an ortholog of otopetrins such as human OTOP1 (OMIM:607806), OTOP2 (OMIM:607827), and OTOP3 (OMIM:607828), along with other homologs in fish and insects (such as crustacean OTOPLb); otopetrins are predicted to have 12 transmembrane helices; by homology with OTOP1 and OTOPLb, OTPL-6 might regulate intracellular calcium fluxes during extracellular stimuli, promote extracellular calcium carbonate biomineralization, or both; OTPL-6's paralogs include OTPL-1 through OTPL-8; OTPL-6 has no function in mass RNAi assays.
otpl-7	WBGene00018480		otpl-7 encodes an ortholog of otopetrins such as human OTOP1 (OMIM:607806), OTOP2 (OMIM:607827), and OTOP3 (OMIM:607828), along with other homologs in fish and insects (such as crustacean OTOPLb); otopetrins are predicted to have 12 transmembrane helices; by homology with OTOP1 and OTOPLb, OTPL-7 might regulate intracellular calcium fluxes during extracellular stimuli, promote extracellular calcium carbonate biomineralization, or both; OTPL-7's paralogs include OTPL-1 through OTPL-8; OTPL-7 has no function in mass RNAi assays.
otpl-8	WBGene00010403		otpl-8 encodes an ortholog of otopetrins such as human OTOP1 (OMIM:607806), OTOP2 (OMIM:607827), and OTOP3 (OMIM:607828), along with other homologs in fish and insects (such as crustacean OTOPLb); otopetrins are predicted to have 12 transmembrane helices; by homology with OTOP1 and OTOPLb, OTPL-8 might regulate intracellular calcium fluxes during extracellular stimuli, promote extracellular calcium carbonate biomineralization, or both; OTPL-8's paralogs include OTPL-1 through OTPL-7; OTPL-8 has no function in mass RNAi assays.
pat-10	WBGene00003934	EF Hand	pat-10 encodes body wall muscle troponin C, the calcium-binding component of the troponin complex of actin thin filaments; PAT-10 is essential for muscle contraction and thus for completion of embryonic morphogenesis and elongation; by homology, PAT-10 likely functions to regulate body wall muscle contraction in response to changes in intracellular calcium; PAT-10 is expressed in body wall muscle but is also detected in vulval and anal muscles; expression in body wall muscle begins during early embryonic morphogenesis, concurrent with expression of other body wall muscle structural components
pdi-3	WBGene00003964		pdi-3 encodes a protein disulfide isomerase (PDI) required for normal cuticle collagen deposition and, subliminally, for maintenance of normal body shape; PDI-3 has both PDI and calcium-dependent transglutaminase activity in vitro, crosslinking proteins through a gamma-glutamyl epsilon-lysine dual residue; PDI-3 is homologous to the human PDI ERp60, as well as a homolog in the dog heartworm Dirofilaria immitis that is required for larval molting; the joint PDI/transglutaminase activity of PDI-3 is biochemically conserved in nematodes and mammals; pdi-3 is constitutively and broadly expressed, particularly in gut and hypodermis.
pef-1	WBGene00003969	EF Hand	pef-1 encodes an ortholog of the rhodopsin-specific serine/threonine phosphatase RdgC in D. melanogaster and of RdgC-related proteins in humans (PPEF1 and PPEF2; OMIM:300109 and OMIM:602256); PEF-1 has a EF-hand calcium-binding domain, and is peripherally bound to the plasma membrane by myristoylation and palmitoylation; pef-1 has no known mutant phenotype, but is expressed in several anterior amphid neurons, from what appears to be the 3-fold embryo stage onward; on the basis of its expression and its orthology, PEF-1 may negatively regulate signalling through serpentine receptors such as STR-2.
R04F11.1	WBGene00011014		R04F11.1 encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; R04F11.1 is worm-specific, with obvious homologs only in C. elegans; R04F11.1 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.

Table E.1.1. Worm core protein list - cont.

rgs-7	WBGene00004350	C2	rgs-7 encodes three isoforms of an atypical regulator of G protein signaling (RGS) protein that contains both a C2 and an RGS domain, and that binds EGL-30 via both domains, similar in organization to mammalian PDZ-RGS3; the RGS domain has GTPase activator protein (GAP) activity in vitro on mammalian G[alphai]3; RGS-7 thus may stably bind EGL-30 at all stages of EGL-30's GTPase cycle, might organize EGL-30 into a stable multiprotein signalling complex (with GTP[gamma]S), and might persistently inhibit EGL-30 when triggered by calcium or phospholipids; rgs-7 has no obvious function in mass RNAi assays
sca-1	WBGene00004736		The sca-1 gene encodes a sarco-endoplasmic reticulum Ca[$^{2+}$] ATPase homolog that is required for development and muscle function; sca-1 is also a ortholog of human ATP2A1 (OMIM:108730), which when mutated leads to Darier-White disease (OMIM:124200).
ser-1	WBGene00004776		ser-1 encodes a putative ortholog of mammalian 5-HT2 metabotropic serotonin receptors; SER-1 is required in both vulval muscle and neurons for the stimulation of egg-laying by serotonin (5-HT), but is completely dispensable for stimulation by the uptake inhibitor fluoxetine, and mostly dispensable for stimulation by the tricyclic antidepressant imipramine; SER-1 and SER-7 are redundantly required for normal egg-laying; SER-1 is required for normal turning during male mating, and ser-1 mutants show reduced male tail curling in exogenous 5-HT, but ser-1 males retain mating ability in the laboratory; SER-1 is weakly required for pharyngeal pumping; SER-1 is expressed in diverse neurons (head, nerve ring, vulval, ventral cord motoneurons, tail, and many others), in diverse muscles (pharyngeal, vulval, and male-specific diagonal), and in uterine cells; stimulation of heterologously expressed SER-1 induces a rise in free intracellular calcium; SER-1 has low affinity for 5-HT, and a mixture of pharmacological similarities to mammalian 5-HT1 and 5-HT2 receptors; SER-1 is stimulated by alpha-methyl-5HT, and probably antagonized by methiotheptin; SER-1 is coexpressed with MPZ-1, has a PDZ binding motif (ETFL) that aids its signalling, and binds PDZ domain 10 of MPZ-1 in vitro; SER-1's stimulation of egg-laying is impeded by mpz-1(RNAi), if and only if SER-1's ETFL motif is intact; mod-1;ser-1 double mutants subtly overbend their bodies while moving backward.
slo-1	WBGene00004830		slo-1 encodes a voltage- and calcium-activated potassium (BK) channel; SLO-1 activity is required for muscle inactivation: mutants have wild-type levels of motor activity, but have less smooth movement and tend to stop and reverse direction; SLO-1 is expressed in neurons and in body wall and vulval muscle; SLO-1 localization in proximity to the EGL-19 L-type calcium channel in muscle membrane is mediated by the dystrophin-associated protein complex (DAPC) and the ISLO-1 adapter protein with which SLO-1 physically interacts and also colocalizes with in muscle membrane.
slo-2	WBGene00004831		slo-2 encodes a high conductance, sodium-activated potassium channel; when expressed in Xenopus oocytes, SLO-2 exhibits a synergistic dependence on both chloride and calcium ions for activation; expression of a dominant negative SLO-2 mutant can interfere with activity of SLO-1, a second high conductance potassium channel, suggesting that the two channels can form functional heteromultimers; a SLO-2::GFP fusion protein is expressed in body wall, vulval, and pharyngeal/intestinal valve muscles, nerve ring processes, and some motor neurons in the ventral nerve cord; SLO-2 expression largely overlaps with that of SLO-1
spe-41	WBGene00004976		spe-41 encodes one of three C. elegans calcium-permeable cation channels that are members of the TRPC (transient receptor potential canonical) subfamily of TRP channels; spe-41 activity is required in sperm for productive sperm-oocyte interactions during fertilization, after the point of gamete contact; SPE-41 is expressed solely in sperm and upon sperm activation, translocalizes from the membranous organelles of spermatids to the plasma membranes of the cell body and pseudopod of mature sperm, consistent with the observation that the activity of store-operated calcium channels in C. elegans is much higher in mature sperm than in spermatids; when expressed in mammalian tissue culture cells, SPE-41 activity is sufficient to enhance calcium influx

Table E.1.1. Worm core protein list - cont.

stg-1	WBGene00007670		C18D1.4 encodes a claudin homolog that may regulate ion channels; C18D1.4 is similar to mammalian voltage-dependent calcium channel gamma subunits that are known or suspected to prevent epilepsy in vivo (e.g., stargazin; MGI:1316660); C18D1.4 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water
stg-2	WBGene00017400		mammalian calcium channel homolog
stim-1	WBGene00021910		stim-1 encodes an ortholog of STIM1, a putative calcium sensor of the endoplasmic reticulum thought to regulate store-operated calcium entry; stim-1 is required for LIN-3-induced sheath cell contractions, and thus for ovulation and fertility; like stim-1(RNAi) animals, animals expressing constitutively active transgenic STIM-1 are sterile, but unlike stim-1(RNAi), they also defecate abnormally (with an arrhythmic pBoc step); hyperactivated STIM-1 may also perturb osmoregulation and induce neuronal necrosis; stim-1(RNAi) has no effect on the fertility of rrf-1 mutants, indicating that STIM-1's role in fertility is purely somatic; STIM-1 is expressed in spermatheca, gonad sheath cells, uterine epithelial cells, intestine, and head neurons (probably amphid, inner labial, or both).
str-2	WBGene00006070		str-2 encodes a seven-transmembrane protein that is a candidate olfactory receptor; STR-2::GFP reporter fusions are expressed in chemosensory neurons including the AWC neuron, in which STR-2 is expressed asymmetrically, in either the AWCL or AWCR neuron, but not both; asymmetric STR-2 expression in AWC is regulated by axon contact, innexins and a calcium and MAPKKK signaling pathway; STR-2 localizes to cilia; loss of str-2 function via RNAi has been reported to result in extended life span.
T05E11.2	WBGene00011479		T05E11.2 encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; T05E11.2 is worm-specific, with a with highly divergent sequence lacking obvious homologs; T05E11.2 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.
T28B4.4	WBGene00020887		T28B4.4 encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; T28B4.4 is worm-specific, weakly similar to CLC-1; T28B4.4 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.
tax-6	WBGene00006527		tax-6 encodes an ortholog of calcineurin A that is required for inhibition and adaptation of several sensory neurons and for the normal regulation of egg-laying by serotonin; TAX-6 stimulates transcription of rcn-1, and binds RCN-1 protein if free $Ca[^{2+}]$ is present; in general, calcineurin positively regulates calcium-dependent signaling; tax-6 loss-of-function mutants exhibit an extended lifespan which is characterized by enhanced autophagy and dependent upon the presence of functional bec-1
tnc-2	WBGene00006583	EF Hand	tnc-2 encodes pharyngeal muscle troponin C, the calcium-binding component of the troponin complex of actin thin filaments; TNC-2 is essential for postembryonic development, and by homology, likely functions to regulate pharyngeal muscle contraction in response to changes in intracellular calcium; TNC-2 is expressed exclusively in pharyngeal muscle

Table E.1.1. Worm core protein list - cont.

tra-3	WBGene00006606	C2	tra-3 encodes an atypical calpain regulatory protease that lacks calcium-binding EF hand domains; during development, TRA-3 likely promotes female development in XX hermaphrodites by cleaving the membrane-associated TRA-2A protein to generate a TRA-2 peptide likely to have feminizing activity; in addition, TRA-3 is one of several proteases that mediate necrotic cell death during neurodegeneration.
unc-2	WBGene00006742		unc-2 encodes a calcium channel alpha subunit required for desensitization to dopamine, normal flexion and speed during movement, normally low sensitivity of whole animals to serotonin, and neuronal migrations promoted by humoral serotonin; UNC-2 is orthologous to human CACNA1A (OMIM:601011, mutated in familiar hemiplegic migraine or episodic ataxia 2) and to Drosophila CACOPHONY; UNC-2 is expressed primarily in motor neurons, several sensory neurons, and HSN and VC neurons controlling egg-laying.
unc-31	WBGene00006767		unc-31 encodes a pleckstrin homology (PH) domain-containing protein that is the C. elegans ortholog of human CADPS/CAPS (calcium-dependent activator protein for secretion OMIM:604667); UNC-31 functions in post-docking calcium-regulated dense-core vesicle fusion that is required for egg laying, locomotion, pharyngeal pumping, and recovery from the dauer larval stage; in addition, UNC-31 functions in the insulin receptor signaling pathway that regulates adult life span where it may control $Ca[^{2+}]$-regulated secretion of an insulin-like ligand; UNC-31 is not required for synaptic vesicle exocytosis; unc-31::gfp transcriptional reporters are expressed in most, if not all, neurons, vulval muscles, vulval cells, the spermatheca, and secretory cells such as the uterine UV1 cells; UNC-31::GFP translational fusions localize to neuronal cell bodies, axonal projections and to sites of synaptic contact, consistent with other dense-core vesicle proteins
unc-43	WBGene00006779		unc-43 encodes the C. elegans type II calcium/calmodulin-dependent protein kinase (CaMKII) ortholog; unc-43 activity is required for a number of processes, including locomotion, neuronal cell fate specification and regulation of synaptic density, egg laying, defecation, and meiotic maturation; in regulating some of these processes, unc-43 acts in concert with MAPKKK and G protein signaling pathways; UNC-43 is expressed in neurons, oocytes, and gonadal sheath cells
unc-54	WBGene00006789		unc-54 encodes a muscle myosin class II heavy chain (MHC B); UNC-54 is the major myosin heavy chain expressed in C. elegans and is required for locomotion and egg-laying; in vivo and in vitro, UNC-54 interacts with ITR-1, an IP3 receptor, and this interaction may provide a link between calcium signaling and the muscle cytoskeleton; UNC-54 is a thick filament component that is expressed in multiple muscle cell classes including body wall, intestinal, anal depressor, and sphincter muscles as well as sex-specific muscles that develop postembryonically.
unc-64	WBGene00006798		unc-64 encodes syntaxin, a plasma membrane receptor for intracellular vesicles that is orthologous to vertebrate syntaxin 1A (OMIM:186590) and Drosophila Syx1A; UNC-64 is required for normal locomotion and possibly also for insulin secretion; as an essential component of the core synaptic vesicle fusion machinery, UNC-64 interacts with UNC-13, a diacylglycerol-binding protein, and SNB-1/synaptobrevin; UNC-64 trafficking from the endoplasmic reticulum to the plasma membrane is mediated by UNC-18, an SM (Sec1, Munc18) family member; unc-64 mutations can be suppressed by mutations in slo-1, a calcium-activated potassium channel; UNC-64 is expressed ubiquitously in the nervous system and in secretory cells such as the vulval uv1 cell and the excretory gland cells.
unc-68	WBGene00006801	EF Hand	unc-68 encodes a ryanodine receptor ortholog that is expressed in body-wall muscle cells and is required for normal body tension and locomotion; unc-68 mutants are flaccid, sluggish, and resistant to ryanodine, but have normal muscle ultrastructure; UNC-68 is dispensable for excitation-contraction coupling itself, but may amplify its calcium signals; the unc-68/kra-1(kh30) mutation is an S1444N substitution at a putative protein kinase C phosphorylation site; UNC-68 reduced unc-103(sy557)-induced spicule protraction by one half; unc-68 is orthologous to the human genes RYR1 (OMIM:180901, mutated in malignant hyperthermia and central core disease), RYR2 (OMIM:180902, mutated in stress-induced polymorphic ventricular tachycardia), and RYR3 (OMIM:180903).

Table E.1.1. Worm core protein list - cont.

vav-1	WBGene00006887		vav-1 encodes a Rho/Rac-family guanine nucleotide exchange factor orthologous to the Vav proto-oncogene; VAV-1 is a complex protein with several domains, from N- to C-terminus: a calponin-like actin-binding domain; a RhoGEF/DH domain; a pleckstrin-like domain; a SH2 motif domain; and an SH3 domain; in C. elegans vav-1 is an essential gene and functions to regulate the concentration of intracellular calcium and thus, plays a key role in regulating rhythmic behaviors such as pharyngeal pumping, ovulation, and defecation; in addition, vav-1 has been shown to be a target of the mir-61 miRNA in the secondary vulval precursor cells, where it functions to negatively regulate LIN-12 activity; VAV-1::GFP reporters are expressed in the pharynx, the contractile sheath cells of the gonad, the four posterior-most intestinal cells, and the three rectal epithelial cells; VAV-1::GFP is also seen in the distal gonad, body wall muscle, and in the vulval precursor cells and their descendants
Y67A10A.9	WBGene00013458		Y67A10A.9 encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia; Y67A10A.9 is worm-specific, with obvious homologs only in C. elegans; Y67A10A.9 has no obvious function in mass RNAi assays; claudins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (TJs), induce TJs when transgenically expressed in cells normally lacking them, and can mediate the specific conductance of of specific ions (e.g., magnesium or calcium) through TJs while blocking the flow of water.
zyg-1	WBGene00006988		zyg-1 encodes a homolog of polo-like and calcium-dependent kinases that is required for daughter centriole formation; ZYG-1 localizes transiently to centrosomes and acts at least one cell cycle prior to each spindle assembly event; ZYG-1 activity is required for proper localization of the SAS-5 and SAS-6 coiled-coil proteins that are also required for centriole duplication
cmk-1	WBGene00000553		cmk-1 encodes a Ca+2/calmodulin-dependent protein kinase I (CaMK1); CMK-1 activity is required, cell autonomously and downstream of the cyclic nucleotide-gated channel TAX-4, for several aspects of AFD thermosensory neuron differentiation, including expression of the gcy-8 guanylyl cyclase and nhr-38 nuclear hormone receptor genes and morphology of the AFD sensory endings; cmk-1 activity is thus also required for normal thermosensory behavior; when expressed in COS-7 cells, CMK-1 and CKK-1 positively regulate the transcriptional activity of endogenous CREB; a cmk-1::gfp reporter is expressed in head sensory and interneurons as well as in the ventral nerve cord; expression is seen specifically in the neurons of the thermosensory circuit, AFD, AIY, and AIZ; in the AFD neurons, CMK-1 localizes exclusively to the cytoplasm, however when expressed in COS-7 cells, CMK-1 is seen in the nucleus.
csq-1	WBGene00000822		csq-1 encodes a calsequestrin that is required to normal survival in media containing either excess or deficient $Ca[^{2+}]$, and that also binds UNC-68's intraluminal domain; CSQ-1 binds $Ca[^{2+}]$ in vitro, and is thought to buffer $Ca[^{2+}]$ stores in sarcoplasmic reticulum; CSQ-1 is expressed in bodywall muscle cells from mid-embryogenesis to adulthood, as well as in vulval muscles, and in the isthmus and terminal bulb regions of the pharynx; within bodywall muscle cells, CSQ-1 is distributed in a meshlike pattern, which ultrastructurally corresponds with apical sarcoplasmic reticulum, and which is very much like that of the UNC-68 ryanodine receptor; a 17-residue C-terminal domain of CSQ-1 binds the intraluminal loops of UNC-68 in vitro, and deletion of this domain disrupts CSQ-1 subcellular localization in vivo, as do K111A mutations; an N-terminal domain (residues 32-43) is required for CSQ-1's normal subcellular distribution, with a deletion mutant showing aggregate rather than meshlike distribution; in unc-68(e540) or unc-68(r1161) mutants, CSQ-1 is instead distributed randomly within muscle cells, indicating that the subcellular localization of CSQ-1 requires UNC-68; csq-1(RNAi) animals and csq-1(jh109) mutants show no grossly obvious phenotypes, do not enhance unc-68(e540) phenotypes, and display normal muscle cell development and locomotion; however, csq-1(jh109) mutants have abnormally low viability in media containing either excess or deficient $Ca[^{2+}]$; sca-1 mRNA levels are reduced in csq-1(jh109) mutants; CSQ-1 is orthologous to human CASQ1 (OMIM:114250) and CASQ2 (OMIM:114251, mutated in catecholamine-induced polymorphic ventricular tachycardia).

Table E.1.1. Worm core protein list - cont.

itr-1	WBGene00002173		itr-1 encodes a putative inositol (1,4,5) trisphosphate receptor that affects the defecation cycle and pharyngeal pumping, and also affects ovulation in a pathway downstream of LET-23; interacts with UNC-54 in vivo, and is expressed in the adult intestine, pharynx, excretory cell, germ line, and spermatheca, with limited neuronal expression
pkd-2	WBGene00004035		pkd-2 encodes an ortholog of human PKD2 (OMIM:173910; mutated in autosomal dominant polycystic kidney disease) that is expressed in the cilia of three types of male-specific sensory neurons and that is required for two aspects of male mating behavior; PKD-2 acts with LOV-1; EGL-44 and EGL-46 regulate cell-specific expression of lov-1 and pkd-2 to specify the behavioral function of the HOB neuron
pmr-1	WBGene00004063		The pmr-1 gene encodes a Golgi P-type ATPase $Ca^{\wedge 2+}/Mn^{\wedge 2+}$-pump; mutations in its human ortholog, ATP2C1, cause Hailey-Hailey disease (OMIM:169600).
rcn-1	WBGene00004321		rcn-1 encodes a calcipressin, which binds and inhibits calcineurin; RCN-1's orthologs include yeast Rcn1p and human DSCR1 (OMIM:602917; overexpressed in Down syndrome and Alzheimer disease); rcn-1 is expressed in hypodermal seam cells, various neurons, vulval epithelial and muscle cells, marginal pharyngeal cells, and in the male tail; rcn-1 transcription is regulated by TAX-6, and RCN-1 binds TAX-6 protein if free $Ca[^{2+}]$ is present; RCN-1 overexpression, like calcineurin deficiency, causes defects in body size, cuticle thickness, fertility, growth rate, and serotonin-resistance in egg-laying
stn-1	BGene00006062		Adapter protein that binds to and probably organizes the subcellular localization of a variety of membrane proteins. Component of the dystrophin glycoprotein complex (DGC). Interacts with dyb-1, dys-1 and snf-6 to form the DGC. May link various receptors to the actin cytoskeleton and the dystrophin glycoprotein complex (DGC). May also act by slowing calcium channel activity via a direct or indirect mechanism potentially involving other second messengers. Plays an early role in the formation of the neuromuscular junction and is necessary for muscle maintenance
unc-105	WBGene00006832		unc-105 encodes a muscle degenerin, a mechanosensory membrane channel that is a member of the degenerin ion channel superfamily and similar to subunits of the mammalian amiloride-sensitive epithelial sodium channel (ENaC, OMIM:600761); UNC-105 is required for normal growth and for normal contraction and organization of muscle, where it is expressed; UNC-105 interacts with LET-2, an alpha2 chain of type IV collagen that is also expressed by muscle and localized to the basement membrane between muscle and hypodermis
deg-1	WBGene00000950		deg-1 encodes a putative ion channel for which altered function induces necrotic cell death of neurons required for mechanosensory behavior; expressed in the ASH, IL1, AVD, AVG, and PVC neurons in addition to some muscle cells
egl-8	WBGene00001177	EF Hand	egl-8 encodes a phospholipase C beta homolog that affects pharyngeal pumping, defecation, and activity levels; it is genetically downstream of egl-30 with respect to aldicarb-induced paralysis, and is expressed in most or all neurons, with the strongest staining in the head and tail ganglia, and in the intestine
ipp-5	WBGene00002146		ipp-5 encodes a type I inositol 5-phosphatase homolog; ipp-5 acts downstream of let-23 to negatively regulate IP3 signaling and is involved in spermathecal contractions during ovulation; an ipp-5::gfp transcriptional reporter is expressed in the adult distal spermatheca and weakly in the proximal sheath
let-23	WBGene00002299		let-23 encodes an EGF-receptor-family transmembrane tyrosine kinase that affects viability, inductive signaling during development of the vulva, male spicule formation, posterior development of the epidermis, ovulation and behavioral quiescience during lethargus; it is genetically upstream of the let-60/RAS pathway with respect to viability and vulval development, is upstream of phosopholipase C gamma with respect to fertility and quiesence, and is expressed in vulval precursor cells and the ALA neuron
lin-3	WBGene00002992	EGF	lin-3 encodes a member of the EGF family of peptide growth factors that affects induction of vulval development, viability, ovulation, behavioral quiescence during lethargus, and male spicule development; it acts genetically upstream of let-23, and is expressed in the anchor cell of the developing gonad

Table E.1.1. Worm core protein list - cont.

mec-10	WBGene00003174		The mec-10 gene encodes an amiloride-sensitive Na+ channel protein (degenerin) required to sense gentle mechanical stimuli (e.g. touch) along the body wall; it is paralogous to mec-4, which has similar function in vivo
mec-4	WBGene00003168		mec-4 encodes an amiloride-sensitive Na+ channel protein (degenerin) required to sense gentle mechanical stimuli (e.g. touch) along the body wall; MEC-4 is paralogous to, and coexpressed with, MEC-10, in all six touch cell neurons where they form a heteromeric mechanotransduction channel
osm-9	WBGene00003889		The osm-9 gene encodes a capsaicin receptor homolog involved in sensory responses to a subset of chemical stimuli and to ASH neuron-mediated osmotic and mechanical stimuli; OSM-9 is also involved in adaptation to volatile odorants and salts.
plc-3	WBGene00004038	C2	plc-3 encodes a phospholipase C gamma homolog; PLC-3 activity is required for regulation of a number of biological processes, including ovulatory sheath contractions, nose touch, behavioral quiescence, oscillatory Ca^{2+} signaling in the intestine, and innate immunity; plc-3 reporter fusions are expressed in sheath cells and the spermatheca, the intestine, embryonic cells including the epidermal cells during morphogenesis, and the ALA neuron
unc-36	WBGene00006772		L-type voltage-dependent Ca^{2+} channel, alpha2/delta subunit

E.1.2 Annotations Tables for Worm

Table E.1.2 Cellular Component Terms

Go id	Cellular Component Term
5624	membrane fraction
5634	nucleus
5737	cytoplasm
5783	endoplasmic reticulum
5794	Golgi apparatus
5840	ribosome
5886	plasma membrane
5887	integral to plasma membrane
5891	voltage-gated calcium channel complex
5929	cilium
5954	calcium- and calmodulin-dependent protein kinase complex
8076	voltage-gated potassium channel complex
16020	membrane
16021	integral to membrane
16023	cytoplasmic membrane-bounded vesicle
16324	apical plasma membrane
30018	Z disc
30424	axon
30425	dendrite
31260	pseudopodium membrane
31430	M band
42175	nuclear membrane-endoplasmic reticulum network
42383	sarcolemma
43025	neuronal cell body
45202	synapse

Table E.1.3. Molecular Function Terms

Go id	Molecular Function Term
3735	structural constituent of ribosome
3824	catalytic activity
4672	protein kinase activity
4674	protein serine/threonine kinase activity
4683	calmodulin-dependent protein kinase activity
4993	serotonin receptor activity
5216	ion channel activity
5220	inositol 1,4,5-trisphosphate-sensitive calcium-release channel activity
5245	voltage-gated calcium channel activity
5249	voltage-gated potassium channel activity
5262	calcium channel activity
5388	calcium-transporting ATPase activity
5509	calcium ion binding
5515	protein binding
5516	calmodulin binding
5524	ATP binding
8095	inositol-1,4,5-trisphosphate receptor activity
15085	calcium ion transmembrane transporter activity
15269	calcium-activated potassium channel activity
15279	store-operated calcium channel activity
15410	manganese-transporting ATPase activity
15662	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism
16773	phosphotransferase activity, alcohol group as acceptor
16787	hydrolase activity
16820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances
17022	myosin binding
30346	protein phosphatase 2B binding
31490	chromatin DNA binding
43176	amine binding
51082	unfolded protein binding
51378	serotonin binding

Table E.1.4. Biological Process Terms

Goid	Biological Process Term
3	reproduction
1556	oocyte maturation
2119	nematode larval development
6350	transcription
6412	translation
6457	protein folding
6468	protein amino acid phosphorylation
6754	ATP biosynthetic process
6811	ion transport
6812	cation transport
6813	potassium ion transport
6816	calcium ion transport
6828	manganese ion transport
6874	cellular calcium ion homeostasis
6897	endocytosis
6898	receptor-mediated endocytosis
6935	chemotaxis
6936	muscle contraction
6950	response to stress
6972	hyperosmotic response
6979	response to oxidative stress
7186	G-protein coupled receptor protein signaling pathway
7204	elevation of cytosolic calcium ion concentration
7210	serotonin receptor signaling pathway
7320	insemination
7338	single fertilization
7622	rhythmic behavior
7635	chemosensory behavior
8152	metabolic process
8340	determination of adult lifespan
8355	olfactory learning
9103	lipopolysaccharide biosynthetic process
9790	embryonic development
9792	embryonic development ending in birth or egg hatching
10042	response to manganese ion
10468	regulation of gene expression
12501	programmed cell death
18991	oviposition
18996	molting cycle, collagen and cuticulin-based cuticle
19722	calcium-mediated signaling
19915	lipid storage
30104	water homeostasis

Table E.1.4. Biological Process Terms - cont.

30334	regulation of cell migration
30421	defecation
30728	ovulation
30968	endoplasmic reticulum unfolded protein response
32513	negative regulation of protein phosphatase type 2B activity
32793	positive regulation of CREB transcription factor activity
34607	turning behavior involved in mating
40007	growth
40010	positive regulation of growth rate
40011	locomotion
40017	positive regulation of locomotion
40018	positive regulation of multicellular organism growth
40024	dauer larval development
40035	hermaphrodite genitalia development
40040	thermosensory behavior
42048	olfactory behavior
42493	response to drug
42713	sperm ejaculation
43050	pharyngeal pumping
43051	regulation of pharyngeal pumping
43052	thermotaxis
45214	sarcomere organization
45664	regulation of neuron differentiation
45807	positive regulation of endocytosis
45860	positive regulation of protein kinase activity
45944	positive regulation of transcription from RNA polymerase II promoter
46777	protein amino acid autophosphorylation
46928	regulation of neurotransmitter secretion
48149	behavioral response to ethanol
48598	embryonic morphogenesis
48609	reproductive process in a multicellular organism
48812	neuron projection morphogenesis
50804	regulation of synaptic transmission
51481	reduction of cytosolic calcium ion concentration
51489	regulation of filopodium assembly
51592	response to calcium ion
55074	calcium ion homeostasis
55085	transmembrane transport
60179	male mating behavior

E.2 Homology Search of 17 Worm core Proteins

Table E.2.1. Homologs of 17 Worm Core Proteins through Species

C.ELEGANS	D.MELANOGASTER (Fruit Fly)				D.RERIO (Zebra Fish)				X.LAEVIS (Frog)				M.MUSCULUS (Mouse)				H.SAPIENS (Human)			
Worm	% identity	e-val	prot name	ID	% identity	e-val	prot name	ID	% identity	e-val	prot name	ID	% identity	e-val	prot name	ID	% identity	e-val	prot name	ID
ckk-1	44,91	1,01E-107	CG17698, isoform C calmodulin-dependent protein kinase activity	NP_001036633	46,76	4,91E-107	calcium/calmodulin-dependent protein kinase kinase 1	NP_001014361	45,14	2,43E-112	calcium/calmodulin-dependent protein kinase kinase 1, alpha	NP_001082316	46,06	2,10E-105	calcium/calmodulin-dependent protein kinase kinase 2	NP_663333	48,15	1,76E-105	calcium/calmodulin-dependent protein kinase kinase 2 isoform 2	NP_705719
cmk-1	68,10	2,01E-137	Calcium/calmodulin-dependent protein kinase I, isoform G	NP_524622	59,20	1,12E-122	calcium/calmodulin-dependent protein kinase type 1D	NP_001116532	62,93	2,53E-124	calcium/calmodulin-dependent protein kinase I	NP_001080279	61,78	1,39E-123	calcium/calmodulin-dependent protein kinase type 1	NP_598687	61,49	5,50E-123	calcium/calmodulin-dependent protein kinase type 1	NP_003647
cnb-1	80,12	7,12E-76	calcineurin B2	NP_524874	41,52	9,27E-28	calcium-binding protein p22	NP_956009	41,52	6,53E-29	calcium binding protein P22	NP_001085375	73,10	1,73E-67	calcineurin subunit B type 2	NP_001004025	78,36	1,06E-74	calcineurin subunit B type 1	NP_000936
crt-1	58,23	1,18E-130	calreticulin	NP_524293	60,25	1,71E-139	calreticulin	NP_956007	59,24	2,55E-137	calreticulin	NP_001080765	57,72	5,28E-135	calreticulin precursor	NP_031617	57,72	2,78E-135	calreticulin precursor	NP_004334
csq-1	14,63	9,38E-03	protein disulfide isomerase, isoform A	NP_524079	27,58	1,11E-41	calsequestrin-1	NP_001070192	27,58	3,49E-45	calsequestrin 1	NP_001080227	29,98	1,72E-44	calsequestrin-1	NP_033943	29,02	4,72E-44	calsequestrin-1	NP_001222
egl-19	57,88	0,00E+00	$Ca[^{2+}]$-channel protein alpha[[1]] subunit D, isoform B	NP_525104	55,13	0,00E+00	voltage-dependent L-type calcium channel subunit alpha-1D	NP_982351					57,38	0,00E+00	voltage-dependent L-type calcium channel subunit alpha-1D isoform 1	NP_083257	57,49	0,00E+00	voltage-dependent L-type calcium channel subunit alpha-1D isoform c	NP_001122311

Table E.2.1. Homologs of 17 Worm Core Proteins through Species - cont.

itr-1	23,31	0,00E+00	inositol 1,4,5,-tris-phosphate receptor, isoform A	NP_730942	29,60	0,00E+00	inositol 1,4,5-triphosphate receptor, type 1a	XP_696007	16,87	0,00E+00	inositol 1,4,5-triphosphate receptor, type 1	NP_001084015	37,59	0,00E+00	inositol 1,4,5-trisphosphate receptor type 3	NP_542120	37,76	0,00E+00	inositol 1,4,5-trisphosphate receptor type 3	NP_002215
mca-1	50,00	0,00E+00	plasma membrane calcium ATPase, isoform J	NP_001033803	49,92	0,00E+00	plasma membrane calcium-transporting ATPase 2	NP_001116710	49,52	0,00E+00	plasma membrane calcium ATPase 3	NP_001087020	50,24	0,00E+00	plasma membrane calcium ATPase 3	NP_796210	50,48	0,00E+00	plasma membrane calcium-transporting ATPase 3 isoform 3b	NP_001001344
mca-3	63,45	0,00E+00	plasma membrane calcium ATPase, isoform I	NP_726564	59,31	0,00E+00	plasma membrane calcium-transporting ATPase 3	NP_001121714	59,48	0,00E+00	plasma membrane calcium ATPase 3	NP_001087020	61,21	0,00E+00	plasma membrane calcium ATPase 3	NP_796210	61,47	0,00E+00	plasma membrane calcium-transporting ATPase 3 isoform 3b	NP_001001344
pmr-1	56,27	0,00E+00	secretory pathway calcium atpase, isoform C	NP_730744	36,63	1,65E-125	sarcoplasmic/endoplasmic reticulum calcium ATPase 2	NP_957259	55,72	0,00E+00	ATPase, Ca++ transporting, type 2C, member 2	NP_001089971	58,05	0,00E+00	calcium-transporting ATPase type 2C member 1	NP_778190	58,27	0,00E+00	calcium-transporting ATPase type 2C member 1 isoform 1a	NP_055197
rcn-1	31,88	9,56E-24	calcipressin, sarah	NP_524378	33,33	8,92E-31	calcipressin-3	NP_001005392	28,50	1,45E-28	regulator of calcineurin 1	NP_001079860	36,23	3,25E-32	calcipressin-3	NP_075356	34,30	2,92E-32	calcipressin-3	NP_038469
sca-1	69,69	0,00E+00	calcium ATPase at 60A, isoform A	NP_476832	68,27	0,00E+00	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2b	NP_001025448	68,08	0,00E+00	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2	NP_001080404	68,46	0,00E+00	sarcoplasmic/endoplasmic reticulum calcium ATPase 1	NP_031530	66,95	0,00E+00	sarcoplasmic/endoplasmic reticulum calcium ATPase 2 isoform 1	NP_733765

Table E.2.1. Homologs of 17 Worm Core Proteins through Species - cont.

ser-1	18,01	4,70E-37	serotonin receptor 7	NP_524599	19,91	1,15E-49	5-hydroxytryptamine receptor 2B	NP_001038208	19,62	2,19E-48	5-hydroxytryptamine (serotonin) receptor 2B	NP_001082744	20,35	8,54E-46	D(2) dopamine receptor	NP_034207	19,62	2,07E-44	D(2) dopamine receptor isoform long	NP_000786
slo-1	65,00	0,00E+00	slowpoke, isoform C	NP_996289	31,84	0,00E+00	calcium-activated potassium channel subunit alpha-1	NP_001139072	54,74	0,00E+00	calcium-activated potassium channel subunit alpha-1	NP_001079159	56,05	0,00E+00	calcium-activated potassium channel subunit alpha-1	NP_034740	56,14	0,00E+00	calcium-activated potassium channel subunit alpha-1 isoform c	NP_001154824
spe-41	18,50	7,80E-37	trp-like, isoform A	NP_476895	18,27	7,55E-38	short transient receptor potential channel 6	NP_001025453	14,87	2,78E-23	transient receptor potential cation channel, subfamily C, member 1	NP_001083819	17,56	3,94E-33	short transient receptor potential channel 7	NP_036165	17,80	9,08E-33	short transient receptor potential channel 3 isoform a	NP_001124170
tax-6	71,59	0,00E+00	calcineurin A at 14F, isoform A	NP_727985	67,10	0,00E+00	serine/threonine-protein phosphatase 2B catalytic subunit alpha	NP_001017856	67,10	0,00E+00	protein phosphatase 3, catalytic subunit, alpha isoform	NP_001080813	68,04	0,00E+00	serine/threonine-protein phosphatase 2B catalytic subunit beta isoform	NP_032940	68,60	0,00E+00	serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform isoform 2	NP_001124163
unc-36	20,82	4,18E-73	straightjacket	NP_610902	21,62	9,99E-58	voltage-dependent calcium channel subunit alpha-2/delta-1	NP_001038425	9,77	3,99E-34	calcium channel, voltage-dependent, alpha 2/delta 3 subunit	NP_001079180	22,74	4,31E-54	voltage-dependent calcium channel subunit alpha-2/delta-1 isoform b	NP_001104314	21,86	2,42E-57	voltage-dependent calcium channel subunit alpha-2/delta-1	NP_000713

E.3 Modules for Worm

E.3.1. Module List Obtained from 17 worm core PIN

Cluster	Score (Density*#Nodes)	Nodes	Edges	Node IDs
1	2,043	23	47	WBGene00003090, WBGene00001589, WBGene00000900, WBGene00019893, WBGene00004174, WBGene00011935, WBGene00018808, WBGene00018742, WBGene00003241, WBGene00010222, WBGene00007486, WBGene00001377, WBGene00002273, WBGene00015823, WBGene00006870, WBGene00021495, WBGene00011449, WBGene00012664, WBGene00004099, WBGene00000031, WBGene00000034, WBGene00010830, WBGene00004139
2	1,8	5	9	WBGene00009179, WBGene00003576, WBGene00021648, WBGene00020064, WBGene00022357
3	1,333	6	8	WBGene00001561, WBGene00006993, WBGene00015899, WBGene00003936, WBGene00019167, WBGene00011444
4	1,25	4	5	WBGene00020142, WBGene00018638, WBGene00022042, WBGene00018637
5	1	3	3	*WBGene00004153, WBGene00000024, WBGene00013481*
6	1	3	3	*WBGene00015791, WBGene00009678, WBGene00019102*
7	1	4	4	WBGene00007103, WBGene00010673, WBGene00000033, WBGene00017937

E.3.2. Module Figures for Worm

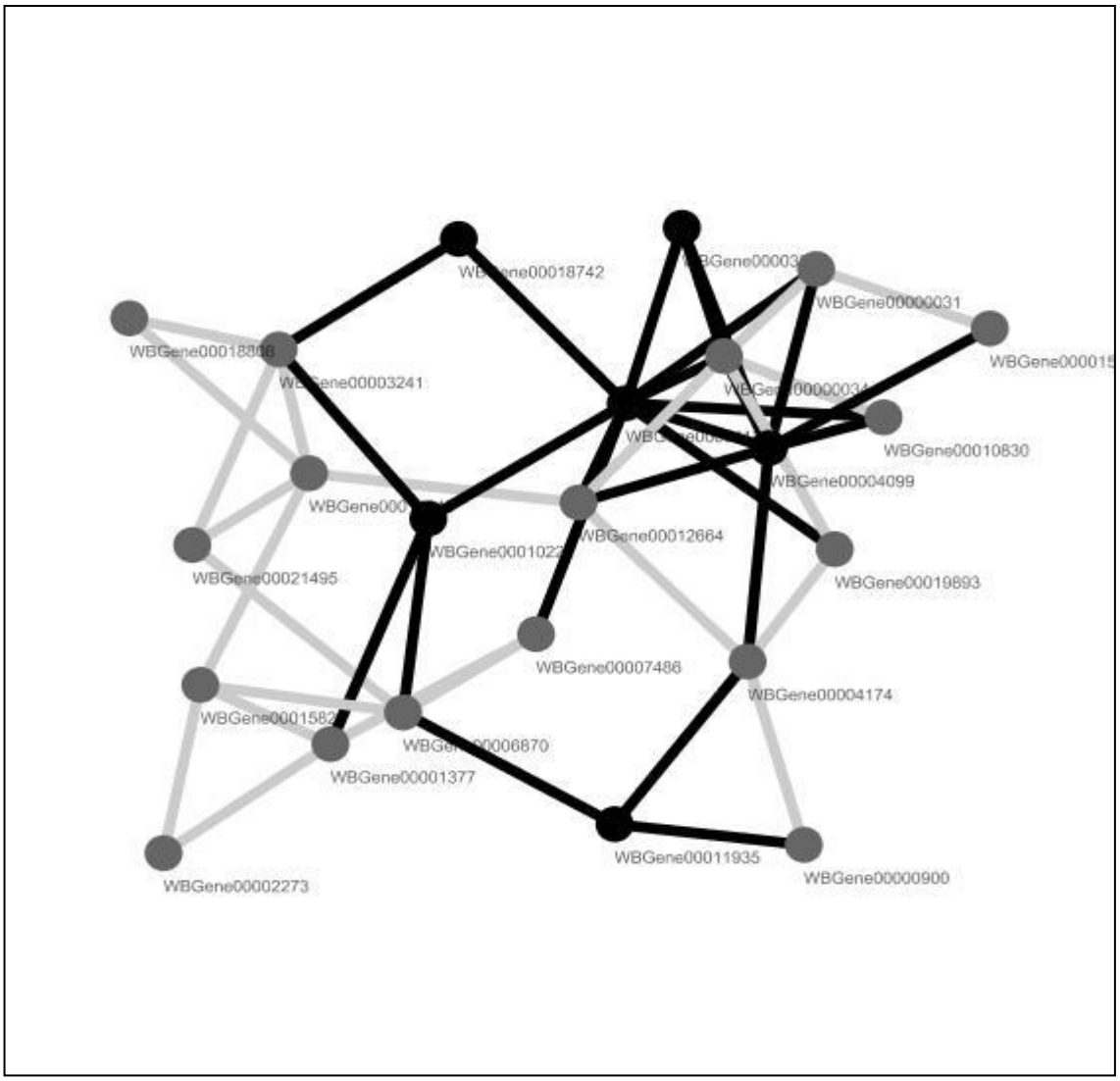


Figure E.3.2.1. Module 1 for worm

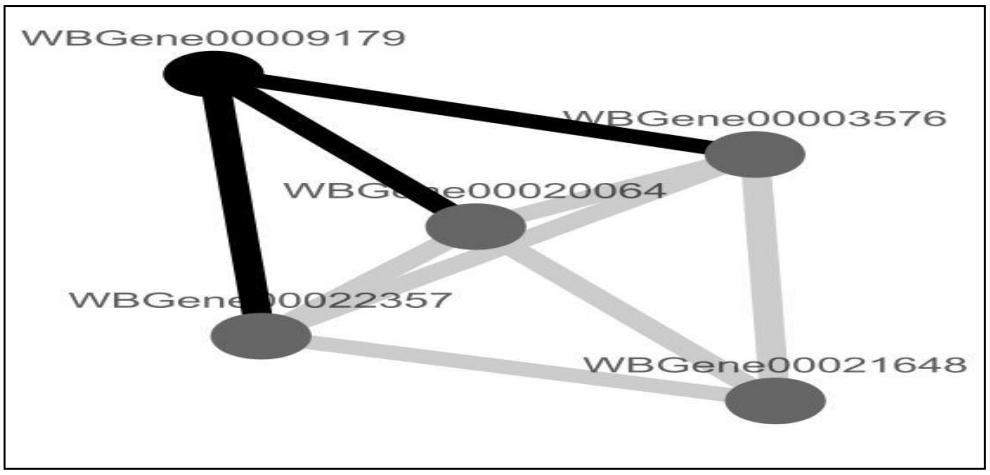


Figure E.3.2.2. Module 2 for worm

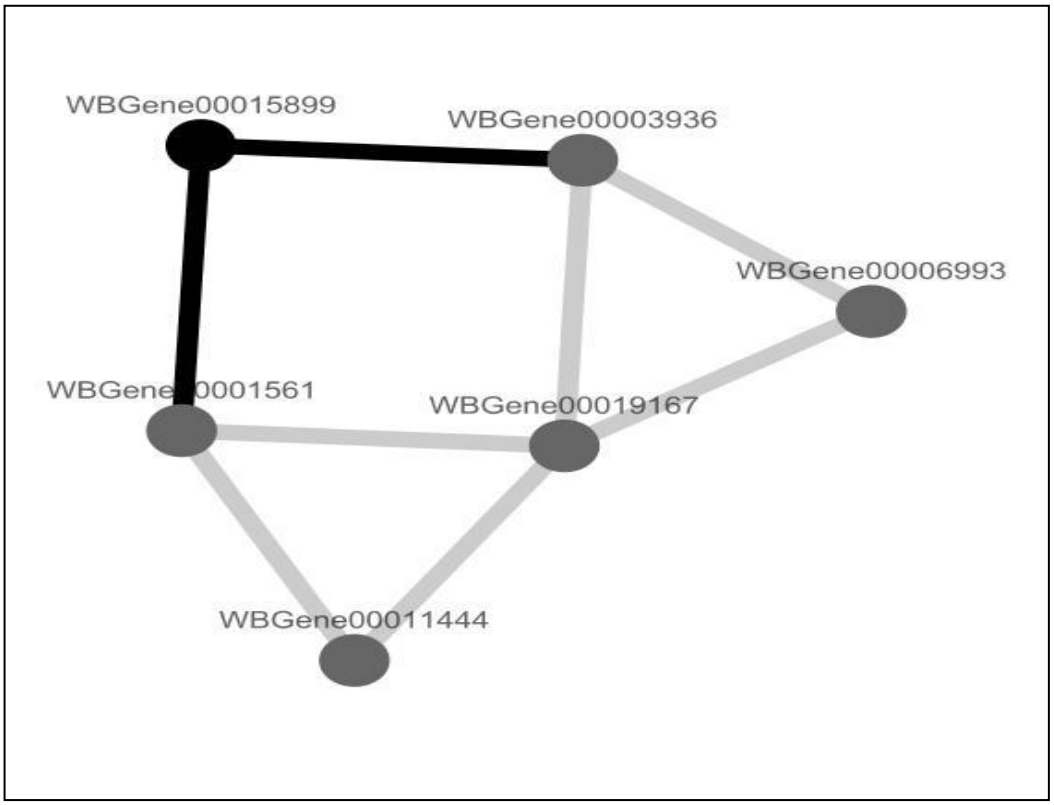


Figure E.3.2.3. Module 3 for worm

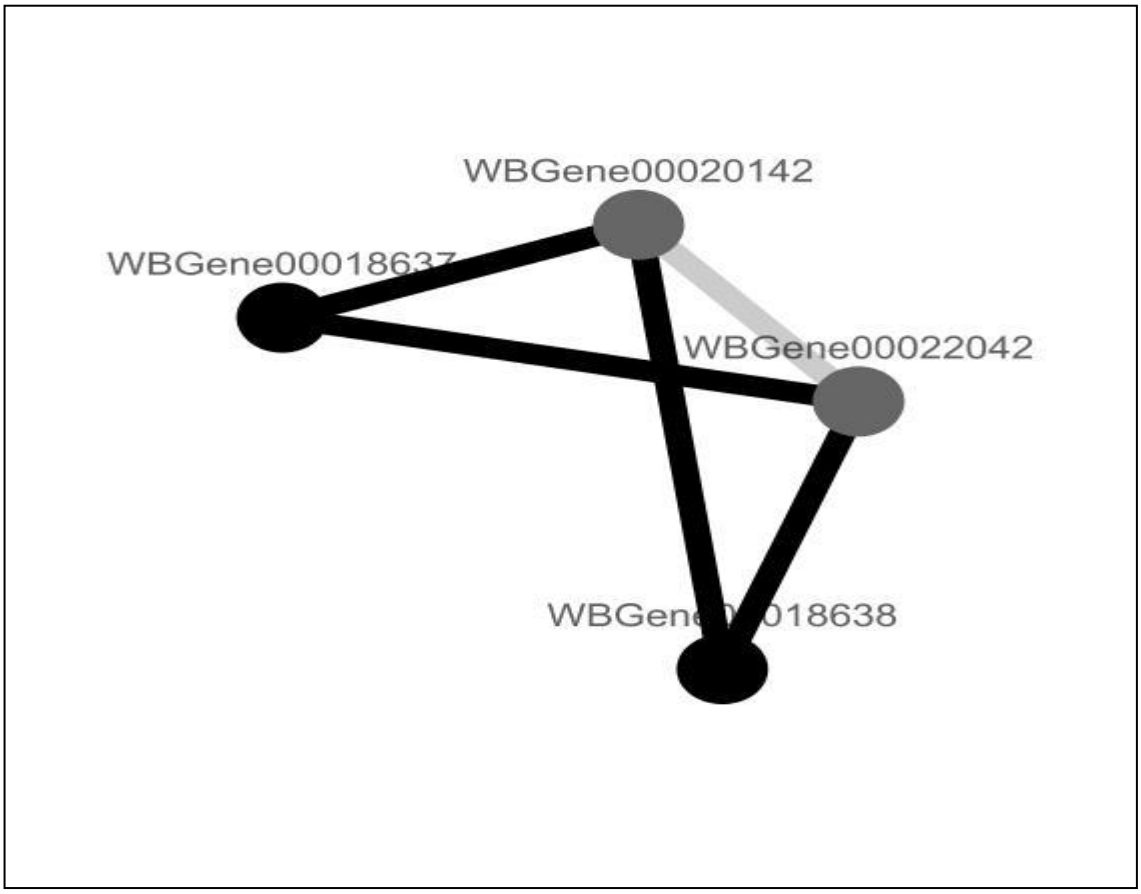


Figure E.3.2.4. Module 4 for worm

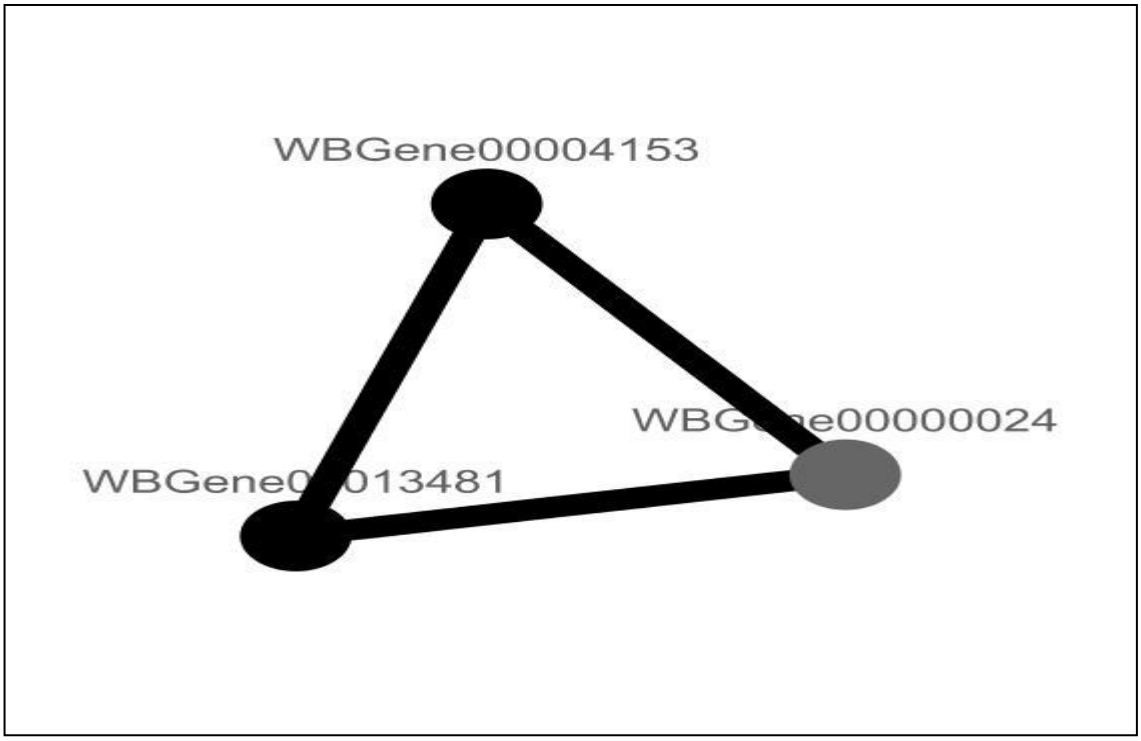


Figure E.3.2.5. Module 5 for worm

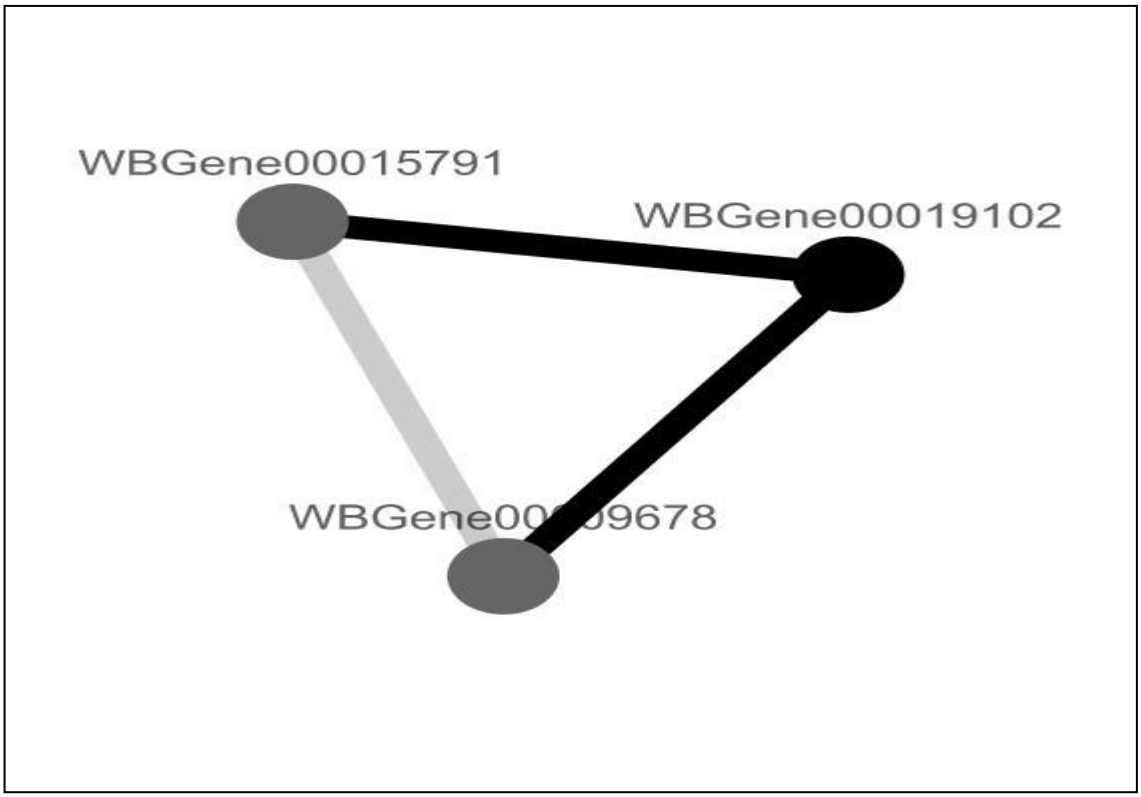


Figure E.3.2.6. Module 6 for worm

APPENDIX F

F.1 Core and KEGG Proteins for Human

Table F.1.1. Core Proteins for human

Grey shaded ones are the core proteins

	HPRD Id	Uniprot Id	Name	A.a	Tissue expressed	Description
PMCA	159	P23634	PMCA4	1241	Brain, Cerebellum, Cerebral cortex, Hippocampus	ATPase, Ca++ transporting, plasma membrane 4
	160	Q01814	PMCA2	1243	Brain	ATPase, Ca++ transporting, plasma membrane 2
NMDAR	339	P36544	CHRNA7	502	Brain, Hippocampus, Substantia nigra, Thalamus	Neuronal acetylcholine receptor subunit alpha-7
	698	Q12879	NMDAR2A	1464	Brain, Hippocampus	Glutamate [NMDA] receptor subunit epsilon-1
	699	Q14957	NMDAR2C	1236	Amygdala, Caudate nucleus, Corpus Callosum, Hippocampus, Hypothalamus, Substantia nigra, Subthalamic nucleus, Thalamus	Glutamate [NMDA] receptor subunit epsilon-3
	3977	Q99572	P2X7	595	Brain	P2X purinoceptor 7
	4095	O15399	NMDAR2D	1336	Brain, Cerebellar cortex, Cerebellum, Cerebral cortex, Purkinje Cells	Glutamate [NMDA] receptor subunit epsilon-4
	15926	Q05586	NMDAR1	938	Brain, Caudate nucleus, Cerebellum, Dentate gyrus, Frontal cortex, Hippocampus, Locus coeruleus, Nucleus accumbens, Pontine nucleus, Putamen, Temporal cortex, Temporal lobe	Glutamate [NMDA] receptor subunit zeta-1
	15984	Q9UBL9	P2X2	471	Brain, Amygdala	P2X purinoceptor 2

Table F.1.1. Core Proteins for human – cont.

VOC	247	Q01668	Cav1.3	2161	Brain	Voltage-dependent L-type calcium channel subunit alpha-1D
	3004	O00555	Cav2.1	2505	Brain, Cerebellum	Voltage-dependent P/Q-type calcium channel subunit alpha-1A
	3005	Q00975	Cav2.2	2339	Dentate nucleus, Hippocampus, Parahippocampal gyrus, Purkinje cell, Pyramidal cell	Voltage-dependent N-type calcium channel subunit alpha-1B
	3006	Q15878	Cav2.3	2313	Brain, Cerebellum, Dentate gyrus, Entorhinal cortex, Hippocampus, Parahippocampal gyrus	Voltage-dependent R-type calcium channel subunit alpha-1E
	7442	O95180	Cav3.2	2353	Brain, Amygdala, Caudate nucleus, Cerebellum, Cerebral cortex, Corpus callosum, Frontal lobe, Hippocampus, Medulla oblongata, Occipital lobe, Putamen, Substantia nigra, Subthalamic Nucleus, Temporal lobe, Thalamus	Voltage-dependent T-type calcium channel subunit alpha-1H
SOC	11894	P48995	TrpC1	793	Brain	Short transient receptor potential channel 1
RYR	1618	P21817	RYR-1	5038	Cerebellum, Hippocampus	Ryanodine receptor 1
	1619	Q92736	RYR-2	4965	Brain	Ryanodine receptor 2
	1620	Q15413	RYR-3	4870	Brain, Caudate nucleus, Cerebellum, Hippocampus, Substantia nigra, Subthalamus, Thalamus	Ryanodine receptor 3
IP3R	925	Q14643	IP3R 1	2758	Caudate nucleus, Cerebellum, Hippocampus, Purkinje cell	Inositol 1,4,5-trisphosphate receptor type 1
SERCA	161	P16615	SERCA2	1042	Brain	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2
	3568	Q93084	SERCA3	1043	Brain, Amygdala, Caudate nucleus, Cerebellum, Hippocampus, Medulla oblongata, Occipital lobe, Putamen, Substantia nigra, Subthalamic Nucleus, Temporal lobe, Thalamus	Sarcoplasmic/endoplasmic reticulum calcium ATPase 3

Table F.1.1. Core Proteins for human - cont.

NCX	1659	P32418	NCX1	973	Hippocampus, Temporal lobe	Sodium/calcium exchanger 1
	9723	P57103	NCX3		Brain	Sodium/calcium exchanger 3
	15990	Q9UPR5	NCX2	921	Brain, Amygdala, Caudate nucleus, Cerebellum, Corpus callosum, Hippocampus, Substantia nigra, Subthalamic Nucleus, Thalamus	Sodium/calcium exchanger 2
ANT	58	P12235	ANT 1	298	Brain	ADP/ATP translocase 1, mitochondrial carrier
	2147	P05141	ANT 2	298	Brain	ADP/ATP translocase 2, mitochondrial carrier
PPID	11875	Q08752	PPID	370	Brain	Peptidyl-prolyl cis-trans isomerase D
VDAC	1904	P45880	VDAC2	294	Cerebral cortex, Hippocampus	Voltage-dependent anion-selective channel protein 2
	5137	P21796	VDAC1	283	Brain, Cerebellum, Cerebral cortex	Voltage-dependent anion-selective channel protein 1
	6750	Q9Y277	VDAC3	283	Brain	Voltage-dependent anion-selective channel protein 3
CAM	242	P62158	CALM2	149	Brain	Calmodulin
	243	P62158	CALM3	149	Brain	Calmodulin 3
CD38	116	P28907	CD38	300	Brain	ADP-ribosyl cyclase 1
PTK	579	P00533	EGFR	1210	Brain, Cerebral cortex, Hippocampus	Epidermal growth factor receptor
	1281	P04626	ERBB2	1255	Cerebellum, Cerebral cortex, Hippocampus	Receptor tyrosine-protein kinase erbB-2
	1423	P09619	PDGFRB	1106	Brain	Beta-type platelet-derived growth factor receptor
	1429	P16234	PDGFRA	1089	Brain	Alpha-type platelet-derived growth factor receptor
	1820	P21860	ERBB3	1342	Brain, Cerebellum, Cerebral cortex, Hippocampus	Receptor tyrosine-protein kinase erbB-3
	2767	Q15303	ERBB4	1308	Brain, Cerebellum, Hippocampus	Receptor tyrosine-protein kinase erbB-4
GPCR	79	P25100	ADRA1D	572	Hippocampus	Alpha-1D adrenergic receptor
	81	P35348	ADRA1A	466	Brain, Cerebrum, Cerebellum, Hippocampus	Alpha-1A adrenergic receptor
	107	P30556	AGTR1	359	Nervous System	Type-1 angiotensin II receptor
	322	P32238	CCKAR	428	Brain, Cerebrum, Cerebellum,	Cholecystokinin receptor type A

Table F.1.1. Core Proteins for human - cont.

					Hippocampus	
	323	P32239	CCKBR	447	Brain	Cholecystokinin type B receptor
	328	P08172	CHRM2	466	Brain	Muscarinic acetylcholine receptor M2
	329	P20309	CHRM3	590	Brain	Muscarinic acetylcholine receptor M3
	338	P11229	CHRM1	460	Brain	Muscarinic acetylcholine receptor M1
	1208	P25103	TACR1	407	Amygdala, Caudate nucleus, Cerebellum, CNS, Dentate gyrus, Frontal cortex, Globus pallidus, Hippocampus, Hypothalamus, Locus coeruleus, Medulla oblongata, Middle frontal gyrus, Nucleus accumbens, Parahippocampal gyrus, Putamen, Striatum, Substantia nigra, Superior frontal gyrus, Temporal cortex, Thalamus	Substance-P receptor
	1328	P30559	OXTR	389	Nervous System	Oxytocin receptor
	1638	P28223	HTR2A	471	Cerebral cortex	5-hydroxytryptamine receptor 2A
	1763	P25116	F2R	425	Brain	Proteinase-activated receptor 1
	2429	P28335	HTR2C	458	Brain	5-hydroxytryptamine receptor 2C
	2600	P47901	AVPR1B	424	Nervous System	Vasopressin V1b receptor
	2639	P46663	BDKRB1	353	Brain, Cerebrum, Cerebellum, Hippocampus, Hypothalamus	B1 bradykinin receptor
	2894	P37288	AVPR1A	418	Brain	Vasopressin V1a receptor
	3077	P41595	HTR2B	481	Brain	5-hydroxytryptamine receptor 2B
	4980	P41594	GRM5	1212	Brain, Caudate nucleus, Cerebellum, Cerebral cortex, Dentate gyrus, Frontal cortex, Hippocampus, Locus coeruleus, Occipital cortex, Parietal cortex, Pontine nucleus, Purkinje cell,	Metabotropic glutamate receptor 5

Table F.1.1. Core Proteins for human - cont.

					Putamen, Pyramidal cell, Subthalamic nucleus, Thalamic nucleus	
	5129	Q13255	GRM1	1194	Brain, Amygdala, Caudate nucleus, Cerebellum, Cerebral cortex, Hippocampus, Putamen, Substantia nigra, Subthalamic Nucleus, Thalamus	Metabotropic glutamate receptor 1
	8834	P24530	EDNRB	442	Brain	Endothelin B receptor
	11767	P34981	TRHR	398	Brain	Thyrotropin-releasing hormone receptor
GS	758	P38405	GNAL	381	Brain	Guanine nucleotide-binding protein G(olf) subunit alpha
	761	P63092	GNAS	394	Brain, Occipital lobe, Temporal lobe	Guanine nucleotide-binding protein G(s) subunit alpha isoforms short
ADCY	51	P40145	ADCY8	1251	Brain	Adenylate cyclase type 8
PKA	1482	P22694	PRKACB	351	Brain	cAMP-dependent protein kinase catalytic subunit beta
	1483	P22612	PRKACG	351	Brain, Cerebrum, Cerebellum, Hippocampus	cAMP-dependent protein kinase catalytic subunit gamma
GQ	759	P29992	GNA11	359	Brain	Guanine nucleotide-binding protein subunit alpha-11
	2998	P50148	GNAQ	359	Brain, Frontal cortex	Guanine nucleotide-binding protein G(q) subunit alpha
	5095	O95837	GNA14	355	Brain	Guanine nucleotide-binding protein subunit alpha-14
PLCB	6177	Q9NQ66	PLCB1	1216	Brain, Amygdala, Caudate nucleus, Cerebellum, Corpus callosum, Hippocampus, Thalamus	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta-1
PLCG	1398	P19174	PLCG1	1290	Brain, Cerebellum, Cerebral cortex, Hippocampus	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-1
SPHK	4768	Q9NYA1	SPHK1	384	Brain	Sphingosine kinase 1
CaN	234	Q08209	PPP3CA	521	Brain	Calmodulin-dependent calcineurin A subunit alpha isoform
	235	Q8N1F0	PPP3CB	525	Brain	Serine/threonine-protein phosphatase
CAMK	7169	Q16566	CAMK4	473	Brain, Cerebellum, Cerebral cortex,	Calcium/calmodulin-dependent protein kinase type IV

Table F.1.1. Core Proteins for human - cont.

					Hippocampus	
	12127	Q13554	CAMK2B	664	Brain, Cerebral cortex	Calcium/calmodulin-dependent protein kinase type II subunit beta
NOS	1224	P29474	NOS3	1203	Brain	Nitric oxide synthase, endothelial
	1225	P35228	NOS2	1153	Brain, Cerebellum, Cerebral cortex	Nitric oxide synthase, inducible
	1226	P29475	NOS1	1434	Brain, Cerebellum, Cerebral cortex, Hippocampus	Nitric oxide synthase, brain
FAK2	3131	Q14289	PTK2B	1009	Brain, Amygdala, Caudate nucleus, Cerebral cortex, Corpus callosum, Hippocampus, Hypothalamus, Substantia nigra, Subthalamic nucleus, Thalamus	Calcium-dependent tyrosine kinase
IP33K	941	P23677	ITPKA	461	Brain, Hippocampus	Inositol-trisphosphate 3-kinase A
PKC	1498	P17252	PRKCA	672	Brain, Cerebellum, Cerebral cortex, Hippocampus	Protein kinase C alpha type
	1499	P05771	PRKCB	671	Brain, Cerebellum, Cerebral cortex, Hippocampus	Protein kinase C beta type
	1502	P05129	PRKCG	697	Brain	Protein kinase C gamma type

F.2 Annotation Tables for Human

Table F.2.1 Cellular Component Terms

Go id	Go Name
1950	plasma membrane enriched fraction
5624	membrane fraction
5634	nucleus
5635	nuclear envelope
5637	nuclear inner membrane
5639	integral to nuclear inner membrane
5730	nucleolus
5737	cytoplasm
5739	mitochondrion
5741	mitochondrial outer membrane
5743	mitochondrial inner membrane
5783	endoplasmic reticulum
5789	endoplasmic reticulum membrane
5790	smooth endoplasmic reticulum
5792	microsome
5886	plasma membrane
5887	integral to plasma membrane
5891	voltage-gated calcium channel complex
5892	nicotinic acetylcholine-gated receptor-channel complex
5901	caveola
5911	cell-cell junction
5938	cell cortex
5955	calcineurin complex
8021	synaptic vesicle
9897	external side of plasma membrane
14069	postsynaptic density
14701	junctional sarcoplasmic reticulum membrane
14802	terminal cisterna
16020	membrane
16021	integral to membrane
16323	basolateral plasma membrane
16324	apical plasma membrane
16529	sarcoplasmic reticulum
17146	N-methyl-D-aspartate selective glutamate receptor complex
19717	synaptosome
30017	sarcomere
30018	Z disc

Table F.2.1 Cellular Component Terms - cont.

30054	cell junction
30288	outer membrane-bounded periplasmic space
30314	junctional membrane complex
30315	T-tubule
30425	dendrite
31088	platelet dense granule membrane
31094	platelet dense tubular network
31095	platelet dense tubular network membrane
31594	neuromuscular junction
31674	I band
31965	nuclear membrane
32059	bleb
33017	sarcoplasmic reticulum membrane
34704	calcium channel complex
42645	mitochondrial nucleoid
42734	presynaptic membrane
42995	cell projection
43005	neuron projection
43025	neuronal cell body
43034	costamere
43195	terminal button
43197	dendritic spine
43204	perikaryon
43234	protein complex
45121	membrane raft
45202	synapse
45211	postsynaptic membrane
46930	pore complex

Table F.2.2. Molecular Function Terms

Go id	Go Name
166	nucleotide binding
287	magnesium ion binding
1530	lipopolysaccharide binding
1540	beta-amyloid binding
1614	purinergic nucleotide receptor activity
3755	peptidyl-prolyl cis-trans isomerase activity
4872	receptor activity
4889	nicotinic acetylcholine-activated cation-selective channel activity
4931	extracellular ATP-gated cation channel activity
4972	N-methyl-D-aspartate selective glutamate receptor activity
5102	receptor binding
5215	transporter activity
5216	ion channel activity
5219	ryanodine-sensitive calcium-release channel activity
5220	inositol 1,4,5-trisphosphate-sensitive calcium-release channel activity
5230	extracellular ligand-gated ion channel activity
5234	extracellular-glutamate-gated ion channel activity
5244	voltage-gated ion channel activity
5245	voltage-gated calcium channel activity
5261	cation channel activity
5262	calcium channel activity
5388	calcium-transporting ATPase activity
5432	calcium
5471	ATP
5507	copper ion binding
5509	calcium ion binding
5515	protein binding
5516	calmodulin binding
5524	ATP binding
8022	protein C-terminus binding
8095	inositol-1,4,5-trisphosphate receptor activity
8270	zinc ion binding
8308	voltage-gated anion channel activity
8331	high voltage-gated calcium channel activity
8332	low voltage-gated calcium channel activity
15081	sodium ion transmembrane transporter activity
15085	calcium ion transmembrane transporter activity
15207	adenine transmembrane transporter activity
15267	channel activity
15270	dihydropyridine-sensitive calcium channel activity
15276	ligand-gated ion channel activity

Table F.2.2. Molecular Function Terms - cont.

15278	calcium-release channel activity
15279	store-operated calcium channel activity
15288	porin activity
15297	antiporter activity
15464	acetylcholine receptor activity
15643	toxin binding
16018	cyclosporin A binding
16594	glycine binding
16595	glutamate binding
16787	hydrolase activity
16820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances
16853	isomerase activity
17081	chloride channel regulator activity
19905	syntaxin binding
22843	voltage-gated cation channel activity
30165	PDZ domain binding
31072	heat shock protein binding
31775	lutropin-choriogonadotropic hormone receptor binding
34236	protein kinase A catalytic subunit binding
34237	protein kinase A regulatory subunit binding
35091	phosphoinositide binding
42165	neurotransmitter binding
42166	acetylcholine binding
42802	identical protein binding
42803	protein homodimerization activity
43539	protein serine/threonine kinase activator activity
43924	suramin binding
46872	metal ion binding
46982	protein heterodimerization activity
47485	protein N-terminus binding
48155	S100 alpha binding
50839	cell adhesion molecule binding

Table F.2.3. Biological Process Terms

Go id	Go Name
2	mitochondrial genome maintenance
96	sulfur amino acid metabolic process
187	activation of MAPK activity
902	cell morphogenesis
1508	regulation of action potential
1662	behavioral fear response
1666	response to hypoxia
1845	phagolysosome assembly
1916	positive regulation of T cell mediated cytotoxicity
1934	positive regulation of protein amino acid phosphorylation
1964	startle response
1975	response to amphetamine
2026	regulation of the force of heart contraction
2028	regulation of sodium ion transport
5513	detection of calcium ion
6006	glucose metabolic process
6091	generation of precursor metabolites and energy
6457	protein folding
6509	membrane protein ectodomain proteolysis
6649	phospholipid transfer to membrane
6754	ATP biosynthetic process
6800	oxygen and reactive oxygen species metabolic process
6810	transport
6811	ion transport
6812	cation transport
6814	sodium ion transport
6816	calcium ion transport
6820	anion transport
6874	cellular calcium ion homeostasis
6884	cell volume homeostasis
6900	membrane budding
6915	apoptosis
6936	muscle contraction
6941	striated muscle contraction
6954	inflammatory response
6984	ER-nucleus signaling pathway
7005	mitochondrion organization
7009	plasma membrane organization
7154	cell communication
7155	cell adhesion
7165	signal transduction

Table F.2.3. Biological Process Terms - cont.

7166	cell surface receptor linked signaling pathway
7204	elevation of cytosolic calcium ion concentration
7214	gamma-aminobutyric acid signaling pathway
7215	glutamate signaling pathway
7268	synaptic transmission
7270	nerve-nerve synaptic transmission
7274	neuromuscular synaptic transmission
7416	synapse assembly
7517	muscle organ development
7520	myoblast fusion
7584	response to nutrient
7605	sensory perception of sound
7611	learning or memory
7612	learning
7613	memory
7628	adult walking behavior
8016	regulation of heart contraction
8104	protein localization
8152	metabolic process
8219	cell death
8284	positive regulation of cell proliferation
8344	adult locomotory behavior
8542	visual learning
8544	epidermis development
8627	induction of apoptosis by ionic changes
9611	response to wounding
9612	response to mechanical stimulus
9617	response to bacterium
9749	response to glucose stimulus
9791	post-embryonic development
10043	response to zinc ion
10467	gene expression
10524	positive regulation of calcium ion transport into cytosol
10881	regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ion
14049	positive regulation of glutamate secretion
14051	gamma-aminobutyric acid secretion
14054	positive regulation of gamma-aminobutyric acid secretion
14056	regulation of acetylcholine secretion
14070	response to organic cyclic substance
14808	release of sequestered calcium ion into cytosol by sarcoplasmic reticulum

Table F.2.3. Biological Process Terms - cont.

15853	adenine transport
15866	ADP transport
15867	ATP transport
16049	cell growth
16079	synaptic vesicle exocytosis
16485	protein processing
17158	regulation of calcium ion-dependent exocytosis
19228	regulation of action potential in neuron
19233	sensory perception of pain
21522	spinal cord motor neuron differentiation
21590	cerebellum maturation
21679	cerebellar molecular layer development
21680	cerebellar Purkinje cell layer development
21702	cerebellar Purkinje cell differentiation
21750	vestibular nucleus development
22008	neurogenesis
30182	neuron differentiation
30431	sleep
30501	positive regulation of bone mineralization
30644	cellular chloride ion homeostasis
31000	response to caffeine
31668	cellular response to extracellular stimulus
32026	response to magnesium ion
32060	bleb assembly
32308	positive regulation of prostaglandin secretion
32353	negative regulation of hormone biosynthetic process
32469	endoplasmic reticulum calcium ion homeostasis
32496	response to lipopolysaccharide
32720	negative regulation of tumor necrosis factor production
32755	positive regulation of interleukin-6 production
32963	collagen metabolic process
33058	directional locomotion
33198	response to ATP
34405	response to fluid shear stress
35094	response to nicotine
35235	ionotropic glutamate receptor signaling pathway
35249	synaptic transmission, glutamatergic
42098	T cell proliferation
42133	neurotransmitter metabolic process
42177	negative regulation of protein catabolic process
42221	response to chemical stimulus
42391	regulation of membrane potential

Table F.2.3. Biological Process Terms - cont.

42417	dopamine metabolic process
42428	serotonin metabolic process
42445	hormone metabolic process
42493	response to drug
42542	response to hydrogen peroxide
42593	glucose homeostasis
42981	regulation of apoptosis
43029	T cell homeostasis
43065	positive regulation of apoptosis
43085	positive regulation of catalytic activity
43113	receptor clustering
43132	NAD transport
43409	negative regulation of MAPKKK cascade
43410	positive regulation of MAPKKK cascade
43524	negative regulation of neuron apoptosis
44254	multicellular organismal protein catabolic process
44419	interspecies interaction between organisms
45332	phospholipid translocation
45471	response to ethanol
45766	positive regulation of angiogenesis
45779	negative regulation of bone resorption
45822	negative regulation of heart contraction
45919	positive regulation of cytolysis
45944	positive regulation of transcription from RNA polymerase II promoter
46513	ceramide biosynthetic process
46541	saliva secretion
46931	pore complex assembly
48167	regulation of synaptic plasticity
48169	regulation of long-term neuronal synaptic plasticity
48266	behavioral response to pain
48511	rhythmic process
48705	skeletal system morphogenesis
48791	calcium ion-dependent exocytosis of neurotransmitter
48813	dendrite morphogenesis
48873	homeostasis of number of cells within a tissue
50715	positive regulation of cytokine secretion
50717	positive regulation of interleukin-1 alpha secretion
50718	positive regulation of interleukin-1 beta secretion
50770	regulation of axonogenesis
50830	defense response to Gram-positive bacterium
50850	positive regulation of calcium-mediated signaling
50882	voluntary musculoskeletal movement

Table F.2.3. Biological Process Terms - cont.

50883	musculoskeletal movement, spinal reflex action
50885	neuromuscular process controlling balance
50890	cognition
51209	release of sequestered calcium ion into cytosol
51259	protein oligomerization
51260	protein homooligomerization
51281	positive regulation of release of sequestered calcium ion into cytosol
51480	cytosolic calcium ion homeostasis
51495	positive regulation of cytoskeleton organization
51592	response to calcium ion
51602	response to electrical stimulus
51709	regulation of killing of cells of another organism
51775	response to redox state
51789	response to protein stimulus
51899	membrane depolarization
51901	positive regulation of mitochondrial depolarization
51924	regulation of calcium ion transport
51930	regulation of sensory perception of pain
51935	glutamate uptake involved in synaptic transmission
55074	calcium ion homeostasis
55085	transmembrane transport
60024	rhythmic synaptic transmission
60048	cardiac muscle contraction
60079	regulation of excitatory postsynaptic membrane potential
60401	cytosolic calcium ion transport
60402	calcium ion transport into cytosol
60547	negative regulation of necrotic cell death
70296	sarcoplasmic reticulum calcium ion transport

F.3. Modules for Human

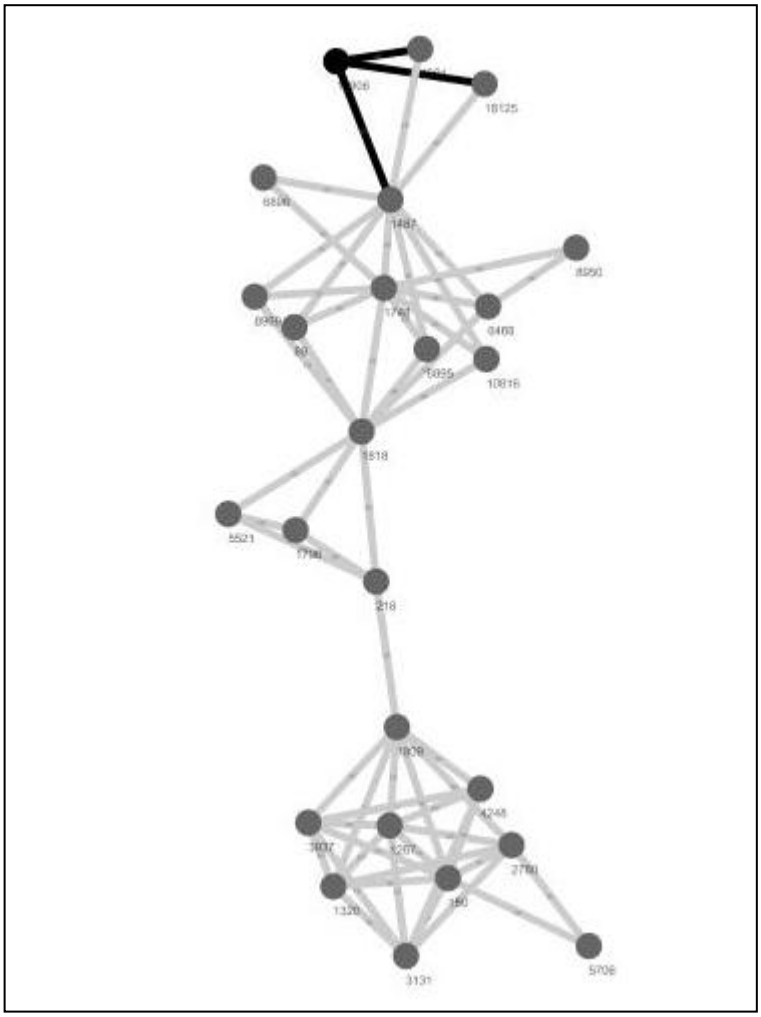


Figure F.3.1. Module 1 for human

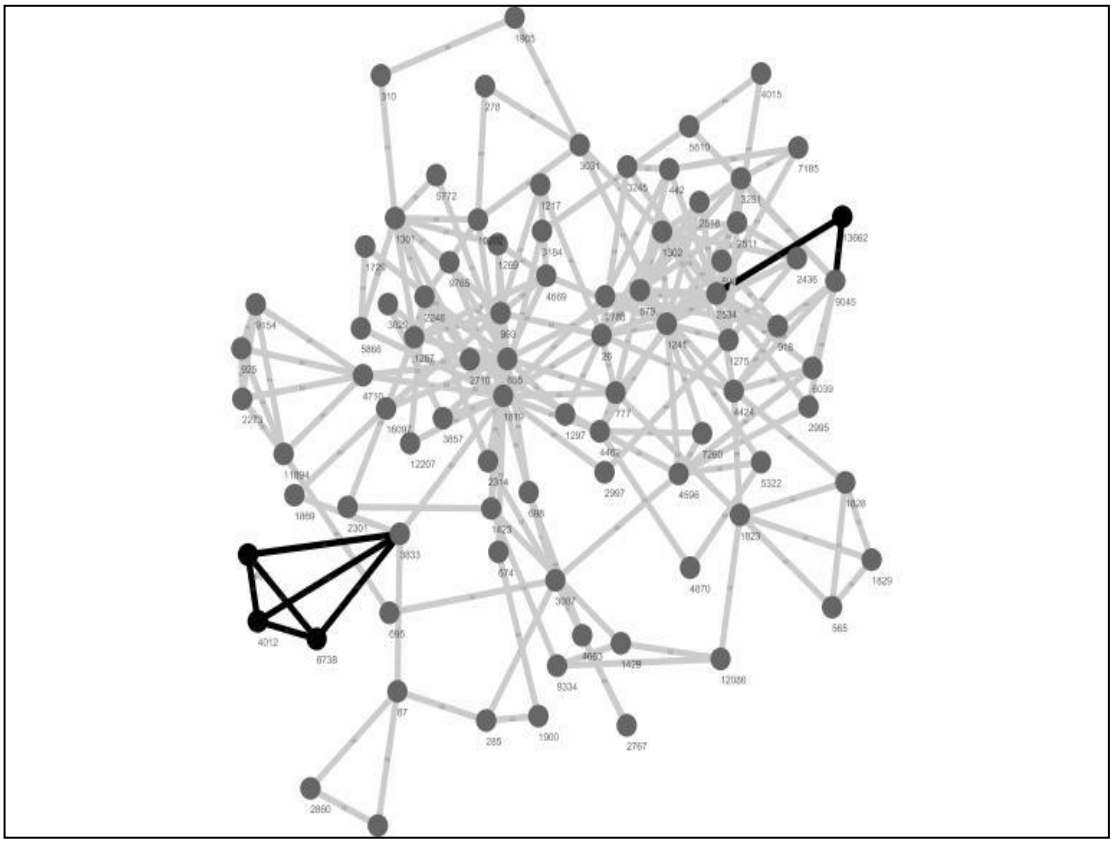


Figure F.3.2. Module 2 for human

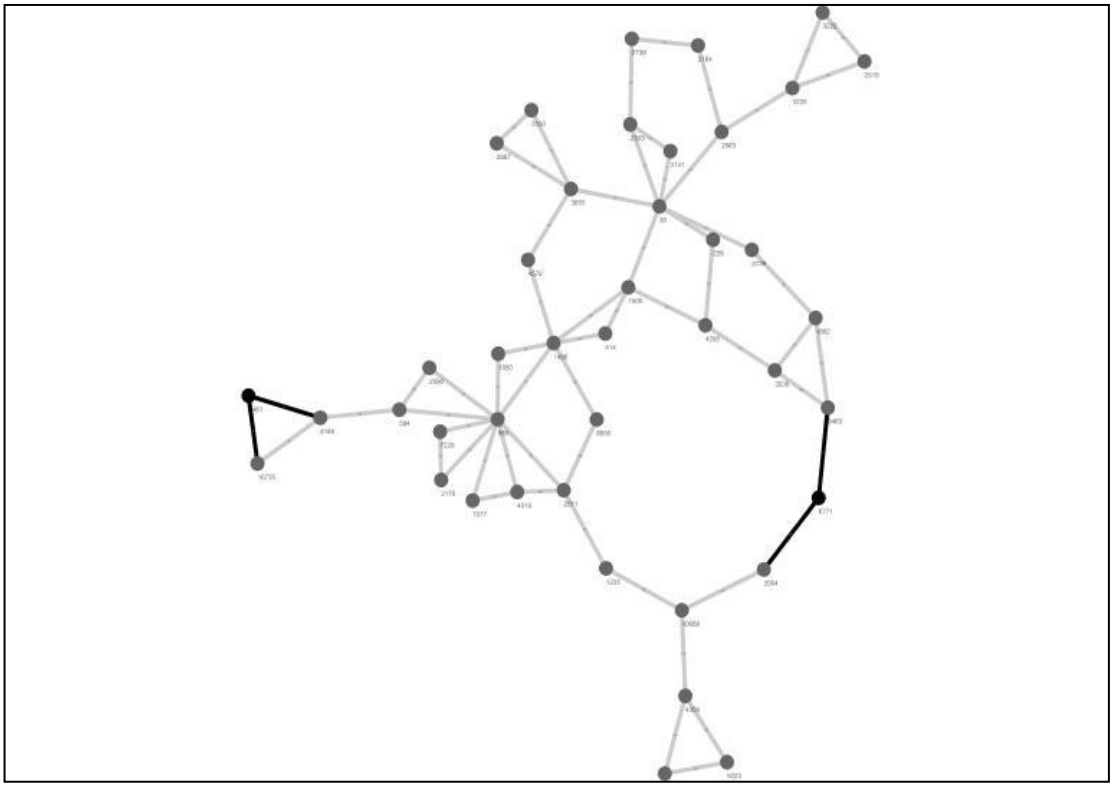


Figure F.3.3. Module 3 for human

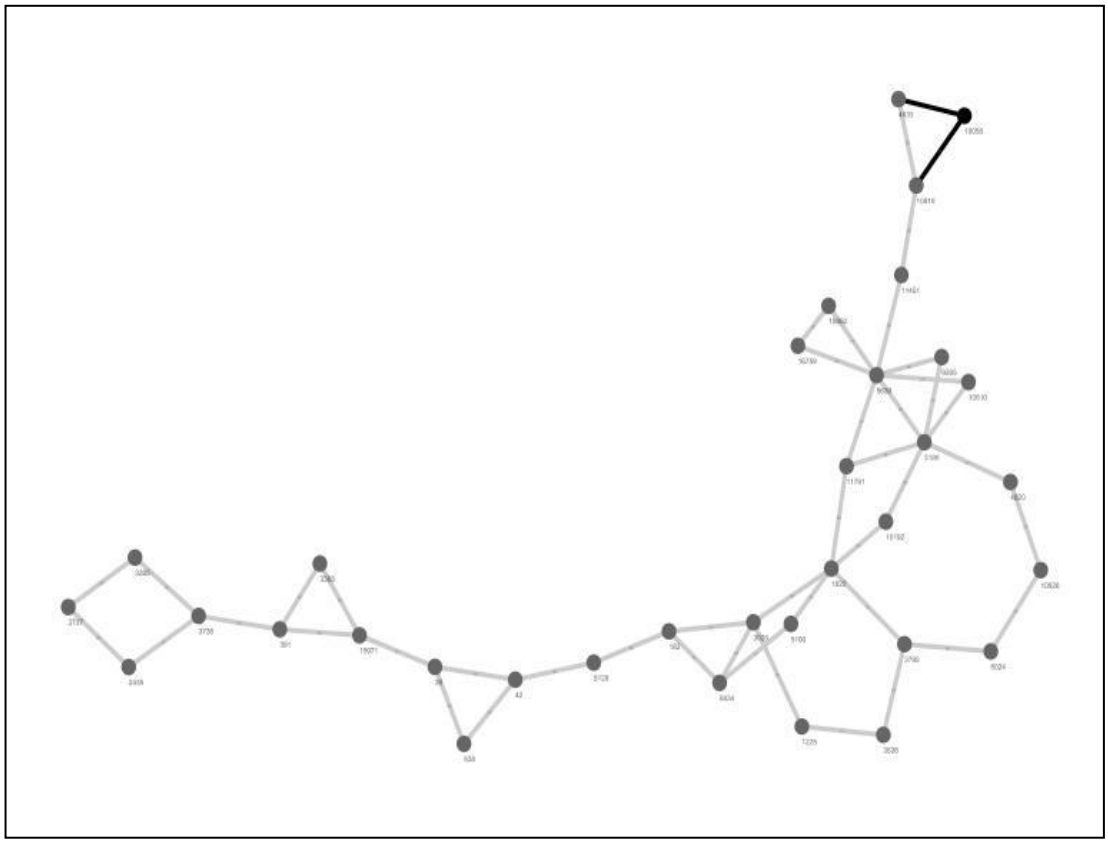


Figure F.3.4. Module 4 for human

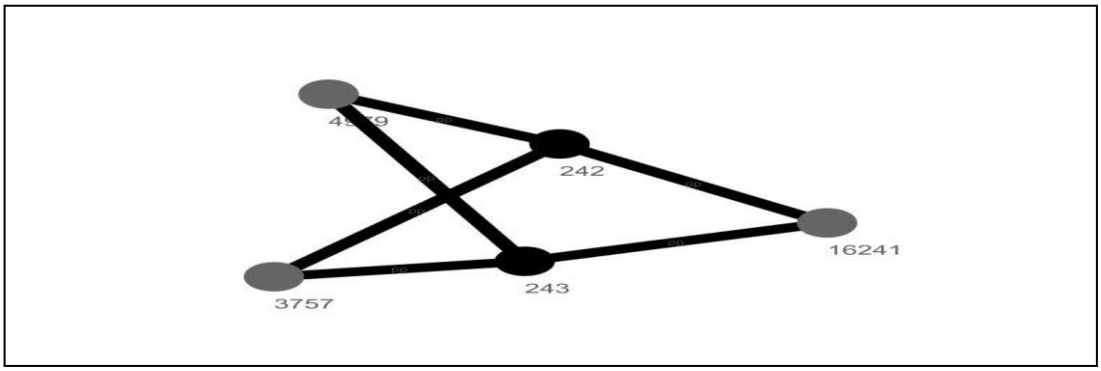


Figure F.3.5. Module 5 for human

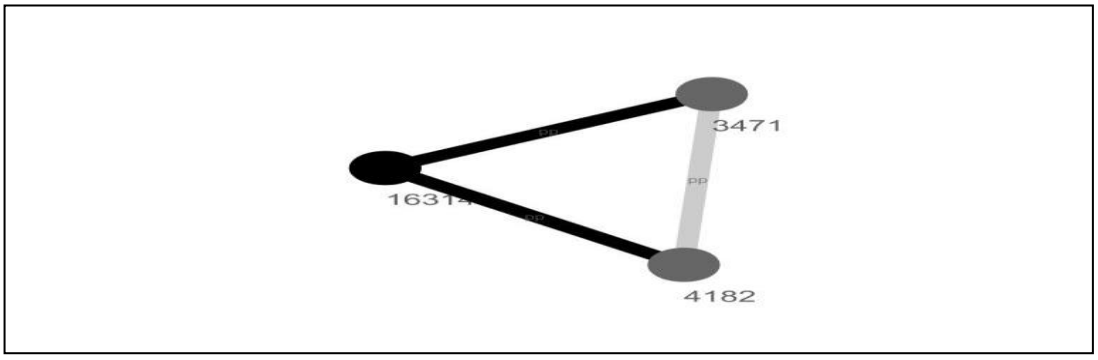


Figure F.3.6. Module 6 for human

F.4. Neurodegenerative Diseases

The Grey shaded proteins are the ones found in human PIN. The bold proteins are the ones that are not expressed in brain/brain regions.

F.4.1. Alzheimer's Disease Pathway

Table F.4.1. Alzheimer's Disease Pathway

Alzheimer		
Name	**HPRD id**	**Group in KEGG**
HSD17B10	2223	ABAD
ADAM10	3723	ADAM1017
ADAM17	4703	ADAM1017
APAF1	3755	APAF1
APH1	6360	APH-1
APOE	135	APOE
APP	100	APP
APPBP1	4545	APP-BP1
ATF6	***16118***	***ATF6***
BACE1	7255	BACE
BACE2	5744	BACE
BAD	4409	BAD
BID	3590	BID
CAPN1	253	CALPAIN
CAPN2	254	CALPAIN
CALM	***12997***	***CAM***
CALM1	241	CAM
CALM2	242	CAM
CALM3	243	CAM
CALML3	244	CAM
CALML5	5538	CAM
CASP3	2799	CASP3
CASP7	***3457***	***CASP7***
CASP8	3459	CASP8
CASP9	3756	CASP9
CDK5	449	CDK5
CHP2	***14259***	***CN***
PPP3R2	***10164***	***CN***
CHP	6101	CN
PPP3C	234	CN
PPP3CC	236	CN
PPP3R	235	CN

Table F.4.1. Alzheimer's Disease Pathway - cont.

NDUFA10	***11950***	***CX I***
NDUFA2	***11883***	***CX I***
NDUFA3	***4830***	***CX I***
NDUFA4L2	***14249***	***CX I***
NDUFA6	***11884***	***CX I***
NDUFA7	***11885***	***CX I***
NDUFA9	***4832***	***CX I***
NDUFB1	***11952***	***CX I***
NDUFB10	***11958***	***CX I***
NDUFB2	***11953***	***CX I***
NDUFB3	***11954***	***CX I***
NDUFB4	***11955***	***CX I***
NDUFB5	***11956***	***CX I***
NDUFB7	***11957***	***CX I***
NDUFB8	***11886***	***CX I***
NDUFC2	***11960***	***CX I***
NDUFS2	***4285***	***CX I***
NDUFS4	***4073***	***CX I***
NDUFS5	***11961***	***CX I***
NDUFS6	***10356***	***CX I***
NDUFS7	***3491***	***CX I***
NDUFA1	2093	CX I
NDUFA4	4831	CX I
NDUFA5	3398	CX I
NDUFA8	4525	CX I
NDUFAB1	11951	CX I
NDUFB6	4504	CX I
NDUFB9	3264	CX I
NDUFC1	11959	CX I
NDUFS1	1145	CX I
NDUFS3	10355	CX I
NDUFS8	3683	CX I
NDUFV1	1191	CX I
NDUFV2	2757	CX I
NDUFV3	3716	CX I
SDHC	***3878***	***CX II***
SDHD	***4069***	***CX II***
SDHA	2914	CX II
SDHB	1707	CX II
RIP1	***1874***	***CX III***
QCR7	***11770***	***CX III***
QCR8	***17941***	***CX III***
QCR9	***13691***	***CX III***
CYTB	7427	CX III

Table F.4.1. Alzheimer's Disease Pathway - cont.

QCR1	1875	CX III
QCR10	15628	CX III
QCR2	1876	CX III
QCR6	15629	CX III
COX2	***7421***	***CX IV***
COX6A2	***3598***	***CX IV***
COX6B2	***13088***	***CX IV***
COX7A2L	***5774***	***CX IV***
COX7B	***4812***	***CX IV***
COX7B2	***13089***	***CX IV***
COX7C	***4800***	***CX IV***
COX4	466	CX IV
COX4I2	12143	CX IV
COX5A	4799	CX IV
COX5B	467	CX IV
COX6A	3644	CX IV
COX6B	493	CX IV
COX6C	494	CX IV
COX7A	481	CX IV
COX7A2	482	CX IV
COX8	468	CX IV
COX8C	13090	CX IV
ATP5G1	***4428***	***CX V***
ATP5G2	***4429***	***CX V***
ATP5G3	***4112***	***CX V***
ATPeF0A	***2472***	***CX V***
ATP5A1	1258	CX V
ATP5B	44	CX V
ATP5C1	156	CX V
ATP5D	4397	CX V
ATP5E	5856	CX V
ATP5F1	4467	CX V
ATP5H	16519	CX V
ATP5J	4399	CX V
ATP5O	2899	CX V
CYC	479	CYTC
CYC1	480	CYTC
ERK1	3479	ERK1/2
ERK2	1496	ERK1/2
FADD	3909	FADD
TNFRSF1A	1861	FAS/TNFR
TNFRSF6	609	FAS/TNFR
APBB1	4087	FE65
GAPDH	713	GAPD

Table F.4.1. Alzheimer's Disease Pathway - cont.

GNAQ	2998	GQ
GSK3B	5418	GSK3B
IDE	894	IDE
IL1B	985	IL-1
ITPR2	***2536***	***IP3R***
ITPR3	***926***	***IP3R***
ITPR1	925	IP3R
ERN1	4943	IRE
LPL	1999	LPL
LRP1	138	LRP
NCSTN	***5584***	***NCSTN***
MME	***392***	***NEP***
GRIN1	15926	NMDAR
GRIN2A	698	NMDAR
GRIN2B	697	NMDAR
GRIN2C	699	NMDAR
GRIN2D	4095	NMDAR
NOS1	1226	NOS
CDK5R1	4586	P35
PEN2	6362	PEN2
EIF2AK3	***4942***	***PERK***
PLCB2	***4985***	***PLC***
PLCB4	***2885***	***PLC***
PLCB	6177	PLC
PLCB3	2577	PLC
PSEN1	87	PSEN
PSEN2	2860	PSEN
RYR3	1620	RYR
ATP2A	157	SERCA
ATP2A2	161	SERCA
ATP2A3	3568	SERCA
SNCA	1227	SNCA
MAPT	1142	TAU
TNFA	***1855***	***TNF***
CACNA1C	***246***	***VDCC***
CACNA1S	***248***	***VDCC***
CACNA1D	247	VDCC
CACNA1F	2119	VDCC

F.4.2. Parkinson's Disease Pathway

Table F.4.2. Parkinson's Disease Pathway

Parkinson		
Name	**HPRD id**	**Group in KEGG**
APAF1	3755	APAF1
CASP3	2799	CASP3
CASP9	3756	CASP9
PNUTL1	4100	CDCRE11
ND1	***8365***	***CX I***
ND2	***8952***	***CX I***
ND3	***8953***	***CX I***
ND4	***8954***	***CX I***
ND4L	***8955***	***CX I***
ND5	***2467***	***CX I***
ND6	***8956***	***CX I***
NDUFA10	***11950***	***CX I***
NDUFA2	***11883***	***CX I***
NDUFA3	***4830***	***CX I***
NDUFA6	***11884***	***CX I***
NDUFA7	***11885***	***CX I***
NDUFA9	***4832***	***CX I***
NDUFB10	***11958***	***CX I***
NDUFB2	***11953***	***CX I***
NDUFB3	***11954***	***CX I***
NDUFB4	***11955***	***CX I***
NDUFB5	***11956***	***CX I***
NDUFB7	***11957***	***CX I***
NDUFB8	***11886***	***CX I***
NDUFC2	***11960***	***CX I***
NDUFS2	***4285***	***CX I***
NDUFS4	***4073***	***CX I***
NDUFS5	***11961***	***CX I***
NDUFS6	***10356***	***CX I***
NDUFS7	***3491***	***CX I***
NDUFA1	2093	CX I
NDUFA4	4831	CX I
NDUFA5	3398	CX I
NDUFA8	4525	CX I
NDUFAB1	11951	CX I
NDUFB6	4504	CX I
NDUFB9	3264	CX I
NDUFC1	11959	CX I
NDUFS1	1145	CX I
NDUFS3	10355	CX I

Table F.4.2. Parkinson's Disease Pathway - cont.

NDUFS8	3683	CX I
NDUFV1	1191	CX I
NDUFV2	2757	CX I
NDUFV3	3716	CX I
SDHC	***3878***	***CX II***
SDHD	***4069***	***CX II***
SDHA	2914	CX II
SDHB	1707	CX II
RIP1	***1874***	***CX III***
QCR7	***11770***	***CX III***
QCR8	***17941***	***CX III***
QCR9	***13691***	***CX III***
CYTB	7427	CX III
QCR1	1875	CX III
QCR10	15628	CX III
QCR2	1876	CX III
QCR6	15629	CX III
COX2	***7421***	***CX IV***
COX7B	***4812***	***CX IV***
COX7C	***4800***	***CX IV***
COX4	466	CX IV
COX5A	4799	CX IV
COX5B	467	CX IV
COX6A	3644	CX IV
COX6B	493	CX IV
COX6C	494	CX IV
COX7A	481	CX IV
COX8	468	CX IV
ATPeF0A	***2472***	***CX V***
ATP5A1	1258	CX V
ATP5B	44	CX V
ATP5C1	156	CX V
ATP5D	4397	CX V
ATP5E	5856	CX V
ATP5F1	4467	CX V
ATP5H	16519	CX V
ATP5J	4399	CX V
ATP5O	2899	CX V
CYC	479	CYTC
SLC6A3	543	DAT
PARK7	3961	DJ-1
HTRA2	5919	HTRA2
LRRK2	10613	LRRK2
ANT	2147	MPTP

Table F.4.2. Parkinson's Disease Pathway - cont.

ANT2	58	MPTP
CYPD	11875	MPTP
VDAC1	5137	MPTP
VDAC3	6750	MPTP
GPR37	3992	PAE1R
PARK2	3967	PARKIN
PINK1	10514	PINK1
SNCA	1227	SNCA
SNCAIP	4804	SYNPHILIN1
TH	1865	TH
UBB	6771	UB
UBE1L	***1873***	***UBA1***
UBE1	2440	UBA1
UBE2G1	***3340***	***UBC6/7***
UBE2J1	***15599***	***UBC6/7***
UBE2J2	***15600***	***UBC6/7***
UBE2G2	4387	UBC6/7
UBE2L6	***4867***	***UBCH7/8***
UBE2L3	4762	UBCH7/8
UCHL1	1877	UCHL1

F.4.3. ALS Pathway

Table F.4.3. ALS Pathway

ALS		
Name	**HPRD id**	**Group in KEGG**
ALS2	5893	ALS2
APAF1	3755	APAF1
MAP3K5	3904	ASK1
BAD	4409	BAD
BAX	2498	BAX
BCL2	***1045***	***BCL2***
BCL2L1	2497	BCLXL
BID	3590	BID
CHP2	***14259***	***CAN***
PPP3R2	***10164***	***CAN***
CHP	6101	CAN
PPP3CA	234	CAN
PPP3CB	235	CAN
PPP3CC	236	CAN
PPP3R1	3195	CAN
CASP1	977	CASP1
CASP3	2799	CASP3
CASP9	3756	CASP9
CAT	282	CAT
CCS	4841	CCS
CYC	479	CYTC
DAXX	4424	DAXX
DERL1	***12302***	***DERLIN***
SLC1A2	2625	EAAT2
GRIA1	695	GLUR
GRIA2	4726	GLUR
GRIN1	15926	GLUR
GRIN2A	698	GLUR
GRIN2B	697	GLUR
GRIN2C	699	GLUR
GRIN2D	4095	GLUR
GPX1	11749	GPX1
MAP2K3	***3816***	***MKK3***
MAP2K6	3155	MKK6
NEFH	1204	NEFH
NEFL	1206	NEFL
NEF3	1205	NEFM
NOS1	1226	NOS
MAPK12	***3868***	***p38***
MAPK11	4208	p38

Table F.4.3. ALS Pathway - cont.

MAPK13	9113	p38
MAPK14	2619	p38
TP53	1859	P53
PRPH	1365	PRPH
PRPH2	1553	PRPH
RAB5A	1542	RAB5
RAC1	3627	RAC1
SOD1	937	SOD1
TNFA	***1855***	***TNFA***
TNFRSF1A	1861	TNFR
TNFRSF1B	1862	TNFR
TOM40	***16274***	***TOM***
TOMM40L	***13354***	***TOM***

F.4.4. Huntington's Disease Pathway

Table F.4.4. Huntington's Disease Pathway

Huntington		
Name	**HPRD id**	**Group in KEGG**
AP2A	3014	AP2
AP2B1	3015	AP2
AP2S1	3762	AP2
APAF1	3755	APAF1
BAX	2498	BAX
BDNF	214	BDNF
CASP3	2799	CASP3
CASP8	3459	CASP8
CASP9	3756	CASP9
CREBBP	2534	CBP
EP300	4078	CBP
CLTA	351	CLATHRIN
CLTB	352	CLATHRIN
CLTC	350	CLATHRIN
CLTCL1	3172	CLATHRIN
RCOR1	9645	COREST
CREB3L1	***16751***	***CREB***
CREB3L3	***16752***	***CREB***
CREB3L4	***6323***	***CREB***
CREB1	442	CREB
CREB3	7339	CREB
CREB3L2	16391	CREB
CREB5	9896	CREB
NDUFA10	***11950***	***CX I***
NDUFA2	***11883***	***CX I***
NDUFA3	***4830***	***CX I***
NDUFA6	***11884***	***CX I***
NDUFA7	***11885***	***CX I***
NDUFA9	***4832***	***CX I***
NDUFB1	***11952***	***CX I***
NDUFB10	***11958***	***CX I***
NDUFB2	***11953***	***CX I***
NDUFB3	***11954***	***CX I***
NDUFB4	***11955***	***CX I***
NDUFB5	***11956***	***CX I***
NDUFB7	***11957***	***CX I***
NDUFB8	***11886***	***CX I***
NDUFC2	***11960***	***CX I***
NDUFS2	***4285***	***CX I***
NDUFS4	***4073***	***CX I***

Table F.4.4. Huntington's Disease Pathway - cont.

NDUFS5	***11961***	***CX I***
NDUFS6	***10356***	***CX I***
NDUFS7	***3491***	***CX I***
NDUFA1	2093	CX I
NDUFA4	4831	CX I
NDUFA5	3398	CX I
NDUFA8	4525	CX I
NDUFAB1	11951	CX I
NDUFB6	4504	CX I
NDUFB9	3264	CX I
NDUFC1	11959	CX I
NDUFS1	1145	CX I
NDUFS3	10355	CX I
NDUFS8	3683	CX I
NDUFV1	1191	CX I
NDUFV2	2757	CX I
NDUFV3	3716	CX I
SDHC	***3878***	***CX II***
SDHD	***4069***	***CX II***
SDHA	2914	CX II
SDHB	1707	CX II
RIP1	***1874***	***CX III***
QCR7	***11770***	***CX III***
QCR8	***17941***	***CX III***
QCR9	***13691***	***CX III***
CYTB	7427	CX III
QCR1	1875	CX III
QCR10	15628	CX III
QCR2	1876	CX III
QCR6	15629	CX III
COX2	***7421***	***CX IV***
COX7B	***4812***	***CX IV***
COX7C	***4800***	***CX IV***
COX4	466	CX IV
COX5A	4799	CX IV
COX5B	467	CX IV
COX6A	3644	CX IV
COX6B	493	CX IV
COX6C	494	CX IV
COX7A	481	CX IV
COX8	468	CX IV
ATPeF0A	***2472***	***CX V***
ATP5A1	1258	CX V
ATP5B	44	CX V

Table F.4.4. Huntington's Disease Pathway - cont.

ATP5C1	156	CX V
ATP5D	4397	CX V
ATP5E	5856	CX V
ATP5F1	4467	CX V
ATP5H	16519	CX V
ATP5J	4399	CX V
ATP5O	2899	CX V
CYC	479	CYTC
DCTN1	7206	DYNACTIN
DCTN2	9568	DYNACTIN
DCTN4	13126	DYNACTIN
DNAI1	***5080***	***DYNEIN***
DNAI2	***12019***	***DYNEIN***
DNAL1	***16616***	***DYNEIN***
DNAH1	11790	DYNEIN
DNAH2	17214	DYNEIN
DNAH3	4513	DYNEIN
DNAL4	13239	DYNEIN
DNALI1	3681	DYNEIN
GNAQ	2998	G
GPX1	11749	GPX1
HAP1	2972	HAP1
HDAC1	3143	HDACS
HDAC2	5521	HDACS
HIP1	3461	HIP1
ESRRBL1	9430	HIPPT
HD	883	HTT
ITPR1	925	IP3R1
GRM5	4980	MGLUR5
ANT	58	MPTP
ANT2	2147	MPTP
ANT3	2461	MPTP
ANT4	13160	MPTP
CYPD	11875	MPTP
VDAC1	5137	MPTP
VDAC2	1904	MPTP
VDAC3	6750	MPTP
TFAM	2702	MTTFA
GRIN1	15926	NMDAR
GRIN2B	697	NMDAR
NRF1	***2930***	***NRF***
TP53	1859	P53
PGC1	5155	PGC-1A
PLCB2	***4985***	***PLC***

Table F.4.4. Huntington's Disease Pathway - cont.

PLCB4	***2885***	***PLC***
PLCB	6177	PLC
PLCB3	2577	PLC
RPB1	***8916***	***POL II***
RPB10	***1155***	***POL II***
RPB12	***5826***	***POL II***
RPB2	***15943***	***POL II***
RPB5	***15946***	***POL II***
RPB8	***9082***	***POL II***
RPB9	***15944***	***POL II***
RPB11	17878	POL II
RPB3	15945	POL II
RPB4	16191	POL II
RPB6	5105	POL II
RPB7	15995	POL II
NR1C3	3288	PPARY
DLG4	4199	PSD95
BBC3	***16165***	***PUMA***
REST	8990	REST
SIN3A	9690	SIN3A
SOD1	937	SOD1
SOD2	938	SOD2
SP1	1796	SP1
TAF4B	***19270***	***TAF II 13C***
TAF4	11797	TAF II 13C
TBPL2	***10607***	***TBP***
TBP	2511	TBP
TBPL1	9272	TBP
TGM2	1825	TTG
UCP1	***221***	***UCP1***

REFERENCES

Albert, T. K., H. Hanzawa, Y. I. Legtenberg, M. J. de Ruwe, F. A. van den Heuvel, M. A. Collart, R. Boelens and H. T. Timmers, 2002, "Identification of a ubiquitin-protein ligase subunit within the CCR4-NOT transcription repressor complex", *The EMBO journal,* Vol. 21, Iss. 3, pp. 355-364.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, 1997, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic acids research,* Vol. 25, Iss. 17, pp. 3389-3402.

Ames, J. B., K. B. Hendricks, T. Strahl, I. G. Huttner, N. Hamasaki and J. Thorner, 2000, "Structure and calcium-binding properties of Frq1, a novel calcium sensor in the yeast Saccharomyces cerevisiae", *Biochemistry,* Vol. 39, Iss. 40, pp. 12149-12161.

Ando, A. and C. Suzuki, 2005, "Cooperative function of the CHD5-like protein Mdm39p with a P-type ATPase Spf1p in the maintenance of ER homeostasis in Saccharomyces cerevisiae", *Molecular genetics and genomics : MGG,* Vol. 273, Iss. 6, pp. 497-506.

Apweiler, R., A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi and L. S. Yeh, 2004, "UniProt: the Universal Protein knowledgebase", *Nucleic acids research,* Vol. 32, Iss. Database issue, pp. D115-119.

Arga, K. Y., Z. I. Onsan, B. Kirdar, K. O. Ulgen and J. Nielsen, 2007, "Understanding signaling in yeast: Insights from network analysis", *Biotechnology and bioengineering,* Vol. 97, Iss. 5, pp. 1246-1258.

Ashkenazy, H., E. Erez, E. Martz, T. Pupko and N. Ben-Tal, 2010, "ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids", *Nucleic acids research,* Vol. 38, Iss. Web Server issue, pp. W529-533.

Avery, L. and H. R. Horvitz, 1990, “Effects of starvation and neuroactive drugs on feeding in Caenorhabtidis elegans”, *Journal of Experimental Zoology*, Vol. 253, pp. 263-270.

Azzouz, N., O. O. Panasenko, C. Deluen, J. Hsieh, G. Theiler and M. A. Collart, 2009, "Specific roles for the Ccr4-Not complex subunits in expression of the genome", *RNA,* Vol. 15, Iss. 3, pp. 377-383.

Babst, M., T. K. Sato, L. M. Banta and S. D. Emr, 1997, "Endosomal transport function in yeast requires a novel AAA-type ATPase, Vps4p", *The EMBO journal,* Vol. 16, Iss. 8, pp. 1820-1831.

Batiza, A. F., T. Schulz and P. H. Masson, 1996, "Yeast respond to hypotonic shock with a calcium pulse", *The Journal of biological chemistry,* Vol. 271, Iss. 38, pp. 23357-23362.

Beeler, T., K. Gable, C. Zhao and T. Dunn, 1994, "A novel protein, CSG2p, is required for Ca2+ regulation in Saccharomyces cerevisiae", *The Journal of biological chemistry,* Vol. 269, Iss. 10, pp. 7279-7284.

Berridge, M. J., 1998, "Neuronal calcium signaling", *Neuron,* Vol. 21, Iss. 1, pp. 13-26.

Bezprozvanny, I., 2009, "Calcium signaling and neurodegenerative diseases", *Trends in molecular medicine,* Vol. 15, Iss. 3, pp. 89-100.

Birchwood, C. J., J. D. Saba, R. C. Dickson and K. W. Cunningham, 2001, "Calcium influx and signaling in yeast stimulated by intracellular sphingosine 1-phosphate accumulation", *The Journal of biological chemistry,* Vol. 276, Iss. 15, pp. 11712-11718.

Blake, J. A. and M. A. Harris, 2002, "The Gene Ontology (GO) project: structured vocabularies for molecular biology and their application to genome and expression analysis", *Current protocols in bioinformatics,* Ch. 7.

Borkovich, K. A., F. W. Farrelly, D. B. Finkelstein, J. Taulien and S. Lindquist, 1989, "hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures", *Molecular and cellular biology,* Vol. 9, Iss. 9, pp. 3919-3930.

Bouquin, N., A. L. Johnson, B. A. Morgan and L. H. Johnston, 1999, "Association of the cell cycle transcription factor Mbp1 with the Skn7 response regulator in budding yeast", *Molecular biology of the cell,* Vol. 10, Iss. 10, pp. 3389-3400.

Bryant, N. J. and T. H. Stevens, 1998, "Vacuole biogenesis in Saccharomyces cerevisiae: protein transport pathways to the yeast vacuole", *Microbiology and molecular biology reviews : MMBR,* Vol. 62, Iss. 1, pp. 230-247.

Casamayor, A. and M. Snyder, 2003, "Molecular dissection of a yeast septin: distinct domains are required for septin interaction, localization, and function", *Molecular and cellular biology,* Vol. 23, Iss. 8, pp. 2762-2777.

Chen, X. J., 2004, "Sal1p, a calcium-dependent carrier protein that suppresses an essential cellular function associated With the Aac2 isoform of ADP/ATP translocase in Saccharomyces cerevisiae", *Genetics,* Vol. 167, Iss. 2, pp. 607-617.

Cherry, J. M., C. Adler, C. Ball, S. A. Chervitz, S. S. Dwight, E. T. Hester, Y. Jia, G. Juvik, T. Roe, M. Schroeder, S. Weng and D. Botstein, 1998, "SGD: Saccharomyces Genome Database", *Nucleic acids research,* Vol. 26, Iss. 1, pp. 73-79.

Chou, S., S. Lane and H. Liu, 2006, "Regulation of mating and filamentation genes by two distinct Ste12 complexes in Saccharomyces cerevisiae", *Molecular and cellular biology,* Vol. 26, Iss. 13, pp. 4794-4805.

Clandinin, T. R., J. A. DeModena and P. W. Sternberg, 1998, "Inositol trisphosphate mediates a RAS-independent response to LET-23 receptor tyrosine kinase activation in C. elegans", *Cell,* Vol. 92, Iss. 4, pp. 523-533.

Clapham, D. E., 1995, "Calcium signaling", *Cell,* Vol. 80, Iss. 2, pp. 259-268.

Collart, M. A., 2003, "Global control of gene expression in yeast by the Ccr4-Not complex", *Gene,* Vol. 313, pp. 1-16.

Cronin, S. R., R. Rao and R. Y. Hampton, 2002, "Cod1p/Spf1p is a P-type ATPase involved in ER function and Ca2+ homeostasis", *The Journal of cell biology,* Vol. 157, Iss. 6, pp. 1017-1028.

Cui, J. and J. A. Kaandorp, 2006, "Mathematical modeling of calcium homeostasis in yeast cells", *Cell calcium,* Vol. 39, Iss. 4, pp. 337-348.

Cui, J., J. A. Kaandorp, P. M. Sloot, C. M. Lloyd and M. V. Filatov, 2009, "Calcium homeostasis and signaling in yeast cells and cardiac myocytes", *FEMS yeast research,* Vol. 9, Iss. 8, pp. 1137-1147.

Cunningham, K. W. and G. R. Fink, 1996, "Calcineurin inhibits VCX1-dependent H+/Ca2+ exchange and induces Ca2+ ATPases in Saccharomyces cerevisiae", *Molecular and cellular biology,* Vol. 16, Iss. 5, pp. 2226-2237.

Cyert, M. S., 2001, "Genetic analysis of calmodulin and its targets in Saccharomyces cerevisiae", *Annual review of genetics,* Vol. 35, pp. 647-672.

Cyert, M. S., R. Kunisawa, D. Kaim and J. Thorner, 1991, "Yeast has homologs (CNA1 and CNA2 gene products) of mammalian calcineurin, a calmodulin-regulated phosphoprotein phosphatase", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 88, Iss. 16, pp. 7376-7380.

Cyert, M. S. and J. Thorner, 1992, "Regulatory subunit (CNB1 gene product) of yeast Ca2+/calmodulin-dependent phosphoprotein phosphatases is required for adaptation to pheromone", *Molecular and cellular biology,* Vol. 12, Iss. 8, pp. 3460-3469.

Dailey, L., S. B. Roberts and N. Heintz, 1987, "RNA polymerase II transcription factors H4TF-1 and H4TF-2 require metal to bind specific DNA sequences", *Molecular and cellular biology,* Vol. 7, Iss. 12, pp. 4582-4584.

Dal Santo, P., M.A. Logan, A.D. Chisholm and E.M. Jorgensen, 1999, “The inositol trisphosphate receptor regulates a 50-second behavioral rhythm in *C. elegans*”, *Cell,* Vol. *98,* pp. 757-767.

de Nadal, E., P. M. Alepuz and F. Posas, 2002, "Dealing with osmostress through MAP kinase activation", *EMBO reports,* Vol. 3, Iss. 8, pp. 735-740.

Degand, I., P. Catty, E. Talla, D. Thines-Sempoux, A. de Kerchove d'Exaerde, A. Goffeau and M. Ghislain, 1999, "Rabbit sarcoplasmic reticulum Ca(2+)-ATPase replaces yeast PMC1 and PMR1 Ca(2+)-ATPases for cell viability and calcineurin-dependent regulation of calcium tolerance", *Molecular microbiology,* Vol. 31, Iss. 2, pp. 545-556.

Denis, V. and M. S. Cyert, 2002, "Internal Ca(2+) release in yeast is triggered by hypertonic shock and mediated by a TRP channel homologue", *The Journal of cell biology,* Vol. 156, Iss. 1, pp. 29-34.

Flick, J. S. and J. Thorner, 1993, "Genetic and biochemical characterization of a phosphatidylinositol-specific phospholipase C in Saccharomyces cerevisiae", *Molecular and cellular biology,* Vol. 13, Iss. 9, pp. 5861-5876.

Fu, L., A. Miseta, D. Hunton, R. B. Marchase and D. M. Bedwell, 2000, "Loss of the major isoform of phosphoglucomutase results in altered calcium homeostasis in Saccharomyces cerevisiae", *The Journal of biological chemistry,* Vol. 275, Iss. 8, pp. 5431-5440.

Fuller, R. S., A. Brake and J. Thorner, 1989, "Yeast prohormone processing enzyme (KEX2 gene product) is a Ca2+-dependent serine protease", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 86, Iss. 5, pp. 1434-1438.

Futschik, M. E., G. Chaurasia and H. Herzel, 2007, "Comparison of human protein-protein interaction maps", *Bioinformatics,* Vol. 23, Iss. 5, pp. 605-611.

Ghosh, A. and M. E. Greenberg, 1995, "Calcium signaling in neurons: molecular mechanisms and cellular consequences", *Science,* Vol. 268, Iss. 5208, pp. 239-247.

Gokhale, K. C., G. P. Newnam, M. Y. Sherman and Y. O. Chernoff, 2005, "Modulation of prion-dependent polyglutamine aggregation and toxicity by chaperone proteins in the yeast model", *The Journal of biological chemistry,* Vol. 280, Iss. 24, pp. 22809-22818.

Gorbsky, G. J., M. Kallio, J. R. Daum and L. M. Topper, 1999, "Protein dynamics at the kinetochore: cell cycle regulation of the metaphase to anaphase transition", *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology,* Vol. 13 Suppl 2, pp. S231-234.

Gotoh, O., 1999, "Multiple sequence alignment: algorithms and applications", *Advances in biophysics,* Vol. 36, pp. 159-206.

Grimshaw, S. J., H. R. Mott, K. M. Stott, P. R. Nielsen, K. A. Evetts, L. J. Hopkins, D. Nietlispach and D. Owen, 2004, "Structure of the sterile alpha motif (SAM) domain of the Saccharomyces cerevisiae mitogen-activated protein kinase pathway-modulating protein STE50 and analysis of its interaction with the STE11 SAM", *The Journal of biological chemistry,* Vol. 279, Iss. 3, pp. 2192-2201.

Guldener, U., M. Munsterkotter, G. Kastenmuller, N. Strack, J. van Helden, C. Lemer, J. Richelles, S. J. Wodak, J. Garcia-Martinez, J. E. Perez-Ortin, H. Michael, A. Kaps, E. Talla, B. Dujon, B. Andre, J. L. Souciet, J. De Montigny, E. Bon, C. Gaillardin and H. W. Mewes, 2005, "CYGD: the Comprehensive Yeast Genome Database", *Nucleic acids research,* Vol. 33, Iss. Database issue, pp. D364-368.

Gursoy, A., O. Keskin and R. Nussinov, 2008, "Topological properties of protein interaction networks from a structural perspective", *Biochemical Society transactions,* Vol. 36, Iss. Pt 6, pp. 1398-1403.

Hahn, M. W. and A. D. Kern, 2005, "Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks", *Molecular biology and evolution,* Vol. 22, Iss. 4, pp. 803-806.

Hartwell, L. H., J. J. Hopfield, S. Leibler and A. W. Murray, 1999, "From molecular to modular cell biology", *Nature,* Vol. 402, Iss. 6761 Suppl, pp. C47-52.

Heath, V. L., S. L. Shaw, S. Roy and M. S. Cyert, 2004, "Hph1p and Hph2p, novel components of calcineurin-mediated stress responses in Saccharomyces cerevisiae", *Eukaryotic cell,* Vol. 3, Iss. 3, pp. 695-704.

Holyoak, C. D., S. Thompson, C. Ortiz Calderon, K. Hatzixanthis, B. Bauer, K. Kuchler, P. W. Piper and P. J. Coote, 2000, "Loss of Cmk1 Ca(2+)-calmodulin-dependent protein kinase in yeast results in constitutive weak organic acid resistance, associated with a post-transcriptional activation of the Pdr12 ATP-binding cassette transporter", *Molecular microbiology,* Vol. 37, Iss. 3, pp. 595-605.

Hook, S. S., A. Orian, S. M. Cowley and R. N. Eisenman, 2002, "Histone deacetylase 6 binds polyubiquitin through its zinc finger (PAZ domain) and copurifies with deubiquitinating enzymes", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 99, Iss. 21, pp. 13425-13430.

Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori and Y. Sakaki, 2001, "A comprehensive two-hybrid analysis to explore the yeast protein interactome", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 98, Iss. 8, pp. 4569-4574.

Jeong, H., S. P. Mason, A. L. Barabasi and Z. N. Oltvai, 2001, "Lethality and centrality in protein networks", *Nature,* Vol. 411, Iss. 6833, pp. 41-42.

Jimeno, S., A. G. Rondon, R. Luna and A. Aguilera, 2002, "The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability", *The EMBO journal,* Vol. 21, Iss. 13, pp. 3526-3535.

Joy, M. P., A. Brock, D. E. Ingber and S. Huang, 2005, "High-betweenness proteins in the yeast protein interaction network", *Journal of biomedicine & biotechnology,* Vol. 2005, Iss. 2, pp. 96-103.

Kafadar, K. A. and M. S. Cyert, 2004, "Integration of stress responses: modulation of calcineurin signaling in Saccharomyces cerevisiae by protein kinase A", *Eukaryotic cell,* Vol. 3, Iss. 5, pp. 1147-1153.

Kalyoncu, S., O. Keskin and A. Gursoy, 2010, "Interaction prediction and classification of PDZ domains", *BMC bioinformatics,* Vol. 11, pp. 357.

Kambouris, N. G., D. J. Burke and C. E. Creutz, 1993, "Cloning and genetic characterization of a calcium- and phospholipid-binding protein from Saccharomyces cerevisiae that is homologous to translation elongation factor-1 gamma", *Yeast,* Vol. 9, Iss. 2, pp. 151-163.

Kane, P. M., 2006, "The where, when, and how of organelle acidification by the yeast vacuolar H+-ATPase", *Microbiology and molecular biology reviews : MMBR,* Vol. 70, Iss. 1, pp. 177-191.

Kerr, R., V. Lev-Ram, G. Baird, P. Vincent, R. Y. Tsien, W. R. Schafer, 2002, "Optical imaging of calcium transients in neurons and pharyngeal muscle of C. elegans*", Neuron*, Vol. 26, pp. 583–594.

Kirischuk, S. and A. Verkhratsky, 1996, "Calcium homeostasis in aged neurones", *Life sciences,* Vol. 59, Iss. 5-6, pp. 451-459.

Kitazono, A. A. and S. J. Kron, 2002, "An essential function of yeast cyclin-dependent kinase Cdc28 maintains chromosome stability", *The Journal of biological chemistry,* Vol. 277, Iss. 50, pp. 48627-48634.

Klipp, E. and W. Liebermeister, 2006, "Mathematical modeling of intracellular signaling pathways", *BMC neuroscience,* Vol. 7 Suppl 1, pp. S10.

LaCount, D. J., M. Vignali, R. Chettier et al., 2005, "A protein interaction network of the malaria parasite Plasmodium falciparum", *Nature,* Vol. 438, Iss. 7064, pp. 103-107.

Laskowski, R. A., E. G. Hutchinson, A. D. Michie, A. C. Wallace, M. L. Jones and J. M. Thornton, 1997, "PDBsum: a Web-based database of summaries and analyses of all PDB structures", *Trends in biochemical sciences,* Vol. 22, Iss. 12, pp. 488-490.

Li, C. W. and B. S. Chen, 2010, "Identifying functional mechanisms of gene and protein regulatory networks in response to a broader range of environmental stresses", *Comparative and functional genomics,* pp. 408705.

Li, S., C. M. Armstrong, N. Bertin et al., 2004, "A map of the interactome network of the metazoan C. elegans", *Science,* Vol. 303, Iss. 5657, pp. 540-543.

Lin, C. Y., C. H. Chin, H. H. Wu, S. H. Chen, C. W. Ho and M. T. Ko, 2008, "Hubba: hub objects analyzer--a framework of interactome hubs identification for network biology", *Nucleic acids research,* Vol. 36, Iss. Web Server issue, pp. W438-443.

Locke, E. G., M. Bonilla, L. Liang, Y. Takita and K. W. Cunningham, 2000, "A homolog of voltage-gated Ca(2+) channels stimulated by depletion of secretory Ca(2+) in yeast", *Molecular and cellular biology,* Vol. 20, Iss. 18, pp. 6686-6694.

Lopez, F., M. Leube, R. Gil-Mascarell, J. P. Navarro-Avino and R. Serrano, 1999, "The yeast inositol monophosphatase is a lithium- and sodium-sensitive enzyme encoded by a non-essential gene pair", *Molecular microbiology,* Vol. 31, Iss. 4, pp. 1255-1264.

Maere, S., K. Heymans and M. Kuiper, 2005, "BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks", *Bioinformatics,* Vol. 21, Iss. 16, pp. 3448-3449.

Makino, T. and T. Gojobori, 2006, "The evolutionary rate of a protein is influenced by features of the interacting partners", *Molecular biology and evolution,* Vol. 23, Iss. 4, pp. 784-789.

Markovich, S., A. Yekutiel, I. Shalit, Y. Shadkchan and N. Osherov, 2004, "Genomic approach to identification of mutations affecting caspofungin susceptibility in Saccharomyces cerevisiae", *Antimicrobial agents and chemotherapy,* Vol. 48, Iss. 10, pp. 3871-3876.

Matheos, D. P., T. J. Kingsbury, U. S. Ahsan and K. W. Cunningham, 1997, "Tcn1p/Crz1p, a calcineurin-dependent transcription factor that differentially regulates gene expression in Saccharomyces cerevisiae", *Genes & development,* Vol. 11, Iss. 24, pp. 3445-3458.

Mayer, M. L., B. E. Caplin and M. S. Marshall, 1992, "CDC43 and RAM2 encode the polypeptide subunits of a yeast type I protein geranylgeranyltransferase", *The Journal of biological chemistry,* Vol. 267, Iss. 29, pp. 20589-20593.

Miyamoto, S., Y. Ohya, Y. Sano, S. Sakaguchi, H. Iida and Y. Anraku, 1991, "A DBL-homologous region of the yeast CLS4/CDC24 gene product is important for Ca(2+)-modulated bud assembly", *Biochemical and biophysical research communications,* Vol. 181, Iss. 2, pp. 604-610.

Muller, E. M., E. G. Locke and K. W. Cunningham, 2001, "Differential regulation of two Ca(2+) influx systems by pheromone signaling in Saccharomyces cerevisiae", *Genetics,* Vol. 159, Iss. 4, pp. 1527-1538.

Murchison, D., D. C. Zawieja and W. H. Griffith, 2004, "Reduced mitochondrial buffering of voltage-gated calcium influx in aged rat basal forebrain neurons", *Cell calcium,* Vol. 36, Iss. 1, pp. 61-75.

Nagaraj, S., A. Rajendran, C. E. Jackson and M. S. Longtine, 2008, "Role of nucleotide binding in septin-septin interactions and septin localization in Saccharomyces cerevisiae", *Molecular and cellular biology,* Vol. 28, Iss. 16, pp. 5120-5137.

Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono and M. Kanehisa, 1999, "KEGG: Kyoto Encyclopedia of Genes and Genomes", *Nucleic acids research,* Vol. 27, Iss. 1, pp. 29-34.

Ohya, Y. and D. Botstein, 1994, "Diverse essential functions revealed by complementing yeast calmodulin mutants", *Science,* Vol. 263, Iss. 5149, pp. 963-966.

Ortutay, C. and M. Vihinen, 2009, "Identification of candidate disease genes by integrating Gene Ontologies and protein-interaction networks: case study of primary immunodeficiencies", *Nucleic acids research,* Vol. 37, Iss. 2, pp. 622-628.

Pearson, W. R., 1995, "Comparison of methods for searching protein sequence databases", *Protein science : a publication of the Protein Society,* Vol. 4, Iss. 6, pp. 1145-1160.

Peri, S., J. D. Navarro, T. Z. Kristiansen et al., 2004, "Human protein reference database as a discovery resource for proteomics", *Nucleic acids research,* Vol. 32, Iss. Database issue, pp. D497-501.

Portman, D. S., J. P. O'Connor and G. Dreyfuss, 1997, "YRA1, an essential Saccharomyces cerevisiae gene, encodes a novel nuclear protein with RNA annealing activity", *RNA,* Vol. 3, Iss. 5, pp. 527-537.

Pozos, T. C., I. Sekler and M. S. Cyert, 1996, "The product of HUM1, a novel yeast gene, is required for vacuolar Ca2+/H+ exchange and is related to mammalian Na+/Ca2+ exchangers", *Molecular and cellular biology,* Vol. 16, Iss. 7, pp. 3730-3741.

Prasad, T. S., K. Kandasamy and A. Pandey, 2009, "Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology", *Methods in molecular biology,* Vol. 577, pp. 67-79.

Prezant, T. R., W. E. Chaltraw, Jr. and N. Fischel-Ghodsian, 1996, "Identification of an overexpressed yeast gene which prevents aminoglycoside toxicity", *Microbiology,* Vol. 142 (Pt 12), pp. 3407-3414.

Procter, J. B., J. Thompson, I. Letunic, C. Creevey, F. Jossinet and G. J. Barton, 2010, "Visualization of multiple alignments, phylogenies and gene family evolution", *Nature methods,* Vol. 7, Iss. 3 Suppl, pp. S16-25.

Przytycka, T. M., M. Singh and D. K. Slonim, 2010, "Toward the dynamic interactome: it's about time", *Briefings in bioinformatics,* Vol. 11, Iss. 1, pp. 15-29.

Raitt, D. C., F. Posas and H. Saito, 2000, "Yeast Cdc42 GTPase and Ste20 PAK-like kinase regulate Sho1-dependent activation of the Hog1 MAPK pathway", *The EMBO journal,* Vol. 19, Iss. 17, pp. 4623-4631.

Ravasz, E., A. L. Somera, D. A. Mongru, Z. N. Oltvai and A. L. Barabasi, 2002, "Hierarchical organization of modularity in metabolic networks", *Science,* Vol. 297, Iss. 5586, pp. 1551-1555.

Reinoso-Martin, C., C. Schuller, M. Schuetzer-Muehlbauer and K. Kuchler, 2003, "The yeast protein kinase C cell integrity pathway mediates tolerance to the antifungal drug caspofungin through activation of Slt2p mitogen-activated protein kinase signaling", *Eukaryotic cell,* Vol. 2, Iss. 6, pp. 1200-1210.

Reiser, V., D. C. Raitt and H. Saito, 2003, "Yeast osmosensor Sln1 and plant cytokinin receptor Cre1 respond to changes in turgor pressure", *The Journal of cell biology,* Vol. 161, Iss. 6, pp. 1035-1040.

Reiser, V., S. M. Salah and G. Ammerer, 2000, "Polarized localization of yeast Pbs2 depends on osmostress, the membrane protein Sho1 and Cdc42", *Nature cell biology,* Vol. 2, Iss. 9, pp. 620-627.

Rual, J. F., K. Venkatesan, T. Hao et al., 2005, "Towards a proteome-scale map of the human protein-protein interaction network", *Nature,* Vol. 437, Iss. 7062, pp. 1173-1178.

Ruiz-Roig, C., C. Vieitez, F. Posas and E. de Nadal, 2010, "The Rpd3L HDAC complex is essential for the heat stress response in yeast", *Molecular microbiology,* Vol. 76, Iss. 4, pp. 1049-1062.

Rutledge, E., L. Bianchi, M. Christensen, C. Boehmer, R. Morrison, A. Broslat, A.M. Beld, A. George, D. Greenstein, and K. Strange, 2001, “CLH-3, a ClC-2 anion channel ortholog activated during meiotic maturation in C. Elegans oocytes”, *Current Biology*, Vol. 11, pp. 161–170.

Saito, H. and K. Tatebayashi, 2004, "Regulation of the osmoregulatory HOG MAPK cascade in yeast", *Journal of biochemistry,* Vol. 136, Iss. 3, pp. 267-272.

Sato, K. and A. Nakano, 2002, "Emp47p and its close homolog Emp46p have a tyrosine-containing endoplasmic reticulum exit signal and function in glycoprotein secretion in Saccharomyces cerevisiae", *Molecular biology of the cell,* Vol. 13, Iss. 7, pp. 2518-2532.

Schulz, T. A. and C. E. Creutz, 2004, "The tricalbin C2 domains: lipid-binding properties of a novel, synaptotagmin-like yeast protein family", *Biochemistry,* Vol. 43, Iss. 13, pp. 3987-3995.

Schwikowski, B., P. Uetz and S. Fields, 2000, "A network of protein-protein interactions in yeast", *Nature biotechnology,* Vol. 18, Iss. 12, pp. 1257-1261.

Sengupta, S. M., M. VanKanegan, J. Persinger, C. Logie, B. R. Cairns, C. L. Peterson and B. Bartholomew, 2001, "The interactions of yeast SWI/SNF and RSC with the nucleosome before and after chromatin remodeling", *The Journal of biological chemistry,* Vol. 276, Iss. 16, pp. 12636-12644.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, 2003, "Cytoscape: a software environment for integrated models of biomolecular interaction networks", *Genome research,* Vol. 13, Iss. 11, pp. 2498-2504.

Sharan, R., I. Ulitsky and R. Shamir, 2007, "Network-based prediction of protein function", *Molecular systems biology,* Vol. 3, pp. 88.

Sherbet, G.V., 2001, *Calcium Signaling in Cancer*, CRC Press.

Shimozono, S., T. Fukano, K. D. Kimura, I. Mori, Y. Kirino, A. Miyawaki, 2004,"Slow calcium dynamics in pharyngeal muscle in Caenorhabditis elegans during fast pumping", *EMBO Reports*, Vol. 5, pp. 521– 526.

Sonnhammer, E. L., S. R. Eddy and R. Durbin, 1997, "Pfam: a comprehensive database of protein domain families based on seed alignments", *Proteins,* Vol. 28, Iss. 3, pp. 405-420.

Sopko, R., D. Huang, N. Preston et al., 2006, "Mapping pathways and phenotypes by systematic gene overexpression", *Molecular cell,* Vol. 21, Iss. 3, pp. 319-330.

Soutourina, J., V. Bordas-Le Floch, G. Gendrel et al., 2006, "Rsc4 connects the chromatin remodeler RSC to RNA polymerases", *Molecular and cellular biology,* Vol. 26, Iss. 13, pp. 4920-4933.

Stark, C., B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz and M. Tyers, 2006, "BioGRID: a general repository for interaction datasets", *Nucleic acids research,* Vol. 34, Iss. Database issue, pp. D535-539.

Strange, K., 2003, "From Genes to Integrative Physiology: Ion Channel and Transporter Biology in Caenorhabditis elegans", *Physiological Reviews*, Vol. 83, pp 377-415.

Sullivan, D. S., S. Biggins and M. D. Rose, 1998, "The yeast centrin, cdc31p, and the interacting protein kinase, Kic1p, are required for cell integrity", *The Journal of cell biology,* Vol. 143, Iss. 3, pp. 751-765.

Tang, H. Y. and M. Cai, 1996, "The EH-domain-containing protein Pan1 is required for normal organization of the actin cytoskeleton in Saccharomyces cerevisiae", *Molecular and cellular biology,* Vol. 16, Iss. 9, pp. 4897-4914.

Tatebayashi, K., K. Yamamoto, K. Tanaka, T. Tomida, T. Maruoka, E. Kasukawa and H. Saito, 2006, "Adaptor functions of Cdc42, Ste50, and Sho1 in the yeast osmoregulatory HOG MAPK pathway", *The EMBO journal,* Vol. 25, Iss. 13, pp. 3033-3044.

Thibault, O. and P. W. Landfield, 1996, "Increase in single L-type calcium channels in hippocampal neurons during aging", *Science,* Vol. 272, Iss. 5264, pp. 1017-1020.

Toda, T., S. Cameron, P. Sass, M. Zoller and M. Wigler, 1987, "Three different genes in S. cerevisiae encode the catalytic subunits of the cAMP-dependent protein kinase", *Cell,* Vol. 50, Iss. 2, pp. 277-287.

Vasilyeva, E., Q. Liu, K. J. MacLeod, J. D. Baleja and M. Forgac, 2000, "Cysteine scanning mutagenesis of the noncatalytic nucleotide binding site of the yeast V-ATPase", *The Journal of biological chemistry,* Vol. 275, Iss. 1, pp. 255-260.

Verkhratsky, A. and E. C. Toescu, 2003, "Endoplasmic reticulum Ca(2+) homeostasis and neuronal death", *Journal of cellular and molecular medicine,* Vol. 7, Iss. 4, pp. 351-361.

Vernarecci, S., G. Colotti, P. Ornaghi, E. Schiebel, E. Chiancone and P. Filetici, 2007, "The yeast penta-EF protein Pef1p is involved in cation-dependent budding and cell polarization", *Molecular microbiology,* Vol. 65, Iss. 4, pp. 1122-1138.

von Plehwe, U., U. Berndt, C. Conz, M. Chiabudini, E. Fitzke, A. Sickmann, A. Petersen, D. Pfeifer and S. Rospert, 2009, "The Hsp70 homolog Ssb is essential for glucose sensing via the SNF1 kinase network", *Genes & development,* Vol. 23, Iss. 17, pp. 2102-2115.

Wolff, A. M., J. G. Petersen, T. Nilsson-Tillgren and N. Din, 1999, "The open reading frame YAL048c affects the secretion of proteinase A in S. cerevisiae", *Yeast,* Vol. 15, Iss. 5, pp. 427-434.

Xenarios, I., D. W. Rice, L. Salwinski, M. K. Baron, E. M. Marcotte and D. Eisenberg, 2000, "DIP: the database of interacting proteins", *Nucleic acids research,* Vol. 28, Iss. 1, pp. 289-291.

Yamashiro, C. T., P. M. Kane, D. F. Wolczyk, R. A. Preston and T. H. Stevens, 1990, "Role of vacuolar acidification in protein sorting and zymogen activation: a genetic analysis of the yeast vacuolar proton-translocating ATPase", *Molecular and cellular biology,* Vol. 10, Iss. 7, pp. 3737-3749.

Yin, X., N.J. Gower, H.A. Baylis, and K. Strange, 2004,“Inositol 1,4,5- trisphosphate signaling regulates rhythmic contractile activity of smooth muscle-like sheath cells in the nematode Caenorhabditis elegans”, *Molecular Biology of The Cell*, Vol. 15, pp. 3938–3949.

Yook, S. H., Z. N. Oltvai and A. L. Barabasi, 2004, "Functional and topological characterization of protein interaction networks", *Proteomics,* Vol. 4, Iss. 4, pp. 928-942.

Yooseph, S., G. Sutton, D. B. Rusch et al., 2007, "The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families", *PLoS biology,* Vol. 5, Iss. 3, pp. e16.

Zelter, A., M. Bencina, B. J. Bowman, O. Yarden and N. D. Read, 2004, "A comparative genomic analysis of the calcium signaling machinery in Neurospora crassa, Magnaporthe grisea, and Saccharomyces cerevisiae", *Fungal genetics and biology : FG & B,* Vol. 41, Iss. 9, pp. 827-841.

Zhang, Y., 2003, "Transcriptional regulation by histone ubiquitination and deubiquitination", *Genes & development,* Vol. 17, Iss. 22, pp. 2733-2740.

Zotenko, E., J. Mestre, D. P. O'Leary and T. M. Przytycka, 2008, "Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality", *PLoS computational biology,* Vol. 4, Iss. 8, pp. e1000140.