ABSTRACT

A Novel Computational Method for Predicting Tissue-specific Disease-associated Signaling Pathways in Human Utilizing *Caenorhabditis elegans* Reference Data

Xuan Peng, M.S.

Mentors: Myeongwoo Lee, Ph.D. and Young-Rae Cho, Ph.D.

Signal transduction is a hot topic as molecular biology grows because it directly relates to cellular processes, supporting function of the organism as a whole. A dysfunctional signal transduction will cause uncoordinated cellular behaviors. For humans, these uncoordinated cellular behaviors will cause diseases. To study mechanism of signal transduction, diverse approaches have been applied, including traditional experimental and computational methods. Compared to traditional experimental approaches, computational methods are better in analyzing large amounts of data and predicting results from limited data. In this research, a novel computational method is built to predict tissue-specific disease-associated signaling pathways in human by referring to *C. elegans* data. Tissue-specificity and disease association data are utilized to perform this prediction, with a support of a novel pathway finding algorithm. Lists of candidate pathways associated with certain selected diseases are successfully generated from the results.

A Novel Computational Method for Predicting Tissue-specific Disease-associated Signaling Pathway in Human Utilizing *Caenorhabditis elegans* Reference Data

by

Xuan Peng, B.S.

A Thesis

Approved by the Department of Biology

Robert D. Doyle, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Master of Science

Approved by the Thesis Committee

Myeongwoo Lee, Ph.D., Chairperson

Young-Rae Cho, Ph.D., Co-Chairperson

Cheolho Sim, Ph.D.

Accepted by the Graduate School
August 2013

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

Copyright © 2013 by Xuan Peng

All rights reserved

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First, I would like to acknowledge Dr. Myeongwoo Lee, who guided me towards a bioinformatics project as my master thesis. He helped a lot in the biological interpretation of this project helping me with categorizing tissue specific data of *C. elegans* by providing ground truth signaling pathways. I also acknowledge Dr. Young-Rae Cho for suggesting this project and also attending his course of bioinformatics which was very informational. Also I would like to thank my other committee member, Dr. Cheolho Sim, for offering guidance and support. I thank Baylor University for teaching assistantship awards which made my days in Baylor enjoyable. And finally, I extend thanks to my parents and friends who supported me always.

CHAPTER ONE

Introduction

1.1 Literature Review: Why We Are Interested in Meta-analysis of Signaling Pathways?

Signal transduction is a cell communication that extracellular ligands usually bind to cell surface receptors and cause a relay of reactions such as changes in cellular metabolism, function, or development triggered by the receptor-signal complex (1). It contributes extensively to developmental programs and adaptation responses of different tissues in metazoans. Therefore, proper signal transduction network is essential for the maintenance of tissue functions in different organisms. On the other hand, anomalies in signaling transduction may lead to defective functions of the tissues. In humans, defects in cell signaling can cause various diseases, for example, various types of cancers (2, 3), diabetes (4) or Parkinson's disease (5). Many human diseases are caused by defective cell signaling throughout all tissues, while some are caused by defects in specific tissues. For example, tissue-specific insulin signaling is related to cardiovascular disease (6). Therefore, it's important to profile disease-associated signaling pathways in both genome-wide and tissue-specific levels, to provide the information for the selection of drug targeted to each disease.

One of the challenges in profiling human disease-associated signaling pathways is the incompleteness of biological information in human gene function and protein-protein interaction networks. An alternative approach is to use information of simpler but with more

complete data in model organisms as reference to give an insight to disease-associated signaling pathway profiling in human. The nematode *Caenorhabditis elegans* is a suitable model organism, as its gene functions and signaling pathways are relatively well studied (7). Compared to simpler organisms that widely used in bioinformatics study such as *Saccharomyces cerevisiae*, the nematode *C. elegans* has also a highly differentiated structure comparable to human. Thus, it's more probable to find tissue-specific disease-associated pathway similar to human in *C. elegans* than from yeast (8). By utilizing the genome information, including gene expression profiles, protein-protein interactions and functional information, like gene ontology data, from related databases, we may predict signaling pathways tissue-specifically, especially for those consist of genes that are similar to disease-associated genes in human.

For tissue-specificity, Xiao SJ et al. (9) stated that "the tissue-specific genes are a group of genes whose function and expression are preferred in one or several tissues/cell types". But in order to obtain clearer results, a narrow and explicit definition for tissue-specificity is required. Two simple ideas of tissue-specificity can be applied. The first is to define genes only expressed in a certain tissue as tissue-specific; the second is to define two interacting gene pairs only coexpressed in a certain tissue as tissue-specific. The second approach is a broader concept that includes the first approach of tissue-specificity. For the first approach, candidate pathways should only include tissue-specific genes; for the other, it only covers tissue-specific gene interactions. Using these two criteria, tissue-specific pathways can be predicted.

In order to predict signaling pathways, bioinformatics researchers have developed different methods, mostly based on the statement that the strongest path between source and sink nodes, the first gene and the last gene in a linear pathway, in an interaction network has been often expected to be a functional pathway (10). In recent years, as the knowledge in the area of pathway prediction increased and pathway prediction demands more details. Pathway prediction has already been expanded from genome-wide to both genome-wide and tissue-specific pathway prediction.

For example, Livnat Jerby and Tomer Shlomi (11) used a computational approach to reconstruct a tissue-specific metabolic model and applied in the study of human liver metabolism. They derived the model from a generic model based on network integration with various molecular data sources by machine learning process. Similarly, Tomer Shlomi (12, 13) also studied prediction of human tissue-specific metabolic pathways and successfully proposed a computational method that describes the tissue specificity of human metabolism. For signaling pathway prediction, Tamás Korcsmáros (14) and his co-workers developed an algorithm revealing tissue-specific signal pathway networks and cross-talk patterns between different tissues using human, yeast, and *C. elegans* interaction data. They built up a database contained 8 major signaling pathways, and derived from published information and compiled pathways with semi-automatic searches and uniform curation rules. As long as algorithms were developed, database (for example, TiSGeD, tissue-specific database built by Sheng-jian Xiao (9) etc.) of tissue-specific genes was also built for the study of gene functions in a tissue-specific level. Also tissue-specific functional elements in a certain category were studied in a computational method. For example, Len A. Pennacchio helped predicting tissue-specific enhancers in human genome (15). These studies supported that it is needed to study pathway prediction in a tissue-specific level along with a genome-wide level. Although many researchers have focused on the tissue-specific gene interactions in metabolic pathways, the studies about cell signaling pathways were rarely performed. Therefore, we undertook a bioinformatics approach to define tissue-specific gene interaction network of cell signaling pathways in this research.

1.2. Aims of the Project: What is Planned to Accomplish in This Project?

The aim of this study is to find a bioinformatics method effective for predicting disease-associated tissue-specific pathways in human. As suggested, *C. elegans* was used as a reference model organism to serve this purpose. After finding pathways in *C. elegans* system, data could be matched to human pathways by using homology analysis.

Before the pathway prediction, reference pathway and tissue-specificity data were selected. The reference pathways were obtained from KEGG (Kyoto Encyclopedia of Genes and Genomes) cell signaling database (16). Tissue-specificity data were obtained from the expression profile database in Wormbase (17). To define tissue-specificity, two ideas were tested for this study. The one regarded genes only expressed in a certain tissue as tissue-specific. Therefore, pathways members should come out only from these genes. The other selected gene interactions only existing in a certain tissue as tissue-specific. Therefore, candidate pathways should exclusively include the interactions localized to the same tissue. Upon these conditions, we can propose these two specific aims listed below.

- *Specific Aim 1. Implement the Proposed Pathway Prediction Scoring Model for C.* elegans *Interaction and Expression Data and Validate the Effectiveness of the Model*

As suggested in the introduction, source and sink node functions were restricted to certain functional categories (membrane receptor for source nodes or transcription factor for sink nodes) and a scoring model was considered for functional similarities both ‘between adjacent genes’ and ‘among all the genes’ in a pathway. In order to test the effectiveness of this approach, this model was compared to reference pathways from KEGG database or randomly selected pathways from *C. elegans* interaction map. By comparing mean and variance between reference and random selected pathways, *t*-test, it was tested whether the pathway scoring model is effective or not. If the result showed a great difference in the *p* value of *t*-test between reference and randomly selected pathways, it was concluded that the algorithm is appropriate for pathway prediction.

- *Specific Aim 2. Find Candidate Disease-associated Pathways in Human by Mapping the Predicted C. elegans Pathways Result in Human Network*

By integrating tissue-specificity, homology, and disease-association information, candidate pathways was selected based on the tested pathway prediction algorithm. First, candidate pathways were identified from *C. elegans* gene interaction database by applying the scoring model to different sets of genes such as neuron or gonad specific genes.

Two different levels of tissue-specificity were tested in the study. One selected the gene pairs that only co-expressed in a certain tissue, the other selected the genes only expressed in a certain tissue. For example, if gene A is expressed in neurons and gonad, gene B is expressed only in neurons, the interaction between A and B should exist in the map with the selective criteria of the first level of tissue-specificity, meaning that this interaction was considered as neuron-specific. But for the second level of tissue-specificity, only gene B was selected as neuron-specific. Top scored pathways were selected from the results. Second, by utilizing sequence similarity information between *C. elegans* and human genes, selected pathways were projected to human network. If the members of the pathways did not have human homologs, the pathway was no longer considered. Third, disease-association information was applied to pathways selected from the second step. By queuing to GeneCard disease gene database, pathways with at least one disease-associated gene were selected. After the selection, candidate tissue-specific and disease-associated pathways were collected.

CHAPTER TWO

Preliminary Analysis and Methodology

2.1 Tissue-specific Interaction Map of C. elegans

Genome-wide interaction information was obtained from BioGrid (http://thebiogrid.org/) (18). Repetitive interaction entries were removed from the initial collection. After the selection of gene sets we are interested, we generated interaction maps of these selected genes from the data. In order to show the tissue-specific interaction patterns in *C. elegans*, genes were categorized according to their tissue expression specificity; genes annotated with expression profile are categorized into 10 types of tissues such as coelomocytes, excretory cells, germ cells, gonad, hypodermis, intestine, muscle, neurons, pharynx and seam cells (19).

Interactomes (20) were generated for each of the tissues in three different approaches:

- Approach 1: Isolate an interactome of all genes expressed in a certain tissue and show all gene interactions of the selected genes;
- Approach 2: Isolate an interactome of gene pairs that coexpressed in a certain tissue and show interactions among the selected gene pairs;
- Approach 3: Isolate an interactome of the genes specifically expressed in a certain tissue and show their interactions.

The first approach was considered as a reference not tissue-specific because it includes all the genes that expressed in certain tissues, may not be exclusive. The second and third approaches were considered as two different levels of tissue-specificity. It is apparent that the interactome from the third approach is a sub-map of the second interactome, meaning that the second approach is a broader concept of tissue-specificity (the first level of tissue-specificity) compared to the third approach (the second level of tissue-specificity). Initial maps from neurons and gonad were drawn in Figures 2.1 to 2.6.

Subsequent Analyses were Focused on these Two Tissues because;

1. Neurons and gonad are the two important tissues in *C. elegans* that have comparable tissue/organs in human.
2. These comparable tissues/organs are associated with various important diseases. For example, many mental diseases including Parkinson’s and Alzheimer's diseases are related to neurons and various diseases of reproductive system (21).

Using the Cytoscape 2.8.3 program (22), an interactome was drawn based on *C. elegans* gene interactions present in neurons. The map was displayed in a recommended cluster format. In this approach, total 384 of nodes/genes constitutes 1284 of edges/pairs of gene interactions.

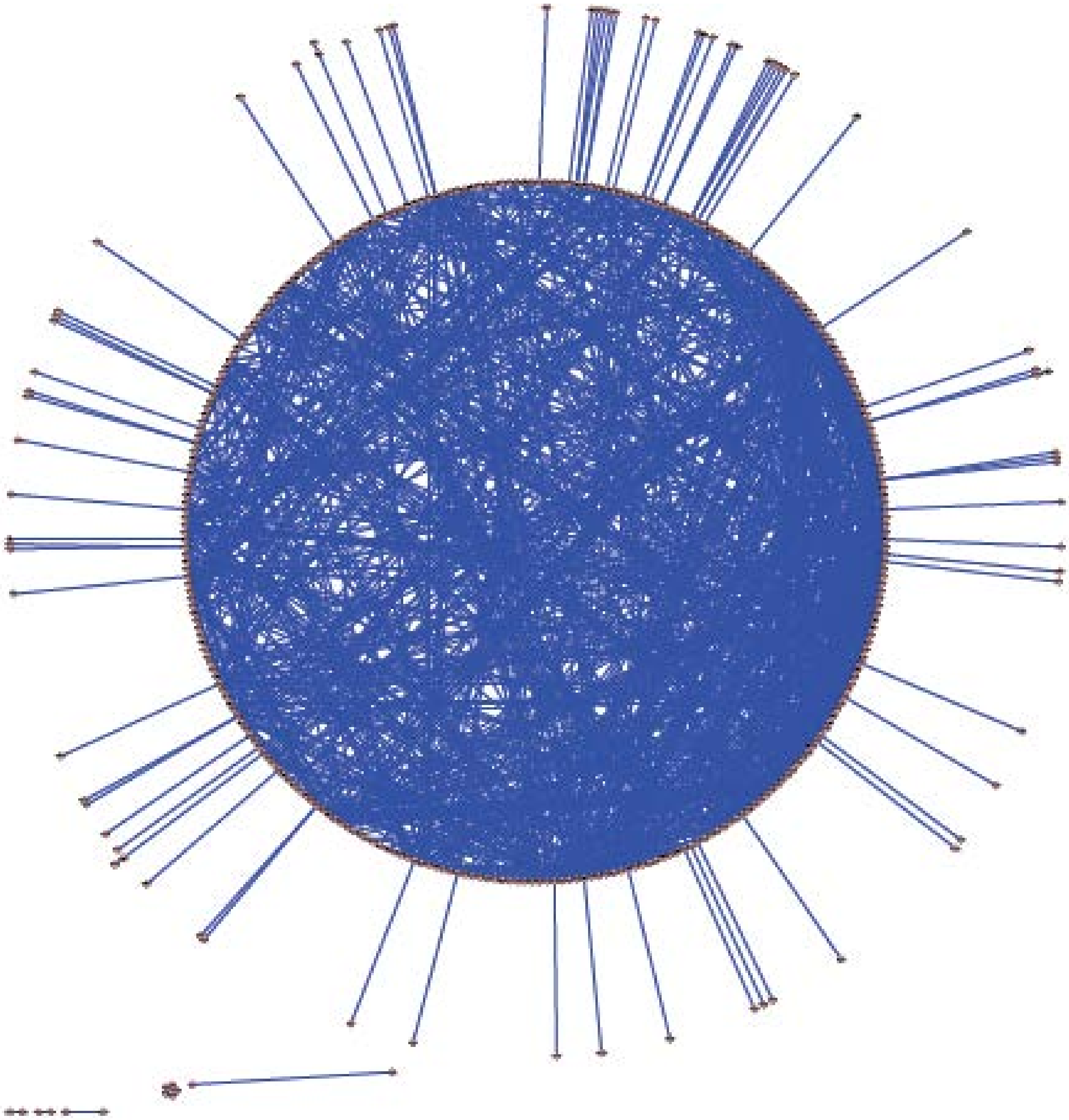


Figure 2.1. Interaction map showing gene interactions between all genes expressed in neurons, Approach 1.

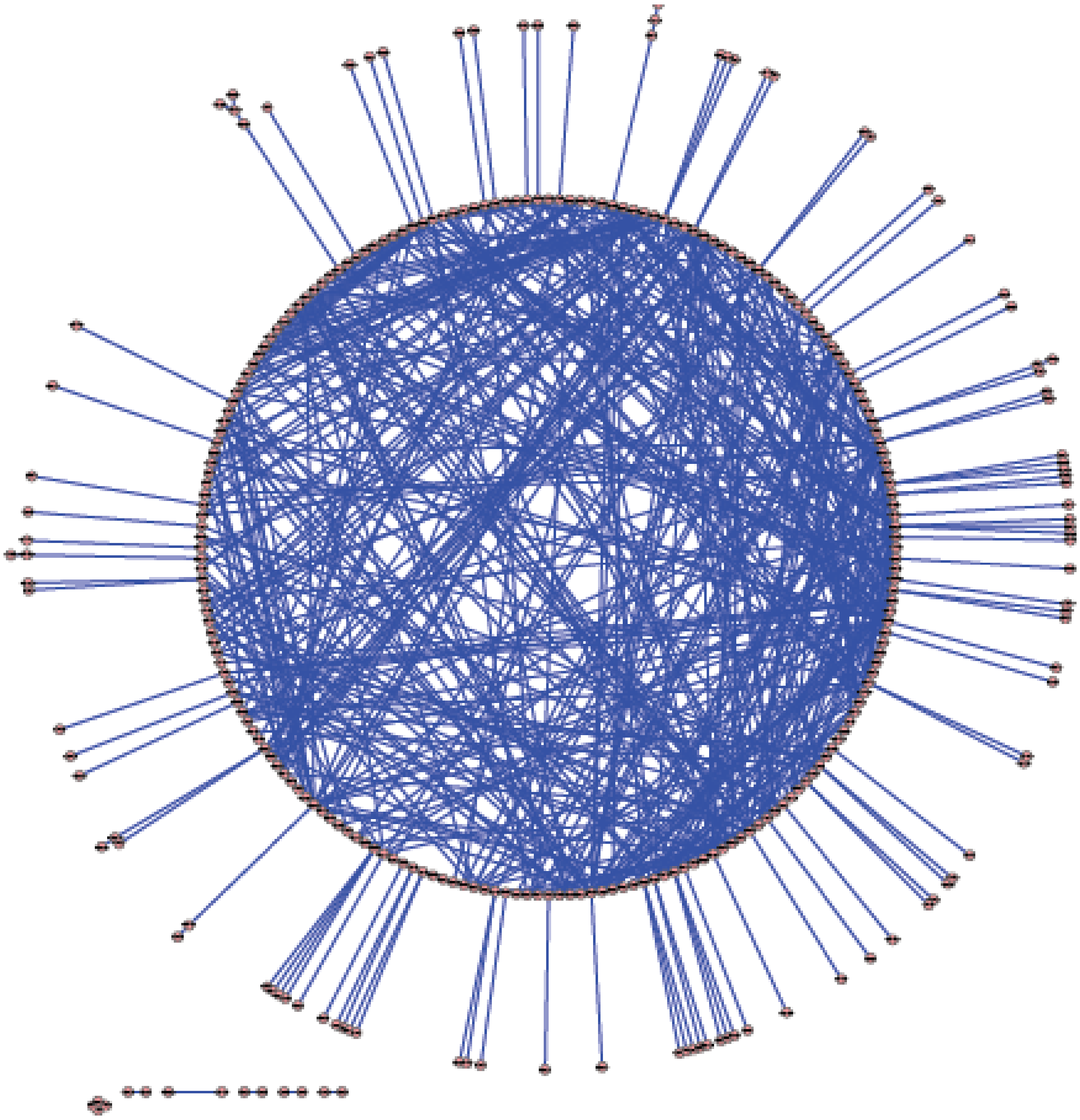


Figure 2.2. Interaction map showing the gene pairs co-expressed in neurons, Approach 2.

To improve the result from Figure 2.1., selection conditions were changed to isolate gene pairs only co-expressed in the neurons, both genes are co-expressed in neuron but may not be exclusive. This approached resulted in 319 of nodes/gene, 686 edges/interactions. Comparing to the result from Approach 1 in Fig 2.1., the number of interaction was decreased by 47%.

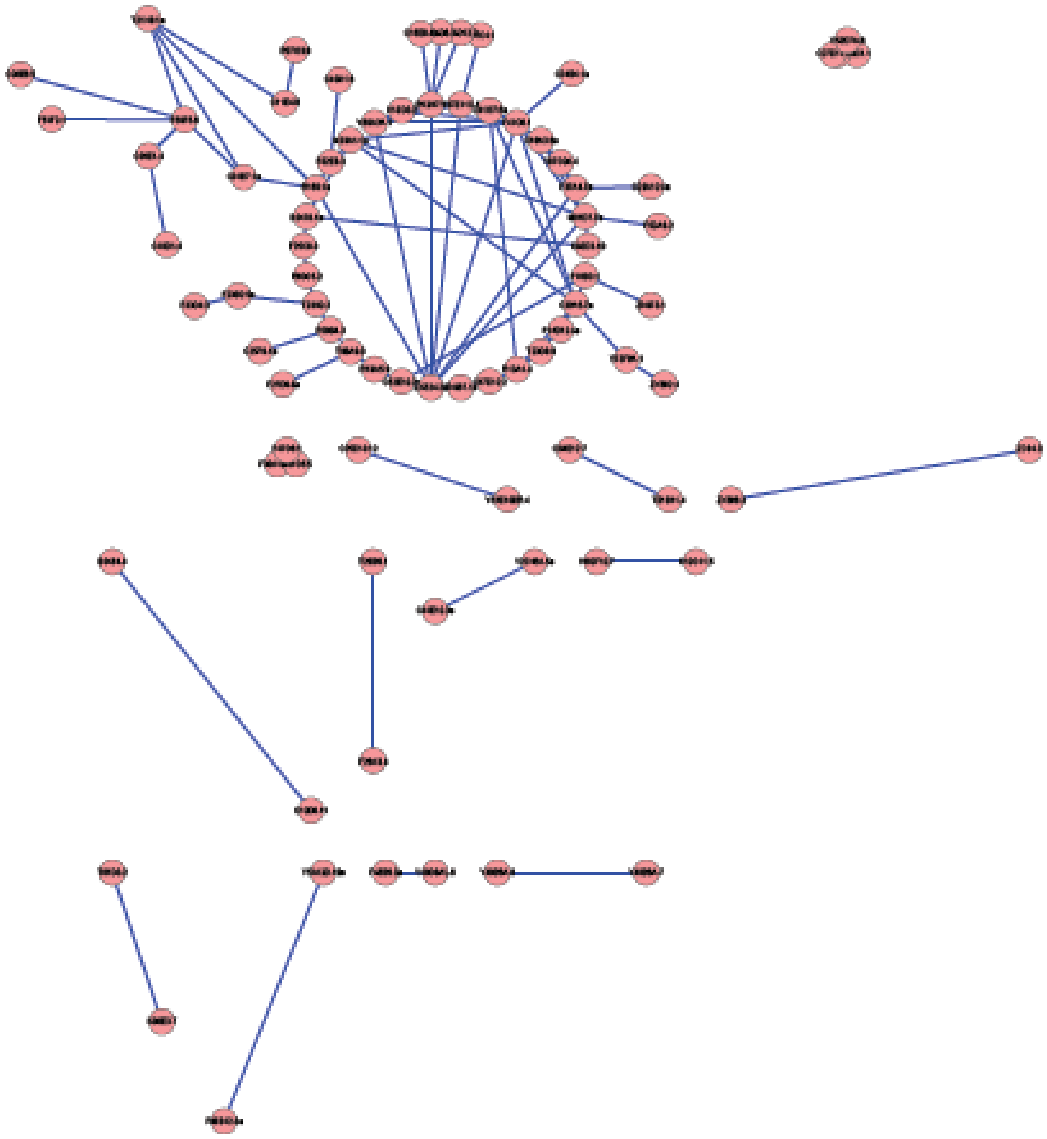


Figure 2.3. Interaction map showing interactions between genes expressed exclusively in neurons, Approach 3.

In this approach, selection conditions were further narrowed down to show interactions between genes that only expressed in neurons exclusively. This approached resulted in 83 of nodes/genes, 92 edges/gene interactions. Comparing to the result from Approach 1 in Fig 2.1., the number of interaction was decreased by 93%.

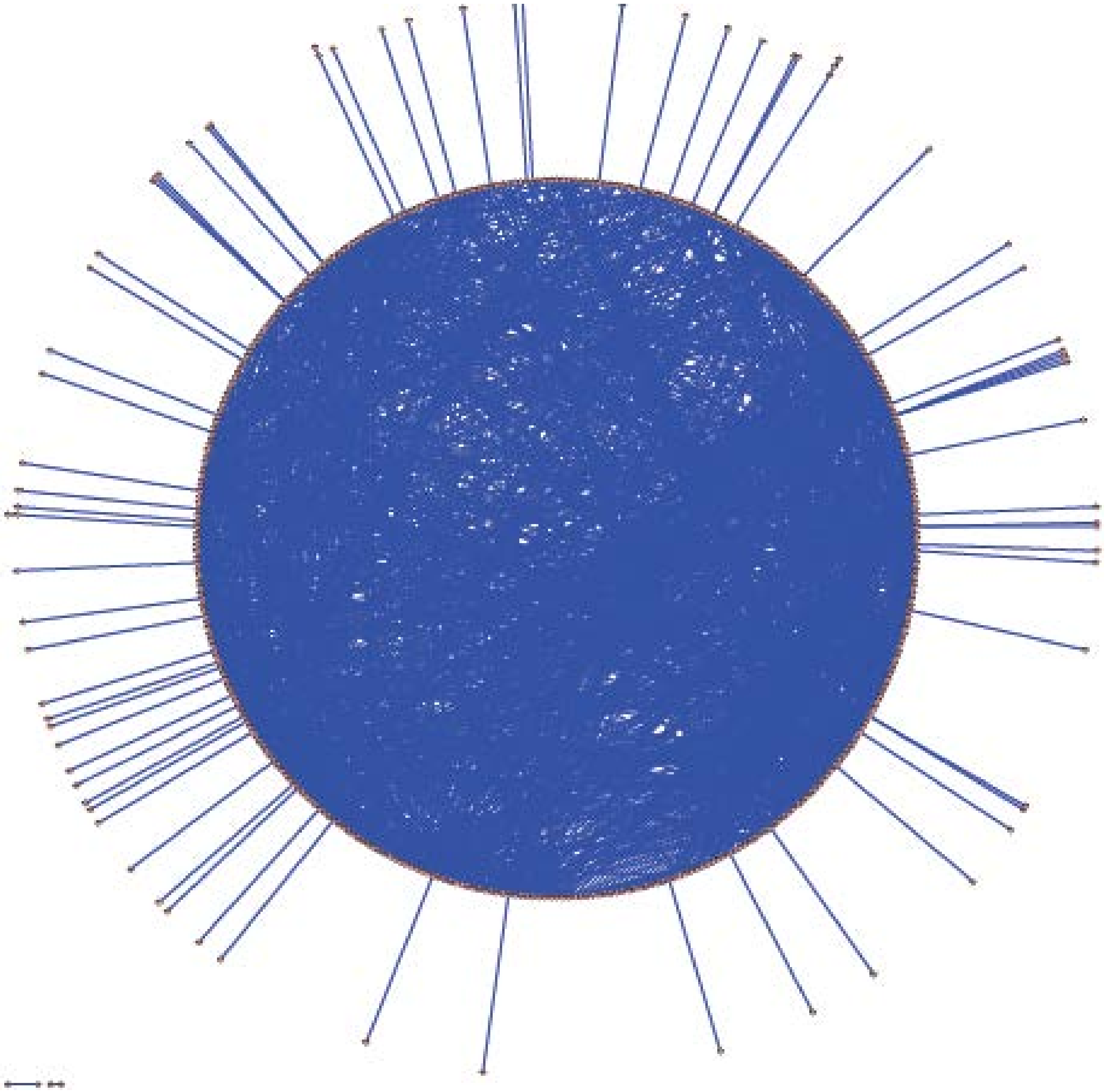


Figure 2.4. Interaction map of genes expressed in gonad showing interactions between all the genes expressed in gonad, Approach 1.

Using the Cytoscape 2.8.3 program (22), interaction map was drawn based on C. *elegans* interactions between genes present in gonad. The map was displayed in a recommended cluster format. In this approach, total 453 of nodes/genes constitutes 2029 of edges/pairs of gene interactions.

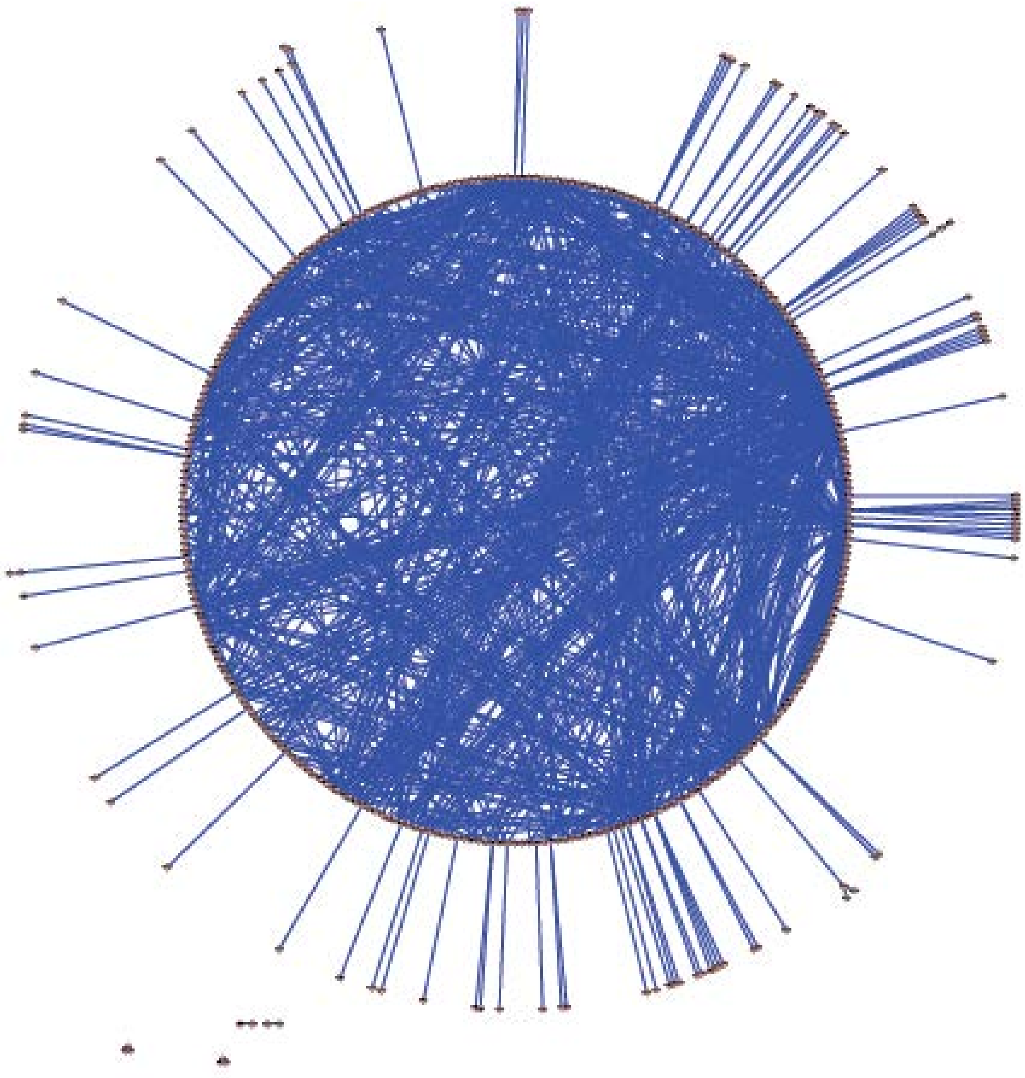


Figure 2.5. Interaction map showing interactions of the gene pairs only co-expressed in gonad, Approach 2.

To improve the result from Figure 2.4., section conditions were changed to isolate gene pairs co-expressed in the gonad. Both genes are co-expressed in neuron but may not be exclusive. This approached resulted in 397 of nodes/gene, 1227 edges/interactions. Comparing to the result from Approach 1 in Fig 2.4., the number of interaction was decreased 40%.

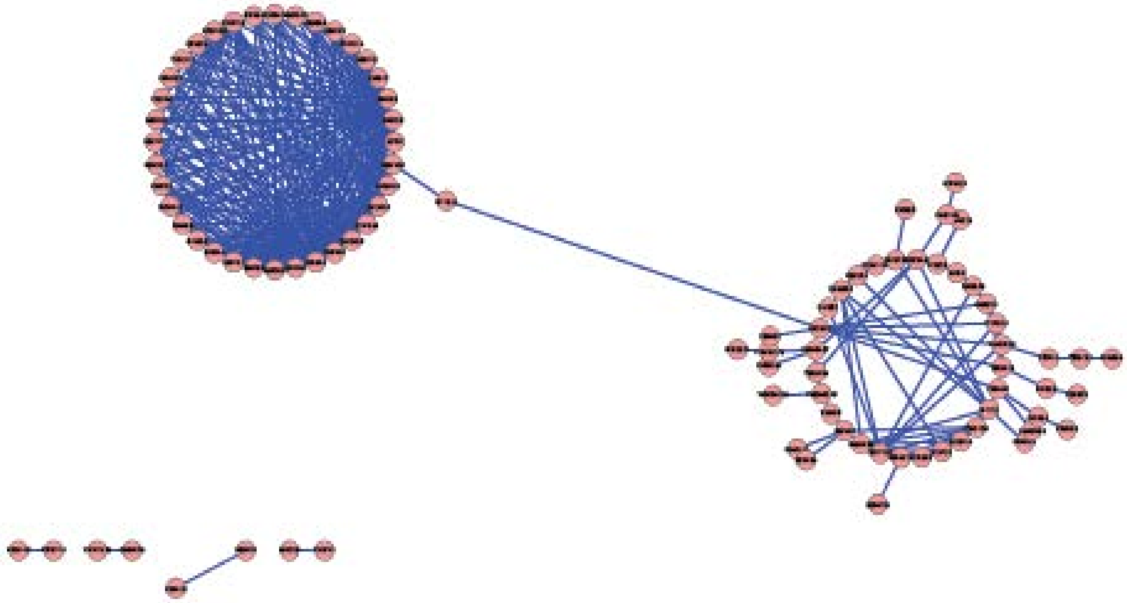


Figure 2.6. Interaction maps of genes expressed in gonad showing interactions between genes only expressed in gonad, Approach 3.

In this approach, selection conditions were further narrowed down to show interactions between genes that only expressed in gonad exclusively. This approached resulted in 94 of nodes/genes, 288 edges/gene interactions. Comparing to the result from Approach 1 in Fig 2.4., the number of interaction was decreased 86%.

Figures 2.3 and 2.6 demonstrated that genes may function together, for approach 3 of tissue-specificity. Compared among three different maps, we can clearly see that the gene interactions are narrowed down when more rigorous standard of tissue-specificity was applied, but still they are connected to each other when Approach 3 applied, which means there may be some pathways in the map. Because of that, it's meaningful to approach pathway prediction method tissue-specifically so that the inner relationship of tissue-specific genes can be revealed.

2. 2. Reference Pathways from KEGG Signaling Database

In order to test pathways prediction method is proper or not, the reference pathways are needed to provide information as a positive control. Reference pathways are selected from KEGG (Kyoto Encyclopedia of Genes and Genomes) signaling pathways database (http://www.genome.jp/kegg/pathway.html), evaluated with information in published articles. Signaling pathways are listed in five different categories, including:

MAP Kinase Pathways

The MAP kinase pathway includes many proteins, including MAPK (Mitogen-activated protein kinases), which communicate by adding phosphate groups to a neighbor protein, which acts as an on/off switch.

Calcium Pathways

Calcium signaling is a common signaling mechanism that mainly involves calcium ion in signaling transduction, related to various basic cellular activities. Calcium can act in signal transduction after influx resulting from activation of ion-channels or as a second messenger.

ErbB Pathways

The ErbB pathways mainly include EGFR (epidermal growth factor), regulating diverse biologic responses, including proliferation, differentiation, cell motility, and survival.

Wnt Pathways

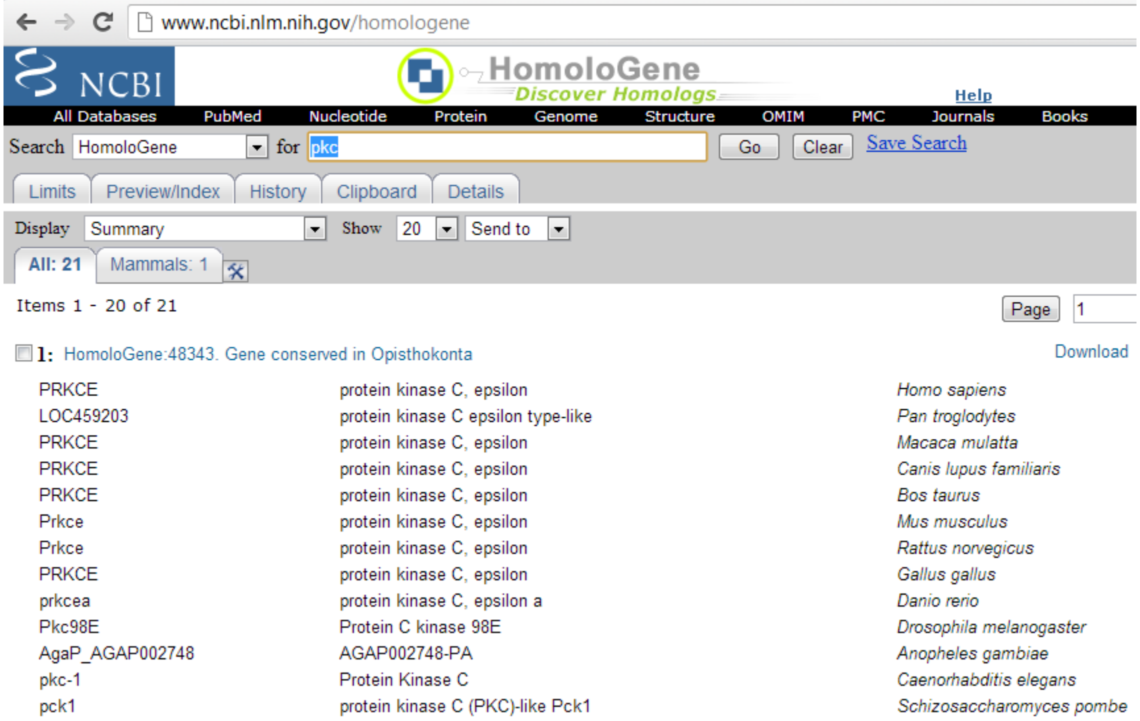
The Wnt signaling pathway is a network of proteins, mainly include Wnt etc, that passes signals from receptors on the surface of the cell to the cell's nucleus where the signaling cascade leads to the expression of target genes. It controls various cellular activities such as growth and proliferation.

mTOR Pathways

The mTOR pathways mainly include mTOR (mammalian target of rapamycin), a serine/threonine protein kinase. mTOR pathways regulate cellular activity including growth, proliferation, motility, and survival. Full lists of the reference pathways are shown in Appendix.

2.3. Homology Information to Map the Predicted Pathways of C. elegans *to Humasn*

After generating the tissue-specific pathways in *C. elegans* system, we went further to map these predicted pathways to human. Upon usage of homology relationship database of *C. elegans* and human, pathways predicted in *C. elegans* could be mapped to human. Homology information can be generated by homologene from NCBI (http://www.ncbi.nlm.nih.gov/homologene) (23). An example of *pkc* homology information is shown below. PKC (protein kinase C) plays important roles in different cascades of signaling transduction.



Search HomoloGene for pkc

Display Summary Show 20 Send to

All: 21 Mammals: 1

Items 1 - 20 of 21

Page 1

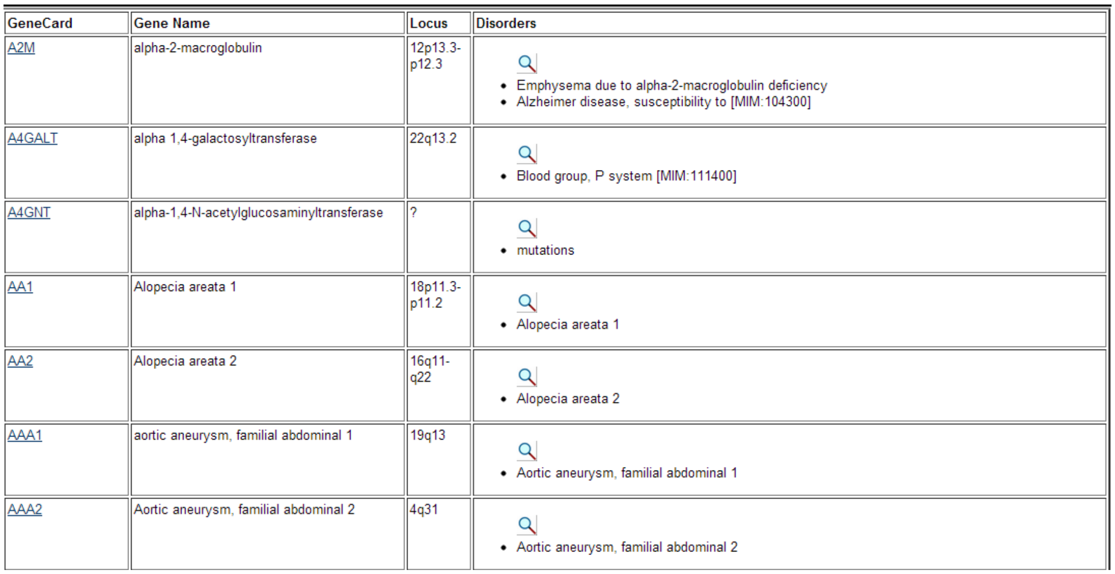
☐ 1: HomoloGene:48343. Gene conserved in Opisthokonta

PRKCE	protein kinase C, epsilon	*Homo sapiens*
LOC459203	protein kinase C epsilon type-like	*Pan troglodytes*
PRKCE	protein kinase C, epsilon	*Macaca mulatta*
PRKCE	protein kinase C, epsilon	*Canis lupus familiaris*
PRKCE	protein kinase C, epsilon	*Bos taurus*
Prkce	protein kinase C, epsilon	*Mus musculus*
Prkce	protein kinase C, epsilon	*Rattus norvegicus*
PRKCE	protein kinase C, epsilon	*Gallus gallus*
prkcea	protein kinase C, epsilon a	*Danio rerio*
Pkc98E	Protein C kinase 98E	*Drosophila melanogaster*
AgaP_AGAP002748	AGAP002748-PA	*Anopheles gambiae*
pkc-1	Protein Kinase C	*Caenorhabditis elegans*
pck1	protein kinase C (PKC)-like Pck1	*Schizosaccharomyces pombe*

Figure 2.7. Homology information for PKC (Protein Kinase C) in homologene database

2.4. Disease Association Information

Disease-associated gene database is available at GeneCard (http://www.genecards.org) (24): 6237 diseases genes in GeneCard database as the date of March 29th. An example list is displayed below. We can see disease/disorder information and also the gene names and locus of A2M, A3GALT, A4GNT, AA1 and AA2.



GeneCard	Gene Name	Locus	Disorders
A2M	alpha-2-macroglobulin	12p13.3-p12.3	🔍 • Emphysema due to alpha-2-macroglobulin deficiency • Alzheimer disease, susceptibility to [MIM:104300]
A4GALT	alpha 1,4-galactosyltransferase	22q13.2	🔍 • Blood group, P system [MIM:111400]
A4GNT	alpha-1,4-N-acetylglucosaminyltransferase	?	🔍 • mutations
AA1	Alopecia areata 1	18p11.3-p11.2	🔍 • Alopecia areata 1
AA2	Alopecia areata 2	16q11-q22	🔍 • Alopecia areata 2
AAA1	aortic aneurysm, familial abdominal 1	19q13	🔍 • Aortic aneurysm, familial abdominal 1
AAA2	Aortic aneurysm, familial abdominal 2	4q31	🔍 • Aortic aneurysm, familial abdominal 2

Figure 2.8. Example list of disease-associated gene in GeneCard database

By identifying a pathway member in the disease-associated genes database, tissue-specific pathway can be analyzed in disease-associatedness. If one candidate pathway contains one of these disease-associated genes, this pathway could be considered as a tissue-specific disease-associated pathway.

2.5. Pathway Prediction Algorithm

2.5.1. Basic Idea of Algorithm

For the prediction of signaling pathways, most previous measures just chose randomly from the whole genome of studied organism as candidate members, which give many false positives because only certain genes have high possibility serving as a source/sink node, or function in a signaling pathway. So setting up two nodes per pathway would be useful to make an appropriate algorithm with higher accuracy, also this approach reduced the runtime of the algorithm to make this algorithm viable for large set of data, because the number of tested pathways was reduced. Previous studies (10) and pathway reference pathways listed in Appendix suggested that source and sink nodes should have the molecular function of receptor ligand and transcription factor, respectively. By filtering genes based on their molecular function, the candidate pathways would be narrowed down for both running speed and accuracy.

Next is to score candidate pathways. Previous studies stated that all members in a candidate pathway should have similar attributes, which varied using different scoring methods, like utilizing expression level or gene ontology (25). But this might be incomplete because a pathway could have multiple functions and each node could have different functions. Better way is to score the pathway chain pair by pair to look for the function of these gene pairs and then combine together to obtain the functional score of the whole pathways.

For pathway selection algorithm, in addition, simple exhausted search (test every possible combination of pathway members) could be sufficient if the data size is small. A heuristic search, which means strategies using readily accessible, though loosely applicable, information to control problem solving in human beings and machines (26), should be applied if needed. At last, we needed to map these results from *C. elegans* to human disease related genes, so disease-associated and homology data is required to be utilized to map *C. elegans* result to human result.

2.5.2. Software, Program Language and Database Resources

Python 3.2.3 is used throughout the research for data output and analysis as main programming language. Cytoscape 2.8.3 (22) is used for generating interaction maps from interaction data. As stated in previous sections, interaction data of *C. elegans* genes is obtained from BioGrid; homology information is obtained from NCBI homologenes database; disease association information is obtained from GeneCard disease gene database. (18, 23, 24)

2.5.3. Source and Sink Nodes Restriction Based on Gene Ontology Terms

As it is suggested, sink nodes can be restricted according to gene ontology terms. In this case, we can filter genes with molecular function of transmembrane signaling receptor activity (GO: 0004888) for source nodes candidates and genes with molecular function of nucleic acid binding transcription factor activity (GO: 0001071) for sink nodes candidates. From the *C. elegans* gene ontology, using these criteria, candidate genes were filtered out.

2.5.4. Pathway Scoring Method

Scoring was relied on ontology data. Genes without ontology information were ruled out. Basic idea of this scoring model is to test the relationship between each pair of adjacent genes, instead of only considering the overall relationship of all members in a pathway. For example, if every pairs of adjacent genes in a candidate pathway were functionally close to each other, the score of this candidate pathway would be high, in other word, it was considered as a highly reliable candidate pathway.

Scoring for a length of 4 candidate pathway (consist of 5 genes) of A-B-C-D-E =

$\frac{(Pa\cap Pb)\cup (Pb\cap Pc)\cup(Pc\cap Pd)\cup(Pd\cap Pe)}{Pa\cup Pb\cup Pc\cup Pd\cup Pe}$ P means the set of ontology terms assigned to a certain gene. Both numerator and denominator took the sum of annotated times for all genes within the set. Because the numerator set is always a subset of denominator set, the range of the score will be from 0 to 1.

2.5.5. Evaluation of Pathway Scoring Method

In order to test the feasibility of the pathway scoring model, it was compared to both reference and randomly selected pathways. By comparing scoring distribution of reference and randomly selected pathways, it was tested that whether pathway scoring model is feasible or not. If the result showed a significant difference between reference and randomly selected pathways, it can be concluded that this algorithm is appropriate for pathway prediction. If the result showed that difference between them is insignificant, this algorithm is proved to be not good for pathway prediction.

Randomly selected pathways are selected based on the criterion that source and sink node gene should be assigned to molecular function of receptor and transcription factors, respectively. Pathway length is limited to a range of 2 to 8 according to ground truth pathways listed in Appendix. For each different pathway length, 5 randomly selected pathways were chosen with the criterion mentioned above. Scoring model mentioned above was compared to reference pathways (total 24) and randomly selected pathways (total 35).

Results are shown in the Figure 2.9. From the results, further statistical analysis is shown in Table 2.1.

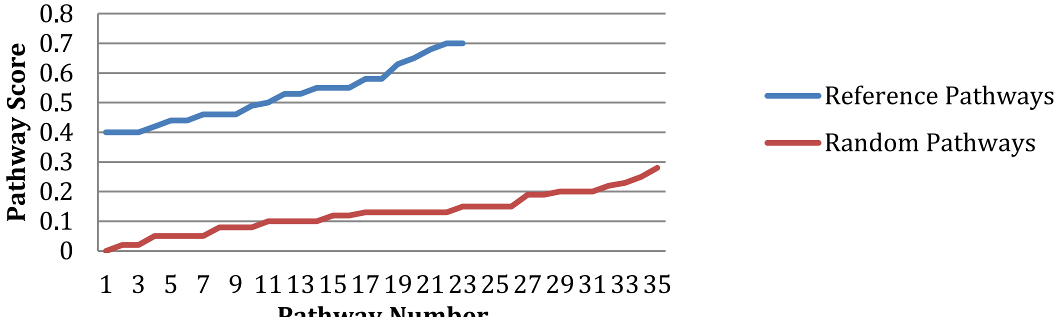


Figure 2.9. Example list of disease-associated gene in GeneCard database.

Table 2.1. Statistical analysis of difference between score of reference pathways and randomly selected pathways.

Pathway	Average Score	Standard Deviation
Reference Pathway	0.526086957	0.526086957
Randomly selected pathways	0.127429	0.127429
P-value of T-test	<0.0001	

Using the proposed algorithm, the above results showed that the score of reference pathways and randomly selected pathways showed a great difference between each other. This supported that the proposed algorithm is appropriate for pathway prediction using *C. elegans* gene interaction data.

2.6. Tissue-specific Disease Associated Pathways Prediction

Candidate pathways were selected based on the tested pathway prediction algorithm, by utilizing tissue-specificity, homology, gene ontology and disease-association information. Specific steps are laid out in the list below.

1. Neuron and gonad tissue-specific genes are selected based on the two criteria mentioned in the previous sections.

2. In order to locate *C. elegans* pathways in human, genes with human homologs were identified in the tissue-specific genes generated from #1.

3. In order to apply the scoring model to these genes, genes correlated to biological process of signaling are selected to meet the criteria that all the genes should involve in a signaling process in a candidate signaling pathway.

4. The scoring model and a certain heuristic manner, which means applying certain searching strategy to reduce the amounts to be calculated, to speed up the process are applied, the top 20 scored tissue-specific pathways are generated. Also all the pathways meet the minimum criteria of the heuristic manner are considered as tissue-specific pathways, the reliability is reflected in the score of the pathways.

5. By queuing to important disease gene list, pathway contains the selected gene associated with certain important diseases will be selected from the tissue-specific pathways generated in step 4, and afterwards candidate tissue-specific pathways associated with that certain disease can be selected.

CHAPTER THREE

Results

3.1. Tissue-specific Pathway Detection

3.1.1. Statistical analysis of *C. elegans* neuron and gonad specific genes

In order to directly show the percentage of genes we can test in this research, tissue-specific genes in gonad are referred to NCBI and GeneCard as suggested in previous sections, to see which genes have human homologs. Also the percentage of these genes compared to the total number of *C. elegans* genes with known expression pattern is also listed. The total number of *C. elegans* genes with known expression is 11465, obtained from Wormbase, Statistical analysis results are shown below in Table 3.1.

Although there were more than ten thousand genes with known expression, tissue-specific ones in neurons and gonad only included 3%-4% in a broad concept of tissue-specificity and less than 1% in a narrow concept of tissue-specificity, meaning that tissue-specific genes were not widespread in the whole *C. elegans* genome. In terms of homology, we found more than a half of the *C. elegans* gonad and neuron specific genes are conserved in human. As expected, these two tissues are comparable to human tissues, neurons appeared to have a more conserved genes compared to that of gonad.

Table 3.1. Statistical Analysis of *C. elegans* neuron and gonad specific genes related to homology and disease association information

Gene	Total Number of genes	Percentage out of total number of *C. elegans* Genes with known expression pattern	Number of genes have human homologs	Percentage out of total number of *C. elegans* Genes with known expression pattern
Neuron specific genes in approach 2	318	2.77%	207	1.81%
Neuron specific genes in approach 3	83	0.72%	51	0.44%
Gonad specific genes in approach 2	397	3.46%	203	1.77%
Gonad specific genes in approach 3	89	0.78%	23	0.20%

3.1.2. Statistical Analysis for Ontology Terms of Tissue-specific Genes in C. elegans

As stated in the methodology chapter of this thesis, selected pathways should have all the members annotated to the ontology of biological process, signaling (GO:0023052), with source nodes annotated to ontology of molecular function, transmembrane signaling receptor activity (GO: 0004888) and sink nodes annotated to ontology of molecular function, nucleic acid binding transcription factor activity (GO:0001071). Table 3.2 shows the number of the tissue-specific genes with human homologs in the two tissues, and how many genes are annotated to the ontology terms. Figures 3.1 - 3.4 show the interactions between these tissue-specific genes involved in signaling. Membrane receptors and transcription factors, candidate source and sink nodes, are also shown in green and yellow colors in the maps, respectively.

Table 3.2 Statistical Analysis for ontology terms of tissue-specific genes in *C. elegans*

Gene	Total number of genes with human homologs	Number of genes annotated to signaling	Number of genes annotated to membrane receptors	Number of genes annotated to transcription factors
Neuron-specific genes in approach 2	207	95	6	5
Neuron-specific genes in approach 3	51	51	2	0
Gonad-specific genes in approach 2	203	92	4	6
Gonad-specific genes in approach 3	23	17	0	0

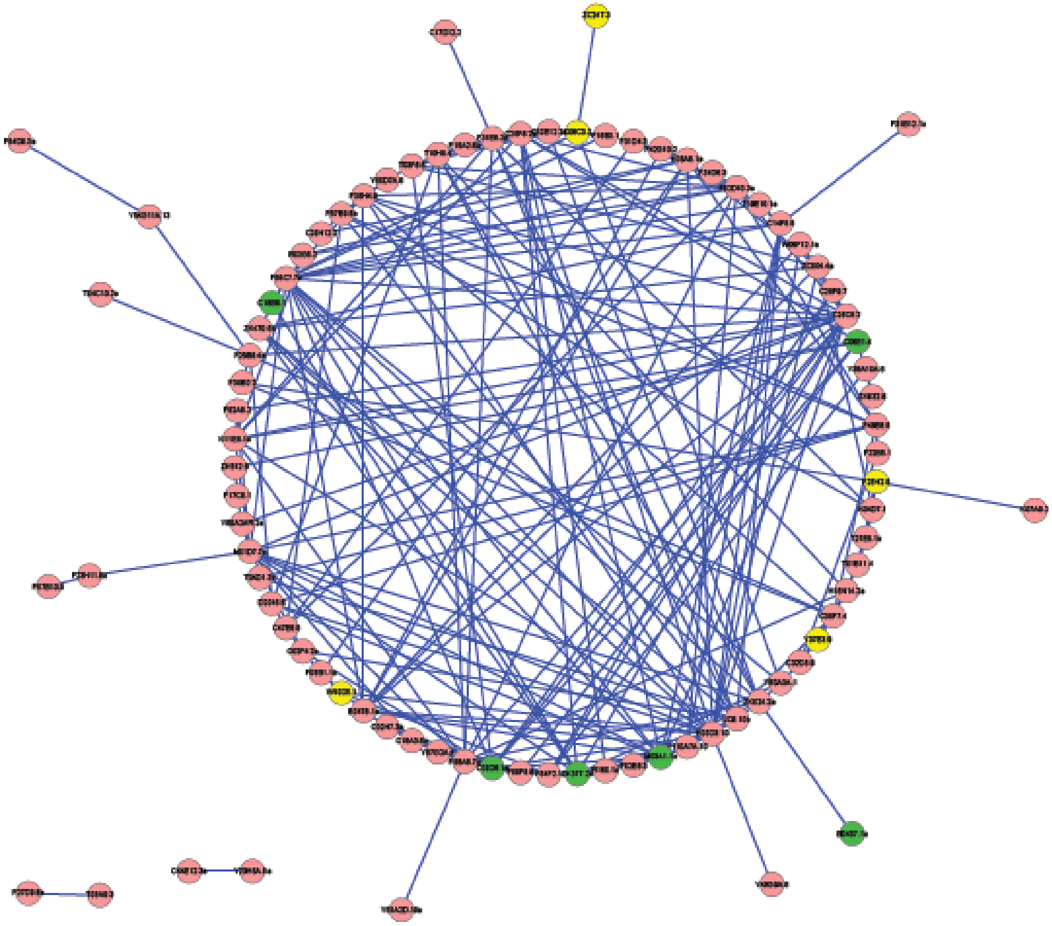


Figure 3.1. Interaction map showing interactions of the gene pairs with human homologs and annotated to signaling that coexpressed only in neurons, it's a sub-network of Figure 2.2.

This interaction map shows interaction of the gene pairs with human homologs and annotated to signaling that coexpressed only in neurons, containing 85 nodes/genes and 308 edges/interactions. Among the 85 genes, 6 can serve as candidate source nodes and 4 can serve as candidate sink nodes.

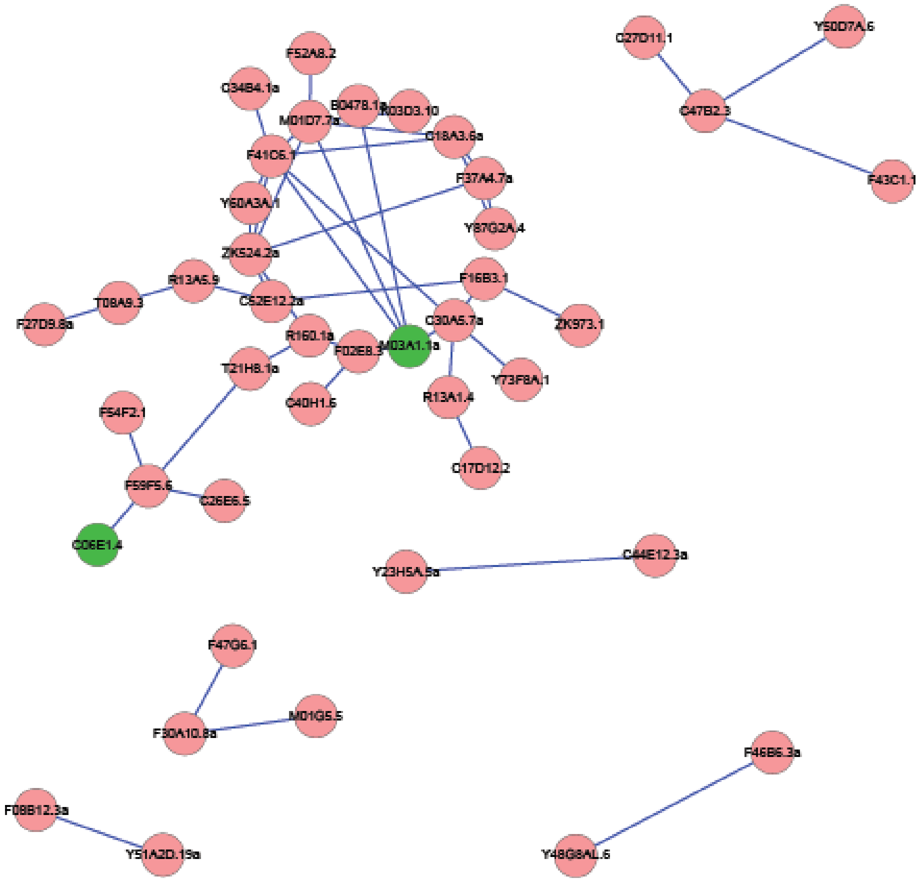


Figure 3.2. Interaction map showing interactions between genes with human homologs and annotated to signaling that only expressed in neurons, it's a sub-network of Figure 2.3.

This interaction map shows interactions between genes with human homologs and annotated to signaling that only expressed in neurons, containing 43 nodes/genes and 47 edges/interactions. Among the 43 genes, 2 can serve as candidate source nodes and 0 can serve as candidate sink nodes.

This interaction map shows interactions of the gene pairs with human homologs and annotated to signaling that only coexpressed in gonad, containing 43 nodes/genes and 47 edges/interactions. Among the 43 genes, 2 can serve as candidate source nodes and 0 can serve as candidate sink nodes.

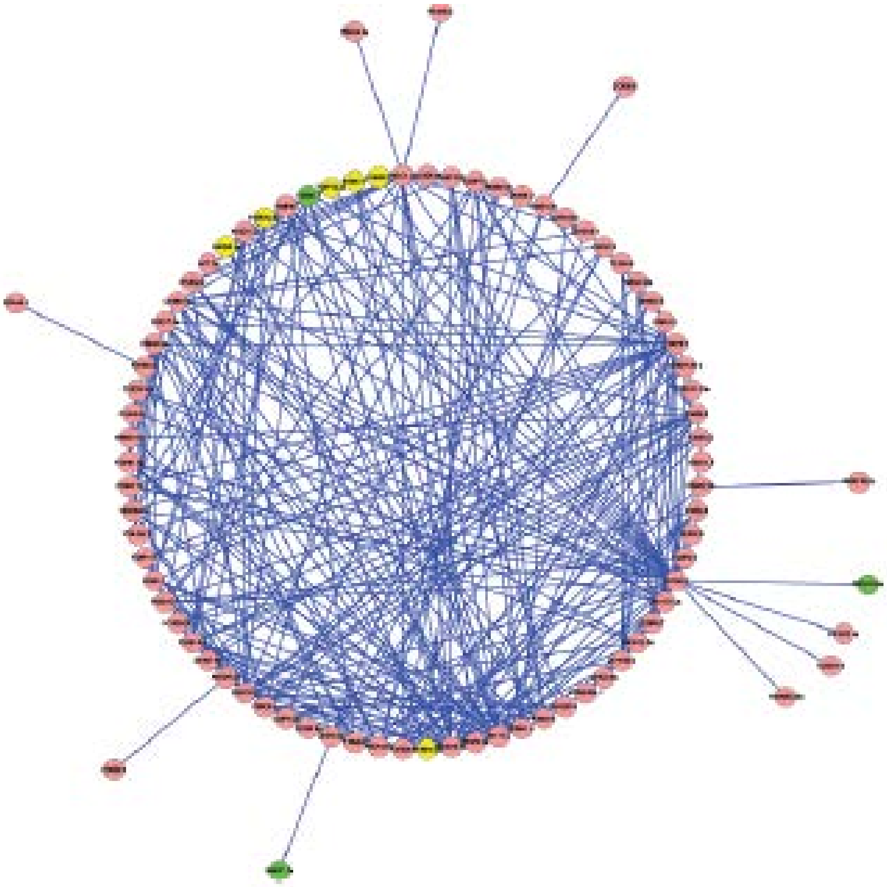


Figure 3.3. Interaction map showing interactions of the gene pairs with human homologs and annotated to signaling that only coexpressed in gonad, it's a sub-network of Figure 2.5.

This interaction map shows interactions of the gene pairs with human homologs and annotated to signaling that only coexpressed in gonad, containing 43 nodes/genes and 4 edges/interactions. Among the 43 genes, 2 can serve as candidate source nodes and 0 can serve as candidate sink nodes.

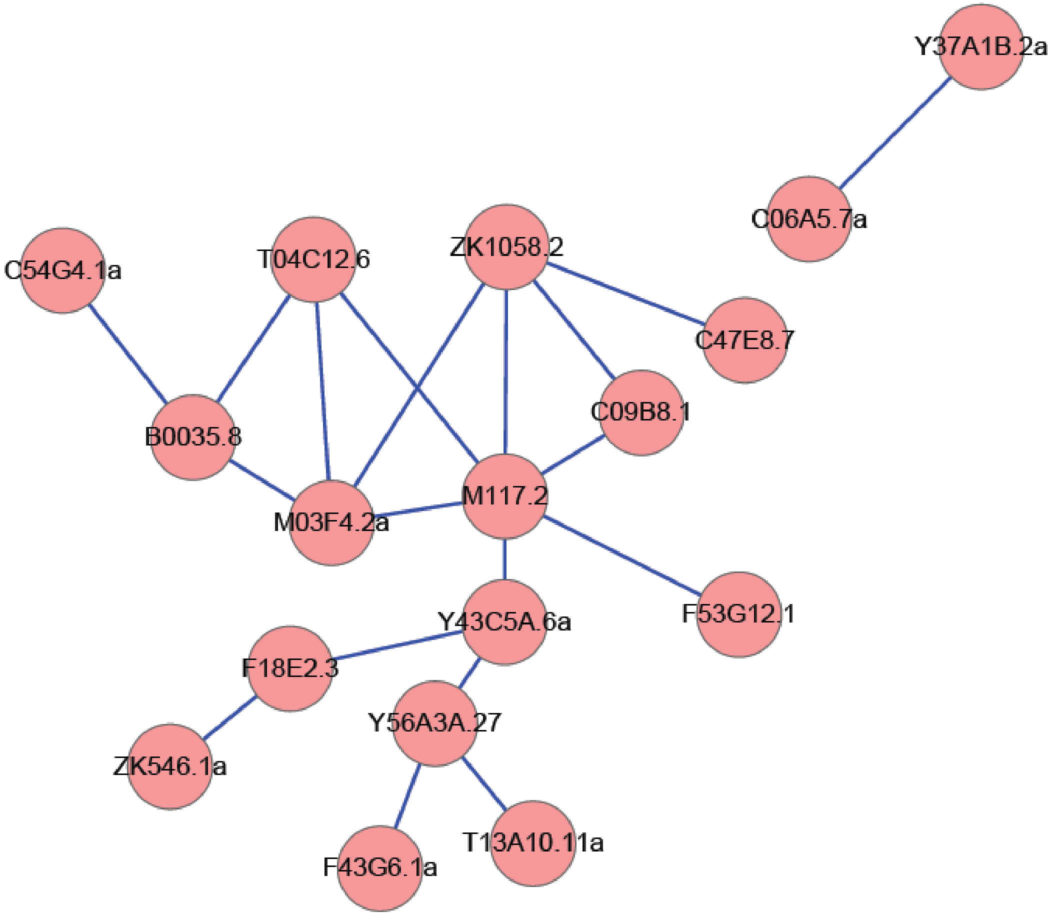


Figure 3.4. Interaction map showing interactions between genes with human homologs and annotated to signaling that only coexpressed in gonad, it's a sub-network of Figure 2.6.

This interaction map shows interactions between genes with human homologs and annotated to signaling that only coexpressed in gonad, containing 17 nodes/genes and 19 edges/interactions. Among the 17 genes, 0 can serve as candidate source nodes and 0 can serve as candidate sink nodes. Shown in approach 3, no candidate sink node exists for both of the tissues. The reason is not known, but one possible reason for this is that upon evolution, those highly tissue-specific genes have important function lose their tracks in human (have human homologs) but those more general expressed genes will keep their tracks in human.

Also it may be due to lack of complete information for these genes. Upon this result, later analysis will only focus on approach 2 tissue-specificity since the decent numbers of candidate source and sink nodes exist.

3.1.3. Results of Prediction for Tissue-specific Pathways

From figure 3.2 and 3.4, we failed to find proper sink or source nodes for the approach 3 tissue-specificity, so we analyzed the approach 2 tissue-specificity only. As we can see in Figure 3.1 and 3.3, the interaction map is still complex, so a heuristic manner is implemented. Basic idea is to start from candidate source nodes and search for their neighbor interacting genes, and the e. And afterwards, this process was repeated, neighbors' neighbors were searched for and the score from candidate source nodes to these nodes was calculated and filtered with the same threshold. This process will be repeated until it reaches a candidate sink node, and in this case the path was the output as a candidate tissue-specific pathway; or if it reached a node that already appeared previously, this path search was terminated; of it already contained 10 members (path length of 9), this path search was terminated.

Threshold score is tested from 0.28 (maximum score of tested randomize pathways in section 3.1.4) to 0.4 (minimum score of tested reference pathways in section 3.1.4). As it can be seen in Figure 3.5, 3.2 yielded the best balance of runtime and generated result. So here we decided to use 3.2 as a threshold score. As stated, all pathway members should be annotated to GO: 0023052 (signaling, biological process), source and sink nodes are selected as a previous stated method. Top 20 pathways were selected for both tissues; pathway prediction result is shown in Table 3.3 and 3.4.

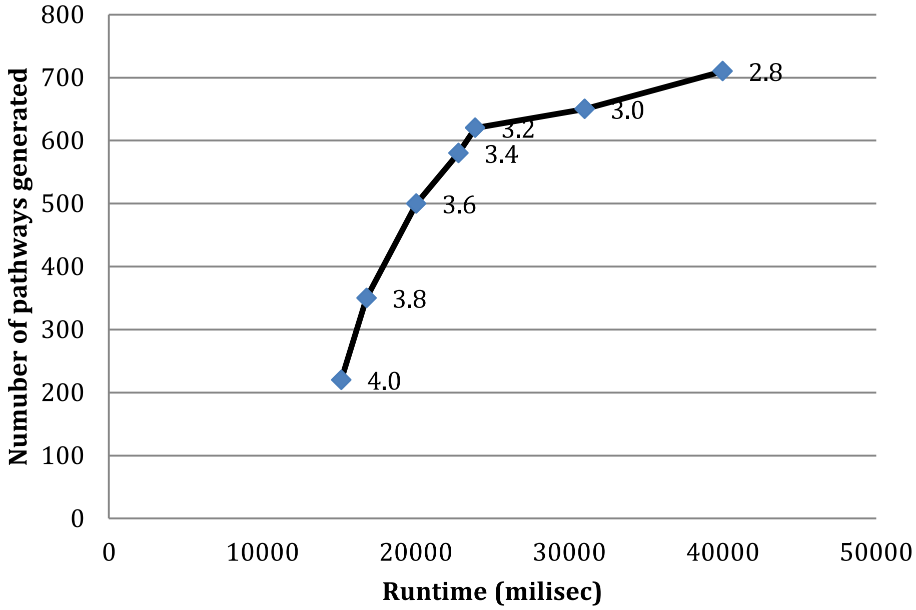


Figure 3.5. The test result of threshold score set in the algorithm.

Among the top-20 tissue-specific pathways in neurons, 13 out of 20 are in length lower than or equal to 3, which means they contain 3 or 4 members. Only 3 out of 20 are in length longer than or equal to 5. Similarly among top-20 tissue-specific pathways in gonad, 9 out of 20 are in length lower than 3, and only 2 out of 20 are in length longer than or equal to 5. For all the pathways meets the minimum threshold score criteria, 171 out of 262 candidate neuron-specific pathways are in length lower than or equal to 3, 284 out of 357 candidate gonad-specific pathways are in such length. These results show that candidate pathways predicted using proposed scoring model are biased in shorter length. On one hand, this may show the actual patterns of the signaling pathways; on the other hand, the scoring model itself may make shorter pathways obtain higher scores in general. In the future work, attributes concerning path lengths may be needed to add to scoring model due to this result.

Table 3.3. Top 20 predicted neuron-specific pathways

Source node	Intermediate nodes	Sink node	Score
M03A1.1a[27]	F55C7.7a[28]	C08C3.3[29]	0.70
M03A1.1a	F55C7.7a;Y60A3A.1[30]	C08C3.3	0.66
C09D8.1a[31]	F55C7.7a	C08C3.3	0.66
M03A1.1a	F55C7.7a;F16B3.1[32]	C08C3.3	0.65
ZK377.2a[33]	F55C7.7a	C08C3.3	0.63
B0457.1a[34]	ZK524.2a[35]; F55C7.7a	C08C3.3	0.63
C06E1.4[36]	ZK632.6[37];Y38A10.A5[38]; F54F2.1[39]	W10D5.1[40]	0.59
M03A1.1a	C14F5.5a[41];F55C7.7a	C08C3.3	0.59
C09D8.1a	C29F9.1[42];F54F2.1	W10D5.1	0.57
C09D8.1a	K03D3.10[43]; F55C7.7a	C08C3.3	0.56
B0457.1a	ZK524.2a;T10H9.4[44];F31E8.2a[45];K04D7.1[46]	F25H2.5[47]	0.54
C09D8.1a	F595F5.6[48]; F54F2.1	W10D5.1	0.54
ZK377.2a	K03D3.10; F55C7.7a	C08C3.3	0.53
B0457.1a	K03D3.10;K08A8.1a[49];ZC504.4a[50];K04D7.1	F25H2.5	0.52
ZK377.2a	K03D3.10; C14F5.5a;F55C7.7a	C08C3.3	0.52
ZK377.2a	Y60A3A.1;C32D5.9[51]	Y37E3.9[52]	0.50
C16B8.1[53]	F55C7.7a	C08C3.3	0..49
C09D8.1a	K03D3.10; C14F5.5a;F55C7.7a	C08C3.3	0.47
C16B8.1	Y60A3A.1;C32D5.9	Y37E3.9	0.46
C09D8.1a	K03D3.10;K08A8.1a;ZC504.4a;K04D7.1	F25H2.5	0.46
Average Score			0.56

Table 3.4. Top 20 predicted gonad-specific pathways

Source node	Intermediate nodes	Sink node	Score
F07A11.6a(54)	ZK792.6(55);AC72.a(56);F43C1.2a(57)	K10G6.1(58)	0.67
F07A11.6a	ZK792.6	T28F12.2a(59)	0.66
T23D8.1(60)	C03C10.1;ZK792.6	T28F12.2a	0.63
F07A11.6a	ZK792.6	K10G6.1	0.62
F07A11.6a	ZK792.6;AC72.a;F43C1.2a	C48D51.a(61)	0.60
T23D8.1	ZK792.6	T28F12.2a	0.59
B0457.1a	ZK370.3a(62);C09B87.a(63);ZK792.6	T28F12.2a	0.59
F07A11.6a	ZK792.6;AC72.a;F43C1.2a	B0547.1(64)	0.57
F07A11.6a	ZK792.6;C26C6.2a	F47D12.4a(65)	0.55
F07A11.6a	ZK792.6	B0547.1	0.54
B0457.1a	ZK370.3a;C09B87.a;ZK792.6	K10G6.1	0.54
B0457.1a	ZK370.3a;C07G1.5(66);ZK1010.1(67);C47E8.5(68)	T28F12.2a	0.52
T23D8.1	ZK792.6	K10G6.1	0.50
B0457.1a	ZK370.3a; B0523.5(69); ZK792.6	B0547.1	0.49
T23D8.1	F31E10.4(70)	T21B10.2a	0.48
B0457.1a	ZK370.3a;T10H9.4; ZK792.6	B0547.1	0.46
B0457.1a	ZK370.3a;B0523.5; ZK792.6	K10G6.1	0.46
T23D8.1	F31E10.4;R08D67.a(67);ZK637.8a(71);C47E8.5(72)	T21B10.2a	0.45
T23D8.1	C03C10.a	F47D12.4a	0.44
B0457.1a	ZK370.3a;T10H9.4; ZK792.6	K10G6.1	0.44
Average Score			0.54

3.2. Disease Associated Pathway Detection

3.2.1. Disease association of C. elegans *Tissue-specific Genes' Human Homologs*

Disease association information is summarized out of information in GeneCard disease genes database (24). The number of these disease associated genes and the percentage out of the total number of research genes is listed in Table 3.5. We investigated the relationship between these genes and diseases, the diseases were categorized based on The International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10). Category list is shown in Appendix B. In order to some pathways related to some important diseases, five genes associated with important diseases for each tissues was selected and shown in Table 4.6 and 4.7, showing information for neuron and gonad specific genes, respectively. Full list is shown in Appendix B. Those genes not listed in Appendix but counted in Table 4.5 were regarded as disease-associated genes but do not have information which disease/disorder is correlated with them.

Although it is restricted to tissue-specific genes, there are still many genes related to certain diseases/disorders, and some of them are really important ones. It's meaningful to find out the pathways with these disease associated genes.

Table 3.5. Disease association information of *C. elegans* neuron and gonad tissue-specific genes

Approach	Total number of genes with human homologs	Number of genes associated with diseases/disorders	Number of genes associated with specific diseases/disorders
Neuron-specific genes in approach 2	207	73	33
Neuron-specific genes in approach 3	51	19	12
Gonad-specific genes in approach 2	203	79	48
Gonad-specific genes in approach 3	23	4	4

3.2.2. Tissue-specific Disease Associated Pathway Detection

Upon utilizing the disease-association information in Section 3.2.1, we were able to filter out disease-associated pathways from the predicted tissue-specific pathways. Maximum top 5 scored pathways are selected for each gene. The filtered results were shown in Tables 4.9 and 4.10. Although we find some links to important diseases such as Parkinson’s disease, there are no pathways detected. However, some links to cancers including leukemia or prostate cancer associated certain pathways were detected with our algorithm.

Table 3.6. Disease association information of *C. elegans* neuron-specific genes' human homologs associated with important diseases

C. elegans gene name	Human homolog's gene name	The most speific tissue-specificity	Disease Category (ICD-10)	Diseases/disorders associated

F54D8.3a[66]	ALDH2	Approach 2	II;XXI	Alcohol sensitivitye;
				Hangover;
				Esophageal cancer, alcohol-related

T03F6.5[72]	PAFAH1B1	Approach 2	XVII	Lissencephaly-1;Subcortical laminar heterotopia (Brain formation disorder)

K08E3.7[73]	PARK2	Approach 3	VI;II;I	Parkinson disease, juvenile,type 2;
				Adenocarcinoma of lung;
				Adenocarcinoma,ovarian;
				Leprosy

T04D1.3a[74]	SH3GL1	Approach 2	II	Leukemia, acute myeloid
C44B11.3[75]	TUBA1A	Approach 2	XVII	Lissencephaly 3 (Brain formation disorder)

Table 3.7. Disease association information of *C. elegans* gonad-specific genes' human homologs

C. elegans gene name	Human homolog's gene name	The most speific tissue-specificity	Disease Category (ICD-10)	Diseases/disorders associated

C12D8.10a[77]	8AKT1	Approach 2	II	Breast cancer, somatic;
				Colorectal cancer, somatic;
				Ovarian cancer, somatic;
				Schizophrenia

F28H6.1a[78]	AKT1	Approach 2	II	Breast cancer, somatic;
				Colorectal cancer, somatic;
				Ovarian cancer, somatic;
				Schizophrenia

ZK370.3a	HIP1	Approach 2	II	Prostate cancer, progression of

ZK792.6	KRAS	Approach 2	II	Lung cancer;
				Bladder cancer;
				Breast cancer, somatic;
				Pancreatic carcinoma, somatic;
				Gastric cancer ;Leukemia, acute myelogenous;
				Noonan syndrome 3;
				Cardiofaciocutaneous syndrome

T04D1.3a	SH3GL1	Approach 2	II	Leukemia, acute myeloid

Table 3.8. Neuron-specific important disease associated pathways

Table 3.8.1. T03F6.5 associated pathways

Source node	Intermediate nodes	Sink node
M03A1.1a	C25F6.2a(79);T03F6.5;C52E12.6(67);F16B3.1	C08C3.3
ZK377.2a	C25F6.2a;T03F6.5;C52E12.6;F16B3.1	C08C3.3
C06E1.4	C25F6.2a;T03F6.5;C52E12.6;F16B3.1	C08C3.3
Disease	Lissencephaly-1;Subcortical laminar heterotopia (Brain formation disorder)	

Table 3.8.2. T04D1.3a associated pathways

Source node	Intermediate nodes	Sink node
B0457.1a	ZK524.2a;JC8.10a(80);T04D1.3a;F48E8.5(81);F38H4.9(82);F55C7.7a)	C08C3.3
Disease	Leukemia, acute myeloid	

No specific pathways shown as F54D8.3a, K08E3.7 or C44B11.3 associated pathways.

Table 3.9. Gonad-specific important disease associated pathways

Table 3.9.1. C12D8.10a associated pathways (continues)

Source node	Intermediate nodes	Sink node
F07A11.6a	ZK792.6; C12D8.10a;F56A8.7a(83);C26C6.2a(84)	F47D12.4a
T23D8.1	ZK792.6; C12D8.10a;F56A8.7a;C26C6.2a	F47D12.4a

F07A11.6a	ZK792.6; C12D8.10a; F28H6.1a; F56A8.7a;C26C6.2a	F47D12.4a
T23D8.1	ZK792.6; C12D8.10a;F56A8.7a;C26C6.2a	F47D12.4a

Disease	Breast cancer, somatic;
	Colorectal cancer, somatic;
	Ovarian cancer, somatic	

Table 3.9.2. F28H6.1a associated pathways

Source node	Intermediate nodes	Sink node
F07A11.6a	ZK792.6; F28H6.1a;F56A8.7a;C26C6.2a	F47D12.4a
T23D8.1	ZK792.6; F28H6.1a;F56A8.7a;C26C6.2a	F47D12.4a
F07A11.6a	ZK792.6; C12D8.10a; F28H6.1a; F56A8.7a;C26C6.2a	F47D12.4a
T23D8.1	ZK792.6; C12D8.10a;F56A8.7a;C26C6.2a	F47D12.4a

Disease	Breast cancer, somatic;
	Colorectal cancer, somatic;
	Ovarian cancer, somatic	

Table 3.9.3. ZK370.3a associated pathways

Source node	Intermediate nodes	Sink node
B0457.1a	ZK370.3a;C09B87.a;ZK792.6	T28F12.2a
B0457.1a	ZK370.3a;C09B87.a;ZK792.6	K10G6.1
B0457.1a	ZK370.3a;C07G1.5;ZK1010.1;C47E8.5	T28F12.2a
B0457.1a	ZK370.3a; B0523.5; ZK792.6	B0547.1
B0457.1a	ZK370.3a;T10H9.4; ZK792.6	B0547.1
Disease	Prostate cancer	

Table 3.9.4. ZK792.6 associated pathways

Source node	Intermediate nodes	Sink node
F07A11.6a	ZK792.6;AC72.a;F43C1.2a	K10G6.1
F07A11.6a	ZK792.6	T28F12.2a
T23D8.1	C03C10.1(85);ZK792.6	T28F12.2a
F07A11.6a	ZK792.6	K10G6.1
F07A11.6a	ZK792.6;AC72.a;F43C1.2a	C48D51.a
Disease	Lung cancer; Bladder cancer; Breast cancer, somatic; Pancreatic carcinoma, somatic; Gastric cancer ;Leukemia, acute myelogenous; Noonan syndrome 3; Cardiofaciocutaneous syndrome	

No specific pathways shown as T04D1.3a associated pathways.

CHAPTER FOUR

Discussion and Conclusion

In this study, a novel computational algorithm is implemented to predict tissue-specific disease-associated pathways in human by utilizing *C. elegans* gene interaction. Information including interaction data, tissue-specific expression patterns, homology information and disease-association data were gathered from different databases to achieve our goals. First, we proposed a new scoring model to evaluate whether a certain pathway is a candidate or not. To test the feasibility of the scoring method, scores were calculated for both reference and random pathways. Result showed that a great difference was seen in these two categories, proving this method is appropriate. Second, we used this scoring model to predict gonad and neuron specific pathways in *C. elegans*. By generating interaction maps of *C. elegans* genes that have human homolog, it was found that the third approach to generate tissue-specific genes (genes only express in certain tissue) is not appropriate for further tissue-specific pathways prediction because no candidate source and sink nodes in the maps. A heuristic method was applied based on the complexity of the interaction maps instead of a simple exhausted search. At last, we filtered the obtained results for disease associations. If one of the members in a certain pathway is associated with disease, the pathway is considered to associate with that disease. Five important disease associated genes were chosen for each tissue, and all the pathways associated with these diseases were shown in the result (Figures 3.8 and 3.9). The aims are successfully achieved in the project but there are still many aspects to improve. I explained them point by point below.

From previous researches studying tissue-specific pathways, it can be concluded that it's really important to carry on the work on predicting these tissue-specific pathways, especially those associated with diseases. Before an experimental approach is applied, a bioinformatics approach is a potential way to have a general insight in this problem and serves as a guide for further research. By applying an evaluation method on a candidate scoring model, it is proved that this candidate scoring model is appropriate for pathway prediction in the *C. elegans* genetic interactome. To accurately predict these pathways, involving parameters like the reliability level of interactions may also be an appropriate idea. This parameter was sat according to different experimental methods used to confirm the interactions. Other parameters like gene expression level could be applied based on different biological systems. Due to the complexity to make an appropriate scoring model involving these parameters, in this research these are not included, but these may make the scoring model better performance in predicting signaling pathways.

By applying the tested scoring model to tissue-specificity data of *C. elegans* genes, I generated tissue-specific signaling pathways in *C. elegans.* Because there are no generally accepted concepts of tissue-specificity, two concepts of tissue-specificity are tested in my study, approach 2 and approach 3 tissue-specificity. In later analysis, we found no candidate source or sink nodes in the interaction map of approach 3. I focused on analyzing approach 2 tissue-specificity. Considering the data size of tissue-specific genes, a heuristic search method is applied to the algorithm; top 20 pathways were shown in the result but pathways met the minimum criteria is sorted to candidate tissue-specific pathways (around 600 in total numbers). For further analysis, better ideas of tissue-specificity should be tested to see which one performs better than others. For example, quantifying tissue-specificity by applying gene expression level data may be complex but may produce more accurate results. Also the results show higher scores for shorter pathways, for future improvement of scoring model, taking consideration of path lengths is also a good idea.

After generating the candidate tissue-specific pathways, we utilized disease association information from GeneCard database (24). If specific tissue-specific pathways have a member associated with a human disease, the pathway is considered associated with the disease. In the result, we selected five important disease-associated genes for both neurons and gonad, and filtered the candidate tissue-specific pathways with these genes. In this way, I could finally obtain these candidate important disease-associated tissue-specific pathways. In the future, other disease-associated genes can be tested in the same way; also biological experimental ways can be applied to these ways to validate these computational predicted pathways in a biological lab manner. For some important diseases like Parkinson’s disease, I was not able to detect its associated tissue-specific pathways. If it is not due to algorithm limitations, Parkinson’s disease may not be tissue-specific. In order to detect these disease-associated pathways, some broader concept like cross-talks among different tissues should be involved in future research. Also for validation of these predicted pathways, a qualified bioinformatics method like watching their members’ predicted structure may also help to validate a predicted pathway.

In conclusion, it's meaningful to build an algorithm to predict disease-associated pathway tissue-specifically using bioinformatics approaches. By study of tissue-specificity and developing appropriate methods to detect these disease related pathways, we could have a general insight on how these signaling pathways are related to the diseases. In the future, the pathway prediction algorithm as well as combining with experimental approaches should be addressed. First we can provide a guide to study new possible signaling pathways by computational prediction, and second if it’s succeeded to combine with experimental approaches to prove some of the predicted pathways, it can not only support our pathway algorithm as an appropriate one, but also show a way of how to utilize these bioinformatics predicted results to help bench work experiments.

APPENDICES

APPENDIX A

Reference pathways in C. elegans

A.1. Reference MAPK pathways in C. elegans

Source node	Intermediate nodes	Sink node
(membrane receptor)		(transcription factor)
W03H9.4 (CACN)	T01E8.3 (PLC-3) F33D4.2 (ITR-1) F09E5.1 (PKC) Y73B6A.5 (RAF1/LIN-45) Y54E10BL.6 (MEK) F43C1.2 (*mpk-1*/ERK) C37F5.1 (*lin-1*/ELK)	F29G9.4 (FOS)
ZK1067.1 (let-23/EGFR)	C14F5.5 (*sem-5*/GRB2) T28F12.3 (SOS) ZK792.6 (*let-60*/RAS) Y73B6A.5 (RAF1/LIN-45) Y54E10BL.6 (MEK) F43C1.2 (mpk-1/ERK)	C37F5.1 (lin-1/ELK)
ZK1067.1 (let-23/EGFR)	F18G5.3 (*gpa-12*/G12) ZK792.6 (*let-60*/RAS) Y73B6A.5 (RAF1/LIN-45) Y54E10BL.6 (MEK) F43C1.2 (*mpk-1*/ERK)	F29G9.4 (FOS)
ZK1067.1 (let-23/EGFR)	T01E8.3 (PLC-3) F33D4.2 (ITR-1) C09D1.1 (MLCK)	F11C3.3 (unc-54/myob)

A.2. Reference calcium pathways in C. elegans

Source node	Intermediate nodes	Sink node
(membrance receptor)		(transcription factor)
Y40H4A.1 (gar-3/GPCR)	R06A10.2 (*gsa-1*/GS F17C8.1 (ADC-G) ZK909.2 (PKA) C18E9.1 (*cal-2*/CALM)	ZC373.4 (MLCK)
Y40H4A.1 (gar-3/GPCR)	R06A10.2 (*gsa-1*/GSA) F17C8.1 (ADC-G) ZK909.2 (PKA) C18E9.1 (*cal-2*/CALM)	C14B9.8 (PHK)
Y40H4A.1 (gar-3/GPCR)	R06A10.2 (*gsa-1*/GS) F17C8.1 (ADC-G) ZK909.2 (PKA) C18E9.1 (*cal-2*/CALM)	K07A9.2 (cmk-1/CAMK)
ZK1067.1 (PTK)	T01E8.3 (PLC-3)	E01H11.1 (PKC)

A.3. Reference ErbB pathways in C. elegans

Source node (membrance receptor)	Intermediate nodes	Sink node (transcription factor)
ZK1067.1 (let-23/ERBB-1)	T01E8.3 (PLC-3)	F09E5.1 (PKC)
ZK1067.1 (let-23/ERBB-1)	Y41D4B.13 (*ced-2*/CRK)	M79.1 (abl-1/ABL)
ZK1067.1 (let-23/ERBB-1)	ZK470.5(*nck-1*/NCK) C09B8.7 (*pak-1*/PAK) F42G10.2 (*mkk-4*/JNKK) B0478.1 (jnk-1/JNK) C37F5.1 (lin-1/ELK)	T24H10.7(jun-1/JUN)
ZK1067.1 (let-23/ERBB-1)	ZK470.5(nck-1/NCK) C09B8.7 (pak-1/PAK) K08A8.1 (mek-1/JNKK) B0478.1 (jnk-1/JNK) C37F5.1 (lin-1/ELK)	T24H10.7(jun-1/JUN)
ZK1067.1 (let-23/ERB-1)	T01E8.3 (PLC-3)	Y51H4A.17 (sta-1/STAT)
ZK1067.1 (let-23/ERB-1)	Y51H4A.17 (sta-1/STAT) T28F12.3(LET-341/SOS) ZK792.6 (let-60/RAS) Y73B6A.5 (RAF1/LIN-45) Y54E10BL.6 (MEK) F43C1.2 (mpk-1/ERK)	C37F5.1 (lin-1/ELK)
ZK1067.1 (let-23/ERB-1)	Y110A7A.10 (aap-1/PI3K) C12D8.10 (akt-1/PKB) B0261.2 (let-363/mTOR)	Y47D3A.16 (rsks-1/S6K)
ZK1067.1 (let-23/ERBB-1)	Y110A7A.10 (aap-1/PI3K) C12D8.10 (akt-1/PKB)	Y18D10A.5(GSK-3)

A. 4. Reference Wnt Pathways in C. elegans

Source node (membrance receptor)	Intermediate nodes	Sink node (transcription factor)
T23D8.1(mom-5/Frizzled)	Y18D10A.5(GSK-3) R13H4.4 (HMP-1)	W10C8.2 (POP-1)
Y71F9B.5 (lin-17/Frizzled)	Y18D10A.5(GSK-3) R13H4.4 (HMP-1)	W10C8.2 (POP-1)

A. 5. Reference mTOR siganling pathways in C. elegans

Source node (membrance receptor)	Intermediate nodes	Sink node (transcription factor)
T28B8.2 (IGF-1/ INS/IGF)	C54D1.3 (ist-1/IRS1) Y110A7A.10 (PI3K) H42K12.1 (PDK) C12D8.10 (akt-1/AKT) B0261.2 (let-363/mtor)	F38A6.3 (HIF-1)
T28B8.2 (IGF-1/ INS/IGF)	C54D1.3 (ist-1/IRS1) Y110A7A.10 (PI3K) H42K12.1 (PDK) F28H6.1 (akt-2/AKT) B0261.2 (let-363/mtor)	F38A6.3 (HIF-1)
Y59A8B1.4 (par-4/LKB1)	T01C8.1 (aak-2/AMP) B0261.2 (let-363/mtor)	B0348.6 (ife-3/eif4e)
Y59A8B1.4 (par-4/LKB1)	T01C8.1 (aak-2/AMP) B0261.2 (let-363/mtor)	F38A6.3 (HIF-1)
Y59A8B1.4 (par-4/LKB1)	T01C8.1 (aak-2/AMP) B0261.2 (let-363/mtor)	Y71A12B.1 (rps-6/S6)

APPENDIX B

Disease association information of C. elegans tissue-specific genes

B.1. List of disease categories based on ICD-10

Chapter (Category serial number)	Title (Category name)
I	Certain infectious and parasitic diseases
II	Neoplasms
III	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV	Endocrine, nutritional and metabolic diseases
V	Mental and behavioural disorders
VI	Diseases of the nervous system
VII	Diseases of the eye and adnexa
VIII	Diseases of the ear and mastoid process
IX	Diseases of the circulatory system
X	Diseases of the respiratory system
XI	Diseases of the digestive system
XII	Diseases of the skin and subcutaneous tissue
XIII	Diseases of the musculoskeletal system and connective tissue
XIV	Diseases of the genitourinary system
XV	Pregnancy, childbirth and the puerperium
XVI	Certain conditions originating in the perinatal period
XVII	Congenital malformations, deformations and chromosomal abnormalities
XVIII	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
XIX	Injury, poisoning and certain other consequences of external causes
XX	External causes of morbidity and mortality
XXI	Factors influencing health status and contact with health services

B.2. Disease association information of C. elegans *neuron-specific genes' human homologs*

C. elegans	Human	The most specific	Disease	Diseases/disorders associated
gene name	homolog's	tissue-specificity	Category	
	gene name		(ICD-10)	

F54D8.3a	ALDH2	Approach 2	II;XXI	Alcohol sensitivitye;
				Hangover;
				Esophageal cancer, alcohol-related

R01H10.6	BBS5	Approach 2	XVIII	Bardet-Biedl syndrome (Chromosomal disease related to anosmia and diabetes)
M03A1.1a	EPHB2	Approach 3	II	Prostate cancer
B0035.5	G6PD	Approach 2	III	Favism; Hemolytic anemia (related to deafness)
C47B2.6a	GALE	Approach 2	XVII	Galactose epimerase deficiency(related to deafness and impaired growth)
M01D7.7a	GNAQ	Approach 3	III	Bleeding diathesis due to GNAQ deficiency
C06E1.4	GRIA3	Approach 3	XXI	Mental retardation, X-linked 94 (related to intellectual disability)
C44E12.3a	KCNK18	Approach 3	VI	Migraine (Neurological disorder)
Y51A2D.19a	KCNMA1	Approach 3	VI	Generalized epilepsy and paroxysmal dyskinesia (Neurological disorder)
T12C9.3	KCNK18	Approach 2	VI	Migraine(Neurological disorder)
T03F6.5	PAFAH1B1	Approach 2	XVII	Lissencephaly-1;Subcortical laminar heterotopia (Brain formation disorder)

K08E3.7	PARK2	Approach 3	VI;II;I	Parkinson disease, juvenile,type 2;
				Adenocarcinoma of lung;

				Adenocarcinoma,ovarian; Leprosy
F27D9.5	PCCA	Approach 2	IV	Propionicacidemia (Organic acid disorder)
Y37E3.9	PHB	Approach 2	II	Breast cancer, susceptibility to
Y73F8A.1	PKD2	Approach 3	XVII	Polycystic kidney disease 2
F31B12.1a	PLCE1	Approach 2	XIV	Nephrotic syndrome, type 3 (Kidney damage)
C30A5.7a	POU4F3	Approach 3	VIII	Deafness, autosomal dominant 15
Y87G2A.4	RAB27A	Approach 3	IV	Griscelli syndrome, type 2 (Immunodeficiency, lead to early death in childhood)
ZK377.2a	ROBO2	Approach 2	XIV	Vesicoureteral reflux 2 (Abnormal movement of urine)
R13A1.4	SCNN1B	Approach 3	IX	Liddle syndrome (Abnormal kidney function)
T04D1.3a	SH3GL1	Approach 2	II	Leukemia, acute myeloid
M01G5.5	SLC6A14	Approach 3	IV	Obesity, susceptibility to, BMIQ11
F57B10.9	SPG20	Approach 2	VI	Troyer syndrome (Dysfunction of nerves)
F38E9.2	TLL1	Approach 2	XVII	Atrial septal defect 6 (Dysfunction of atrial blood supply)
C44B11.3	TUBA1A	Approach 2	XVII	Lissencephaly 3 (Brain formation disorder)
C36E8.5	TUBB	Approach 2	III	Macrothrombocytopenia, autosomal dominant, TUBB1-related (Genetic disorder of blood platelets)
F46B6.3a	UPF3B	Approach 3	XXI	Mental retardation, X-linked, syndromic 14 (Related to intellectual disability)
F28B12.3	VRK1	Approach 2	VI	Pontocerebellar hypoplasia type 1 (Neurological disorder)

B.2. Disease association information of C. elegans *gonad-specific genes' human homologs*

C. elegans gene name	Human homolog's gene name	The most specific tissue-specificity	Disease Category (ICD-10)	Diseases/disorders associated
T02G5.8	ACAT1	Approach 2	IV	Alpha-methylacetoacetic aciduria (Defect in urine)
M03F4.2a	ACTA2	Approach 3	IX	Aortic aneurysm, familial thoracic 6 (Abnormal aorta formation, cause hemorrhage)
T04C12.4	ACTA2	Approach 2	IX	Aortic aneurysm, familial thoracic 6 (Abnormal aorta formation, cause hemorrhage)
T04C12.5	ACTA2	Approach 2	IX	Aortic aneurysm, familial thoracic 6 (Abnormal aorta formation, cause hemorrhage)
T04C12.6	ACTA2	Approach 3	IX	A Aortic aneurysm, familial thoracic 6 (Abnormal aorta formation, cause hemorrhage)
C12D8.10a	AKT1	Approach 2	II	Breast cancer, somatic; Colorectal cancer, somatic; Ovarian cancer, somatic; Schizophrenia
F28H6.1a	AKT1	Approach 2	II	Breast cancer, somatic; Colorectal cancer, somatic; Ovarian cancer, somatic; Schizophrenia
K11D9.2a	ATP2A1	Approach 2	VI	Brody myopathy (Affect skeletal muscle)
ZK256.1a	ATP2C1	Approach 2	XVII	Hailey-Hailey disease (Genetic disorder of skin)
D2045.1a	ATXN2	Approach 3	VI	Spinocerebellar ataxia-2 (Cause poor coordination of movement)
F28F8.6	ATXN3	Approach 2	VI	Machado-Joseph disease

				(Affect muscle control)
K02A4.1	BCAT2	Approach 2	IV	Hypervalinemia or hyperleucine-isoleucinemia (metabolic disorder related to urine)
F58G6.1	BIN1	Approach 2	VI	Myopathy, centronuclear, autosomal recessive (muscular weakness)
C56C10.3	CHMP4B	Approach 2	VII	Cataract, posterior polar, 3
C43E11.11	COG5	Approach 2	IV	Congenital disorder of glycosylation, type II
F01G12.5a	COL4A5	Approach 2	XVII	Alport syndrome (related to hear loss)
K04H4.1a	COL4A5	Approach 2	XVII	Alport syndrome(related to hear loss)
F41C3.5	CTSA	Approach 2	IV	Galactosialidosis (lysosomal storage disease)
C33D9.1a	FGD3	Approach 2	IV	Glucocorticoid deficiency 3
T06H11.4	GPHN	Approach 2	IV	Molybdenum cofactor deficiency, type C;Hyperekplexia
ZK370.3a	HIP1	Approach 2	II	Prostate cancer, progression of

ZK792.6	KRAS	Approach 2	II	Lung cancer;
				Bladder cancer;
				Breast cancer, somatic;
				Pancreatic carcinoma, somatic;
				Gastric cancer
				;Leukemia, acute myelogenous;
				Noonan syndrome 3;
				Cardiofaciocutaneous syndrome

T22A3.8	LAMA2	Approach 2	VI	Muscular dystrophy, congenital merosin-deficient (weak muscular system)
T10E9.7a	NDUFS3	Approach 2	VI	Leigh syndrome
T03F6.5	PAFAH1B1	Approach 2	XVII	Lissencephaly-1;Subcortical laminar heterotopia (Brain formation disorder)
F31B12.1a	PLCE1	Approach 2	XIV	Nephrotic syndrome, type 3

				(Kidney damage)
F59G1.5	PTPN1	Approach 2	IV	Insulin resistance, susceptibility to
Y43C5A.6a	RAD51	Approach 3	II	Breast cancer, susceptibility to
R05H10.2	RBM28	Approach 2	IV;VI	Alopecia, neurologic defects, and endocrinopathy syndrome
T04D1.3a	SH3GL1	Approach 2	II	Leukemia, acute myeloid
AC7.2a	SHOC2	Approach 2	XVII	Noonan-like syndrome with loose anagen hair
K11G12.4a	SLC11A2	Approach 2	III	Anemia, hypochromic microcytic (Paler red blood cells)
Y46G5A.4	SNRNP200	Approach 2	VII	Retinitis pigmentosa 33 (related to blindness)
C24B5.2a	SPAST	Approach 2	VI	Spastic paraplegia-4 (dysfunction of nerves)
K12C11.2	SUMO1	Approach 2	XVII	Orofacial cleft 10
C09G12.9	TSG101	Approach 2	II	Breast cancer
C06A1.1	VCP	Approach 2	IV	Inclusion body myopathy with early-onset Paget disease and frontotemporal dementia

LITERATURE CITED

[1] Cell-to-Cell Signaling: Hormones and Receptors. May 2013.

[2] Peifer M, Polakis P, (2000) Wnt signaling in oncogenesis and embryogenesis - a look outside the nucleus. Science. 287(5458):1606-9.

[3] Inoue T, Ogawa O (2011) Role of signaling transduction pathways in development of castration-resistant prostate cancer. Prostate Cancer. 2011:647987.

[4] Tornatore L, Thotakura AK, Bennett J, Moretti M, Franzoso G (2012) The nuclear factor kappa B signaling pathway: integrating metabolism with inflammation. Trends Cell Biol. 22(11).

[5] Rochet JC, Hay BA, Guo M (2012) Molecular insights into Parkinson's disease. Prog Mol Biol Transl Sci. 107:125-88.

[6] Rask-Madsen C, Kahn CR (2012) Tissue-specific insulin signaling, metabolic syndrome, and cardiovascular disease. Arterioscler Thromb Vasc Biol. 32(9):2052-9.

[7] Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T etc. (2010) Integrative analysis of the_Caenorhabditis elegans_genome_by the modencode project. Science.330(6012):1775-87

[8] Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, etc. (2000) Comparative genomics of the eukaryotes. Science. 287(5461):2204-15.

[9] Xiao SJ, Zhang C, Zou Q, Ji ZL (2010) tisged: a database for tissue-specific genes. Bioinformatics. 26(9):1273-5.

[10] Joseph MD, Liviu P and Peter DK (2010) Machine learning methods for metabolic pathway prediction. *BMC Bioinformatics* 11:15

[11] Jerby L, Shlomi T, Ruppin E. (2010) Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. Mol Syst Biol. 6:401.

[12] Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E. (2008) Network-based prediction of human tissue-specific metabolism. Nat Biotechnol. 26(9):1003-10.

[13] Mintz-Oron S, Meir S, Malitsky S, Ruppin E, Aharoni A, Shlomi T. (2012) Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. Proc Natl Acad Sci 109(1):339-44.

[14] Korcsmáros T, Farkas IJ, Szalay MS, Rovó P, Fazekas D, Spiró Z, Böde C, Lenti K, Vellai T, Csermely P. (2010) Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery. Bioinformatics. 26(16):2042-50.

[15] Len A. Pennacchio, Gabriela G. Loots, Marcelo A. Nobrega, Ivan Ovcharenko. (2007) Predicting tissue-specific enhancers in the human genome. Genome Res. 17(2): 201–211.

[16] Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012 Jan; 40(Database issue):D109-14

[17] Yook K, Harris TW, Bieri T, Cabunoc A, Chan J, etc (2012) WormBase 2012: more genomes, more data, new website. Nucleic Acids Res. 2012 Jan; 40(Database issue):D735-41

[18] Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, etc (2013) The BioGRID interaction database: 2013 update. Nucleic Acids Res. 2013 Jan; 41(Database issue):D816-23

[19] INTRODUCTION TO *C. elegans* ANATOMY. May.2013 <http://www.wormatlas.org/ver1/handbook/anatomyintro/anatomyintro.htm>

[20] Yu et al (2008) High-quality binary protein interaction map of the yeast interactome network, Science 322:104-110.

[21] Zeng X, Couture LA. (2013) Pluripotent stem cells for Parkinson's disease: progress and challenges. Stem Cell Res Ther. 2013 Apr 15;4(2):25.

[22] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 2003 Nov; 13(11):2498-504

[23] NCBI Resource Coordinators. (2013) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2013 Jan; 41(Database issue):D8-D20.

[24] Rappaport, N, Nativ, N, Stelzer, G, Twik, M, Guan-Golan, Y, Iny Stein, T, Bahir, I, Belinky, F, Morrey, CP, Safran, M and Lancet, D. (2013) MalaCards: an integrated compendium for diseases and their annotation, Database 2013; doi: 10.1093/database/bat018.

[25] Qin T, Tsoi LC, Sims KJ, Lu X, Zheng WJ. (2012) Signaling network prediction by the Ontology Fingerprint enhanced Bayesian network. BMC Syst Biol. 2012;6 Suppl 3:S3. doi: 10.1186/1752-0509-6-S3-S3.

[26] Jon Louis Bentley (1982). Writing Efficient Programs. Prentice Hall. p. 11.

[27] George, S. E., Simokat, K. A., Hardin, J. D., & Chisholm, A. D. (1998). The VAB-1 Eph receptor tyrosine kinase functions in neural and epithelial morphogenesis in C. elegans. Cell, 92, 633-43.

[28] Forrester, W. C., Perens, E., Zallen, J. A., & Garriga, G. (1998). Identification of Caenorhabditis elegans genes required for neuronal differentiation and migration. Genetics, 148, 151-65.

[29] Salser, S. J., & Kenyon, C. J. (1992). Activation of a C. elegans Antennapedia homologue in migrating cells controls their direction of migration. Nature, 355, 255-8.

[30] Ogura, K., & Goshima, Y. (2006). The autophagy-related kinase UNC-51 and its binding partner UNC-14 regulate the subcellular localization of the Netrin receptor UNC-5 in Caenorhabditis elegans. Development, 133, 3441-50.

[31] Chin-Sang, I. D., Moseley, S. L., Ding, M., Harrington, R. J., George, S. E., & Chisholm, A. D. (2002). The divergent C. elegans ephrin EFN-4 functions inembryonic morphogenesis in a pathway independent of the VAB-1 Eph receptor. Development, 129, 5499-510.

[32] LeBoeuf, B., Guo, X., & Garcia, L. R. (2011). The effects of transient starvation persist through direct interactions between CaMKII and ether-a-go-go K+ channels in C. elegans males. Neuroscience, 175, 1-17.

[33] Branda, C. S., & Stern, M. J. (2000). Mechanisms controlling sex myoblast migration in Caenorhabditis elegans hermaphrodites. Dev Biol, 226, 137-51.

[34] Promel, S., Frickenhaus, M., Hughes, S., Mestek, L., Staunton, D., Woollard, A., Vakonakis, I., Schoneberg, T., Schnabel, R., Russ, A. P., & Langenhan, T. (2012). The GPS motif is a molecular switch for bimodal activities of adhesion class G protein-coupled receptors. Cell Rep, 2, 321-31.

[35] Garcia, L. R., & Sternberg, P. W. (2003). Caenorhabditis elegans UNC-103 ERG-like potassium channel regulates contractile behaviors of sex muscles in males before and during mating. J Neurosci, 23, 2696-705.

[36] Vukojevic, V., Gschwind, L., Vogler, C., Demougin, P., de, D. J. Q., Papassotiropoulos, A., & Stetak, A. (2012). A role for -adducin (ADD-1) in nematode and human memory. EMBO J, 31, 1453-66.

[37] Lee, W., Lee, T. H., Park, B. J., Chang, J. W., Yu, J. R., Koo, H. S., Park, H., Yoo, Y. J., & Ahnn, J. (2005). Caenorhabditis elegans calnexin is N-glycosylated and required for stress response. Biochem Biophys Res Commun, 338, 1018-30.

[38] Lee, W., Kim, K. R., Singaravelu, G., Park, B. J., Kim, D. H., Ahnn, J., & Yoo, Y. J. (2006). Alternative chaperone machinery may compensate for calreticulin/calnexin deficiency in Caenorhabditis elegans. Proteomics, 6, 1329-39.

[39] Williams, B. D., & Waterston, R. H. (1994). Genes critical for muscle development and function in Caenorhabditis elegans identified through lethal mutations. J Cell Biol, 124, 475-90.

[40] Choi, K. Y., Ji, Y. J., Jee, C., & Ahnn, J. (2002). Characterization of CeHDA-7, a class II histone deacetylase interacting with MEF-2 in Caenorhabditis elegans. Biochem Biophys Res Commun, 293, 1295-300.

[41] Dixon, S. J., Alexander, M., Fernandes, R., Ricker, N., & Roy, P. J. (2006). FGF negatively regulates muscle membrane extension in Caenorhabditis elegans. Development, 133, 1263-75.

[42] Samuelson AV, Carr CE, Ruvkun G. (2007) Gene activities that mediate increased life span of C. elegans insulin-like signaling mutants. Genes Dev. 21(22):2976-94.

[43] Struckhoff, E. C., & Lundquist, E. A. (2003). The actin-binding protein UNC-115 is an effector of Rac signaling during axon pathfinding in C. elegans. Development, 130, 693-704.

[44] Nonet, M. L., Saifee, O., Zhao, H. J., Rand, J. B., & Wei, L. P. (1998). Synaptic transmission deficits in Caenorhabditis elegans synaptobrevin mutants. J Neurosci, 18, 70-80.

[45] Mathews, E. A., Mullen, G. P., Crowell, J. A., Duerr, J. S., McManus, J. R., Duke, A., Gaskin, J., & Rand, J. B. (2007). Differential expression and function of synaptotagmin 1 isoforms in Caenorhabditis elegans. Mol Cell Neurosci, 34, 642-52.

[46] Demarco, R. S., & Lundquist, E. A. (2010). RACK-1 acts with Rac GTPase signaling and UNC-115/abLIM in Caenorhabditis elegans axon pathfinding and cell migration. PLoS Genet, 6, e1001215.

[47] Napolitano F, D'Angelo F, Bimonte M, Perrina V, D'Ambrosio C, Scaloni A, Russo T, Zambrano N. (2008). A differential proteomic approach reveals an evolutionary conserved regulation of Nme proteins by Fe65 in C. elegans and mouse. Neurochem Res. 33(12):2547-55. doi: 10.1007/s11064-008-9683-z.

[48] Taru, H., & Jin, Y. (2011). The Liprin homology domain is essential for the homomeric interaction of SYD-2/Liprin- protein in presynaptic assembly. J Neurosci, 31, 16261-8.

[49] Koga, M., Zwaal, R. R., K-L, G., Avery, L., & Ohshima, Y. M. (2000). A Caenorhabditis elegans MAP kinase kinase, MEK-1, is involved in stress responses. EMBO J, 19, 5148-56.

[50] Poinat, P., De, A. A., Sookhareea, S., Zhu, X., Hedgecock, E. M., Georges-Labouesse, E., & Labouesse, M. (2002). A conserved interaction between beta1 integrin/PAT-3 and Nck-interacting kinase/MIG-15 that mediates commissural axon navigation in C. elegans. Curr Biol, 12, 622-31.

[51] Melendez, A., Talloczy, Z., Seaman, M., Eskelinen, E. L., Hall, D. H., & Levine, B. (2003). Autophagy genes are essential for dauer development and life-span extension in C. elegans. Science, 301, 1387-91.

[52] Sanz, M. A., Tsang, W. Y., Willems, E. M., Grivell, L. A., Lemire, B. D., van, D. S. H., & Nijtmans, L. G. J. (2003). The mitochondrial prohibitin complex is essential for embryonic viability and germline function in Caenorhabditis elegans. J Biol Chem, 278, 32091-9.

[53] Deshpande, R., Inoue, T., Priess, J. R., & Hill, R. J. (2005). lin-17/Frizzled and lin-18 regulate POP-1/TCF-1 localization and cell type specification during C. elegans vulval development. Dev Biol, 278, 118-29.

[54] Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. (2006). Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. Nat Genet. 38(8):896-903.

[55] Dixon, S. J., Alexander, M., Fernandes, R., Ricker, N., & Roy, P. J. (2006). FGF negatively regulates muscle membrane extension in Caenorhabditis elegans. Development, 133, 1263-75.

[56] Sieburth, D. S., Sun, Q. A., & Han, M. (1998). SUR-8, a conserved Ras-binding protein with leucine-rich repeats, positively regulates Ras-mediated signaling in C. elegans. Cell, 94, 119-30.

[57] Arur, S., Ohmachi, M., Berkseth, M., Nayak, S., Hansen, D., Zarkower, D., & Schedl, T. (2011). MPK-1 ERK controls membrane organization in C. elegans oogenesis via a sex-determination module. Dev Cell, 20, 677-88.

[58] Howard, R. M., & Sundaram, M. V. (2002). C. elegans EOR-1/PLZF and EOR-2 positively regulate Ras and Wnt signaling and function redundantly with LIN-25 and the SUR-2 Mediator component. Genes Dev, 16, 1815-27.

[59] Andachi, Y. (2003). Inactivation of the T-box gene tbx-9 or tbx-8 enhances the phenotype of the Hth/Meis orthologue unc-62 and the Exd/Pbx orthologues ceh-20 and ceh-40 in embryogenesis presented in International Worm Meeting.

[60] Ferrier, A., Charron, A., Sadozai, Y., Switaj, L., Szutenbach, A., & Smith, P. A. (2011). Multiple phenotypes resulting from a mutagenesis screen for pharynx muscle mutations in Caenorhabditis elegans. PLoS One, 6, e26594.

[61] Heard, M., Maina, C. V., Morehead, B. E., Hoener, M. C., Nguyen, T. Q., Williams, C. C., Rowan, B. G., & Gissendanner, C. R. (2010). A functional NR4A nuclear receptor DNA-binding domain is required for organ development in Caenorhabditis elegans. Genesis, 48, 485-91.

[62] Haycraft, C. J., Schafer, J. C., Zhang, Q., Taulman, P. D., & Yoder, B. K. (2002). Identification of CHE-13, a novel IFT protein required for cilia formation presented in West Coast Worm Meeting.

[63] Lucanic, M., Kiley, M., Ashcroft, N., L'etoile, N., & Cheng, H. J. (2006). The Caenorhabditis elegans P21-activated kinases are differentially required for UNC-6/netrin-mediated commissural motor axon guidance. Development, 133, 4549-59.

[64] Wee, S., Hetfeld, B., Dubiel, W., & Wolf, D. A. (2002). Conservation of the COP9/signalosome in budding yeast. BMC Genet, 3, 15.

[65] Lily I. Jiang, Paul W. Sternberg. (1999). An HMG1-like protein facilitates Wnt signaling in *Caenorhabditis elegans*Genes Dev. 13(7): 877–889.

[66] Mayers, J. R., Fyfe, I., Schuh, A. L., Chapman, E. R., Edwardson, J. M., & Audhya, A. (2011). ESCRT-0 assembles as a heterotetrameric complex on membranes and binds multiple ubiquitinylated cargoes simultaneously. J Biol Chem, 286, 9636-45.

[67] C. elegans Sequencing Consortium. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science.282(5396):2012-8.

[68] Morley, J. F., & Morimoto, R. I. (2004). Regulation of longevity in Caenorhabditis elegans by heat shock factor and molecular chaperones. Mol Biol Cell, 15, 657-64.

[69] Campbell, H. D., Schimansky, T., Claudianos, C., Ozsarac, N., Kasprzak, A. B., Cotsell, J. N., Young, I. G., de, H. G. C., & Miklos, G. L. G. (1993). The Drosophila melanogaster flightless-I gene involved in gastrulation and muscle degeneration encodes gelsolin-like and leucine-rich repeat domains and is conserved in Caenorhabditis elegans and humans. Proc Natl Acad Sci U S A, 90, 11386-90.

[70] Geles, K. G., & Adam, S. A. (2001). Germline and developmental roles of the nuclear transport factor importin alpha3 in C. elegans. Development, 128, 1817-30.

[71] Kuroyanagi, H., Watanabe, Y., & Hagiwara, M. (2013). CELF Family RNA-Binding Protein UNC-75 Regulates Two Sets of Mutually Exclusive Exons of the unc-32 Gene in Neuron-Specific Manners in Caenorhabditis elegans. PLoS Genet, 9, e1003337

[72] Birnby, D. A., Link, E. L., Vowels, J. J., Tian, H. Z., Colacurcio, P. L., & Thomas, J. H. (2000). A transmembrane guanylyl cyclase (DAF-11) and Hsp90 (DAF-21) regulate a common set of chemosensory behaviors in caenorhabditis elegans. Genetics, 155, 85-104.

[73] Gil-Krzewska, A. J., Farber, E., Buttner, E. A., & Hunter, C. P. (2010). Regulators of the actin cytoskeleton mediate lethality in a Caenorhabditis elegans dhc-1 mutant. Mol Biol Cell, 21, 2707-20. doi:10.1091/mbc.E09-07-0593

[74] Culetto, E., & Sattelle, D. B. (2000). A role for Caenorhabditis elegans in understanding the function and interactions of human disease genes. Hum Mol Genet, 9, 869-77.

[75] Schuske, K. R., Richmond, J. E., Matthies, D. S., Davis, W. S., Runz, S., Rube, D. A., Van, A. M. D. B., & Jorgensen, E. M. (2003). Endophilin is required for synaptic vesicle endocytosis by localizing synaptojanin. Neuron, 40, 749-62.

[76] Akella, J. S., Wloga, D., Kim, J., Starostina, N. G., Lyons-Abbott, S., Morrissette, N. S., Dougan, S. T., Kipreos, E. T., & Gaertig, J. (2010). MEC-17 is an alpha-tubulin acetyltransferase. Nature, 467, 218-22.

[77] Shen, X., Valencia, C. A., Gao, W., Cotten, S. W., Dong, B., Huang, B. C., & Liu, R. (2008). Ca(2+)/Calmodulin-binding proteins from the C. elegans proteome. Cell Calcium, 43, 444-56.

[78] Wolkow, C. A., Munoz, M. J., Riddle, D. L., & Ruvkun, G. (2002). Insulin receptor substrate and p55 orthologous adaptor proteins function in the Caenorhabditis elegans daf-2/insulin-like signaling pathway. J Biol Chem, 277, 49591-7.

[79] Segbert, C., Johnson, K., Theres, C., van, F. D., & Bossinger, O. (2004). Molecular and functional analysis of apical junction formation in the gut epithelium of Caenorhabditis elegans. Dev Biol, 266, 17-26.

[80] Kuwabara, P. E., & O'Neil, N. (2001). The use of functional genomics in C. elegans for studying human development and disease. J Inherit Metab Dis, 24, 127-38.

[81] Lange, K. I., Heinrichs, J., Cheung, K., & Srayko, M. (2013). Suppressor mutations identify amino acids in PAA-1/PR65 that facilitate regulatory RSA-1/B subunit targeting of PP2A to centrosomes in C. elegans. Biol Open, 2, 88-94.

[82] Griffin, E. E., Odde, D. J., & Seydoux, G. (2011). Regulation of the MEX-5 gradient by a spatially segregated kinase/phosphatase cycle. Cell, 146, 955-68.

[83] McEwen, J. M., & Kaplan, J. M. (2008). UNC-18 promotes both the anterograde trafficking and synaptic function of syntaxin. Mol Biol Cell, 19, 3836-46.

[84] Hofler, C., & Koelle, M. R. (2011). AGS-3 alters Caenorhabditis elegans behavior after food deprivation via RIC-8 activation of the neural G protein G o. J Neurosci, 31, 11553-62.

[85] Banerjee, D., Chen, X., Lin, S. Y., & Slack, F. J. (2010). kin-19/casein kinase I has dual functions in regulating asymmetric division and terminal differentiation in C. elegans epidermal stem cells. Cell Cycle, 9, 4748-65.