Supplementary Information

Inventory of Supplementary Materials

1. Supplementary Figures

2. Supplementary Tables

3. Extended Experimental Procedures

4. Supplementary References

1. Supplementary Figures

Figure S1

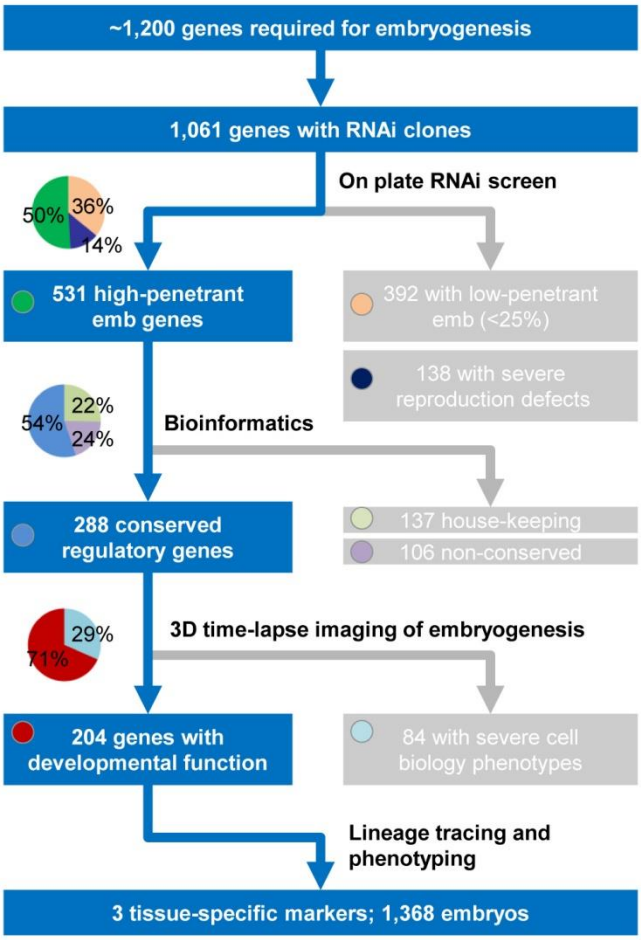


Figure S1 Genome-wide RNAi Screen and Characterization of Conserved Developmental Regulatory Genes, related to Figure 1

~1200 genes that are essential for embryogenesis in the wild-type genetic background identified in previous RNAi-based experiments were extracted from WormBase. First, RNAi screens were performed for 1061 genes with RNAi clones three times to identify penetrant Emb genes. Genes with low penetrance and reproduction phenotypes were excluded. Second, bioinformatic analysis for penetrant genes further classified genes into three categories: conserved regulatory genes, non-conserved genes and house-keeping function genes. Third, 3D time-lapse imaging was applied to all conserved regulatory genes to further exclude genes induced severe cell division phenotypes such as cytokinesis and nuclei morphology and separation defects. The remaining 204 Emb genes with developmental progression effects were used for in-depth lineage-based phenotype analysis. Pie charts show the frequency of each gene categories characterized during series of analyses. Gray boxes highlight excluded gene categories.

Figure S2

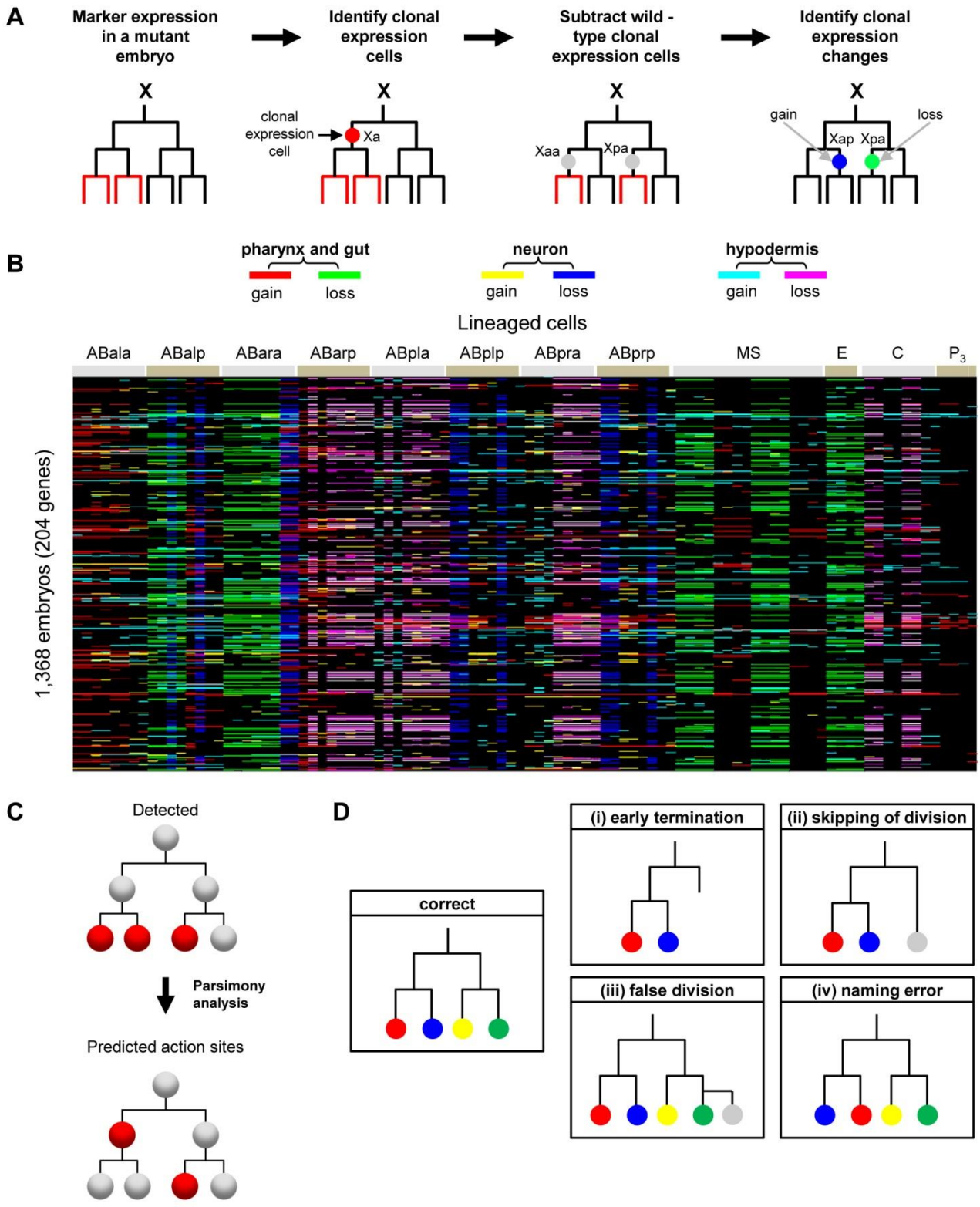

A
Marker expression in a mutant embryo
Identify clonal expression cells
Subtract wild - type clonal expression cells
Identify clonal expression changes
X
clonal expression cell
Xa
Xaa
Xpa
gain
loss
Xap
B
pharynx and gut
neuron
hypodermis
Lineaged cells
ABala
ABalp
ABara
ABarp
ABpla
ABplp
ABpra
ABprp
MS
E
C
P3
1,368 embryos (204 genes)
C
Detected
Parsimony analysis
Predicted action sites
D
correct
(i) early termination
(ii) skipping of division
(iii) false division
(iv) naming error

Figure S2 Phenotype Analysis and Quality Control, Related to Figure 1

(A) Identification of clonal expression changes. A sublineage with 8 terminal cells is used as an example to illustrate the general principle. “X” represents the the founder cell name for the sublineage. Red and black lines represent cells with or without marker expression respectively. Circles represent clonal expressing cells identified by clonal analysis. Details see Extended Experimental Procedure.

(B) Heatmap shows clonal changes of tissue marker expression in 1,368 embryos. Each row shows an embryo and color bars represent the position of fate changes in the cell lineage with names of founder cells indicated above. Different types of changes are shown in different colors. See also Table S2.

(C) Schematic representation of parsimony to predict primary fate changes in progenitor cells. Upper panel shows gene action sites (red) on the 12 founder cells before parsimony. Lower panel shows the results after parsimony in which two daughter cell sites are mapped onto the mother cell.

(D) Schematic representation of typical types of lineage errors. For each panel, colored circles at the bottom represent cell identities. Early termination: lose track of a cell; skipping of division: fail to detect a real cell division event; false division: cell division is assigned to a non-dividing cell; naming error: lineage identity was signed incorrectly. Figure shows a swap of red and blue cycles compared to the corrected one.

Figure S3

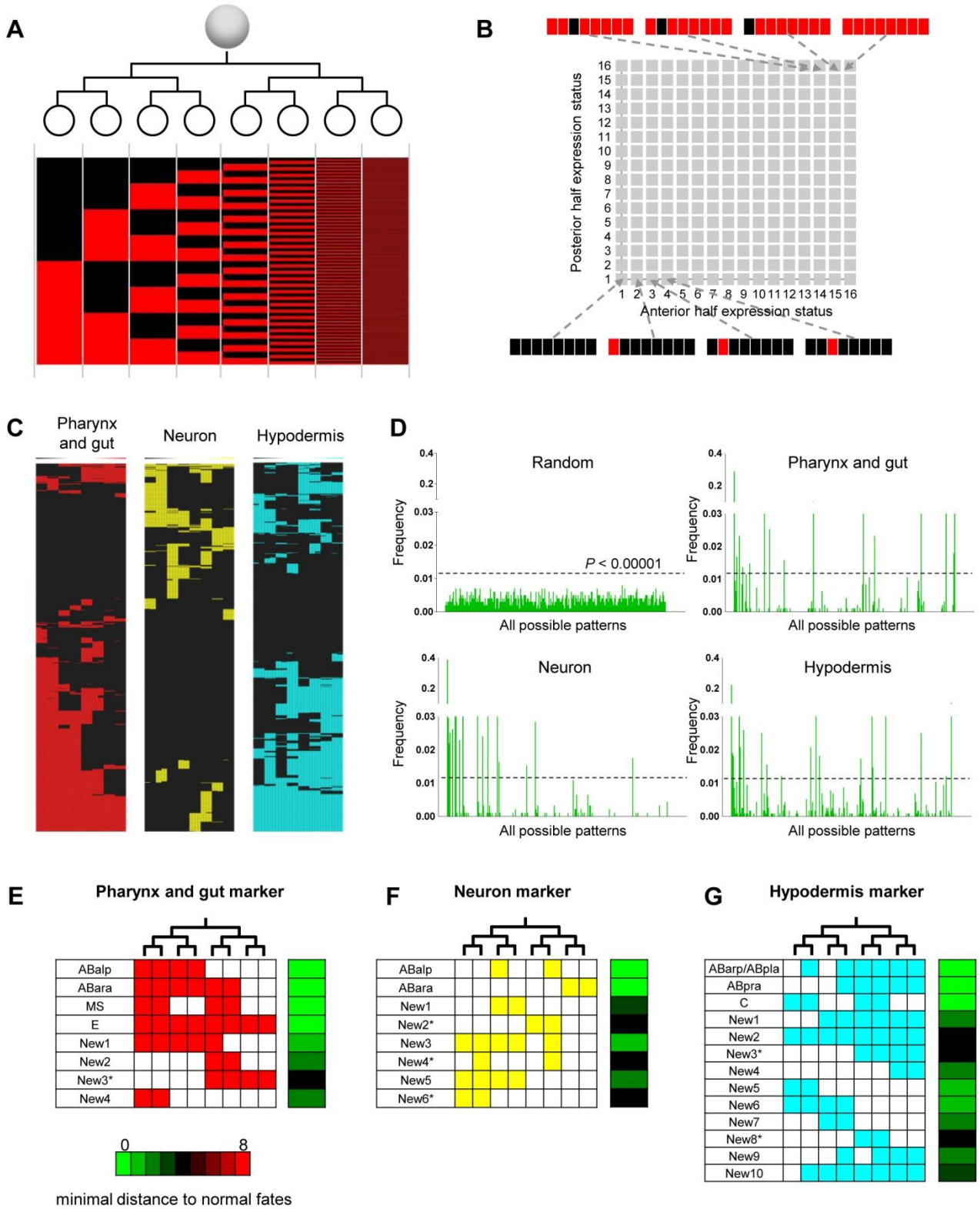

A
B
Posterior half expression status
Anterior half expression status
C
Pharynx and gut
Neuron
Hypodermis
D
Random
Pharynx and gut
Neuron
Hypodermis
Frequency
All possible patterns
P < 0.00001
E
Pharynx and gut marker
ABalp
ABara
MS
E
New1
New2
New3*
New4
F
Neuron marker
ABalp
ABara
New1
New2*
New3
New4*
New5
New6*
G
Hypodermis marker
ABarp/ABpla
ABpra
C
New1
New2
New3*
New4
New5
New6
New7
New8*
New9
New10
0
8
minimal distance to normal fates

Figure S3 Quantitative Analysis of Phenotype Distribution, related to Figure 2

(A) All possible types of marker expression patterns. Lineage structure is shown above with gray circle representing cell and open circles representing the 8 lowest clonal sites (each with 4 terminal cells) of expression changes for a typical cell lineage with 32 terminal cells produced by five rounds of cell division. The heatmap shown below lists all 256 possible types of marker expression patterns with each defines a possible cell fate. Red denotes expression and black denotes non-expression of a given tissue marker.

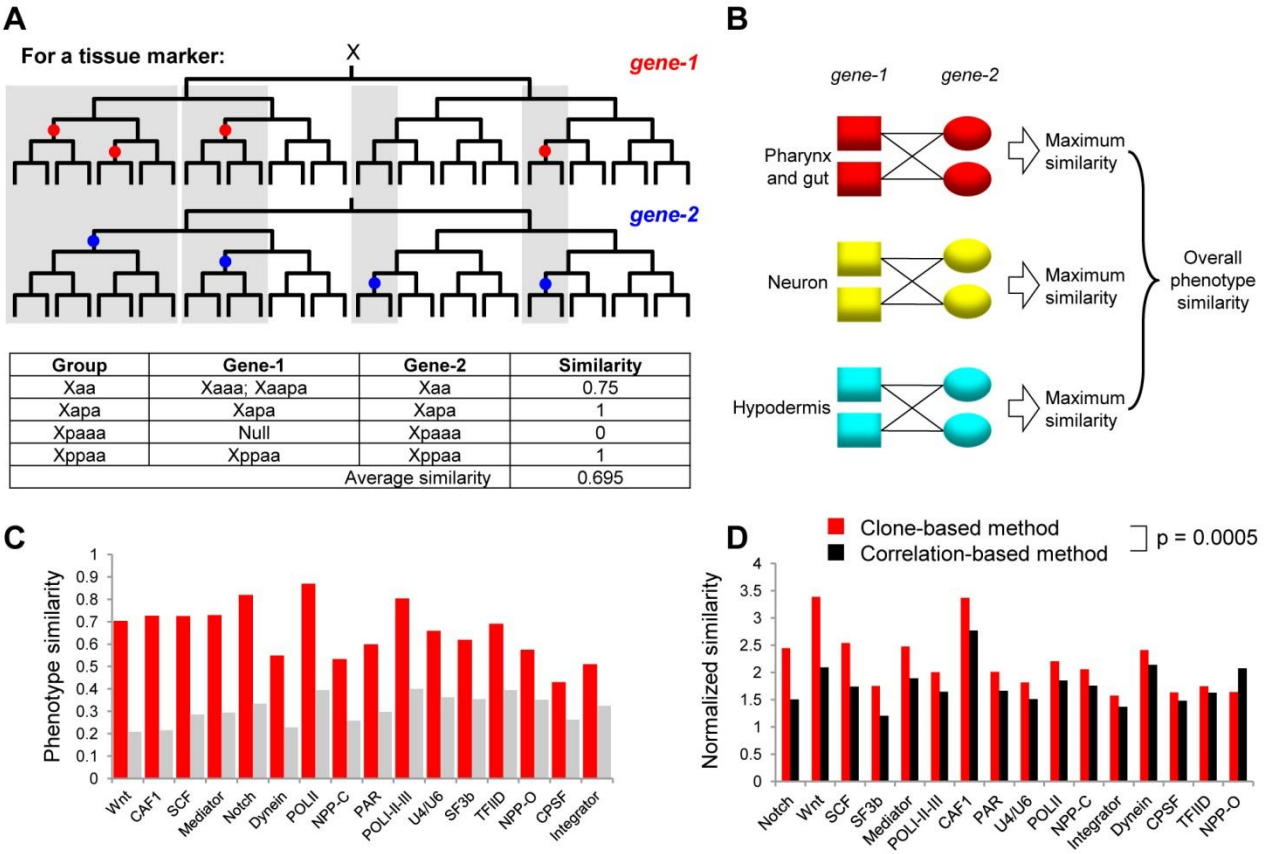
(B) Mapping marker expression patterns onto fate space. We mapped all 256 types of expression patterns onto a 16 by 16 plane with each pattern a unique coordinates based on marker expression pattern. See Extended Experimental Procedures for details.

(C) All detected expression pattern during cell fate changes. Heatmaps show all detected marker expression patterns that are different from the corresponding patterns in the wild type for the 12 founder cells.

(D) Frequency of new expression pattern following fate change. Dashed lines indicate the cut-offs for significantly enriched patterns determined by a binomial test under the assumption that each pattern has an equal opportunity to be observed. P value was calculated by binomial test.

(E-G) Heatmaps showing marker expression pattern for enriched cell fates following cell fate change for all tissue markers. For each tissue marker, the left panel is a heatmap visualization of the expression pattern for eight terminal clones. White denotes non-expression and color denotes expression clones. Those new expression patterns that define normal cell fates used during normal embryogenesis are labeled with the name of the corresponding fate and those that defines new fates that are not used during normal development is labeled as 'New'. Right panel shows the shortest distance of each expression pattern to that of normal cell fates. Magnitude of distance is represented as a color gradient from green to red. Blank expression pattern and blank-like expression patterns (with only one clone difference from blank) were excluded from analysis due to their low interpretability for cell fate.

Figure S4



Group	Gene-1	Gene-2	Similarity
Xaa	Xaaa; Xaapa	Xaa	0.75
Xapa	Xapa	Xapa	1
Xpaaa	Null	Xpaaa	0
Xppaa	Xppaa	Xppaa	1
		Average similarity	0.695

Figure S4 Construction of Gene Network, related to Figure 5

(A) Strategy for measuring phenotypic similarity. Top panel shows an example and bottom table shows the detailed calculation results. Gray shade represents the lineage-based cell grouping. See Extended Experimental Procedures for detail.

(B) Integration of similarity score from embryos and markers. Different tissue markers are shown in different colors, shape represents different genes. For each marker, pairwise phenotypic similarity scores are calculated between all embryos in different gene perturbations and the maximum score is taken to quantify phenotype similarity between two genes. Similarity score from all markers are then averaged to quantify the overall similarity.

(C) Phenotypic similarity scores for genes within a protein complex/ pathway (red) compared to that of background (gray). Background similarity is calculated as the average similarity between members of a complex/pathway to all other genes.

(D) Performance of clone-based method compared to that of correlation method. Bar plot shows the similarity score between genes in a complex/pathway normalized by the background similarity of this complex/pathway to all other genes. P-value was calculated by two-tailed Mann-Whitney U Test.

Figure S5

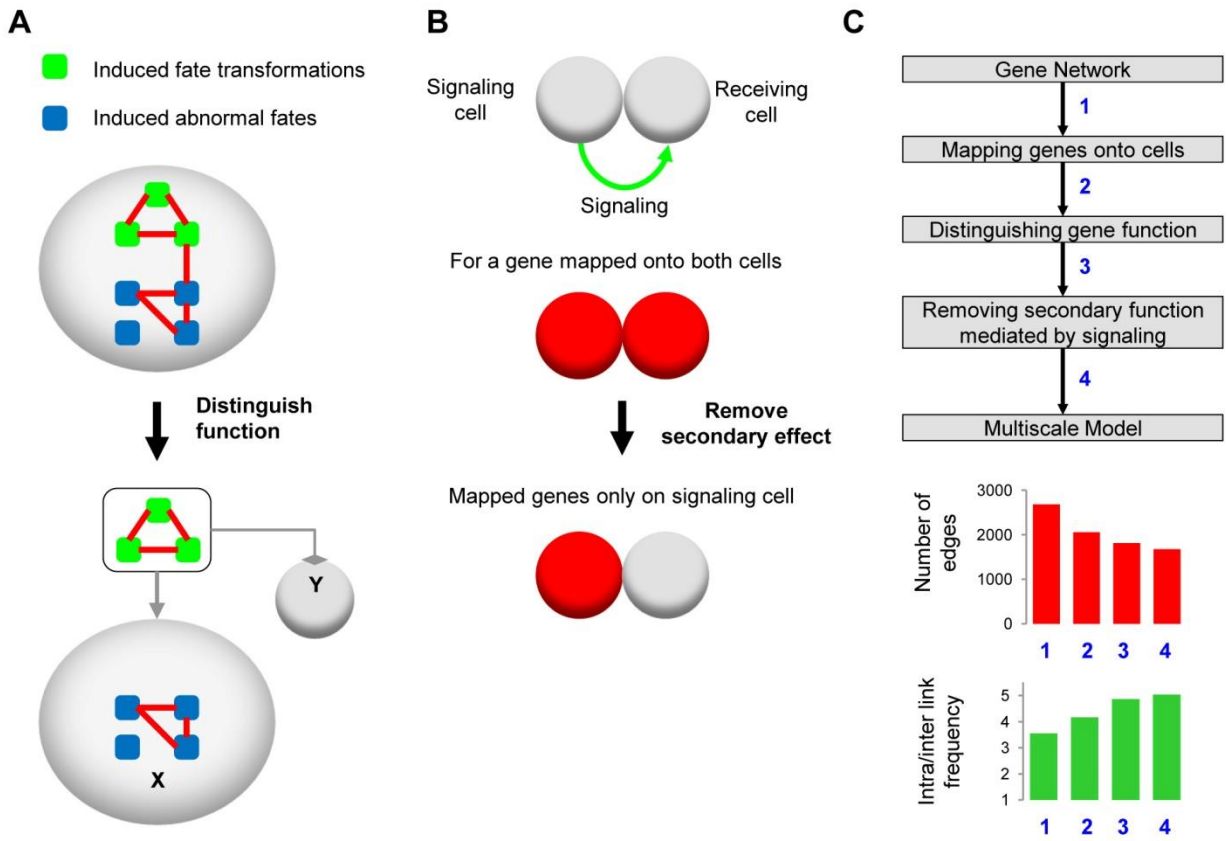


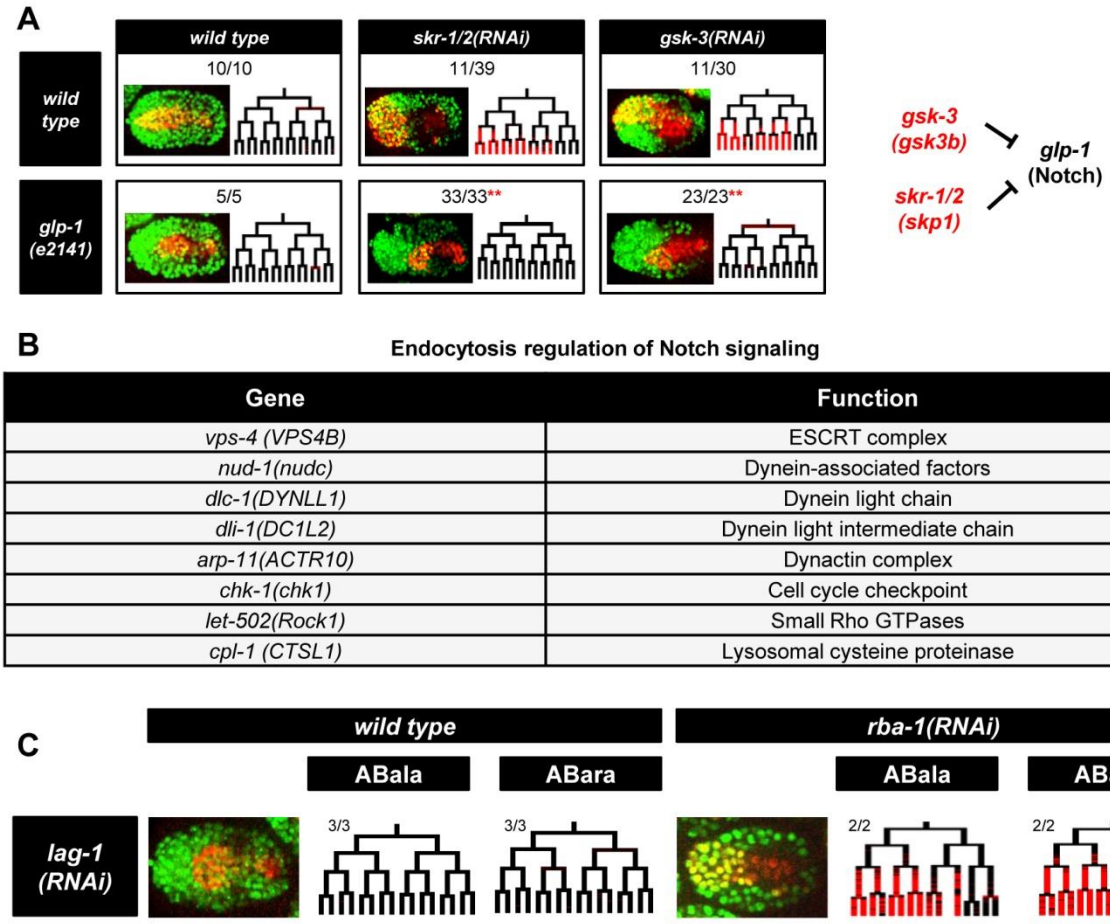
Figure S5 Construction of Multiscale Model of Cell Lineage Differentiation, related to Figure 6

(A) Distinguishing two types of gene function in lineage differentiation. Top panel shows genes (rectangles) that mapped onto a cell (gray circle). Genes that induced fate transformations and induced abnormal fate are shown in different colors. If a gene induced X-to-Y cell fate transformation, the gene is placed outside the cell that normally adopts X fate. Two types of arrows originate from this gene pointing to corresponding cells are used to represent the regulation of cell fate choice. First, a promoting arrow points to cells that normally adopt X fate to show that the gene is required for a cell to choose the original fate; second, a repression arrow points to the cell that normally adopts the Y fate to show that the gene promote X fate by repressing the alternative Y fate. If a gene induces abnormal cell fate, the gene is placed in the cell to show that this gene is required for the differentiation of this cell.

(B) Removal of secondary gene function mediated by cell-to-cell signaling. Top panel shows a cell-to-cell signaling scenario in which fate of the receiving cell's fate is regulated by a signaling from the signaling cell. Middle panel shows a case in which a gene inactivation affects (red) differentiation of both signaling and receiving cells. Bottom panel shows that we treat the phenotype of the signaling cell a primary and phenotype of the receiving cell a secondary effect. Gene effect is removed from the receiving cell.

(C) Improvement of specificity and sensitivity following each step of network analysis during construction of multiscale model. Top: Flow chat shows the four steps of network analyses for multiscale model. See main text for details. Bottom: bar plots shows the reduction of total number of gene links and the increase of specificity (measured as the number of intra-group links normalized by the number of inter-group links). Based on our benchmark of known complexes/pathways (Table S2), 38% of indirect gene links were removed in multiscale model (Figure 6C) compared to the gene network (Figure 5A). The enrichment of intra-group links over the inter-group links increased from 3.5 fold to 5.0.

Figure S6



Gene	Function
vps-4 (VPS4B)	ESCRT complex
nud-1(nudc)	Dynein-associated factors
dlc-1(DYNLL1)	Dynein light chain
dli-1(DC1L2)	Dynein light intermediate chain
arp-11(ACTR10)	Dynactin complex
chk-1(chk1)	Cell cycle checkpoint
let-502(Rock1)	Small Rho GTPases
cpl-1 (CTSL1)	Lysosomal cysteine proteinase

Figure S6 Context-Specific Repression of Notch Signaling, related to Figure 8

(A) Genetic analysis of genes in module II of Figure 8B. Left: Figure shows the lineage differentiation phenotypes of ABala sublineages for *skr-1/2(RNAi) and gsk-3(RNAi)* single and *skr-1/2(RNAi);glp-1(e2142) and gsk-3(RNAi);glp-1(e2142)* double loss of function mutants. Right: summary of epistatic analysis. ** denotes $p < 0.01$ calculated by chi square test.

(B) List of genes involved in the endocytosis regulation of Notch signaling. Module III of Figure 8B identified many genes in the endocytosis and vesicle sorting processes that are known to regulate Notch receptor recycling. Figure shows the genes, homologs and their functions. For example, component of ESCRT complex (sorting complexes required for transport, *vps-4*), Dynein motor proteins (*dlc-1* and *dli-1*) (Driskell et al., 2007), recycling-endosome-positioning regulator (*nud-1*) (Winter et al., 2012) and lysosomal cysteine protease (*cpl-1*). These genes may function as a module to repress Notch signaling.

(C) *rba-1* is epistatic to *lag-1*. Figure shows the lineage differentiation phenotypes of ABala and ABara sublineages for *lag-1* single and *lag-1; rba-1* double loss of function mutants using RNAi. Expression of pharynx marker is shown in red. The effectiveness of *lag-1(RNAi)* in double RNAi was judged by the ABp-to-ABa transformation in the same embryo.

3. Supplemental Tables

Table S1 Summary of dataset, Related to Figure 1

Sheet (1) lists the name, molecular identity and functional classification of 204 identified developmental regulatory genes.

Sheet (2) lists the name and short description of 84 genes that are required for cell division (cell biology genes).

Sheet (3) lists the number of embryos with lethal phenotypes analyzed for each gene.

Sheet (4) lists the functional annotation status for the 204 genes.

Table S2 Lineage Differentiation Phenotypes, Related to Figure 1

Sheet (1) lists the differentiation phenotypes for individual embryos. Each embryo's name contains three fields indicating the perturbed genes, tissue marker strain used (PHA for pharynx and gut; NEU for neuron; HYP for hypodermis) and the randomly assigned embryo ID respectively. Phenotypes for individual cells from different sublineages (with names shown above) are shown as numbers and colors. 0 denotes no change, -1, -2 and -3 denote loss of expression of pharynx and gut, neuron and hypodermis marker respectively; 1, 2 and 3 denote gain of pharynx and gut, neuron and hypodermis marker expression respectively. Cellular phenotypes were propagated from clonal changes through cell lineage. The number of cell in each sublineage was made uniform for the ease of comparison.

Sheet (2) lists number of clonal expression changes induced by each gene inactivation in each tissue marker strain. Phenotypes from different embryo replicates for the same gene inaction were combined (union).

Sheet (3) lists the 40 compiled genes that function in 16 stable protein complexes or core pathways.

Sheet (4) lists cell fate changes in progenitor cells for each gene perturbation. 1 denotes change.

Table S3 Regulation of Progenitor Cell Fate Transformation, Related to Figure 2 and Figure 3

Sheet (1) lists progenitor fate transformations detected for each gene and interpretation.

Sheet (2) lists the distribution of each gene's knockdown phenotypes in specific progenitor cells during lineage differentiation. 1 denotes gene knockdown induced abnormal cell fate, 2 denotes gene knockdown induced fate transformation.

Sheet (3) lists the summary of gene function in differentiation.

Table S4 Gene Regulatory Network, Related to Figure 5

Sheet (1) shows all identified gene pairs and strength of association between them.

Sheet (2) shows the quality assessment results of network using benchmark genes. Frequency of all intra-group links (within a complex/pathway) and inter-group links (between different complexes/pathways) captured in network is shown for individual complex/pathway.

Table S5 Multiscale model of lineage differentiation, Related to Figure 6

Sheet (1) provides high resolution display of multiscale model shown in Figure 6C.

Sheet (2) lists all gene pairs in the multiscale model

Sheet (3)-sheet (25) shows high resolution display of model for each single-cell differentiation context. Gene modules that regulate fate choice are highlighted in blue.

3. Extended Experimental Procedures

C. elegans Strains

Genotypes for all strains used in this study are listed below. Some strains were obtained from Caenorhabditis Genetics Center (CGC). All *C. elegans* strains were grown at 22°C under standard laboratory conditions. Temperature sensitive mutants were maintained at permissive temperature (15 °C) and switched to restrictive temperature (25-26°C) for at least 6 hours before experiment.

RW10425 (pharynx and gut marker): stIs10116 [his-72(promoter)::his-24::mCherry + unc-119(+)]; stIs37 [pie-1(promoter)::mCherry::H2B + unc-119(+)]; stIs10389 [pha-4::GFP::TY1::3xFLAG].

RW10434 (neuron marker): stIs10116 [his-72(promoter)::his-24::mCherry + unc-119(+)]; stIs37 [pie-1(promoter)::mCherry::H2B + unc-119(+)]; stIs10394 [cnd-1::GFP].

RW10348 (major hypodermis marker): stIs10116 [his-72(promoter)::his-24::mCherry + unc-119(+)]. stIs37 [pie1(promoter):: mCherry ::H2B + unc-119(+)]; stIs10318 [nhr-25::GFP; unc-119(+)]; unc-119(+)]; stIs10024 [pie-1::H2B::GFP + unc-119(+)].

BV180 (*glp-1*, hypodermis marker): glp-1(e2141); stIs10116 [his-72(promoter)::his-24::mCherry + unc-119(+)]; stIs37 [pie-1(promoter)::mCherry::H2B + unc-119(+)]; stIs10318 [nhr-25::GFP; unc-119(+)]

BV130 (*glp-1*; pharynx marker): glp-1(e2141); stIs10116 [his-72(promoter)::his-24::mCherry + unc-119(+)]; stIs37 [pie-1(promoter)::mCherry::H2B + unc-119(+)]; stIs10389 [pha-4::TGF(3E3)::GFP::TY1::3xFLAG].

RW10752 (ref-1 transcriptional reporter): stIs10691 [ref-1::H1-Wcherry + unc-119(+)]; zuIs178 [his-72(1kb 5' UTR)::his-72::GFP + unc-119(+)]; stIs10024 [pie-1::H2B::GFP + unc-119(+)].

RNAi Screen and Classification

All genes known to induce Emb phenotypes when knockdown were retrieved from WormBase using WormMart (Harris et al., 2014). Only genes, when knocked down by RNAi, reproducibly induced (with two or more reports) Emb phenotypes in the wild-type genetic background were considered. All RNAi-mediated gene knockdowns were performed by standard feeding method using RNAi clones from commercially available RNAi libraries (Kamath et al., 2003; Rual et al., 2004). RNAi clones corresponding to all previously identified Emb genes were re-arrayed for convenience of screening.

5 to 10 worms synchronized at L1 stage were used for the RNAi screen on plates at an IPTG concentration of 3 mM. The Emb phenotypes were scored by estimating the frequency of dead embryos on the plate 24 hours after the bacteria have been exhausted. Weak RNAi were performed for genes that induced severe phenotypes such as P_0 larval lethal, sterile, and early embryonic arrest by using L4 or young adult worms for 16-24 hours of RNAi treatment.

Results from three RNAi screens were used to classify genes. (i) Gene knockdowns that consistently induced low penetrant Emb (<25%) were defined as low penetrant essential genes. (ii) Gene knockdowns that consistently induced P_0 larval lethal, sterile or produced less than 10 eggs/worm phenotypes were defined as reproduction genes. (iii) Gene knockdowns that induced penetrant Emb (≥25%) were defined as penetrant Emb genes.

Penetrant Emb genes were further classified based on the conservation of protein sequence, predicted molecular function and cell biology phenotypes to identify the 204 gene set for in-depth phenotype analysis (Figure S1). Human orthologs of *C. elegans* genes were based on WormBase conservation annotation and OrthoList (Shaye and Greenwald, 2011). For the 204 target genes, RNAi treatments were performed in three tissue marker strains individually (204X3=612 RNAi treatments) and for most treatments (603, 98.5%) Emb phenotype was induced in two or more embryos which were further lineaged for single-cell phenotype analysis (Table S1).

Estimation of Functional Annotation Status

The functional annotation status of the 204 genes was determined using WormBase (Table S1), a comprehensive biological resource for the nematode (Harris et al., 2014). First, the functional description for each gene was

extracted using WormMart (WS220-bugFix version) and inspected manually. A gene is classified as "without_annotation" if no functional description is available; a gene is classified as "without_embryogenesis_annotation" if no functional annotation regarding embryogenesis is available based on manual examination. Phenotype information of all genes was extracted and a gene was classified as "without_differentiaiton_annotation" if no lineage-specific differentiation phenotypes were reported previously.

3D Time-Lapse Imaging of Embryogenesis

Two color (491 nm and 561 nm) fluorescent 3D time-lapse imaging was acquired using a spinning disk confocal system (Quorum Technologies) and Zeiss Observer Z1 microscopy with Zeiss PlanApo 40x/1.3 Oil objective as previously described (Du et al., 2014). Embryos at or earlier than four-cell stage were collected by dissecting gravid hermaphrodites under a dissecting microscope (Nikon). Embryos were arranged in clusters and mounted between two cover slips in M9 buffer containing 20 μm microspheres (Polyscience, Warrington, PA). A detailed protocol for embryo mounting can be found elsewhere (Bao and Murray, 2011; Murray et al., 2006).

Images were recorded using 30 focal planes spaced 1 μm apart at 20 °C on a temperature-controlled stage at 75 seconds intervals for 6-8 hours (or the first half) of *C. elegans* embryogenesis. Laser power and exposure time of imaging were optimized by considering both the signal to noise ratio of acquired images and laser induced phototoxicity to specimens. We used an imaging parameter under which wild-type embryos can develop normally and hatch at a time comparable to those without laser excitation.

We systematically monitored the efficiency of all RNAi treatments by marking the position of each embryo on the slide and checking viability of all embryos 18 hours after the completion of imaging. Only image series for embryos with Emb phenotypes were processed for in-depth phenotype analysis.

Automated Construction of Cell Lineage

Image series were processed by StarryNite software (http://starrynite.sourceforge.net) for automated cell identification, tracing and lineage construction (Bao et al., 2006; Santella et al., 2014; Santella et al., 2010). Nucleus-localized ubiquitously expressed mCherry is used to identify and trace cells and their divisions. First, every nucleus was identified from 3D image stacks containing 30 Z planes by segmenting nuclear slices within each 2D image, then using a hybrid blob-slice model to assemble these slices into 3D nuclei. Second, a nearest neighbor algorithm is applied to trace cells by matching the identified nuclei at a preceding time point to those at the subsequent time point. During refinement of cell tracing, cell movement, division and death are recognized using a supervised learning framework. Third, the lineage identity of each cell is determined based on the canonical Sulston naming scheme (Sulston et al., 1983). The stereotypical cell configuration of four-cell stage embryos are used to determine body axis and lineage identity of each cell and then propagated to their descendants. In general, the full name of a mother cell is propagated to all daughter cells with an additional letter specifying the position of the daughter cell relative to the division axis of the mother cell. There are three types of cell divisions: anterior-posterior (a-p), left-right (l-r) and dorsal-ventral (d-v). For example, ABa cell is the anterior daughter of AB cell. Detailed information on the naming scheme can be found at (Bao et al., 2006; Santella et al., 2010). StarryNite allows a highly efficient automated construction of a cell lineage till a 350-cell stage embryo from images in 120 minutes using a 2.13 GHz Intel core2 PC computer at an accuracy of 99.5%.

Curation and Quality Control of Cell Lineage

All automatically constructed cell lineages were systematically curated manually to ensure a high level of quality. Interactive lineages and associated 4D image data are visualized by AceTree (http://acetree.sourceforge.net), an intuitive platform to identify and correct lineage errors (Boyle et al., 2006; Murray et al., 2006). Because only a small subset of cells expressing a tissue marker we curated the tracking results for all cells till ~ 200-cell stage and for marker-expressing cells till a 350 cell stage. By this stage, the embryo has completed gastrulation and major tissue fate specification, allowing a reliable assessment of cell differentiation status.

A cell lineage is visualized as a tree structure showing the cell pedigree from the P_0 (zygote) to the end time of cell tracing. Vertical lines represent cells traced over time and horizontal lines represent cell divisions (Figure 1C). Three types of predominant lineage errors will be directly reflected by irregularities in tree topology of cell lineage

which can be easily recognized (Figure S2D) including (i) early termination, (ii) skipping of cell division and (iii) false cell division. In addition, a less frequent error that will not change tree topology is (iv) the swap of cell naming, e.g. calling an anterior daughter posterior. Such errors were identified computationally by determining if the naming of a cell (determined by the tracking program using canonical division patterns in the wild type) matches that based on cell position relative to the body axis. All above types of lineage errors were corrected manually. The curated lineages underwent a second round of computational screening to identify and correct residual errors. After curation, we randomly picked up 100 cell tracks (from P_0 to the traced terminal cell generations) from 1,368 curated cell lineages and followed them manually to estimate the error rate in our dataset and the results showed a high accuracy (Figure 1I). A small subset (20 genes) of this dataset was used previously to construct mechanistic models of fate choice which further validated the data quality (Du et al., 2014).

Determination of Tissue Marker Expression Cells

Each tissue marker strain carries two fluorescently labeled proteins. The ubiquitously expressed mCherry is used for lineage construction and the intensity of an additional nucleus-localized GFP protein fused to a tissue specific protein is used to determine the cell types a cell becomes. Three tissue-specific markers were used: PHA-4 for pharynx and gut (Mango et al., 1994a), CND-1 for a subset of neurons (Hallam et al., 2000) and NHR-25 for major hypodermis cells (Gissendanner and Sluder, 2000; Yanai et al., 2008). The strategy to quantify marker expression is based on a previously described method (Murray et al., 2008). First, average GFP intensity is calculated within each detected nucleus at each time point with the local background signal intensity subtracted. Second, GFP intensity is averaged over a cell's lifetime to measure marker expression level for each cell. Third, a biologically relevant binary cut-off established previously (Du et al., 2014) is applied to determine whether a marker is expressed or not in a cell. Specific cut-offs were applied to each tissue marker strain at a false discovery rate of 0.05. Under these cut-offs the binary marker expression patterns are highly reproducible with a Pearson's Correlation Coefficient r =0.97, 0.93 and 0.95 for pharynx and gut (PHA-4), neuron (CND-1) and hypodermis (NHR-25) markers respectively among 10 wild-type embryos (Du et al., 2014).

Identification of Clonal Expression Changes

Based on the expression status of individual terminal cells we performed clonal expression analysis to identify the earlier intermediate cells on the cell lineage whose terminal cells uniformly express the marker (Du et al., 2014). The intermediate cells are defined as clonal expression cells and the corresponding sublineages are defined as expression clones. Figure 1C shows the clonal expression cells for all markers for a normal embryo. To identify clonal expression cells, first, for each intermediate cell, the frequency of marker expression within its sublineage is calculated iteratively by averaging the frequency of its two daughter cells. Second, binomial test is then applied to assess the statistic significance of the observed frequency compared to that expected by chance. We define a maximum clone with a significantly ($p < 0.01$) high frequency of marker expression (≥0.85) as positive expression clone and the lowest common ancestor cell of this clone as a clonal expression cell. By clonal expression analysis, a complex lineage expression pattern is described as a list of clonal expression cells. As shown previously, clonal analysis provides a robust way to describe lineage expression of marks by minimizing the noise of marker expression in individual cells. For each perturbed embryo we first identify clonal expressing cells and then compared them to that of wild type to compute clonal loss of expression and clonal gain of expression cells according to the lineage relationship between clonal expression cells. A specific example is given in Figure S2A.

Prediction of Gene Action Sites On Progenitor Cells

We used the parsimony strategy to predict a gene's action sites (primary cell fate changes) in 23 progenitor cells. First, a gene is mapped onto corresponding 12 terminal progenitor cells (named cells in Figure 1C) if a gene knockdown induces any phenotypes in the sublineage. Second, parsimony is applied to find the potential earlier sites. If a gene is mapped on both daughter cells we assume its action site is on the mother cell and repeat this process recursively to find the earliest sites of changes (Figure S2C). Third, considering the caveat that one gene could have distinct functions in two daughter cells we further used more decisive action sites to refine the direct parsimony results. A gene that induces cell fate transformation phenotypes (see below) suggests a clear action site in the corresponding cell to regulate cell fate choice. The 174 fate choice action sites (Figure 3B) were used

to refine the results. For those whose inferred action site is on a progenitor cell earlier than the fate choice cell on the lineage, we map the gene function on the fate choice cell and the complementary set of progenitor cells. For those whose inferred action site is on a cell later than the fate choice cell, we mapped the gene function only on the fate choice cell.

ABar Spindle Rotation Defects

During normal development the spindle of ABar cell rotates in response to Wnt signaling, so that the axis of the two ABar daughters is perpendicular to that of the other ABxx daughter cells (Rocheleau et al., 1997; Thorpe et al., 1997; Walston et al., 2004). Failure of spindle rotation changes the fate of the ABar daughters (Du et al., 2014). We used the ABar spindle rotation phenotype to evaluate the accuracy of predicted gene action site by determining whether the predicted gene action sites capture this process.

We calculated the angle between ABara-ABarp and ABala-ABalp cell vectors to approximate ABar spindle rotation. A successfully rotated ABar spindle would generate an angle of 90 degree (Walston et al., 2004). Angle was calculated for three time points: immediately after ABar division (T_0), one time point after (T_1) and two time points after (T_2). T_2 produced the best reproducibility across 71 wild-type embryos (85±6 degrees) and was used to calculate rotation angle for all perturbed embryos. We defined a gene with ABar spindle rotation phenotype if an <30 degree angle was detected in any perturbed embryo. This definition successfully captured genes known to regulate spindle rotation including *mom-2*, *mom-5* and *gsk-3* (Walston et al., 2004).

From the 42 identified genes that affected the ABar spindle rotation process we found that the action site for 36 out of 42 (86%) of these genes was mapped onto ABar cell or its ancestors in our gene-cell map (Figure 1K). Given that cell positions were not explicitly considered during the prediction, this result suggests a high accuracy of the predicted gene action sites.

Benchmark Genes

We systematically examined the molecular identity of 204 genes and compiled a list of genes that are known to function either in a stable protein complex or in a core gene pathway. We identified 40 genes for 16 protein complexes (such as Dynein and Nuclear Pore complex) or pathways (such as Wnt and Notch signaling) (Table S2). These genes were used as benchmark throughout this study to systematically assess phenotypes (Figure 1J), phenotypic similarity (Figures 5B and S4C), efficiency and specificity of global gene regulatory network (Figure S4D) and the multiscale model (Figure S5C).

Quantification of Phenotypic Landscape

We map all 256 possible marker expression patterns onto a 16X16 array to represent the theoretical fate change space. Each marker expression pattern is quantitatively described as eight digits based on the clonal expression status (1 and 0 for expression and nonexpression) of the eight terminal clones for a typical five rounds of cell divisions in sublineages (Figure S3A). We assign a (X,Y) coordinate to each pattern based on the expression pattern of the anterior and posterior half of clones respectively (Figure S3B). For each half, the 16 (2^4) possible patterns are ranked first by the total number and then by the location of positive expression clones, and a unique score (range from 1 to 16) is assigned to each pattern. A pattern with more positive clones that occur more posteriorly is assigned a larger coordinate score. For example, the coordinate for the pattern-00000000 is (1, 1) and the coordinates for the pattern-11111111 is (16, 16).

Simulation of Influence of Lineage Errors

Based on the estimated frequency of errors that affect clonal expression status, we randomly changed the expression status of portions of cell lineage to simulate the potential influence of lineage error on phenotypic landscape. For each simulation, the expression status is assigned either “on” or “off” randomly. 10,000 simulations showed that the error rate did not affect the overall trend of distribution and enrichment of phenotypic landscape (Figure 2C).

Identification of Progenitor Cell Fate Transformations

Cell fate transformations between early progenitor cells were identified using a previously established method (Du et al., 2014). Briefly, lineal expression pattern of tissue markers is used to define the fate of progenitor cells and the wild-type patterns are used to detect the occurrence of normal cell fates in perturbed progenitor cells. A homeotic transformation is identified if a mutant cell adopts a fate normally associated with a different progenitor cell. Due to the limited number of analyzed embryos (2 for most cases) for each tissue marker in each gene perturbation, a transformation was assigned if it is supported by any informative tissue marker (with positive expression for a fate). All detected homeotic transformation phenotypes were further inspected and validated manually (Table S3).

Definition of Lineage Distance

Lineage distance between two cells is defined as the total number of cell divisions that separate the two cells from their lowest common ancestor (Figure 2G). To characterize intrinsic regulation of fate transformability, fate transformations that are known to be regulated by cell-to-cell signaling were excluded from analysis for Figure 2H. In addition, atypical mother-daughter transformations were also excluded from analysis.

Measurement of Phenotypic Similarity

We developed a method that uses clonal expression change as the unit of measurement to quantify phenotypic similarity between embryos (Figure S4A). First, cells with clonal expression changes from two embryos are segregated into different groups based on their lineal identity. Lineal-related cells, for which one cell is a descendent of the other, are grouped together. Clonal gain and loss phenotypes are analyzed separately. Second, for each cell group the phenotypic similarity is measured as lineal similarity between cells from two embryos. For example, similarity between a mother and a daughter cell is 0.5 and similarity between a cell and its granddaughter is 0.25. Finally, phenotypic similarity scores for all groups and for gain and loss phenotypes are averaged to quantify the overall similarity score.

~20% of loss of one marker was associated with gain of expression of other markers, which provides information to assay cell fate changes. Considering the gain phenotypes would be more specific than that of loss phenotypes we weighted them differently according to their frequency. The ratio between gain and loss phenotype is 1:4, we therefore assigned 0.8 and 0.2 weights to them respectively when calculating the overall similarity for a marker.

Theoretically, the clone-based method would more capture the primary phenotypic similarity. We compared the clone-based method to that of widely used correlation-based method in which vectors of marker expression changes in terminal cells are compared. Indeed, we found the clone based method outperformed correlation based method in distinguishing positive gene interactors from the background (Figure S4D).

Phenotypic Similarity Between Genes

We examined differentiation phenotypes covering the entire cell lineage using three tissue markers and analyzed multiple embryos for each marker. The phenotypic similarity between genes was determined by calculating phenotypic similarity score in different sublineages for different tissue markers using the above method. We divided the entire cell lineage into 23 sublineages and for each sublineage, for each marker, we calculated the maximum phenotypic similarity between all embryos for two genes (Figure S4B). The score is then averaged for all tissue markers and for all sublineages weighted by their distance to the root (P_0) of the cell lineage. By doing so, the contribution of all sublineages and all tissue markers are considered when calculating the similarity.

Based on the distribution of gene similarity score (0.256 ± 0.213) for all pairwise gene comparisons we defined gene pairs with a score over 0.606 ($p < 0.05$ based on Z-score) as strong relationship (genes with similar function) and gene pairs with a score over 0.5 (empirically determined) as weak relationship (Table S4). The gene network successfully distinguished known interactors (components in known protein complexes or pathways) from random gene pairs (Figure 5B).

Determination of Secondary Effects Through Cell-Cell Signaling

Six known signaling events were considered including P_2-to-ABp (Mango et al., 1994b; Mello et al., 1994; Moskowitz et al., 1994); P_2-to-EMS(E) (Goldstein, 1992; Lin et al., 1995; Rocheleau et al., 1997; Thorpe et al., 1997); MS-to-ABalp (Hutter and Schnabel, 1994; Moskowitz et al., 1994), MS-to-ABara (Hutter and Schnabel, 1994; Moskowitz et al., 1994); ABala-to-ABpla (Hutter and Schnabel, 1995) and C-to-ABar (Walston et al., 2004). For each signaling-to-receiving cell pair, we removed a gene from the receiving cell if this gene also affects the differentiation of the signaling cell or its ancestors (Figure S5B). This is under the assumptions that (i) cell fate is upstream of the signal-sending ability of the signaling cell and that (ii) the differentiation defects of the receiving cell is a secondary effects of perturbed signaling sending.

Landscape Simulations

Fate tendency simulation based on real landscape: the distribution of terminal fate choices in developmental landscapes was characterized by Monte Carlo simulation. Starting from the zygote, a cell division is modeled by randomly selecting from all possible fate trajectories with equal likelihood assigned to each (Figure 7A). Each daughter cell continues dividing until all cells have reached a terminal progenitor fate. The simulation was run repeatedly (n = 1,000) and the relative prevalence of all terminal fates was measured.

Landscape randomization: a set of randomly generated landscapes was assessed in a similar fashion by randomly assigning the same number of fate trajectories to connect terminal and intermediate fates within the lineage. A constraint was placed on possible trajectories to prevent a differentiation event from leading to a fate higher in the lineage tree than the originating cell in the wild-type lineage, otherwise all trajectories were considered including recursive transformations and trajectories horizontal within the lineage. This simulation was run repeatedly (n = 1,000) over a large number of randomly generated landscapes (n = 1,000).

Simulation of regulation complexity: a Monte Carlo simulation was used to conceptualize the relationship between the number of progenitor cells (N), the number of terminal cell fates (T) and the likelihood of random differentiation producing a relatively uniform distribution of all cell types (defined here as ±50% of the expected value of each cell type: N/T). For each combination of N and T, each cell was randomly assigned one of T terminal fates and the average success rate tallied over a large number of repetitions (n = 10,000).

Programming and Software

Identification of marker-expressing cells, identification of expression clones and clonal changes, identification of fate transformations, measurement of phenotype similarities between genes, parsimony to map gene function onto cells, calculation of angles, network comparison and inference were implemented with Perl and MATLAB scripts. Network visualization was performed using Cytoscape (Shannon et al., 2003).

4. Supplemental References

Bao, Z., and Murray, J.I. (2011). Mounting Caenorhabditis elegans embryos for live imaging of embryogenesis. Cold Spring Harbor protocols *2011*.

Bao, Z., Murray, J.I., Boyle, T., Ooi, S.L., Sandel, M.J., and Waterston, R.H. (2006). Automated cell lineage tracing in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America *103*, 2707-2712.

Boyle, T.J., Bao, Z., Murray, J.I., Araya, C.L., and Waterston, R.H. (2006). AceTree: a tool for visual analysis of Caenorhabditis elegans embryogenesis. BMC bioinformatics *7*, 275.

Driskell, O.J., Mironov, A., Allan, V.J., and Woodman, P.G. (2007). Dynein is required for receptor sorting and the morphogenesis of early endosomes. Nature cell biology *9*, 113-120.

Du, Z., Santella, A., He, F., Tiongson, M., and Bao, Z. (2014). De novo inference of systems-level mechanistic models of development from live-imaging-based phenotype analysis. Cell *156*, 359-372.

Gissendanner, C.R., and Sluder, A.E. (2000). nhr-25, the Caenorhabditis elegans ortholog of ftz-f1, is required for epidermal and somatic gonad development. Developmental biology *221*, 259-272.

Goldstein, B. (1992). Induction of gut in Caenorhabditis elegans embryos. Nature *357*, 255-257.

Hallam, S., Singer, E., Waring, D., and Jin, Y. (2000). The C. elegans NeuroD homolog cnd-1 functions in multiple aspects of motor neuron fate specification. Development *127*, 4239-4252.

Harris, T.W., Baran, J., Bieri, T., Cabunoc, A., Chan, J., Chen, W.J., Davis, P., Done, J., Grove, C., Howe, K.*, et al.* (2014). WormBase 2014: new views of curated biology. Nucleic acids research *42*, D789-793.

Hutter, H., and Schnabel, R. (1994). glp-1 and inductions establishing embryonic axes in C. elegans. Development *120*, 2051-2064.

Hutter, H., and Schnabel, R. (1995). Establishment of left-right asymmetry in the Caenorhabditis elegans embryo: a multistep process involving a series of inductive events. Development *121*, 3417-3424.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M.*, et al.* (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature *421*, 231-237.

Lin, R., Thompson, S., and Priess, J.R. (1995). pop-1 encodes an HMG box protein required for the specification of a mesoderm precursor in early C. elegans embryos. Cell *83*, 599-609.

Mango, S.E., Lambie, E.J., and Kimble, J. (1994a). The pha-4 gene is required to generate the pharyngeal primordium of Caenorhabditis elegans. Development *120*, 3019-3031.

Mango, S.E., Thorpe, C.J., Martin, P.R., Chamberlain, S.H., and Bowerman, B. (1994b). Two maternal genes, apx-1 and pie-1, are required to distinguish the fates of equivalent blastomeres in the early Caenorhabditis elegans embryo. Development *120*, 2305-2315.

Mello, C.C., Draper, B.W., and Priess, J.R. (1994). The maternal genes apx-1 and glp-1 and establishment of dorsal-ventral polarity in the early C. elegans embryo. Cell *77*, 95-106.

Moskowitz, I.P., Gendreau, S.B., and Rothman, J.H. (1994). Combinatorial specification of blastomere identity by glp-1-dependent cellular interactions in the nematode Caenorhabditis elegans. Development *120*, 3325-3338.

Murray, J.I., Bao, Z., Boyle, T.J., Boeck, M.E., Mericle, B.L., Nicholas, T.J., Zhao, Z., Sandel, M.J., and Waterston, R.H. (2008). Automated analysis of embryonic gene expression with cellular resolution in C. elegans. Nature methods *5*, 703-709.

Murray, J.I., Bao, Z., Boyle, T.J., and Waterston, R.H. (2006). The lineaging of fluorescently-labeled Caenorhabditis elegans embryos with StarryNite and AceTree. Nature protocols *1*, 1468-1476.

Rocheleau, C.E., Downs, W.D., Lin, R., Wittmann, C., Bei, Y., Cha, Y.H., Ali, M., Priess, J.R., and Mello, C.C. (1997). Wnt signaling and an APC-related gene specify endoderm in early C. elegans embryos. Cell *90*, 707-716.

Rual, J.F., Ceron, J., Koreth, J., Hao, T., Nicot, A.S., Hirozane-Kishikawa, T., Vandenhaute, J., Orkin, S.H., Hill, D.E., van den Heuvel, S.*, et al.* (2004). Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. Genome Res *14*, 2162-2168.

Santella, A., Du, Z., and Bao, Z. (2014). A semi-local neighborhood-based framework for probabilistic cell lineage tracing. BMC bioinformatics *15*, 217.

Santella, A., Du, Z., Nowotschin, S., Hadjantonakis, A.K., and Bao, Z. (2010). A hybrid blob-slice model for accurate and efficient detection of fluorescence labeled nuclei in 3D. BMC bioinformatics *11*, 580.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

Shaye, D.D., and Greenwald, I. (2011). OrthoList: a compendium of C. elegans genes with human orthologs. PloS one *6*, e20085.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Developmental biology *100*, 64-119.

Thorpe, C.J., Schlesinger, A., Carter, J.C., and Bowerman, B. (1997). Wnt signaling polarizes an early C. elegans blastomere to distinguish endoderm from mesoderm. Cell *90*, 695-705.

Walston, T., Tuskey, C., Edgar, L., Hawkins, N., Ellis, G., Bowerman, B., Wood, W., and Hardin, J. (2004). Multiple Wnt signaling pathways converge to orient the mitotic spindle in early C. elegans embryos. Developmental cell *7*, 831-841.

Winter, J.F., Hopfner, S., Korn, K., Farnung, B.O., Bradshaw, C.R., Marsico, G., Volkmer, M., Habermann, B., and Zerial, M. (2012). Caenorhabditis elegans screen reveals role of PAR-5 in RAB-11-recycling endosome positioning and apicobasal cell polarity. Nature cell biology *14*, 666-676.

Yanai, I., Baugh, L.R., Smith, J.J., Roehrig, C., Shen-Orr, S.S., Claggett, J.M., Hill, A.A., Slonim, D.K., and Hunter, C.P. (2008). Pairing of competitive and topologically distinct regulatory modules enhances patterned gene expression. Molecular systems biology *4*, 163.