

# Chromosomal clustering of muscle-expressed genes in *C. elegans*

Peter J. Roy\*, Joshua M. Stuart†, Jim Lund\*, Stuart K. Kim\*

*\*Departments of Developmental Biology and Genetics, Stanford University Medical Center, Stanford, CA, 94305*

*†Stanford Medical Informatics, 251 Campus Drive, MSOB X-215, Stanford, California, 94305*

**Chromosomes are divided into domains of open chromatin, where genes have the potential to be expressed, and domains of closed chromatin where genes are not expressed<sup>1</sup>. Classic examples of open chromatin domains include puffs on polytene chromosomes in *Drosophila* and extended loops from lampbrush chromosomes<sup>2,3</sup>. If multiple genes were typically expressed together from a single open chromatin domain, the position of co-expressed genes along the chromosomes would appear clustered. To investigate if co-expressed genes are clustered, we examined the chromosomal positions of the genes expressed in first larval stage (L1) muscle of *C. elegans*. Here we report that co-expressed genes in *C. elegans* are clustered in groups of 2-5 along the chromosomes, suggesting that expression from a chromatin domain can extend over several genes. These observations reveal a higher-order organization of the structure of the genome in which the order of genes along the chromosome is correlated with their expression in specific tissues.**

We developed a method called mRNA-tagging to isolate muscle mRNA since this tissue is difficult to isolate in *C. elegans*. The basis of the technique is to use a characterized promoter to express an epitope-tagged mRNA-binding protein, such as

poly(A)-binding protein (PAB-1), in cells or tissues of interest (Fig 1). Because poly(A)-binding proteins bind tightly to the poly(A) tail of mRNAs<sup>4</sup>, mRNAs from specific tissues can be enriched by cross-linking them to the tagged PAB-1, and co-immunoprecipitating the mRNA/tagged-PAB-1 complex using an anti-epitope monoclonal antibody. DNA microarrays can then be used to identify which mRNAs have been enriched by co-immunoprecipitation, indicating that the corresponding gene is expressed in the same cells as the tagged PAB-1.

To isolate the mRNA expressed in muscle, we first generated animals that express FLAG::PAB-1 in non-pharyngeal muscles from an integrated transgene using the *myo-3* promoter<sup>5</sup> (see Methods) (Fig. 2a). The mRNA/FLAG::PAB-1 complex was co-immunoprecipitated from cell lysate using anti-FLAG monoclonal antibodies. About 55% of the FLAG::PAB-1 was immunoprecipitated from the lysate (Fig. 2b). Dot blot and RT-PCR experiments show that *unc-54*, which is specifically expressed in muscle<sup>5</sup>, was co-immunoprecipitated with the muscle-expressed FLAG::PAB-1 but that *gld-1*, which is specifically expressed in the germ line<sup>6</sup>, was not (Fig. 2c and data not shown).

Next, we used DNA microarrays to analyze the ratio of the mRNA enriched by co-immunoprecipitation with FLAG::PAB-1 relative to the mRNA present in the starting cell-free extract. Fluorescently-labeled probes (see methods) were then hybridized to DNA microarrays<sup>7</sup> containing 90% of the 19,733 genes currently estimated in the *C. elegans* genome<sup>8</sup>. We repeated the mRNA-tagging experiment six times in order to statistically assess which genes are significantly enriched.

We found that the rank order of genes that are enriched in each immunoprecipitation experiment is more consistent than their absolute level of enrichment. This indicates that the immunoprecipitation procedure enriches genes consistently relative to each other, but that the efficiency of immunoprecipitation is variable. Hence, the percentile rank of enrichment for every gene from the six repeats

was averaged together. Genes that are not enriched by mRNA-tagging should have an average percentile rank of about 50%, while genes expressed in muscle should have a rank significantly higher. A Student's t-test identified 1364 genes that are significantly enriched in the muscle mRNA-tagging experiments ( $p < 0.001$ ) (Fig. 3; see Supplemental Table 1 at <http://cmgm.stanford.edu/~kimlab/muscle/>). In control experiments using worms that do not express FLAG::PAB-1 (see methods), only 85 genes are enriched ( $p < 0.001$ ; Supplemental Table 1).

We verified the mRNA-tagging approach by first showing that the list of 1364 genes contains 16 of 18 muscle positive controls (89%) (Supplemental Table 2). The two positive controls that were not identified by mRNA-tagging (*hlh-1* and *etr-1*) also do not show muscle expression by RNA *in situ* analysis (<http://dolphin.lab.nig.ac.jp/>; Supplemental table 2). Second, the expression pattern for 47 of the 1364 genes was known from prior work, and 44 are expressed in body wall muscle (94%) (Supplemental Table 3). In contrast, only 13 of 51 genes selected at random are expressed in muscle (25.5%; <http://dolphin.lab.nig.ac.jp/>; Supplemental Table 4). Third, we compared the list of 1364 genes to two groups of negative controls. The first group is 559 genes that are known not to be expressed in L1 muscle. Of these, 13 are in the list of 1364 genes (2.3%) (Supplemental Table 5). The second group is 7681 genes that show undetectable hybridization signals in DNA microarray experiments using RNA from L1 hermaphrodites. Of these, only 132 are in the list of 1364 muscle-expressed genes (1.7%) (Supplemental Table 6). Finally, the mRNA-tagging approach is not strongly biased against rare mRNAs. The distribution of the signal intensities of the 1364 L1 muscle genes was found to be nearly identical to that of all genes present on the microarray (Supplemental Fig. 1). Taken together, these results demonstrate that the mRNA-tagging technique specifically identified a large fraction of genes expressed in *C. elegans* muscle.

We have shown that mRNA-tagging is a powerful tool to profile gene expression in specific tissues in whole-genome expression studies. There are a large number of previously-characterized promoters that can direct the expression of epitope-tagged PAB-1 in many cell types in *C. elegans*, so that it is now feasible to identify mRNAs expressed in any tissue at different developmental times, under different growth conditions or in different genetic backgrounds. Since poly(A)-binding protein is highly conserved from yeast to humans<sup>9</sup>, mRNA-tagging is applicable to other model organisms to isolate mRNAs expressed in cells or tissues that were previously inaccessible.

The list of 1364 muscle-expressed genes is a global overview of gene expression in *C. elegans* muscle, and provides a foundation for understanding how this tissue functions at the molecular level. There are 5 genes that encode transcription factors that may specify muscle cell function, and 14 genes that encode receptors or other synapse-associated proteins that may function at the neuromuscular junction. As expected, the list also contains housekeeping genes, such as genes involved in energy production, chromatin structure, cytoskeletal function, RNA processing and protein expression (Fig. 3c). This list also contains over 500 genes with no known function. A more complete discussion of the list of muscle-expressed genes will be presented elsewhere.

We next looked for evidence for chromatin domains by investigating the positions of the muscle-expressed genes along the chromosomes. Specifically, we calculated how many muscle-expressed genes had start positions<sup>8</sup> within 10 kb of the start position of another muscle gene (see methods). This number of clustered muscle genes was then compared to the average number of clustered genes from 10,000 randomly-sampled gene lists of the same size as the muscle gene list. Before the number of clustered genes was calculated, however, special consideration was given to neighboring genes that reside in the same operon, and those that may have resulted from a tandem duplication event. Operons are composed of multiple genes that are expressed

from the same promoter and processed by trans-splicing to form separate mRNAs<sup>10,11</sup>. Thus, clustering of different genes from the same operon on the chromosome would not be evidence for chromatin domains, and so we only used the start site of the first gene in an operon. Similarly, tandemly-duplicated genes may be expressed from duplicated regulatory elements, and clustering between these genes was therefore excluded from our calculations. After removing operons and tandemly-repeated genes (see methods and Supplemental Fig. 3 and 4), 1304 genes remain on the list of muscle-expressed genes, and 386 of these (29.6%) are within 10 kb of another muscle-expressed gene. The number of clustered muscle-expressed genes is significantly greater than would be expected due to random chance (310;  $p < 10^{-4}$ ) (Table 1, Fig. 4a). We also showed that the clustering of co-expressed genes is not due to unrecognized operons or read-through transcription; we repeated this analysis considering only those neighbors that are convergently or divergently transcribed and found that they still showed significant levels of clustering ( $p = 2.0 \times 10^{-6}$ ).

The L1 muscle genes are positioned along the chromosomes in small clusters of 2 to 5 genes each (Fig. 4b and c). 85 of the 174 clusters that contain muscle-expressed genes (48.9%) are interrupted by a gene that is not detectably expressed. The genes in clusters contain approximately the same set of biological functions (Supplemental Table 7), are enriched about the same extent by mRNA-tagging, and are expressed at about the same level as the rest of the muscle-expressed genes (Supplemental Note 1). Hence, clustered genes have a wide variety of functions and we found no evidence that they are a specific subset of muscle-expressed genes.

We extended our analysis to determine if other groups of genes that are expressed together exhibited clustering along the chromosomes. Previous microarray studies have identified 650 genes enriched in sperm, 258 genes enriched in oocytes, and 508 germ line intrinsic genes<sup>12</sup>. All three germ line groups showed significant levels of

gene clustering along the chromosomes ( $p < 0.05$ )(Table 1). A different way to find genes that are expressed together is to use expression profiles from a large number of microarray experiments. Specifically, 553 diverse microarray experiments were used to define 44 sets of co-regulated genes reflecting similar expression during development, under different growth conditions and in many different types of mutants<sup>13</sup>. 13 of the 15 largest gene clusters showed significant levels of gene clustering ( $p < 0.05$ )(Table 1).

As a negative control, we performed an analysis of gene clustering using three groups of genes that are predicted to be expressed in a variety of different cell types at different times in development: genes encoding oxidoreductases, transcription factors and proteins involved in lipid, fatty acid, and sterol metabolism (Table 1). None of these groups of genes are clustered along the chromosomes ( $p > 0.2$ ).

Overall, we have found compelling evidence that genes expressed together in the same tissues or co-regulated in diverse microarray experiments are clustered in small groups along the chromosomes. Interestingly, recent work has shown that genes that are highly expressed in a variety of human tissues or that are co-expressed in yeast are clustered along the chromosomes<sup>14,15</sup>. These results suggest that gene clustering of functionally-related genes may occur in all metazoans.

One interpretation of the results presented here is that gene clusters may correspond to regions of active chromatin. Since opening chromatin for one gene can result in the opening of neighboring chromatin<sup>16,17</sup>, transcriptional machinery may access two co-expressed genes more efficiently if they are neighbors than if they are far apart. In this model, the fortuitous juxtaposition of co-expressed genes would be selected over an evolutionary time scale. An alternative interpretation of clustering is that genes expressed in the same tissue appear in clusters because neighboring genes share a single DNA enhancer element. In this model, neighboring genes appear clustered depending on the strength of nearby enhancers and would not require genomic

rearrangements. In conclusion, our observations that co-expressed genes are clustered along the chromosomes of *C. elegans* provide direct evidence that there is a higher order organization of genes within the genome.

## **Methods**

### **Molecular Biology**

Standard molecular biology and worm culture techniques were used to respectively generate the *myo-3p::FLAG::pab-1* construct (*pPRSK9*) and the resulting SD1075 strain (see the supplemental website for details).

### **mRNA-Tagging**

The mRNA tagging protocol is a modification of the protocol by Goodwin et al.<sup>18</sup>. For each sample, 1 ml of packed worms were resuspended in M9, and then fixed in 0.5% formaldehyde in M9 buffer for 1 hour at 4°C. Animals were rinsed in homogenization buffer (HB), and then resuspended in 3 mls of HB. HB is prepared by DEPC treating 300 mM NaCl, 50 mM Hepes pH 7.6, 10 mM MgCl<sub>2</sub>, 1 mM EGTA, 30 mM EDTA, 0.2 mg/ml Heparin (sodium salt), and 10% Glycerol, and then adding DTT to 1 mM, vanadyl ribonucleoside complex to 8 mM (Sigma), 50 U RNasin/ml, and \_ protease inhibitor cocktail tablet (Roche). Animals were lysed with 2-3 passes on a French press at 8000 PSI, and then 25 passes on a Wheaton homogenizer. Large debris was sedimented by centrifugation at 4300 x g for 6 minutes, and then smaller debris was separated from the supernatant by centrifugation at 48000 x g for 20 minutes. At this point, the lysate could be flash frozen and stored for future use.

RNA bound to the FLAG::PAB-1 was enriched using anti-FLAG-M2 affinity gel beads (Sigma). The affinity beads were prepared by rinsing two times in RNase-free glycine

HCl (pH 3.5) and then four times in HB at 4 °C. The beads were collected by centrifugation, and stored at 4 °C. 1 ml of lysate, 50U of RNasin, 8 ul of 200 mM vanadyl ribonucleoside, and 75 ul of pelleted anti-FLAG-M2 affinity gel beads were mixed together for 2-3 hours at 4°C. The affinity beads were washed four times with cold HB, and then the RNA-protein cross links were reversed by incubating the beads in 125 ul of elution buffer (50 mM Tris/HCl pH8.0, 10 mM EDTA, 1.3% SDS; 20U Rnasin) at 65°C for 30 minutes. The supernatant containing the RNA was collected, the elution was repeated once more and then the two supernatants were combined.

RNA was isolated by mixing 250 ul of the immunoprecipitation supernatant (or cell extract) with 1 ml of trizol, and then mixing 250 uls of chloroform. After letting the mixture stand for 10 minutes, the mixture was spun in a microfuge at 14,000 RPM for 15 minutes at 4°C, and then the supernatant was extracted with chloroform. RNA was precipitated by adding 600 ul of isopropanol, incubating at room temperature for 10 minutes and then centrifuging in a microfuge for 20 minute at 14,000 RPM at 4°C. The RNA pellet was rinsed twice with 1 ml of 70% ETOH.

RNA was linearly amplified as previously described<sup>19</sup>. A description of the DNA microarrays, probe preparation, and microarray hybridization are described in Jiang *et al.*, 2000 (see the supplemental website for more details).

### **Data Analysis (mRNA-Tagging)**

To identify genes that are significantly enriched by mRNA tagging, we first normalized the total amount of cy3 and cy5 signal to each other in each hybridization. We measured the ratio of the signals from the co-immunoprecipitated RNA (Cy-5) to total RNA in the cell extract (Cy-3), and then calculated the percentile rank for each gene relative to all of genes in each hybridization. The mRNA tagging experiment was

repeated six times, and the average percentile rank (mean) from all repeats was determined. A Student's t test was then used to determine which genes showed a mean enrichment that was significantly greater than the average enrichment for all genes. Mock mRNA-tagging was done using four repeats with wild-type (N2) worms.

### **Data Analysis (Chromosomal cluster analysis)**

A gene from the muscle gene list was counted as clustered if its start position was within 10 kb of the start position of another muscle gene. We also varied the distance criteria between 1 kb and 1 MB and observed significant clustering ( $p < 0.001$ ) from 1 to 25 kb.

A detailed explanation of the calculations used to measure gene clustering is at the supplemental web site. Briefly, the calculations included only those genes that were present on the microarray and for which we could determine a chromosomal position<sup>8</sup>. The calculations used only the first gene in an operon, and only one gene of tandem repeats. The number of clusters expected due to random chance was calculated separately for each chromosome, and then summed to give the total number. This was done so that the number expected due to random chance reflected any bias in the experimental list. Finally, since the germ line data were obtained from experiments using microarrays containing 11,917 genes<sup>12</sup>, lists of genes were randomly selected from only these genes to avoid bias.

1. Weintraub, H. Tissue-specific gene expression and chromatin structure. *The Harvey Lectures* **79**, 217-244 (1984).
2. Beermann, W. Control of Differentiation at the chromosomal level. *J. Exp. Zool.* **157**, 49-62 (1964).
3. Gall, J. G. & Callan, H. G. H3 Urindine incorporation in lampbrush chromosomes. *Proc Natl Acad Sci U S A* **48**, 562-570 (1962).

4. Grolach, M., Burd, C. G. & Dreyfuss, G. The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. *Exp Cell Res* **211**, 400-7 (1994).
5. Okkema, P. G., Harrison, S. W., Plunger, V., Aryana, A. & Fire, A. Sequence requirements for myosin gene expression and regulation in *Caenorhabditis elegans*. *Genetics* **135**, 385-404 (1993).
6. Jones, A. R., Francis, R. & Schedl, T. GLD-1, a cytoplasmic protein essential for oocyte differentiation, shows stage- and sex-specific expression during *Caenorhabditis elegans* germline development. *Dev Biol* **180**, 165-83 (1996).
7. Jiang, M. et al. Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **98**, 218-23 (2001).
8. Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J. & Spieth, J. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res* **29**, 82-6 (2001).
9. Fukami-Kobayashi, K., Tomoda, S. & Go, M. Evolutionary clustering and functional similarity of RNA-binding proteins. *FEBS Lett* **335**, 289-93 (1993).
10. Blumenthal, T. Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet* **11**, 132-6 (1995).
11. Blumenthal, T. Gene clusters and polycistronic transcription in eukaryotes. *Bioessays* **20**, 480-7 (1998).
12. Reinke, V. et al. A global profile of germline gene expression in *C. elegans*. *Mol Cell* **6**, 605-16 (2000).
13. Kim, S. K. et al. A gene expression map for *Caenorhabditis elegans*. *Science* **293**, 2087-92 (2001).
14. Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet* **26**, 183-6 (2000).
15. Caron, H. et al. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**, 1289-92 (2001).
16. Hebbes, T. R., Clayton, A. L., Thorne, A. W. & Crane-Robinson, C. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *Embo J* **13**, 1823-30 (1994).
17. Stalder, J. et al. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell* **20**, 451-60 (1980).
18. Goodwin, E. B., Okkema, P. G., Evans, T. C. & Kimble, J. Translational regulation of *tra-2* by its 3' untranslated region controls sexual identity in *C. elegans*. *Cell* **75**, 329-39 (1993).
19. Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T. & Marincola, F. M. High-fidelity mRNA amplification for gene profiling. *Nat Biotechnol* **18**, 457-9 (2000).
20. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863-8 (1998).

Supplemental information is available (<http://cmgm.stanford.edu/~kimlab/muscle/>).

**Correspondence and requests for materials should be addressed to S.K.K. (e-mail: [kim@cmgm.stanford.edu](mailto:kim@cmgm.stanford.edu)).**

We thank J. Shaw, K. Mach, M. Laub, J. Lieb, M. Hiller, E. Harmon and J. Wang for helpful discussions, T. Blumenthal and J. Wang for sharing unpublished results, J. Wang for wild-type poly(A) mRNA, and M. Kiraly for assistance with the supplemental website. We also thank the programmers at the Stanford Microarray Database for microarray analysis and database management, and Proteome™ and Wormbase for annotation of *C. elegans* genes. P.J.R. is a Stanford University Beckman Fellow. This work was supported by a Human Frontiers Fellowship (P.J.R.) and grants from the NIH (S.K.K.).

Figure 1. Outline of the mRNA-tagging technique. A. Schematic of *myo-3p::flag::pab-1*. Initiator methionine and stop codon are shown. Dark blue boxes show the predicted *pab-1* exons (underlined and labeled), white boxes show UTRs, and the light blue box represents the FLAG-tag. B. Schematic of the muscle mRNA-tagging strain. *myo-3p* drives expression of FLAG::PAB-1 in non-pharyngeal muscle cells (red), but not in other tissues, such as the intestine (gray). Thus, FLAG::PAB-1 (blue) binds only to muscle-expressed mRNAs, and only muscle-expressed mRNAs co-immunoprecipitate with FLAG::PAB-1.

Figure 2. Muscle mRNA-tagging. A. Anti-FLAG antibody staining of a worm expressing *myo-3p::FLAG::PAB-1* showing specific expression in striated muscle. bwm; body wall muscle. B. Western blotting analysis of immunoprecipitation of FLAG::PAB-1 (black arrow) from cell-free extracts. Western blots were stained with anti-FLAG antibodies. Samples 1 and 2 are extracts prepared from wild-type and FLAG::PAB-1-expressing animals,

respectively. Lysate refers to the cell free lysate, super refers to the supernatant after immunoprecipitation of FLAG::PAB-1, and eluate refers to immunoprecipitated material. Red arrows denote the light and heavy chains of the anti-FLAG antibodies. The eluate lanes contain 1.6x more sample than the lysate and supernatant lanes. C. Dot blot analysis of the tissue specificity of mRNA-tagging. Lane 1 is a 250 ng poly(A) mRNA standard. Lanes 2 and 3 contain RNA that was co-immunoprecipitated from wild-type and *myo-3p::FLAG::PAB-1*-expressing animals, respectively. The left and right blots were hybridized with an *unc-54* probe, which is specifically expressed in body wall muscles, and a *gld-1* probe, which is expressed specifically in the germ line, respectively.

Figure 3. Genes enriched from muscle by mRNA-tagging. A. Histogram of the average percentile rank of enrichment following mRNA-tagging. The X axis shows the average percentile rank of enrichment, and the Y axis shows the number of genes. Outlined bars represent the distribution for all genes. Red bars represent the distribution of the 1364 genes that are significantly enriched (Student's t-test,  $p < 0.001$ ). B. Graphical representation of the percentile ranks of the 1364 muscle enriched genes. Each row represents a single gene. Color represents the percentile rank of enrichment. The first six columns (1-6) are the six repeats and the seventh (A) is the average percentile rank. Genes are listed in descending order of their average percentile rank. See the supplemental website for how the graphical representation was generated. C. Pie chart showing the types of genes enriched in L1 muscle. Numerator denotes the number of genes in the list

that are muscle expressed, and denominator denotes the total number of genes in that biological group.

Figure 4. Muscle gene positions are clustered in small groups of two to five genes along the chromosomes. A. Histogram comparing 10,000 random samples with the observed muscle gene clusters (red arrow). B. Distribution of the number of muscle-expressed genes in each cluster. C. Schematic of 0.7 Mb of chromosome V, showing the positions of all genes (black), and genes expressed in muscle (red).

Table 1. Genes expressed in the same tissue tend to cluster along the chromosomes.

Gene Group <sup>a</sup>	Total Genes Considered <sup>b</sup>	Genes Considered <sup>c</sup>	Observed Clustered <sup>d</sup>	Expected Clustered <sup>e</sup>	P Value <sup>f</sup>
<b>L1 Muscle-Enriched Genes</b>	1364	1304	386	310.1	$7.2 \times 10^{-6}$
<b>Sperm-Enriched Genes</b>	650	616	198	112	$< 10^{-16}$
<b>Oocyte-Enriched Genes</b>	258	242	26	14.4	0.0174
<b>Germline-Intrinsic Genes</b>	508	475	136	67.9	$< 10^{-16}$
<b>mountain 00</b>	2703	2377	962	868.9	$5.8 \times 10^{-4}$
<b>mountain 01</b>	1818	1662	435	465.9	0.8936
<b>mountain 02</b>	1465	1100	360	270.6	$< 10^{-16}$
<b>mountain 03</b>	1363	1166	316	271.8	0.0017
<b>mountain 04</b>	1195	1113	466	268.2	$< 10^{-16}$
<b>mountain 05</b>	1012	944	248	160.7	$< 10^{-16}$
<b>mountain 06</b>	978	915	214	159.8	$1.6 \times 10^{-4}$
<b>mountain 07</b>	909	840	133	130.7	0.3243
<b>mountain 08</b>	810	761	225	119.8	$< 10^{-16}$
<b>mountain 09</b>	786	744	172	102.8	$< 10^{-16}$
<b>mountain 10</b>	635	555	97	64.8	0.0013
<b>mountain 11</b>	587	546	133	71.5	$< 10^{-16}$
<b>mountain 12</b>	462	435	86	35.9	$< 10^{-16}$
<b>mountain 13</b>	396	368	49	27.3	$1.3 \times 10^{-6}$
<b>mountain 14</b>	353	331	46	22.1	$4.4 \times 10^{-6}$
<b>Oxidoreductases</b>	323	317	21	16.7	0.2252

<b>Transcription Factors</b>	255	214	8	8.6	0.5558
<b>Lipid, Fatty Acid, and Sterol Metabolism</b>					
<b>Genes</b>	281	280	12	12.4	0.5329

<sup>a</sup>The genes in each list, except for the mountains, are given in supplemental Table 1. The genes in each mountain can be accessed at [http://cmgm.stanford.edu/~kimlab/topomap/c.\\_elegans\\_topomap.html](http://cmgm.stanford.edu/~kimlab/topomap/c._elegans_topomap.html).

<sup>b</sup>The total number of genes considered

<sup>c</sup> The number of genes considered, after removing operons, putative tandem duplicates, and genes with uncertain genomic positions.

<sup>d</sup>The number of genes in the gene group that are within 10 kb of another gene from the same group.

<sup>e</sup>The number of clusters expected due to random chance. Shown are the average number of clustered genes from 10,000 repetitions of randomly selecting the same number of genes from the genome as the experimental gene group and counting how many genes are within 10 kb of another gene from the same randomly selected list.

<sup>f</sup>The probability that the observed number of genes that cluster could be matched or exceeded by chance.