RESEARCH ARTICLE

Gene Cascade Finder: A tool for identification of gene cascades and its application in *Caenorhabditis elegans*

Yusuke Nomoto[1], Yukihiro Kubota[2], Yuto Ohnishi[1], Kota Kasahara[2], Aimi Tomita[1], Takehiro Oshime[1], Hiroki Yamashita[1], Muhamad Fahmi[2], Masahiro Ito[1,2]*

1 Advanced Life Sciences Program, Graduate School of Life Sciences, Ritsumeikan University, Kusatsu, Shiga, Japan, **2** Department of Bioinformatics, College of Life Sciences, Ritsumeikan University, Kusatsu, Shiga, Japan

* maito@sk.ritsumei.ac.jp

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Obtaining a comprehensive understanding of the gene regulatory networks, or gene cascades, involved in cell fate determination and cell lineage segregation in *Caenorhabditis elegans* is a long-standing challenge. Although RNA-sequencing (RNA-Seq) is a promising technique to resolve these questions, the bioinformatics tools to identify associated gene cascades from RNA-Seq data remain inadequate. To overcome these limitations, we developed Gene Cascade Finder (GCF) as a novel tool for building gene cascades by comparison of mutant and wild-type RNA-Seq data along with integrated information of protein-protein interactions, expression timing, and domains. Application of GCF to RNA-Seq data confirmed that SPN-4 and MEX-3 regulate the canonical Wnt pathway during embryonic development. Moreover, *lin-35*, *hsp-3*, and *gpa-12* were found to be involved in MEX-1-dependent neurogenesis, and MEX-3 was found to control the gene cascade promoting neurogenesis through *lin-35* and *apl-1*. Thus, GCF could be a useful tool for building gene cascades from RNA-Seq data.

Introduction

Spatially and temporally regulated gene expression is essential to precisely modulate cellular behaviors during development in multicellular organisms. Elucidating gene cascades during early embryonic development may improve our understanding of mechanisms of cell fate determination and lineage segregation [1–3]. The nematode *Caenorhabditis elegans*, a model organism of development research, comprises 959 cells in adult hermaphrodites with robustness and reproducibility of the cell lineage [3]. Additionally, over 80% of the *C. elegans* proteome shows homology with human proteins [4], providing a particularly valuable model organism for studies of the developmental system.

PAR proteins, which are expressed immediately after fertilization, are associated with formation of the anterior-posterior polarity of P0 cells and control the localization of polarity mediators such as SPN-4, MEX-1, and MEX-3 in *C. elegans* [5]. Aberrant expression of the

genes encoding these proteins affects cellular and developmental regulation, leading to embryonic lethality in early embryogenesis [6–9]. Specifically, SPN-4 is localized in all blastomeres at the four-cell stage and plays essential roles in axial rotation [8, 10–12]; MEX-1 is expressed in the P1 blastomere, and loss of its function leads to excessive muscle formation [7, 13]; and MEX-3 is expressed in the AB blastomere, and causes excessive muscle formation and hatching failure in mutants [14–16]. Although polarity mediators regulate protein synthesis by binding to the 3′-untranslated region of the target mRNA, it is difficult to directly identify their associated gene cascades.

Conventional genetic and molecular biological approaches have focused on the target gene to be identified and have clarified functions and identified related genes, representing a bottom-up approach. Both functional analysis of individual genes and comprehensive analysis of the genome are indispensable for identification of gene cascades. After determination of the whole genome sequence of *C. elegans* in 1998, genome-wide analysis via a top-down approach was made possible [17], representing the beginning of the post-genome sequencing era. Transcriptomics via DNA microarray analysis [18], proteomics via mass spectrometry analysis [19], and phenomenon analysis by RNA interference [20] have been extensively reported. Furthermore, several methods for comprehensive analyses have been developed, including protein-protein interaction analysis using a yeast two-hybrid system and phage display [21, 22] and multiple mutation analysis using knockdown mutants. At the same time, the WormBase database was constructed to integrate the vast quantities of data obtained from these genome-wide analyses [23]. Accordingly, the development of new technologies and methodologies has enabled the accumulation of detailed genome-wide data.

Next-generation sequencing (NGS) has now replaced conventional Sanger sequencing [24]. Conventional Sanger sequencing can simultaneously analyze 8–96 sequencing reactions, whereas NGS can simultaneously run millions to billions of sequencing reactions in parallel. This technique can dramatically and quickly determine the gene sequences in organisms whose whole-genome sequences have already been determined. Even at the laboratory level, genomic sequencing results can be produced in only a few days, enabling researchers to obtain genome-wide information rapidly. Furthermore, RNA sequencing (RNA-Seq) has recently been developed to measure gene expression levels by counting the number of sequence reads obtained from converting RNA into cDNA [25]. Existing RNA-Seq data analysis tools include RSeQC [26], which measures the quality control of the obtained data; Cufflinks [27], which involves genome mapping; and IsoEM [28], which identifies isoforms within a dataset. These tools can be used to identify gene expression variations from RNA-Seq data. TopGO [29] is an analytical tool used to identify gene function based on RNA-Seq data and can confirm the functions of genes with varying expression levels. In addition, Cascade R was established to identify the gene cascade of a query gene [30]. Cascade R constructs an intergenic network of knockout genes from the results of DNA microarray analysis. However, it requires multiple timeline datasets from microarray analyses.

Genes, especially those expressed in early embryogenesis, function in chronological order rather than having only a single function, and genes responsible for functional expression often exert their effects at the bottom of gene cascades. STRING [31], BIOGRID [32], and WormBase [23] are databases of protein-protein interactions and the gene-dependent regulation of transcription and translation. In order to predict genetic cascades from these databases, researchers currently must perform separate analyses. Moreover, although RNA-Seq can be used to easily acquire large amounts of data via a semi-automatic process, the subsequent analysis must be performed manually and is therefore quite time-consuming. Therefore, the data acquisition capacity currently exceeds the data analysis capacity. Accordingly, automation of the analysis using bioinformatics tools is an important research subject.

In this study, we performed RNA-Seq analysis of the polarity mediator mutants *spn-4*, *mex-1*, and *mex-3* in *C. elegans*. Next, we developed a novel tool, Gene Cascade Finder (GCF), to extract genes with a high probability of being directly or indirectly regulated by these polarity mediators. Finally, the gene cascade and its validity were examined.

Materials and methods

Strains

C. elegans wild type (N2), *mex-1(or286)*, and *mex-3(eu149)* strains, and *Escherichia coli* OP50 strain were provided by the Caenorhabditis Genetics Center (https://cgc.umn.edu/), and the *C. elegans sin-3(tm1276)* and *spn-4(tm291)* strains were provided by National BioResource Project [33].

Culture of *C. elegans* and synchronization at the early embryo stage

All strains except for *mex-1(or286)* were cultured on nematode growth medium agar coated with *E. coli* OP50 at 20˚C. Because *mex-1(or286)* strain is a temperature-sensitive mutant strain, it was cultured at 15˚C to strengthen its phenotype [13]. Furthermore, all strains were transferred to S-Basal solution inoculated with *E. coli* OP50 at 20˚C for large culture. To obtain early embryos from the culture medium, when *C. elegans* adults had only 3–5 eggs, they were synchronized using an alkaline bleaching method, and the early embryos were recovered [34]. These *C. elegans* early embryos were used as the samples for RNA-Seq analysis.

Guideline for RNA-Seq analysis and comparative gene expression analysis

The guideline for RNA-seq data analysis for gene cascade prediction is indicated in the supplementary material (S1 Fig). RNA-seq analysis of wild type (N2) and *sin-3(tm1476)* mutant early embryos (N = 3) was performed by following the Illumina protocols (available on the Illumina, Inc. website). Library preparation for RNA-Seq was performed using the TruSeq Stranded Total RNA LT Sample Prep Kit (Illumina, Inc.). Next, using a MiSeq Reagent Kit v3 (Illumina, Inc.), the DNA samples were denatured, diluted, and paired-end sequencing (75 bp length) was performed in MiSeq sequencer (Illumina). The quality of raw sequence data obtained by RNA-Seq was checked using FastQC. Trimmomatic [35] was employed to trim low quality reads and the sequence data were mapped to a *C. elegans* reference genome (WormBase Version 261) using HISAT2. The count data of wild type and *sin-3* mutant were compared with DESeq2 [36] and differentially expressed genes (DEGs) (p-value < 0.05, $\log_2$ Fold changes between *sin-3* mutant/wild type ($\log_2$ count data of *sin-3* mutant/count data of wild type)) were identified (S1 Table).

RNA-Seq analysis of *mex-1*, *mex-3*, and *spn-4*

The mRNAs of the synchronized *C. elegans* early embryos were purified using RNeasy Minikit (Qiagen NV, Venlo, the Netherlands). Purified mRNAs were reverse-transcribed into cDNAs, amplified by polymerase chain reaction, and fragmented using a TruSeq RNA Sample Prep Kit (Illumina, Inc.). The amplified cDNAs were sequenced using Hi-Seq2000 (Illumina, Inc.) and the sequenced cDNAs were mapped to the *C. elegans* genome sequence and counted according to WormBase (WS190) [23] using DNAnexus. Using this procedure, the mRNA expression levels were obtained as reads per kilobase of exon per million mapped reads (RPKM) [37]. The gene name and RPKM values of wild-type and mutant genes were filed for input data in GCF (S2 Table). Although we used the RPKM method to perform static evaluation, it is better to use DESeq2 or edgeR to define and normalize the fluctuations or differential expression

patterns. However, since RPKM is still utilized by many researchers up until now [38, 39], we initially develop the tool for this classical method and will incorporate other more sophisticated normalization methods to GCF (see Guideline for RNA-Seq analysis and comparative gene expression analysis).

Comparative quantitative gene expression analysis of *mex-1*, *mex-3*, and *spn-4*

Expression levels of gene i in the wild-type and mutant were defined as xi and yi, respectively, and the change rates in these gene expression levels (Ri) were determined as shown in Eq 1.

$$Ri = yi/xi \qquad (i = 1, \ldots\ldots, N) \tag{1}$$

$$Ri < Mi - Qi, \quad Mi + Qi < Ri \tag{2}$$

$$Mi - Qi < Ri < Mi + Qi \tag{3}$$

Further, genes that satisfied the condition of Eq 2 were assumed to show expression level fluctuations, and genes satisfying the condition of Eq 3 were assumed to not show expression level fluctuations. N is the number of genes, and Mi and Qi are the median and quartile deviation, respectively. The expression level of each gene was spatially and temporally regulated in the early embryonic development of *C. elegans*. To analyze the effect of the expression of each gene in mutant for the polarity mediator genes, we compared the relative expression level of each gene between the wild type and deletion mutants for the polarity mediator gene (S2 Table).

Dataset for the software

Information on the expression timing and interactions of all genes in *C. elegans* was extracted from WormBase (Version 256) [23] using the application programming interface. The total transcription factors of *C. elegans* were acquired from the gene ontology database Amigo 2 (http://amigo.geneontology.org/amigo) [40] using the keyword search "GO: 0006351". Furthermore, all gene IDs, protein IDs, and domain information from Pfam (https://pfam.xfam. org/) [41] in *C. elegans* required for functional analysis were extracted from UniProt [42]. We used the embryonic gene expression and gene interaction information from WormBase, and about 20,000 domain information of the gene product from UniProt.

Development of Gene Cascade Finder

The programming language Ruby was used to construct Gene Cascade Finder (GCF). The web interface of GCF was written in Python. The input data for GCF were data from the wild-type and mutant strains, as shown in S2 Table. The output data from GCF included discovered gene cascades (S3–S5 Tables), the data input for Cytoscape (S6 Table), and gene cascade-specific domains and their gene cascades (S7 Table). GCF is available at http://www.gcf.sk. ritsumei.ac.jp.

Direct target prediction by GCF

GCF was developed by the algorithm shown in Fig 1. To identify and verify the candidate for the direct target genes of a query gene, we performed the following tasks:

(1) evaluate expression level fluctuation between wild-type and a mutant of the query gene, (2) confirm that the candidate gene is transcription factor, (3) confirm that the candidate gene
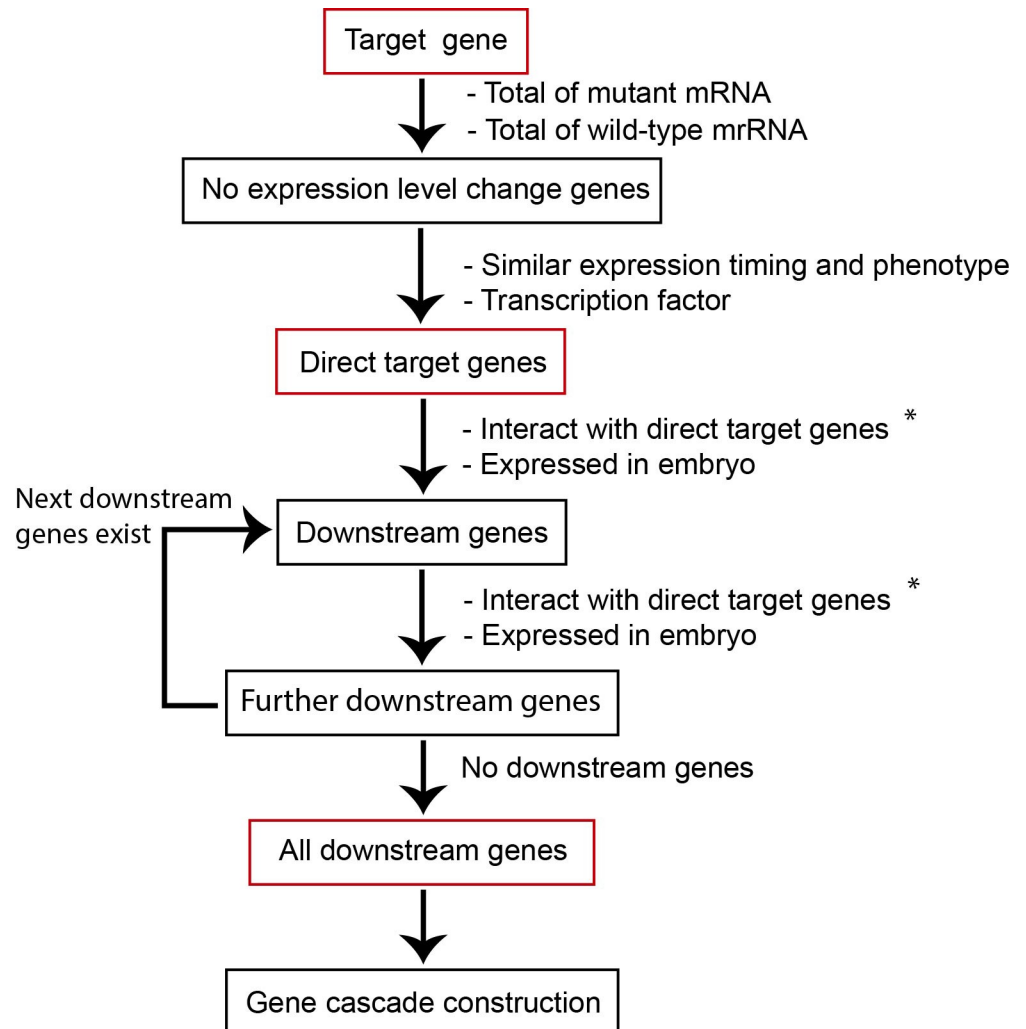
Fig 1. Schematic for prediction of the gene cascade. Application protocol for GCF. Genes in the predicted cascade are indicated by red frames. To identify the entire gene network, we repeatedly identified the downstream genes. The genes surrounded by black frames were required to identify the genes surrounded by red fames. The information of the labels with asterisks was extracted from WormBase.

https://doi.org/10.1371/journal.pone.0215187.g001

and query gene are simultaneously expressed, (4) confirm phenotypic similarity between a mutant of the candidate gene and that of the query gene. For identification of the downstream target of the polarity mediator-mediated gene cascade, we performed the following. First, the candidate genes of the target of the query gene were found from transcription factors and genes with no gene expression level fluctuations, as calculated by Eq 3, with the same cellular localization and phenotype. Query genes bind to target mRNAs to regulate their translation. Thus, the mRNA expression levels of the target genes showed no changes. In addition, the target genes needed to be expressed with the same timing as the query gene because the query gene is directly bound to the target mRNA. Furthermore, the gene cascade of the target genes was mostly consistent with the query gene cascade, suggesting that the target gene may have the same phenotype as the query gene. Therefore, to expand the gene cascade, the target gene should be a transcription factor with downstream genes.

Downstream gene identification by GCF

For identification of genes downstream to the gene cascade, we continuously performed the following. First, the genes from transcription or protein-protein interactions were extracted as downstream gene candidates of the target gene. Second, the expression timing of the candidate genes was checked. Only candidate genes noted as being expressed in early embryos or in embryos in WormBase were defined as downstream genes of the query's target genes. These two steps were repeated to obtain the next downstream genes. Finally, the procedure was repeated until there were no genes left to be extracted to obtain the final gene. Lastly, GCF output the cascade data (S3–S5 Tables). An example showing input obtained from output data from GCF to Cytoscape [43] is provided in S2 Fig.

Specific domain search from the constructed gene cascade

Each direct target gene was rooted, and the functions of their bottom genes were investigated using Pfam [41] in UniProt [42]. The P-values of the domains from the bottom gene products were evaluated using the same formula for Gene Ontology in Panther [44]. If the transcription-related domain was extracted from a cascade, the cascade was no longer considered since it would be functioning only after the early embryo stage.

Results and discussion

In this work, we developed a novel software that is useful to predict the gene cascade during *C. elegans* development. To evaluate the accuracy of the gene cascades that were predicted by GCF, we performed the following. First, we isolated candidate direct target genes of the polarity mediators using RNA-seq data by comparing the data for wild type and each mutant of the polarity mediators, SPN-4, MEX-1 and MEX-3. Next, GCF predicted the gene cascades from the candidate direct target genes localized at the top of the gene cascade. In this step, information that is stored in WormBase was used to predict the gene cascade. Finally, we assumed that the GCF gene cascade prediction might be accurate when the predicted gene cascade may have an essential role in early embryogenesis and may regulate any signaling pathway. Our purpose was assumed to be achieved when the predicted gene cascade was obtained previously or was considered to be useful for further detailed analysis of the cascades involved in *C. elegans* embryonic development. Although we used GCF to predict the polarity mediator gene cascade, it is applicable to analyze any other query gene.

Analysis of mRNA expression by comparative RNA-Seq

To explore polarity mediator-dependent mechanisms, the effects of deficiencies in polarity mediators were analyzed by performing RNA-Seq analysis in early embryos. From the results of comparative RNA-Seq of the wild-type strain and the *spn-4*, *mex-1*, and *mex-3* mutant strains, 15,288, 15,265, and 15,005 genes were identified, respectively (S8 Table). In these gene groups, expression level fluctuations were calculated by examining the median ± quartile deviations. From this analysis, 6,417 genes distributed at −0.65 < log2 (RNA expression level ratio) < 0.69 in the *spn-4* gene, 6,456 genes distributed at −0.74 < log2 (RNA expression level ratio) < 0.82 in the *mex-3* gene, and 6,491 genes distributed at −0.82 < log2 (RNA expression level ratio) < 1.10 in the *mex-1* gene were defined as genes showing no expression level variations (S9 Table).

Gene cascade prediction using Gene Cascade Finder

As shown in S1 Table, gene cascade prediction was performed by inputting data obtained by comparative RNA-Seq into GCF. GCF can predict gene cascades by continuously integrating the results from RNA-Seq along with data on gene expression and intermolecular interactions from WormBase. In total, 127, 180, and 226 gene cascades were predicted from 6,418, 6,457, and 8,513 genes from the comparative analysis of the *spn-4*, *mex-1*, and *mex-3* mutant strains, respectively (Fig 2, S3 Fig and S7–S9 Tables).

Extraction of gene cascade-specific domains using Gene Cascade Finder

The genes and domains located at the bottom of the gene cascade were extracted for functional analysis of the predicted gene cascade (S10 Table). Overall, 53, 146, and 143 genes with 32, 34, and 54 specific domains as the bottom gene were extracted from the gene cascades in the *spn-4*, *mex-1*, and *mex-3* mutants, respectively.

Domain analysis of *spn-4*, *mex-1*, and *mex-3* cascades

To predict the functions of the gene cascades, we focused on the functions of genes localized at the bottom of the gene cascade by analyzing the domains of the gene products using the Pfam protein family database. The functions of the 53 SPN-4-mediated genes were obtained as bottom genes to calculate the functional trends in the *spn-4* cascade. Within the 53 bottom genes, 32 domains were classified based on information from the Pfam database (Table 1) [41]. When we calculated the numbers of these genes, transcription and signal transduction were obtained at high frequency.

Similarly, in the gene cascade of 146 *mex-1*-mediated genes, 27 domains were classified and obtained from the domain analyses to have functions in early embryonic development, cell division, transcription, DNA replication, and signal transduction (Table 1). In contrast, in the gene cascades of the 143 *mex-3*-mediated genes, 54 domains were classified and obtained from the domain analyses to have functions in development, cell cycle, transcription, and signal transduction (Table 1).

Distinct availability between Gene Cascade Finder and a gene cascade prediction package, Cascade R

Among the previously established software packages, Cascade R was established to find out the gene cascade of a query gene [30]. Cascade R [30] constructs an intergenic network of knock-out genes from the results of DNA microarray. The difference in the datasets between Cascade R and GCF lies in that Cascade R constructs gene cascade from the series of microarray data, and GCF needs RNA-seq data and existing database information. Thus, Cascade R needs multiple timeline datasets of microarray. In contrast, using GCF, we can easily visualize the existing information and predict the direct target when a translational regulator is selected as a query gene. Although researchers need to refer to the function of the domain from Pfam, GCF can predict gene cascade by continuously integrating the results from RNA-Seq data and gene expression and intermolecular interaction data from WormBase. Thus, GCF offers inexpensive prediction of gene cascade.

Evaluation of GCF by assessment of gene cascades in the canonical Wnt signaling pathway

Next, we focused on genes involved in SPN-4-mediated signal transduction (Fig 3A). First, we found that MOM-2, a nematode homolog of the Wnt ligand, is involved in the signal
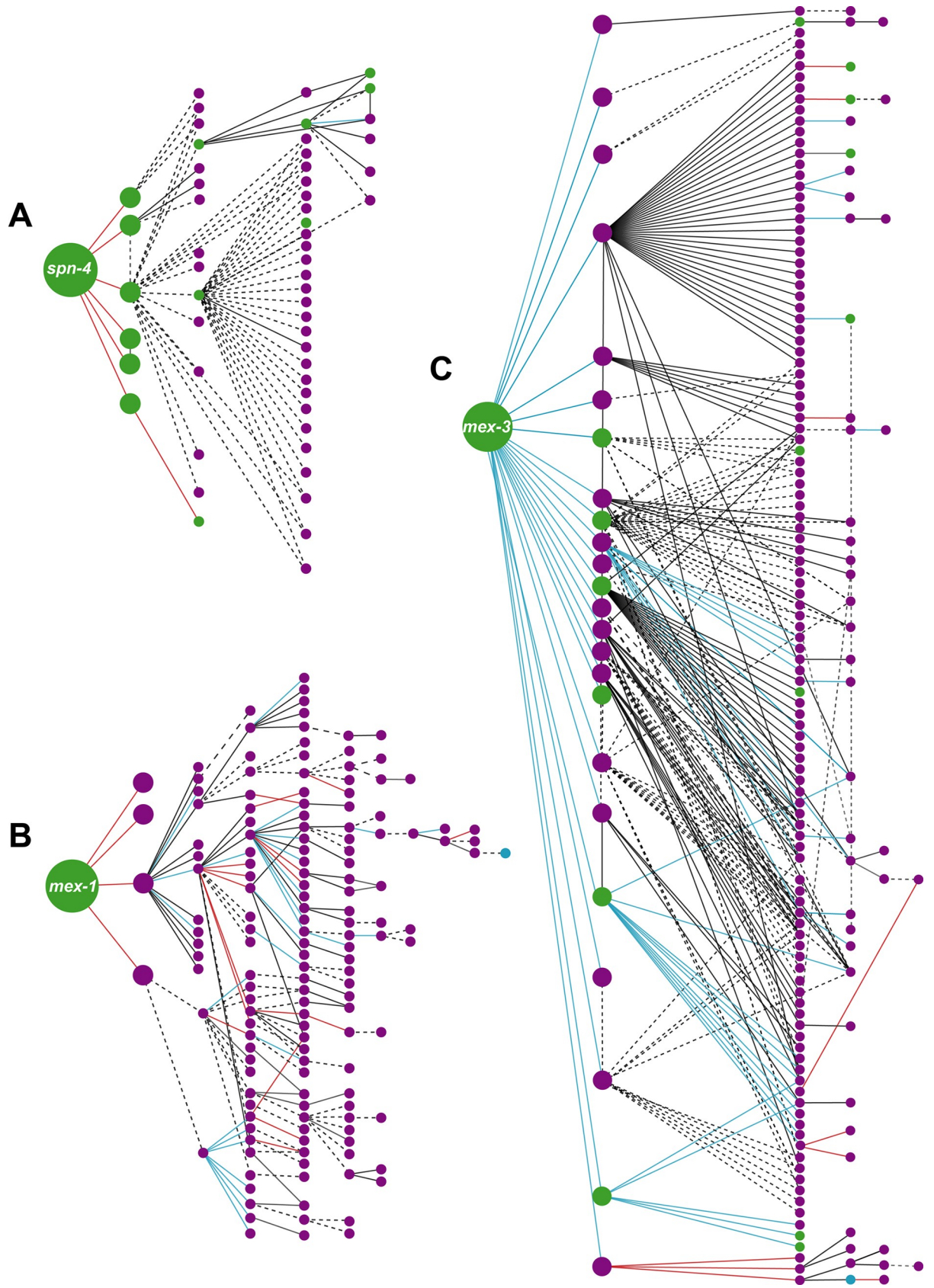
Fig 2. Schematic illustrations of polarity mediator-dependent gene cascades during *C. elegans* embryogenesis. Rendering of each gene cascade was performed using Cytoscape. Nodes indicate each gene in the cascade. Edges indicate the interactions between two genes. Large and intermediate nodes indicate the query genes and the direct target of the query genes, respectively. Other nodes indicate downstream genes. Green nodes indicate genes that are expressed during the early embryonic stage. Purple nodes indicate presumptive early embryonic genes. Red, blue, and black edges indicate positive regulation, negative regulation, and genetic interactions, respectively. Dotted lines indicate protein-protein interactions. (A) *spn-4*-mediated gene cascade. (B) *mex-1*-mediated gene cascade. (C) *mex-3*-mediated gene cascade.

transduction cascade (Table 1). Since both SPN-4 and MOM-2 were previously reported to regulate EMS cell lineage formation and spindle orientation [11, 45], we hypothesized that the SPN-4/MOM-2 gene cascade may have an essential role in early embryogenesis and may regulate the Wnt signaling pathway. Moreover, because OMA-1, MEX-1, and PIE-1, which are known to be essential for MOM-2 expression in embryonic development [45], were also identified in this pathway, the GCF-mediated gene cascade prediction was assumed to be accurate.

Similarly, *unc-37*, which encodes a Groucho/TLE homolog that suppresses Wnt signaling, was found as the "bottom gene" in the *spn-4* and *mex-3* cascades (Fig 3A and 3B). Thus, we propose that GCF-mediated gene cascade prediction may be useful for identification of gene cascades involved in *C. elegans* embryonic development.

Examples of the application of GCF for prediction of new biological functions involved in a gene cascade

Endoplasmic reticulum (ER) stress response pathway in a MEX-1-mediated gene cascade. Because cell division and DNA replication are regulated by a MEX-1-dependent cascade, we further focused on this signal transduction cascade. In the gene cascade related to signal transduction, *paqr-2*, which encodes an adiponectin receptor, was isolated (Fig 3C). Interestingly, a stress response pathway is known to regulate stress responses in both mouse and *C. elegans* embryogenesis [45, 46]. Thus, an evolutionarily conserved gene cascade against environmental stress may be identified using GCF. Moreover, because adiponectin receptor regulates insulin signaling [47], it is likely that MEX-1/PAQR-2-mediated gene cascades may be involved in ER stress tolerance signaling within or parallel to the insulin signaling pathway during embryogenesis.

***lin-35*, *hsp-3*, and *gpa-12* in a MEX-1/DPY-23-mediated gene cascade in neuronal development.** In a MEX-1/DPY-23-mediated gene cascade, the *mex-1*, *lin-14*, *let-60*, *ces-2*, *unc-13*, and *dpy-23* mutants were shown to exhibit specific phenotypes in neuronal development (Fig

Table 1. Scores of the functional characterization of *spn-1*-, *mex-1*-, and *mex-3*-mediated gene cascades by domain analysis.

	spn-4	*mex-1*	*mex-3*
Transcription	25	6	4
Signal transduction	5	3	12
Development	-	2	7
Cell cycle	-	3	5
Cell division	-	-	-
DNA replication	-	2	3
Transport	-	3	-
Others	2	12	18
Unknown	-	3	4

Properties of the gene product of the bottom genes were calculated by domain analysis. The sum of the characteristic features of each cascade (P < 0.05) was then calculated. Domains with a score less than 1 were classified as "Others".
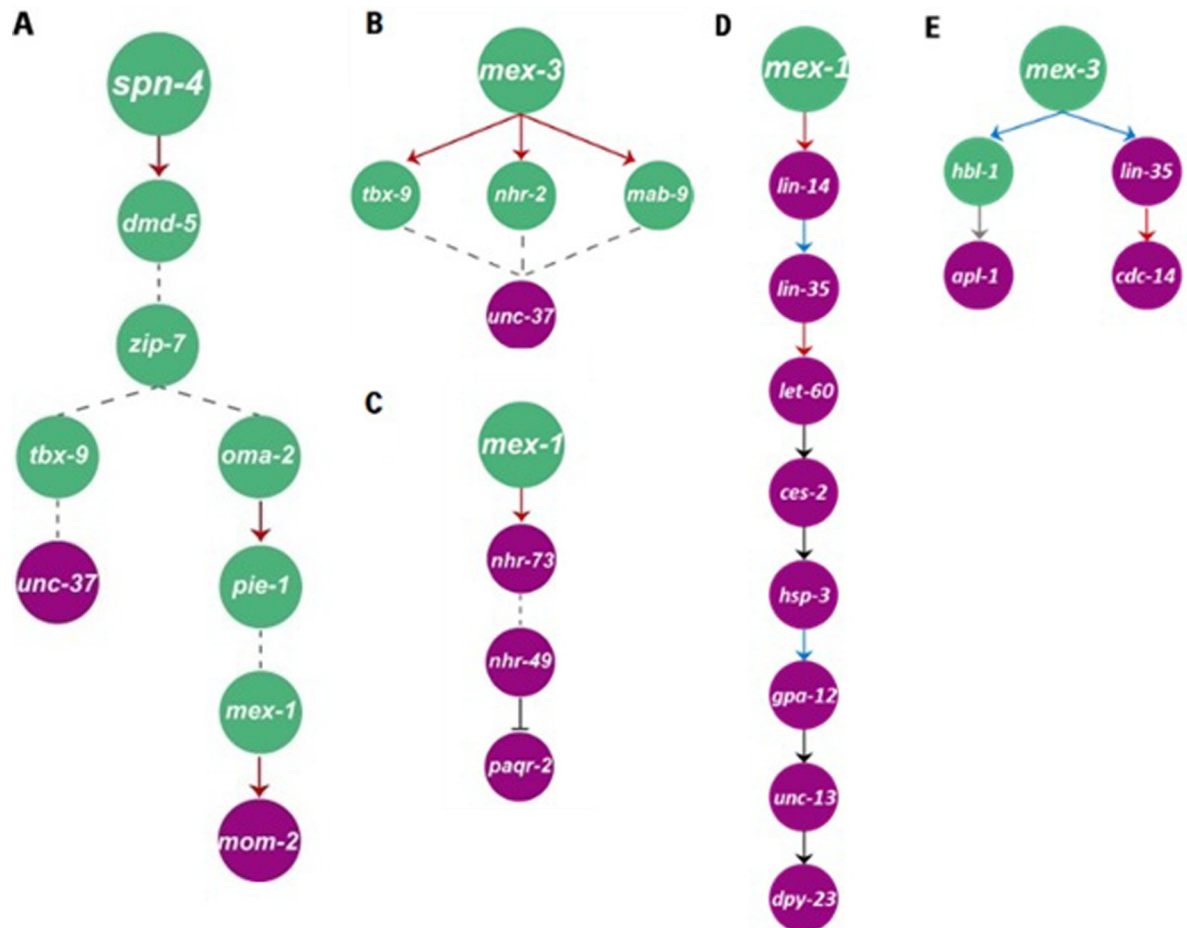
Fig 3. Typical examples of SPN-4, MEX-1, and MEX-3-dependent gene cascades. (A) SPN-4-mediated gene cascade regulates Wnt signaling. (B) MEX-3-mediated gene cascade negatively regulates Wnt signaling. (C) MEX-1-mediated gene cascade regulates endoplasmic reticulum-associated degradation (ERAD) of folding-deficient proteins. (D) MEX-1-mediated gene cascade regulates neuronal development. (E) MEX-3-mediated gene cascade regulates neuronal development.

https://doi.org/10.1371/journal.pone.0215187.g003

3D) [12, 48–50]. Thus, six of the nine genes of this gene cascade were shown to have essential roles in neuronal development, indicating that MEX-1/DPY-23-mediated gene cascades may regulate neuronal function. Although their roles in neuronal development have not yet been investigated, locomotion defects have been reported in *hsp-3* and *gpa-12* mutants [51, 52]. Similarly, the *lin-35(n745)* mutation has been shown to enhance the neuronal phenotype of the neuronal regulator genes *dpy-13* and *unc-104* [53]. Thus, *lin-35*, *hsp-3*, and *gpa-12* may be involved in a DPY-23-mediated gene cascade in neuronal development in embryos [50]. However, further studies are required to examine this possibility.

MEX-3/APL-1-mediated neuronal patterning and MEX-3/CDC-14-mediated cell fate determination in the MEX-3-mediated gene cascade. Because MEX-3 is specifically expressed in AB cells at the four-cell stage, spatiotemporal-regulated synaptic formation defects in *hbl-1* mutants and *apl-1*-dependent embryonic neuronal patterning may be elucidated by identifying MEX-3/APL-1-mediated gene cascades (Fig 3E) [54–56]. In parallel, when we focused on the MEX-3/CDC-14-mediated gene cascade (Fig 3E), CDC-14B, a zebrafish homolog of CDC-14, was shown to be involved in formation of the cilium in sensory neurons [55]. Because sensory neurons have cilia in *C. elegans* [57], CDC-14 may be involved in

an evolutionarily conserved signaling pathway. Similarly, the *lin-35(n745)* mutation was shown to enhance the neuronal phenotype of neuronal regulator genes [53]. Thus, *lin-35* may be involved in a MEX-3/CDC-14-mediated gene cascade in sensory neuron development. Accordingly, our findings suggested that GCF may be useful for predicting the comprehensive functions of query genes and for identification of new genes involved in known gene cascades.

Conclusion

In this study, we created a software program called GCF, which could comprehensively identify genes downstream of the query genes by integrating RNA-Seq data and previously characterized data from WormBase. Using GCF, we analyzed gene cascades of the polarity mediator proteins SPN-4, MEX-1, and MEX-3, and identified 127, 180, and 226 putative gene cascades, respectively. By analyzing the functions of these gene cascades, we confirmed that SPN-4 and MEX-3 regulate the canonical Wnt pathway during embryonic development. Furthermore, we found that the ER stress response and motor neuron development are regulated by MEX-1-dependent cascades, and that neural development is regulated by MEX-3-dependent cascades. Although we used GCF only to evaluate SPN-4, MEX-1, and MEX-3 functions in this study, the method is applicable for other translation or transcription factors involved in early embryogenesis. In addition, GCF provides a general method for predicting the functions of genes involved in a gene cascade during *C. elegans* embryonic development. Taken together, we propose that our strategy using the GCF tool offers a reliable approach for comprehensively identifying networks of embryo-specific gene cascades in *C. elegans*. Importantly, GCF can also be applied to humans and other model organisms such as mice and *Drosophila*.

In the future, by adding threshold for quantitative gene expression analysis of RNA-sequencing data and expanding the algorithm to fit the cell lineage-segregation of *C. elegans* [3], we will be able to predict more reliable quantitative-data-oriented precise gene cascades reflecting four-dimensional (spatial and temporal) regulation [58]. Combinational analysis of improved GCF and molecular biology techniques such as RNA-pull down assays, fluorescent *in situ* hybridization, and phenotypic characterization of the mutants may be required to build a more reliable regulatory network for these gene cascades [59, 60].

Supporting information

S1 Table. List of the genes identified by RNA-seq of wild type and *sin-3* mutant.
(XLSX)

S2 Table. GCF input data example in case of *spn-4*.
(XLSX)

S3 Table. GCF output data example of the *spn-4* cascades.
(XLSX)

S4 Table. GCF output data example of the *mex-1* cascades.
(XLSX)

S5 Table. GCF output data example of the *mex-3* cascades.
(XLSX)

S6 Table. GCF output data example to input cytoscape in case of *spn-4*.
(XLSX)

S7 Table. GCF output data example of the *spn-4* cascade and its bottom genes domain.
(XLSX)

S8 Table. List of the genes identified by RNA-Seq.
(XLSX)

S9 Table. List of the genes that exhibit no change of the expression level.
(XLSX)

S10 Table. List of the gene cascade specific domain.
(XLSX)

S1 Fig. Guideline for RNA-seq analysis and comparative gene expression analysis of wild type and mutants.
(JPG)

S2 Fig. Representative example of the graphic output from GCF. Graphical examples from the GCF software were further processed using Cytoscape, allowing for identification of an output without adding a new input.
(JPG)

S3 Fig. Schematic illustrations of polarity mediator-dependent gene cascades during *C. elegans* embryogenesis. Rendering of each gene cascade was performed using Cytoscape. Nodes indicate each gene in the cascade. Edges indicate the interactions between two genes. Large and intermediate nodes indicate the query genes and the direct target of the query genes, respectively. Other nodes indicate downstream genes. Green nodes indicate genes that are expressed during the early embryonic stage. Purple nodes indicate presumptive early embryonic genes. Red, blue, and black edges indicate positive regulation, negative regulation, and genetic interactions, respectively. Dotted lines indicate protein-protein interactions. (A) *spn-4*-mediated gene cascade. (B) *mex-1*-mediated gene cascade. (C) *mex-3*-mediated gene cascade.
(JPG)

Acknowledgments

Author Contributions

Conceptualization: Yusuke Nomoto, Masahiro Ito.

Data curation: Yusuke Nomoto, Masahiro Ito.

Formal analysis: Yusuke Nomoto, Yukihiro Kubota, Aimi Tomita, Takehiro Oshime, Hiroki Yamashita, Masahiro Ito.

Funding acquisition: Masahiro Ito.

Methodology: Yuto Ohnishi, Muhamad Fahmi.

Project administration: Masahiro Ito.

Resources: Aimi Tomita, Takehiro Oshime, Hiroki Yamashita.

Software: Yusuke Nomoto, Kota Kasahara.

Supervision: Masahiro Ito.

Validation: Yuto Ohnishi, Muhamad Fahmi.

Visualization: Yusuke Nomoto, Kota Kasahara, Muhamad Fahmi.

Writing – original draft: Yusuke Nomoto, Yukihiro Kubota, Masahiro Ito.

Writing – review & editing: Yusuke Nomoto, Yukihiro Kubota, Muhamad Fahmi, Masahiro Ito.

References

1. English J, Pearson G, Wilsbacher J, Swantek J, Karandikar M, Xu S, et al. New insights into the control of MAP kinase pathways. Experimental cell research. 1999; 253(1):255–70. Epub 1999/12/02. https://doi.org/10.1006/excr.1999.4687 PMID: 10579927.

2. Maduro MF, Rothman JH. Making worm guts: the gene regulatory network of the Caenorhabditis elegans endoderm. Developmental biology. 2002; 246(1):68–85. Epub 2002/05/25. https://doi.org/10.1006/dbio.2002.0655 PMID: 12027435.

3. Sulston JE, Schierenberg E, White JG, Thomson JN. The embryonic cell lineage of the nematode Caenorhabditis elegans. Developmental biology. 1983; 100(1):64–119. Epub 1983/11/01. https://doi.org/10.1016/0012-1606(83)90201-4 PMID: 6684600.

4. Lai CH, Chou CY, Ch'ang LY, Liu CS, Lin W. Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics. Genome research. 2000; 10(5):703–13. Epub 2000/05/16. https://doi.org/10.1101/gr.10.5.703 PMID: 10810093; PubMed Central PMCID: PMCPMC310876.

5. Schubert CM, Lin R, de Vries CJ, Plasterk RH, Priess JR. MEX-5 and MEX-6 function to establish soma/germline asymmetry in early C. elegans embryos. Molecular cell. 2000; 5(4):671–82. Epub 2000/07/06. PMID: 10882103.

6. Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J. Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. Nature. 2000; 408 (6810):325–30. Epub 2000/12/01. https://doi.org/10.1038/35042517 PMID: 11099033.

7. Mello CC, Draper BW, Krause M, Weintraub H, Priess JR. The pie-1 and mex-1 genes and maternal control of blastomere identity in early C. elegans embryos. Cell. 1992; 70(1):163–76. Epub 1992/07/10. https://doi.org/10.1016/0092-8674(92)90542-k PMID: 1623520.

8. Ogura K, Kishimoto N, Mitani S, Gengyo-Ando K, Kohara Y. Translational control of maternal glp-1 mRNA by POS-1 and its interacting protein SPN-4 in Caenorhabditis elegans. Development (Cambridge, England). 2003; 130(11):2495–503. Epub 2003/04/19. https://doi.org/10.1242/dev.00469 PMID: 12702662.

9. Tsuboi D, Qadota H, Kasuya K, Amano M, Kaibuchi K. Isolation of the interacting molecules with GEX-3 by a novel functional screening. Biochemical and biophysical research communications. 2002; 292 (3):697–701. Epub 2002/04/02. https://doi.org/10.1006/bbrc.2002.6717 PMID: 11922622.

10. Dorfman M, Gomes JE, O'Rourke S, Bowerman B. Using RNA interference to identify specific modifiers of a temperature-sensitive, embryonic-lethal mutation in the Caenorhabditis elegans ubiquitin-like Nedd8 protein modification pathway E1-activating gene rfl-1. Genetics. 2009; 182(4):1035–49. Epub 2009/06/17. https://doi.org/10.1534/genetics.109.104885 PMID: 19528325; PubMed Central PMCID: PMCPMC2728846.

11. Gomes JE, Encalada SE, Swan KA, Shelton CA, Carter JC, Bowerman B. The maternal gene spn-4 encodes a predicted RRM protein required for mitotic spindle orientation and cell fate patterning in early C. elegans embryos. Development (Cambridge, England). 2001; 128(21):4301–14. Epub 2001/10/31. PMID: 11684665.

12. Hallam SJ, Goncharov A, McEwen J, Baran R, Jin Y. SYD-1, a presynaptic protein with PDZ, C2 and rhoGAP-like domains, specifies axon identity in C. elegans. Nature neuroscience. 2002; 5(11):1137–46. Epub 2002/10/16. https://doi.org/10.1038/nn959 PMID: 12379863.

13. Guedes S, Priess JR. The C. elegans MEX-1 protein is present in germline blastomeres and is a P granule component. Development (Cambridge, England). 1997; 124(3):731–9. Epub 1997/02/01. PMID: 9043088.

14. Huang NN, Mootz DE, Walhout AJ, Vidal M, Hunter CP. MEX-3 interacting proteins link cell polarity to asymmetric gene expression in Caenorhabditis elegans. Development (Cambridge, England). 2002; 129(3):747–59. Epub 2002/02/07. PMID: 11830574.

15. Pagano JM, Farley BM, McCoig LM, Ryder SP. Molecular basis of RNA recognition by the embryonic polarity determinant MEX-5. The Journal of biological chemistry. 2007; 282(12):8883–94. Epub 2007/02/01. https://doi.org/10.1074/jbc.M700079200 PMID: 17264081.

16. Sonnichsen B, Koski LB, Walsh A, Marschall P, Neumann B, Brehm M, et al. Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. Nature. 2005; 434(7032):462–9. Epub 2005/03/26. https://doi.org/10.1038/nature03353 PMID: 15791247.

17. Consortium. CeS. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science (New York, NY). 1998; 282(5396):2012–8. Epub 1998/12/16. https://doi.org/10.1126/science.282.5396.2012 PMID: 9851916.

18. Bumgarner R. Overview of DNA microarrays: types, applications, and their future. Current protocols in molecular biology. 2013;Chapter 22:Unit 22.1. Epub 2013/01/05. https://doi.org/10.1002/0471142727.mb2201s101 PMID: 23288464; PubMed Central PMCID: PMCPMC4011503.

19. Han X, Aslanian A, Yates JR 3rd., Mass spectrometry for proteomics. Current opinion in chemical biology. 2008; 12(5):483–90. Epub 2008/08/23. https://doi.org/10.1016/j.cbpa.2008.07.024 PMID: 18718552; PubMed Central PMCID: PMCPMC2642903.

20. Agrawal N, Dasaradhi PV, Mohmmed A, Malhotra P, Bhatnagar RK, Mukherjee SK. RNA interference: biology, mechanism, and applications. Microbiology and molecular biology reviews: MMBR. 2003; 67(4):657–85. Epub 2003/12/11. https://doi.org/10.1128/MMBR.67.4.657-685.2003 PMID: 14665679; PubMed Central PMCID: PMCPMC309050.

21. Bazan J, Calkosinski I, Gamian A. Phage display—a powerful technique for immunotherapy: 1. Introduction and potential of therapeutic applications. Human vaccines & immunotherapeutics. 2012; 8(12):1817–28. Epub 2012/08/22. https://doi.org/10.4161/hv.21703 PMID: 22906939; PubMed Central PMCID: PMCPMC3656071.

22. Bruckner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful tool for systems biology. International journal of molecular sciences. 2009; 10(6):2763–88. Epub 2009/07/08. https://doi.org/10.3390/ijms10062763 PMID: 19582228; PubMed Central PMCID: PMCPMC2705515.

23. Stein L, Sternberg P, Durbin R, Thierry-Mieg J, Spieth J. WormBase: network access to the genome and biology of Caenorhabditis elegans. Nucleic acids research. 2001; 29(1):82–6. Epub 2000/01/11. https://doi.org/10.1093/nar/29.1.82 PMID: 11125056; PubMed Central PMCID: PMCPMC29781.

24. Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2018; 24(4):335–41. Epub 2017/10/28. https://doi.org/10.1016/j.cmi.2017.10.013 PMID: 29074157; PubMed Central PMCID: PMCPMC5857210.

25. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics. 2009; 10(1):57–63. Epub 2008/11/19. https://doi.org/10.1038/nrg2484 PMID: 19015660; PubMed Central PMCID: PMCPMC2949280.

26. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics (Oxford, England). 2012; 28(16):2184–5. Epub 2012/06/30. https://doi.org/10.1093/bioinformatics/bts356 PMID: 22743226.

27. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols. 2012; 7(3):562–78. Epub 2012/03/03. https://doi.org/10.1038/nprot.2012.016 PMID: 22383036; PubMed Central PMCID: PMCPMC3334321.

28. Nicolae M, Mangul S, Mandoiu, II, Zelikovsky A. Estimation of alternative splicing isoform frequencies from RNA-Seq data. Algorithms for molecular biology: AMB. 2011; 6(1):9. Epub 2011/04/21. https://doi.org/10.1186/1748-7188-6-9 PMID: 21504602; PubMed Central PMCID: PMCPMC3107792.

29. Alexa A, Rahnenfuhrer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics (Oxford, England). 2006; 22(13):1600–7. Epub 2006/04/12. https://doi.org/10.1093/bioinformatics/btl140 PMID: 16606683.

30. Jung N, Bertrand F, Bahram S, Vallat L, Maumy-Bertrand M. Cascade: a R package to study, predict and simulate the diffusion of a signal through a temporal gene network. Bioinformatics (Oxford, England). 2014; 30(4):571–3. Epub 2013/12/07. https://doi.org/10.1093/bioinformatics/btt705 PMID: 24307703.

31. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic acids research. 2015; 43(Database issue):D447–52. Epub 2014/10/30. https://doi.org/10.1093/nar/gku1003 PMID: 25352553; PubMed Central PMCID: PMCPMC4383874.

32. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic acids research. 2006; 34(Database issue):D535–9. Epub 2005/12/31. https://doi.org/10.1093/nar/gkj109 PMID: 16381927; PubMed Central PMCID: PMCPMC1347471.

33. Mitani S., Nematode an experimental animal in the national BioResource project. Experimental animals. 2009; 58(4):351–6. Epub 2009/08/06. https://doi.org/10.1538/expanim.58.351 PMID: 19654432.

34. Motohashi T, Tabara H, Kohara Y. Protocols for large scale in situ hybridization on C. elegans larvae. WormBook: the online review of C elegans biology. 2006:1–8. Epub 2007/12/01. https://doi.org/10. 1895/wormbook.1.103.1 PMID: 18050447; PubMed Central PMCID: PMCPMC4781301.

35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England). 2014; 30(15):2114–20. Epub 2014/04/04. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404; PubMed Central PMCID: PMCPMC4103590.

36. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014; 15(12):550. Epub 2014/12/18. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281; PubMed Central PMCID: PMCPMC4302049.

37. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods. 2008; 5(7):621–8. Epub 2008/06/03. https://doi.org/10.1038/nmeth.1226 PMID: 18516045.

38. Lamarre S, Frasse P, Zouine M, Labourdette D, Sainderichin E, Hu G, et al. Optimization of an RNA-Seq Differential Gene Expression Analysis Depending on Biological Replicate Number and Library Size. Frontiers in plant science. 2018; 9:108. Epub 2018/03/02. https://doi.org/10.3389/fpls.2018.00108 PMID: 29491871; PubMed Central PMCID: PMCPMC5817962.

39. Raabe CA, Voss R, Kummerfeld DM, Brosius J, Galiveti CR, Wolters A, et al. Ectopic expression of Snord115 in choroid plexus interferes with editing but not splicing of 5-Ht2c receptor pre-mRNA in mice. Scientific reports. 2019; 9(1):4300. Epub 2019/03/14. https://doi.org/10.1038/s41598-019-39940-6 PMID: 30862860; PubMed Central PMCID: PMCPMC6414643.

40. Consortium GO. Gene Ontology Consortium: going forward. Nucleic acids research. 2015; 43(Database issue):D1049–56. Epub 2014/11/28. https://doi.org/10.1093/nar/gku1179 PMID: 25428369; PubMed Central PMCID: PMCPMC4383973.

41. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, et al. The Pfam protein families database. Nucleic acids research. 2004; 32(Database issue):D138–41. Epub 2003/12/19. https://doi.org/10.1093/nar/gkh121 PMID: 14681378; PubMed Central PMCID: PMCPMC308855.

42. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. Nucleic acids research. 2004; 32(Database issue):D115–9. Epub 2003/12/19. https://doi.org/10.1093/nar/gkh131 PMID: 14681372; PubMed Central PMCID: PMCPMC308865.

43. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003; 13 (11):2498–504. Epub 2003/11/05. https://doi.org/10.1101/gr.1239303 PMID: 14597658; PubMed Central PMCID: PMCPMC403769.

44. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic acids research. 2013; 41(Database issue):D377–86. Epub 2012/11/30. https://doi.org/10.1093/nar/gks1118 PMID: 23193289; PubMed Central PMCID: PMCPMC3531194.

45. Thorpe CJ, Schlesinger A, Carter JC, Bowerman B. Wnt signaling polarizes an early C. elegans blastomere to distinguish endoderm from mesoderm. Cell. 1997; 90(4):695–705. Epub 1997/08/22. https://doi.org/10.1016/s0092-8674(00)80530-9 PMID: 9288749.

46. Abraham T, Pin CL, Watson AJ. Embryo collection induces transient activation of XBP1 arm of the ER stress response while embryo vitrification does not. Molecular human reproduction. 2012; 18(5):229–42. Epub 2011/12/14. https://doi.org/10.1093/molehr/gar076 PMID: 22155729.

47. Kyriakakis E, Charmpilas N, Tavernarakis N. Differential adiponectin signalling couples ER stress with lipid metabolism to modulate ageing in C. elegans. Scientific reports. 2017; 7(1):5115. Epub 2017/07/13. https://doi.org/10.1038/s41598-017-05276-2 PMID: 28698593; PubMed Central PMCID: PMCPMC5505976.

48. Desai C, Horvitz HR. Caenorhabditis elegans mutants defective in the functioning of the motor neurons responsible for egg laying. Genetics. 1989; 121(4):703–21. Epub 1989/04/01. PMID: 2721931; PubMed Central PMCID: PMCPMC1203655.

49. Kohn RE, Duerr JS, McManus JR, Duke A, Rakow TL, Maruyama H, et al. Expression of multiple UNC-13 proteins in the Caenorhabditis elegans nervous system. Molecular biology of the cell. 2000; 11 (10):3441–52. Epub 2000/10/12. https://doi.org/10.1091/mbc.11.10.3441 PMID: 11029047; PubMed Central PMCID: PMCPMC15005.

50. Pan CL, Baum PD, Gu M, Jorgensen EM, Clark SG, Garriga G. C. elegans AP-2 and retromer control Wnt signaling by regulating mig-14/Wntless. Developmental cell. 2008; 14(1):132–9. Epub 2007/12/28. https://doi.org/10.1016/j.devcel.2007.12.001 PMID: 18160346; PubMed Central PMCID: PMCPMC2709403.

51. Gottschalk A, Almedom RB, Schedletzky T, Anderson SD, Yates JR 3rd, Schafer WR. Identification and characterization of novel nicotinic receptor-associated proteins in Caenorhabditis elegans. The

EMBO journal. 2005; 24(14):2566–78. Epub 2005/07/02. https://doi.org/10.1038/sj.emboj.7600741 PMID: 15990870; PubMed Central PMCID: PMCPMC1176467.

52. Yau DM, Yokoyama N, Goshima Y, Siddiqui ZK, Siddiqui SS, Kozasa T. Identification and molecular characterization of the G alpha12-Rho guanine nucleotide exchange factor pathway in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100 (25):14748–53. Epub 2003/12/06. https://doi.org/10.1073/pnas.2533143100 PMID: 14657363; PubMed Central PMCID: PMCPMC299794.

53. Lehner B, Calixto A, Crombie C, Tischler J, Fortunato A, Chalfie M, et al. Loss of LIN-35, the Caenorhabditis elegans ortholog of the tumor suppressor p105Rb, results in enhanced RNA interference. Genome biology. 2006; 7(1):R4. Epub 2006/03/02. https://doi.org/10.1186/gb-2006-7-1-r4 PMID: 16507136; PubMed Central PMCID: PMCPMC1431716.

54. Bowerman B, Ingram MK, Hunter CP. The maternal par genes and the segregation of cell fate specification activities in early Caenorhabditis elegans embryos. Development (Cambridge, England). 1997; 124 (19):3815–26. Epub 1997/11/21. PMID: 9367437.

55. Clement A, Solnica-Krezel L, Gould KL. Functional redundancy between Cdc14 phosphatases in zebrafish ciliogenesis. Developmental dynamics: an official publication of the American Association of Anatomists. 2012; 241(12):1911–21. Epub 2012/10/03. https://doi.org/10.1002/dvdy.23876 PMID: 23027426; PubMed Central PMCID: PMCPMC3508521.

56. Hornsten A, Lieberthal J, Fadia S, Malins R, Ha L, Xu X, et al. APL-1, a Caenorhabditis elegans protein related to the human beta-amyloid precursor protein, is essential for viability. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104(6):1971–6. Epub 2007/02/ 03. https://doi.org/10.1073/pnas.0603997104 PMID: 17267616; PubMed Central PMCID: PMCPMC1794273.

57. Bae YK, Barr MM. Sensory roles of neuronal cilia: cilia development, morphogenesis, and function in C. elegans. Frontiers in bioscience: a journal and virtual library. 2008; 13:5959–74. Epub 2008/05/30. PMID: 18508635; PubMed Central PMCID: PMCPMC3124812.

58. Hunter CP, Kenyon C. Spatial and temporal controls target pal-1 blastomere-specification activity to a single blastomere lineage in C. elegans embryos. Cell. 1996; 87(2):217–26. Epub 1996/10/18. https:// doi.org/10.1016/s0092-8674(00)81340-9 PMID: 8861906.

59. Iioka H, Macara IG. Detection of RNA-Protein Interactions Using Tethered RNA Affinity Capture. Methods in molecular biology (Clifton, NJ). 2015; 1316:67–73. Epub 2015/05/15. https://doi.org/10.1007/ 978-1-4939-2730-2_6 PMID: 25967053; PubMed Central PMCID: PMCPMC6047865.

60. Langer-Safer PR, Levine M, Ward DC. Immunological method for mapping genes on Drosophila polytene chromosomes. Proceedings of the National Academy of Sciences of the United States of America. 1982; 79(14):4381–5. Epub 1982/07/01. https://doi.org/10.1073/pnas.79.14.4381 PMID: 6812046; PubMed Central PMCID: PMCPMC346675.