Supplementary Methods

Worm profiling

For each analyzed strain, 20 to 30 transgenic worms were placed on a 10cm NGM agar plate seeded with OP50 and left to proliferate at 20ºc for two generations. The mixed stages population was washed out and analyzed using a "Complex Object Parametric Analysis and Sorter" (COPAS) equipped with a profiler system that can analyse up to ~100 animals/sec. This instrument is a flow cytometer with a flow cell that can accommodate objects as large as 1,500 µm. Inside the COPAS profiler flow cell, the laser beam has been optically focused to a 10 µm section, giving rise to fluorescent emission profiles in which details the size of a single cell can be resolved axially along the *C. elegans* body. Profiles were acquired until most worm sizes (from 0.1 mm L1 larvae worm; 50 data-points to ~1mm adult worm; 450 data-points) were represented. This intensity measure is coded as an integer ranging for 0 to 4096. To visually represent this data, we generated a Perl script that transformed those series of profiles into a bitmap image. The value of each data-point of a profile was represented as black-green-yellow-white gradient colour coded pixel on the bitmap image.

Gene Locus / Profiling association

We consolidated all the profiles obtained from several different worm strains bearing a *promoter::GFP* construct corresponding to the same gene locus into a single chronogram. To retrieve the identity of each promoter target we mapped the position of the promoters on our reference genome annotation (wormpep WS150). The functional genomic information used in further analysis (reported RNAi phenotype, Gene Ontology annotations, microarray cluster assignments and protein-protein interactions) was also retrieved from wormbase release WS150.

Profiles orientation

Because worms can enter the profiler presenting their head or their tail first, we re-oriented each profile relative to one another. Profiles acquired from a given transgenic worm population were first ordered by increasing size. An orientation was first arbitrarily chosen for the longest profile. The neighbouring profile was then oriented so that their pairwise Pearson Correlation Coefficient (PCC) was maximized. The PCC was computed as follow:

$$PCC(X,Y) = \frac{\sum_{i}^{n} Xi.Yi - \frac{\sum_{i}^{n} Xi \sum_{i}^{n} Yi}{n}}{\sqrt{(\sum_{i}^{n} Xi^2 - \frac{(\sum_{i}^{n} Xi)^2}{n})(\sum_{i}^{n} Yi^2 - \frac{(\sum_{i}^{n} Yi)^2}{n})}}$$

with Xi and Yi the value of the GFP intensity for the position i for the profiles X and Y respectively and n the length of the profiles

This step was repeated until the 20 first longest profiles were oriented. All subsequent profiles in decreasing order of their size were oriented so that the chosen orientation would maximize its correlation with its 20 preceding profiles. Profiles whose brightest data-point did not exceed the arbitrary threshold of 250 (on the 0 to 4096 scale reported by the COPAS profiler) were oriented but did not participate in the computation of the orientation of the following profiles as they correspond to either animal that don't carry the transgene or do not display expression.

Once all the profiles were oriented relative to one another, two possible chronograms were generated with the two possible orientations. The final orientation (with the animals facing left) was then determined manually based on the position of asymmetric features such as expression in the pharynx, the guts, vulva etc…

Chronogram standardization

Since the size distributions of the population vary between each transgenic strain, we computed a "size standardized" version of each chronogram. Profiles of

identical length were averaged into a single profile, thus generating a unique profile for each worm size. Those averaged profiles were converted into a bitmap image, the value of each data-point, ranging from 0 to 4096, of a profile being represented as a linear black-green-yellow-white gradient color coded pixel of the bitmap image. Would a given size be missing, we introduce in the image a blank line, thus creating chronograms, each representing the activity of a promoter in time and space, that are comparable pixel to pixel.

Chronograms filtering

Not all worm sizes had associated profiles and some recorded profiles contained high local intensity variability. We decided to discard chronograms which had more than 40% missing data-points. This resulted in the removal of 90 chronograms from our dataset.

We applied a two steps noise filtering procedure to the 1520 remaining chronograms. The implementation of this filtering procedure was done using Matlab 7 Image processing toolbox.

The first step was the systematic application of a logical filter to deal with missing sizes. This filter reduces the resolution of a chronogram, using a 7x3 grid to calculate the average of the non-empty data-points of the original image chronogram. Thus, the resulting filtered chronograms are a pixilated version of the original ones, i.e. they contain fewer but more significant data-points. The second step was the systematic application of a linear filter to damp high frequency noise and obtain clearer expression patterns. A linear filter consists in replacing the value of a certain data-point by a linear combination of its value and those of its neighbors. This procedure does not further reduce the resolution of the chronogram.

Mathematically we can express a linear filter as

$$M'_{ij} = \sum_{\substack{\alpha=i-1 \\ \beta=j-1}}^{\substack{\alpha=i+1 \\ \beta=j+1}} M_{\alpha\beta} W_{\alpha\beta} \ ,$$

Where M is the original matrix, W is the weight matrix and M' is the modified matrix.

Chronograms signal level classification

In order to select for further analysis chronograms whose expression was above background, we devised an automated scoring method that was further complemented by manual curation.

An average chronogram, composed of the 68 chronograms for which the average intensity of expression was the lowest, was generated to serve as a quantitative prototype for background expression level.

We then computed the average z-score of each data-point of a chronogram (C) against the background chronogram prototype (B). The z-score was calculated using the following formula:

$$Z = \frac{\sum_{i,j}^{n} \frac{C_{ij} - m(B)_{ij}}{\sigma(B)_{ij}}}{n}$$

Where C_{ij} is the value of the data-point (i,j) of the chronogram under consideration, $m(B)_{ij}$ is the average value of data-point for the background chronogram, $\sigma(B)_{ij}$ is the standard deviation of the data-point(i,j) for the background chronogram, n is the number of data-points considered.

The z-score of each chronogram was used to classify the dataset into two groups: a low confidence set (z-score < 11) containing 790 chronograms and a high confidence set (z-score >= 11) formed by 820 chronograms. This classification was, then, hand curated to include chronograms displaying a weak but observable expression pattern different from the one arising from the autofluorescence of the worm's gut and reclassifying strong but dubious pattern. This resulted in 53 high z-score images being excluded from the high confidence

set and 109 images were promoted from the low confidence to the high confidence set. This resulted in a set of 876 high confidence chronograms and 734 low confidence chronograms.

Sensitivity and reproducibility

The COPAS-profiler technology provides the opportunity to address the biological noise issue by analyzing large number of animals. However to study the noise related to stochastic variations of the promoter activity one would have to get rid of the noise associated with variation of copy numbers of the transgene in the extrachromosomal arrays or associated with mosaicism. To provide an estimate of the homogeneity of expression within a population of worms, graphs are available on the Localizome web site that indicate the percentage of animals that displayed fluorescence intensity above our threshold of expression (> 250).

We evaluated the reproducibility of the chronogram method by measuring the homogeneity of the patterns found where processing the same strain twice (n=14). In 50% of the cases, the obtained PCC is within the top 95^{th} percentile of the general distribution of PCC between run pairs. This corresponds to a tenfold enrichment of extremely strong correlation (p-value=6e-13, Fisher exact test). The same calculation applied to independent strains reporting the activity of promoters associated to the same locus (n= 228) showed a 6.5-fold enrichment of strong correlation (p-value=3e-41, Fisher exact test), indicating the reproducibility of expression patterns obtained by *prom::GFP* fusions.

Average chronograms

A graphical and intuitive way to cross-validate the data is to study the graphical appearance of the average of chronograms associated with a certain functional or anatomical annotation. To prevent chronograms displaying strong expression to dominate weaker ones, each “annotation associated average chronogram” was generated by computing the overall intensity weighted average of each

chronogram associated with a given functional annotation. We then divided the value of each data-point of the obtained average chronograms by the value of the corresponding data-point of the average of all chronograms regardless of their annotations. This allows us to extract the overexpressed pattern associated with a given annotation. Mathematically, the value of a data-point of the average chronogram associated with the i^{th} annotation and located in the (m,n) position will be given by:

$$I_{average_i}(m,n) = \frac{\sum_{j \in Annotation_i} \frac{I_j(m,n)}{\sum_{m,n} I_j(m,n)}}{\sum_{\forall j} I_j(m,n)}$$

Hierarchical clustering of the chronograms

The procedure used to cluster the chronograms consisted in defining and clustering a similarity matrix. The value of the similarity between two images was calculated by using the Pearson's Correlation Coefficient (PCC), computed as follow:

$$PCC(X,Y) = \frac{\sum_{i,j}^{n} Xij.Yij - \frac{\sum_{i,j}^{n} Xji \sum_{i,j}^{n} Yij}{n}}{\sqrt{(\sum_{i,j}^{n} Xij^2 - \frac{(\sum_{i,j} Xij)^2}{n})(\sum_{i,j}^{n} Yij^2 - \frac{(\sum_{i,j} Yij)^2}{n})}}$$

with Xij and Yij the value of the GFP intensity for the position (i,j) for the chronograms X and Y respectively and n the number of data-points considered.

The similarity matrix was clustered by using the Average Linkage Clustering algorithm implemented in Matlab 7 Statistical Toolbox. The heatmap associated with the average linkage tree was also generated using Matlab 7 Statistical Toolbox.

Extraction of functionally enriched subtrees

Chronogram clusters were analyzed for enrichment of a variety of annotation types. Each clade or “cluster” in the dendrogram can be associated with a 876-bit string (each bit indicating membership for a distinct gene/chronogram in the cluster). 586 such cluster strings were generated (one for each non-leaf node in the dendrogram, minus the top-level trivial cluster). 1760 annotation strings (also 876-bit) were similarly constructed, based on GO term association, microarray cluster membership, microscopic anatomical annotation or RNAi phenotype.

To test for enrichment of the annotation types within each cluster, a total of 586□1760=1031360 P-values would need to be computed (one for each cluster/annotation string pair). We used a one-sided cumulative hypergeometric test to detect significant enrichment (Tavazoie et al). Since this test is computationally intensive, we applied a pre-computation filter to remove many of the “uninteresting” results prior to computing the exact P-value. Cluster/annotation strings that share fewer than 3 annotations were removed from the test pool, reducing the number of tested hypotheses to 30748 P-values.

Due to the number of hypotheses tested, we applied a false discovery rate (FDR) controlling procedure to the nominal P-values. Existing FDR controlling procedures (Storey, et al) rely on assumptions that are not directly applicable in our case, in particular that the P-values corresponding to truly null hypotheses are sampled from a continuous probability density function, leading to a uniform null P-value density. The cumulative hypergeometric distribution is discrete (violating the above constraint), the computation-saving filtering skews the null P-value density away from uniformity, and there exist dependencies between annotation strings and cluster strings (dependencies can plague existing FDR estimation methods).

We thus implemented a resampling-based approach to simulate the complete null hypothesis:

1. randomly permute the annotation strings *enmasse* (i.e. each annotation string is shuffled precisely the same way), preserving any dependencies that exist between annotation types (e.g. GO ancestors/descendants);

2. recompute P-values for each annotation/cluster string, applying the same "3 annotation" filter;
3. repeat steps 1-2 for 100 iterations;
4. pool the resulting P-values to generate an empirical null distribution of nominal P-values.

Under the assumption that the observed P-value density is comprised of a mixture of distributions (the null distribution and the alterative distributions), this simulated complete null hypothesis P-value density is overly conservative. To compensate for this, a scaling factor *s* was computed as:

$$s = \frac{|\{p \in O : p > \lambda\}| / |O|}{|\{p \in N : p > \lambda\}| / |N|} \approx \frac{\int_{\lambda}^{1} o(p)dp}{\int_{\lambda}^{1} n(p)dp}, \qquad (1)$$

where *O* and *N* are the collections (and *o*(*p*) and *n*(*p*) are the theoretical density functions) of observed and simulated null P-values, respectively. λ is a tuning parameter in [0,1] that allows one to estimate the smallest P-value such that we do not expect to see any truly alternative hypothesis with P-value $p>\lambda$. (Thus λ is an estimate on the worst type-II error probability for any individual test of a truly alternative hypothesis.) Choosing λ=0.5 leads to *s*=0.868.

Finally, we estimated the FDR f_i for each observed P-value p_i as:

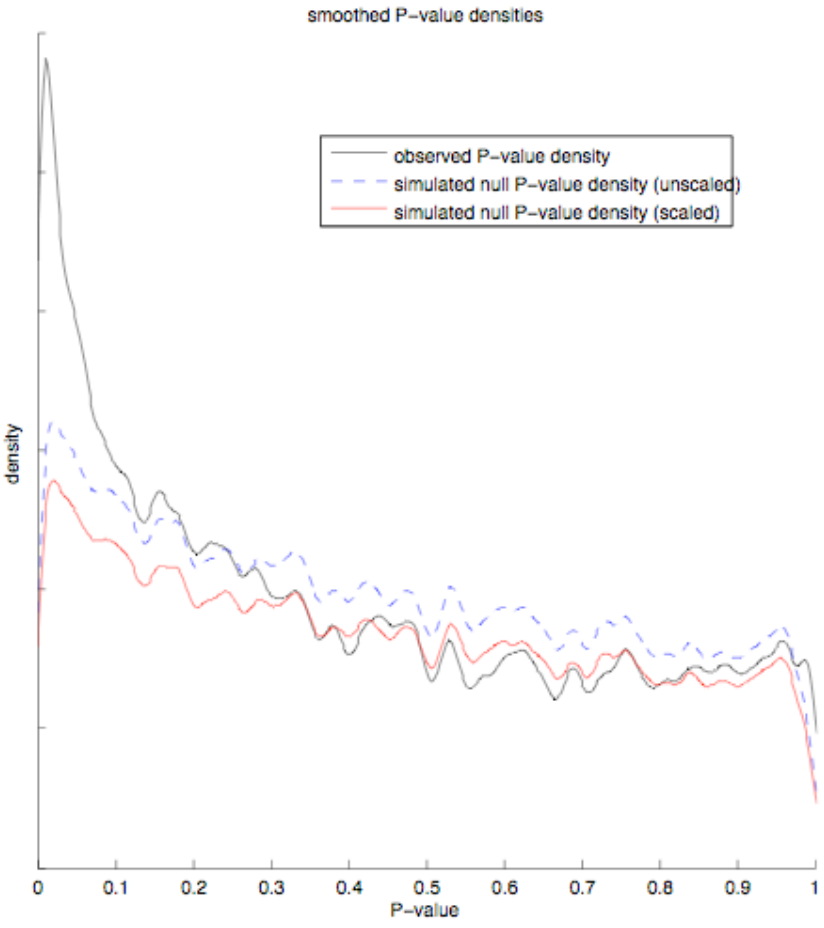
$$f_i = \frac{s \cdot |\{p \in N : p \le p_i\}| / |N|}{|\{p \in O : p \le p_i\}| / |O|} \approx \frac{s \cdot \int_{0}^{p_i} n(p)dp}{\int_{0}^{p_i} o(p)dp}. \qquad (2)$$

Supplemental figure 5 illustrates the smoothed densities of the observed P-values and the simulated null P-values (both before and after scaling) and visually confirms that the simulated null P-values indeed follow the shape of the observed P-values for values greater than the chosen λ.

References

Storey JD and Tibshirani R. (2003) *Statistical significance for genome-wide studies*. Proceedings of the National Academy of Sciences, 100: 9440-9445.

Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. *Systematic determination of genetic network architecture.* Nat Genet. 1999 Jul;22(3):281-5.


smoothed P−value densities
observed P−value density
simulated null P−value density (unscaled)
simulated null P−value density (scaled)
density
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
P−value