nature biotechnology

Genome-scale analysis of *in vivo* spatiotemporal promoter activity in *Caenorhabditis elegans*

Denis Dupuy[1,10], Nicolas Bertin[1,2,10], César A Hidalgo[1,3,10], Kavitha Venkatesan[1], Domena Tu[4], David Lee[4], Jennifer Rosenberg[1], Nenad Svrzikapa[1], Aurélie Blanc[1], Alain Carnec[1], Anne-Ruxandra Carvunis[1], Rock Pulak[5], Jane Shingles[6], John Reece-Hoyes[6], Rebecca Hunt-Newbury[7], Ryan Viveiros[7], William A Mohler[8], Murat Tasan[9], Frederick P Roth[9], Christian Le Peuch[2], Ian A Hope[6], Robert Johnsen[4], Donald G Moerman[7], Albert-László Barabási[1,3], David Baillie[4] & Marc Vidal[1]

Differential regulation of gene expression is essential for cell fate specification in metazoans. Characterizing the transcriptional activity of gene promoters, in time and in space, is therefore a critical step toward understanding complex biological systems. Here we present an *in vivo* spatiotemporal analysis for ~900 predicted *C. elegans* promoters (~5% of the predicted protein-coding genes), each driving the expression of green fluorescent protein (GFP). Using a flow-cytometer adapted for nematode profiling, we generated 'chronograms', two-dimensional representations of fluorescence intensity along the body axis and throughout development from early larvae to adults. Automated comparison and clustering of the obtained *in vivo* expression patterns show that genes coexpressed in space and time tend to belong to common functional categories. Moreover, integration of this data set with *C. elegans* protein-protein interactome data sets enables prediction of anatomical and temporal interaction territories between protein partners.

During development, cell type determination depends on the activation and/or repression of specific subsets of genes[1,2]. One of the requirements to fully understand the molecular networks driving cell differentiation in metazoans is to characterize the *in vivo* expression state of the genome at each differentiation step, that is, to determine what specific set of genes is activated in which specific cell type at what stage of development[3].

C. elegans is a unique multicellular model for such *in vivo* global gene expression studies. Its transparent body and nearly invariant cell lineage[4–6] enable precise, cell-by-cell analysis of promoter activity throughout development in transgenic animals carrying *promoter::GFP* reporter constructs[6–8]. With the completion of the *C. elegans* genome sequence[9], it is conceivable to develop genome-wide analysis of *in vivo* promoter activity for the complete set of promoters. The collection of all obtained patterns, referred to here as a 'localizome' data set, should provide important insight into the functional organization of gene regulation at the genome scale (refs. 10–13, and http://elegans.bcgsc.ca/perl/eprofile/index).

High-magnification fluorescence microscopy has been used to characterize expression at the single-cell level in *C. elegans*. However, this kind of analysis is extremely time consuming and labor intensive[13]. Moreover, because rapid microscopic examination can be performed on only a limited number of animals, the obtained results can be strongly affected by stochastic variations among individuals.

RESULTS

Reconstitution of time-lapse expression patterns

To study *promoter::GFP* expression at the level of a large population of animals with a quantitative read-out, we used a 'complex object parametric analysis and sorter' (COPAS) instrument equipped with a profiler system that analyzes up to ~100 animals/s. This system generates fluorescent emission profiles along the antero-posterior axis of the *C. elegans* body. By analyzing large numbers of animals of all sizes and ages at high throughput, we generated a digitized overview of the promoter activity throughout post-embryonic development.

For each transgenic line analyzed, fluorescence profiles were acquired for thousands of nematodes from a mixed-stage culture. For each worm in the population we then converted the corresponding profile into a color-coded representation of the fluorescence

[1]Center for Cancer Systems Biology (CCSB), and Department of Cancer Biology, Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. [2]Centre de Recherche en Biochimie Macromoléculaire, Centre National de la Recherche Scientifique FRE 2593, 1919 Route de Mende, 34293 Montpellier Cedex 5, France. [3]Center for Complex Network Research, Department of Physics, University of Notre Dame, 225 Nieuwland Science Hall, Notre Dame, Indiana 46556, USA. [4]Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada. [5]Union Biometrica, 84 October Hill Road, Holliston, Massachusetts 01746, USA. [6]Institute of Integrative and Comparative Biology, University of Leeds, Clarendon Way, Leeds LS2 9JT, West Yorkshire, UK. [7]Department of Zoology, The University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada. [8]Department of Genetics and Developmental Biology and Center for Cell Analysis and Modeling, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, Connecticut 06030, USA. [9]Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. [10]These authors contributed equally to this work. Correspondence should be addressed to M.V. (marc_vidal@dfci.harvard.edu), D.B. (baillie@sfu.ca) or A.-L.B. (alb@nd.edu).

Received 14 December 2006; accepted 13 April 2007; published online 7 May 2007; doi:10.1038/nbt1305

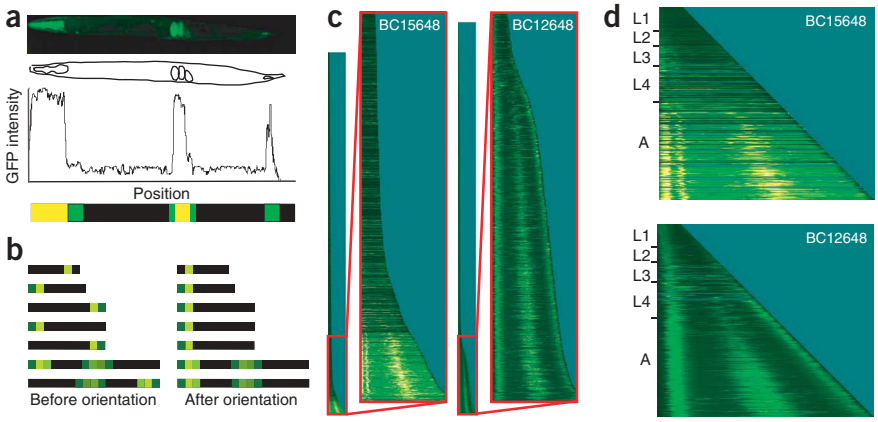

Figure 1 Generating post-embryonic developmental chronograms. (**a**) For each worm, the longitudinal GFP intensity profile is converted into a bar where the length corresponds to the animal's size and the color codes for the fluorescence intensity. (**b**) Profiles are sorted by size and oriented to match their neighbors. (**c**) The complete collection of oriented expression profiles reflects the size distribution of a given population. Represented here are the profiles gathered for the strains BC15648 and BC12648, expressing GFP under the control of *F25B5.1* and *ptl-1* promoters, respectively. The expansions on the right display the profiles of larger animals. (**d**) Averaging profiles of identical size produce chronograms with a standardized shape. Approximate larval stages and adult transitions are indicated on the *y*-axis. The color code represents the absolute GFP intensity measured (increasing values as black-green-yellow-white).

intensity (**Fig. 1a**). After orienting these profiles (**Fig. 1b**), we assembled them so that the short rows at the top represent L1 larvae, whereas the bottom rows correspond to fully-grown adults (**Fig. 1c**). As the distribution of worm sizes varies dramatically between analyzed populations (**Fig. 1c**), we generated images in which each row represents the average of all worms of a given size (**Fig. 1d**). If no animal of a given size was found, the corresponding row was skipped. As these normalized images provide an overview of GFP expression in time throughout post-embryonic development (the length of the worm being a proxy for its age), we refer to them as 'chronograms'.

Several examples of chronograms corresponding to well-characterized GFP-expressing strains are shown in **Figure 2**. Tissue-specific signatures corresponding to major organs are clearly identifiable on the chronograms, even when expression is restricted to a small number of cells, such as olfactory neurons (**Fig. 2b**). An advantage of the chronogram is that it provides information on the temporal expression of the reporter. For example, the onset of *egl-15* expression during the L4 stage is clearly visible in **Figure 2f**.

We acquired chronograms for 1,992 GFP strains which, after accounting for redundancy in transgene content, report the activity of the proximal promoter of 1,610 unique predicted gene loci (**Supplementary Fig. 1** online)[13,14]. We further analyzed 876 chronograms for which the average signal was above background (see "Chronogram signal level classification" in **Supplementary Methods** online). Several factors explain the absence of significant signal in the other 734 chronograms. Most of the strains (79.2%) analyzed carry extrachromosomal arrays of the *promoter::GFP* construct, which may have a transmission rate too low to enable population-scale analysis of expression[15]. Moreover, the presence of extrachromosomal arrays may sometimes cause a deleterious phenotype because, for example, of transcription factor titration effects[16] and may thus be counter-selected. Alternatively, some of the promoters may drive expression at levels below the detection threshold of the instrument under the conditions used here.

Extraction of tissue-specific signatures

We developed an averaging method to extract the characteristic features of sets of chronograms, corresponding either to strains that share individual anatomic annotations defined by microscopic observation[13], or to genes that belong to expression clusters derived from microarray experiments[17]. To prevent chronograms with strong expression from dominating weaker ones, we divided each image by its mean intensity. The average chronogram of all 315 strains annotated as expressed in the pharynx (**Fig. 3a**) displayed a distinctive anterior expression pattern, with a clearly visible separation between the two pharyngeal bulbs, a feature already seen in the *myo-2::GFP* chronogram (**Fig. 2e**), representing the expression of a promoter specific to pharyngeal muscle cells. Furthermore, a late-onset mid-body signal can be seen, corresponding to GFP expression in embryos *in utero*, which is consistent with the timing of pharynx development. The average image corresponding to the 150 strains that scored positive for expression in the nerve ring (**Fig. 3b**) shows a pattern clearly distinct from the pharyngeal one (**Fig. 3a**), with a narrower signal starting at the level of the isthmus of the pharynx. This average nerve ring pattern replicates the signature observed in the chronograms of *unc-119::GFP* (**Fig. 2a**) and *lim-4::GFP* (**Fig. 2b**), both representing expression in the nerve ring.

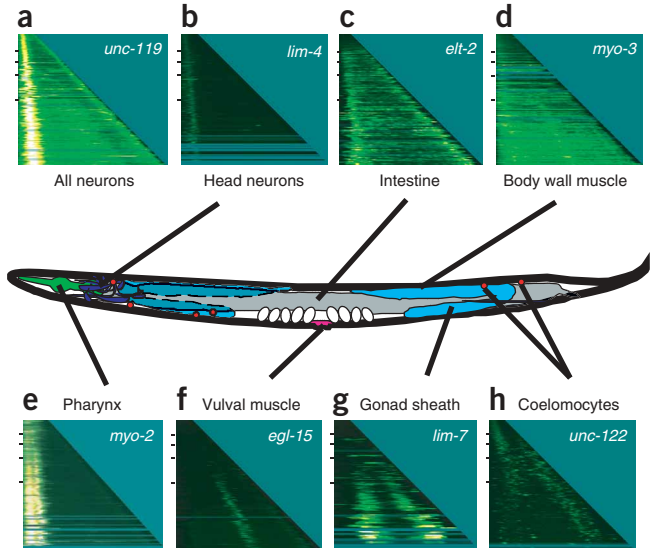

Figure 2 Visualization of tissue-specific expression. (**a**–**h**) Strains known to express GFP in all neurons (*unc-119*, strain IM324 (ref. 36)) (**a**), olfactory neurons (*lim-4*, strain OH96 (ref. 37)) (**b**), intestine (*elt-2*, strain MR142 (ref. 38)) (**c**), body wall and vulval muscles (*myo-3*, strain PD4251 (ref. 39)) (**d**), pharynx (*myo-2*, strain PD4790) (**e**), vulval muscles (*egl-15*, strain NH2447 (ref. 40)) (**f**), gonad sheath cells (*lim-7*, strain OH172 (ref. 41)) (**g**) and coelomocytes (*unc-122*, strain OH910 (ref. 42)) (**h**). For each strain a chronogram is shown in relation to a diagram of the anatomy of a worm.

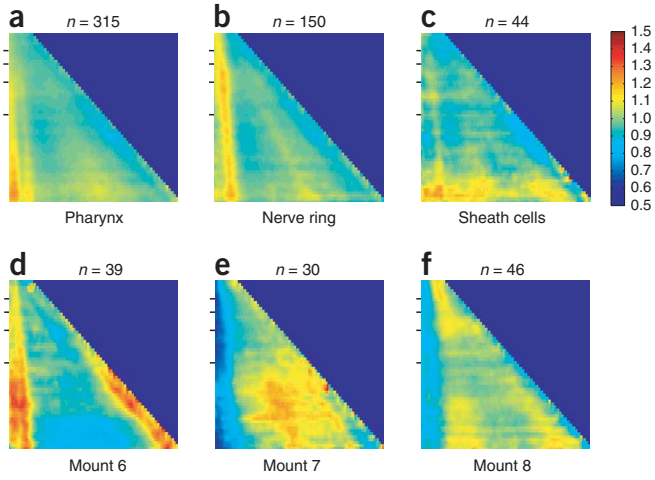

Figure 3 Feature extraction from multiple images. (**a**–**c**) Average chronograms of strains with reporter expression sharing a common anatomic annotation. (**d**–**f**) Strains with reporter fusions for genes corresponding to transcripts clustered in the same topomap mountain[17]. Mounts 6, 7 and 8 are enriched for genes expressed in neurons, germline and intestine respectively. *n* indicates the number of individual chronograms used to generate the average image. The color scale indicates the level of expression of a given position on the chronogram of the group considered relative to the average of all chronograms.

A caveat of this averaging method is that obtaining a robust tissue-specific signature depends on the availability of large numbers of images, because expression is more frequently observed in multiple tissues than in a single one. For example, only 44 strains were annotated with expression in the gonad sheath cells. Because most of them also expressed GFP in a variety of other tissues, the extracted signal (**Fig. 3c**) was not strictly restricted to the tissue-specific signature observed in the *lim-7::GFP* chronogram (**Fig. 2g**).

We also used this averaging approach to extract common features of chronograms corresponding to genes belonging to expression clusters generated based on a compendium of over 500 microarray experiments[17]. The average images associated with certain clusters ('topomap mountains') showed expression patterns that fit with their associated annotations, such as for neurons (mount 6, **Fig. 3d**), germ line (mount 7, **Fig. 3e**) and intestine (mount 8, **Fig. 3f**). However, most topomap mountains, even when sufficiently represented in our data set, did not display any characteristic pattern, suggesting that microarray clusters may not generally constitute a

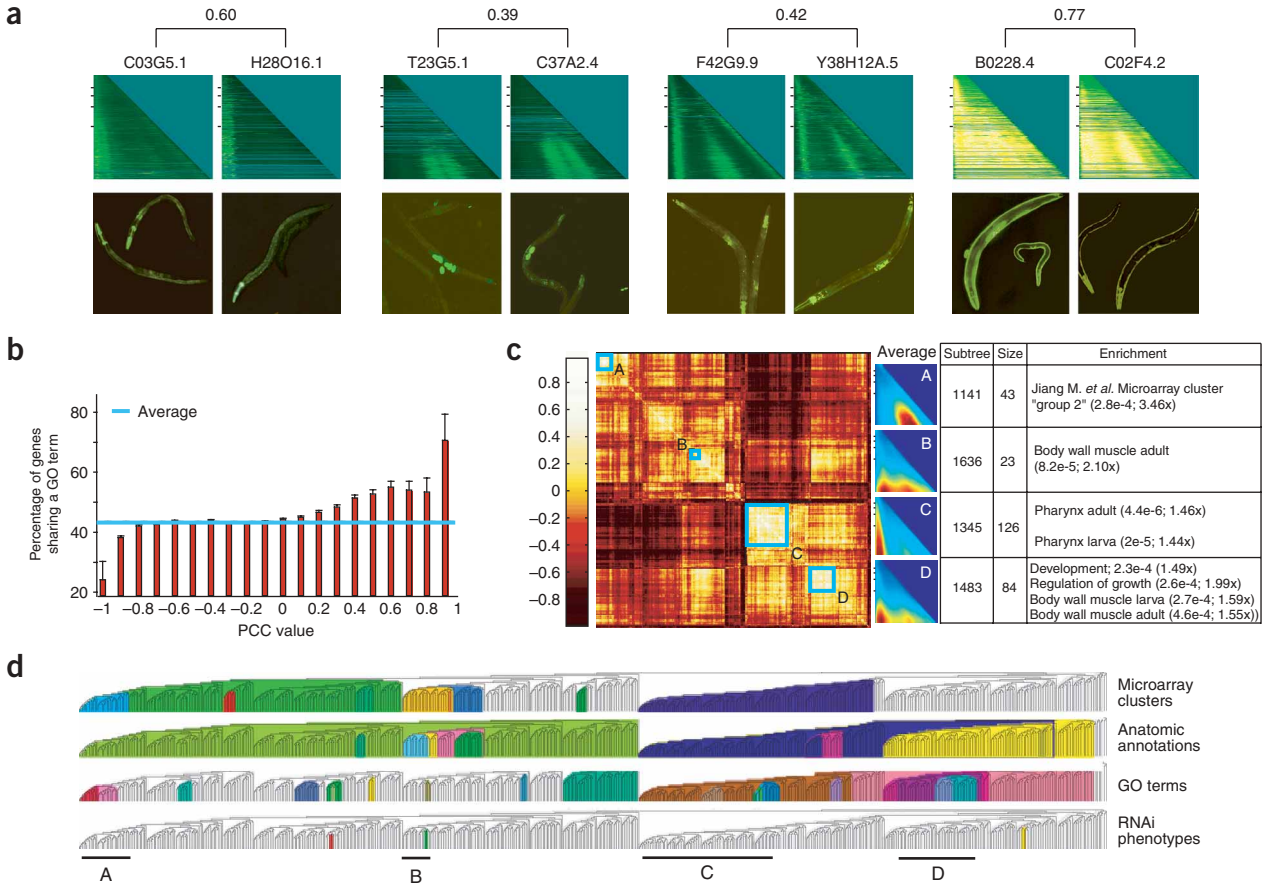

Figure 4 *In vivo* expression pattern clustering. (**a**) Pairs of highly correlated chronograms were selected from the PCC matrix. For each strain, the chronogram and a whole animal fluorescence micrograph are presented. (**b**) Correlation between chronograms PCC and GO terms. (**c**) Clustered matrix of 876 chronograms based on spatial and temporal correlation of expression. Average images of chronograms grouped in some of the neighbor joining tree branches (A–D) are presented as well as their enrichment in specific GO terms, microarray expression cluster[17,23] and/or anatomical annotation. (**d**) Overview of subtrees enrichments in four functional annotation types. Distinct colors indicate distinct categories of enriched terms in each class (**Supplementary Fig. 3** presents an enlarged version of this figure and the list of enriched subtrees is available in **Supplementary Table 1**).

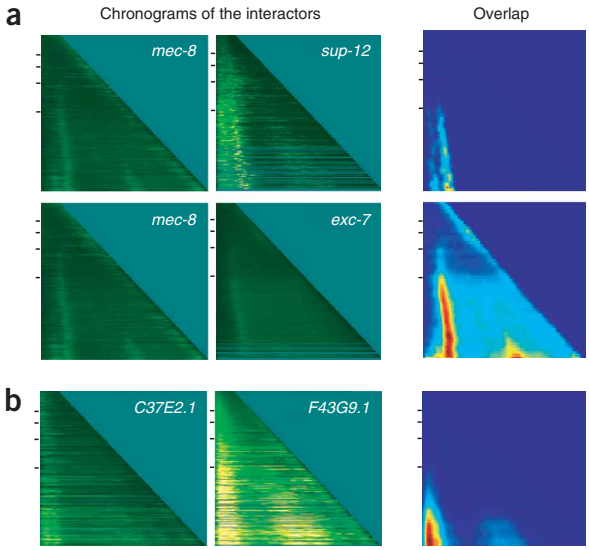

Figure 5 Localization of protein-protein interactions. (**a**) SUP-12 and EXC-7 each physically interact with MEC-8 (ref. 18) in high-throughput yeast two-hybrid. Although all three proteins are implicated in mRNA splicing, their expression patterns are distinct. Overlapping the chronograms of genes encoding interacting proteins can provide insights into where and when the proteins may interact. (**b**) C37E2.1 and F43G9.1 encoding the gamma and alpha subunits of the isocitrate dehydrogenase, respectively, are direct interactors in the WI5 protein-protein interaction map[18] and their chronograms show a high-level of correlation (PCC = 0.53).

good predictor for coexpression in the same tissue. For example, genes grouped in categories such as 'development', 'heat shock', 'aging' or 'Dauer' could accomplish their function in these processes in a wide range of distinct tissues and therefore not be colocalized despite their apparent coexpression.

Chronogram clustering

The digital nature of the chronograms allows quantification of the similarity between them. We calculated Pearson's correlation coefficient (PCC) between all pairs among the 876 chronograms with detectable signal. For a randomly sampled set of pairs within the top 95th percentile of the PCC distribution, we verified that a high PCC value between the chronograms correlated with a strong visual resemblance between the fluorescent micrographs of the corresponding animals (**Fig. 4a** and **Supplementary Fig. 2** online). We also examined whether chronogram similarity correlates with common Gene Ontology (GO) annotations, as has previously been observed between pairs of physically interacting proteins[18], coregulated transcripts[17] (from microarray expression profiling) and genes that share RNAi phenotypes[18–21]. Indeed, two genes whose corresponding chronograms were highly correlated tended to share GO annotations more often than expected by chance ($P = 2e^{-44}$, Fisher exact test; **Fig. 4b**), further demonstrating the power of chronogram comparisons.

We used average-linkage hierarchical clustering[22] to group genes based on the spatiotemporal activity of their promoter throughout post-embryonic development (**Fig. 4c**). As examples, we show four clusters, each corresponding to different branches of the hierarchical tree. For each cluster, we show an average chronogram image and list biological attributes enriched among genes within the cluster. Cluster A displayed significant ($P = 2.8e^{-4}$) enrichment in genes overexpressed in hermaphrodites relative to males and in adults relative to the other stages of development[23]. The average expression pattern observed for cluster A was consistent with these enrichments, with a signal located in the mature uterus and/or developing embryos. Noticeably, cluster A also contained two subtrees that were enriched for the GO terms 'DNA metabolism' and 'mitotic cell cycle' (**Fig. 4d** and **Supplementary Fig. 3** online). This is consistent with expression during early embryogenesis, a developmental stage during which most cell divisions occur. Cluster B was significantly ($P = 8.2e^{-5}$) enriched for genes expressed in adult body wall muscle, consistent with the average image obtained for this cluster, which showed expression appearing in the bigger adult animals. Cluster C enrichment in genes expressed in the pharynx of both larvae and adult animals was reflected by the corresponding average image, which was highly similar to the average chronogram of all strains annotated as expressed in the pharynx (**Fig. 3a**). Finally, cluster D was enriched in genes associated with the regulation of growth rate. It was also enriched in genes expressed in the body wall muscle, like cluster B, which explain the resemblance of their average image. Interestingly, features shared by groups of genes that are distant in the neighbor-joining tree were visible as bright, off-diagonal regions in the matrix (**Fig. 4c**).

Noticeably, the number of subtrees presenting significant ($P < 0.001$) enrichment for a given RNAi phenotype was much lower than for the other functional categories explored (**Fig. 4d**, **Supplementary Table 1** online). This is probably because only 10–12% of *C. elegans*–predicted genes are associated with any RNAi phenotype[18–21,24–26], whereas most are associated with GO terms and/or microarray expression data.

Localizome and interactome

High-throughput interaction mapping techniques, such as yeast two-hybrid, identify physical protein-protein interactions that may occur *in vivo*, but can not determine in what cells or tissues the two proteins in question are coexpressed for the interaction to happen. Comparison of microarray profiles can indicate coregulation between genes but, as most microarray data are derived from whole-animal RNA extraction, it remains possible that their expression occurs in distinct tissues. In contrast, chronograms can be used to define putative common expression territories of interacting proteins by observing the overlap between their expression patterns. MEC-8 was shown to be able to interact with SUP-12 and EXC-7 by high-throughput yeast two-hybrid[18], and all three proteins share the GO annotation 'RNA-binding protein', suggesting that they could function together within a single macromolecular complex. However, the three reporter strains for these genes displayed dissimilar expression patterns with chronograms that overlapped only partially (**Fig. 5a**), suggesting that these proteins function independently of each other *in vivo*, or might interact to function together in only a few cells. This interpretation is consistent with previous studies that indicate independent functions for these proteins. All three proteins have been shown to be involved in regulating tissue-specific alternate splicing, but act in different cells on distinct targets: MEC-8 in the maturation of *unc-52* mRNA in the hypodermis and unknown additional transcripts in neurons[27,28]; SUP-12 on *unc-60* mRNA in muscle cells; and EXC-7 on *sma-1* mRNA in the excretory cells and neurons, respectively[29–31]. On the other hand, strong spatiotemporal expression correlation associated with a protein-protein interaction may indicate a strong functional correlation. For 48 protein pairs from the worm interactome WI5 (ref. 18), the corresponding promoter pairs were assayed in this localizome data set. One of these chronogram pairs (C37E2.1-F43G9.1) scored in the top 95th percentile of the PCC distribution, indicating highly overlapping expression patterns (**Fig. 5b**). Based on the data from these two

unbiased data sets, one could hypothesize a functional association of the two corresponding proteins. Indeed, C37E2.1 and F43G9.1 appear to encode the gamma and alpha subunits of the isocitrate dehydrogenase, respectively.

As the coverage of both worm localizome and interactome maps improve, integration of the two maps will provide more insight into the specific organization of tissue-specific macromolecular networks. For example, we could distinguish two classes of interaction relationships: 'committed' (where the two partners form an obligatory heterodimer and both interactors are always expressed together) and 'uncommitted' (where the interaction is not necessary for the individual proteins to function and both partners may have different expression patterns), a distinction analogous to that proposed between party and date hubs in the yeast interactome[32].

DISCUSSION

The *C. elegans* adult is composed of 959 somatic cells, each identifiable by high-resolution microscopy. However, analyzing individual gene expression patterns at single-cell resolution is currently too cumbersome to be accomplished for all 19,000 genes in *C. elegans*. The high-throughput classification of expression patterns we devised can be viewed as a first step toward more refined annotations and as a complement to traditional microscopic analyses. For example, several clear patterns visible in the chronograms led us to reexamine the corresponding strains and to annotate tissue-specific expression that had been missed in the initial assessment (**Supplementary Fig. 4** online). Chronograms capture the time-lapse expression pattern for post-embryonic development and are quantitative digital data that can be analyzed using mathematical tools similar to those used for microarray data analysis. This approach combines the throughput necessary for genome-wide promoter activity analysis in transgenic worm strains, with a data format suitable for large-scale analyses and for comparisons with other genome-wide data sets.

Chronograms simultaneously provide coarse-grained spatial resolution along the anterior-posterior axis of the animal (length) and high temporal resolution throughout post-embryonic development (time). A limitation of our approach is the lack of the other two spatial dimensions (width and height), which precludes distinguishing between tissues located in the same cross-section (for example, vulval muscles and vulval neurons). We anticipate, however, that future optical developments of the COPAS profiler, or other systems, may provide enhancements in this regard.

Chronograms can be used to identify potential spatiotemporal territories where functional interactions between gene products are most likely to happen. Generating expression annotations for all protein interactions will add spatiotemporal information to the worm interactome network[18]. Localizome mapping will thus help characterize the dynamic aspects of the functional interactions between gene products in *C. elegans*.

METHODS

Worm profiling. For each analyzed strain, 20–30 transgenic animals were placed on 10-cm NGM agar plates seeded with *Escherichia coli* strain OP50 and left to proliferate at 20 °C. Upon exhaustion of the bacterial lawn, the mixed-stage population was washed out and analyzed using a COPAS profiler (Union Biometrica) Individual profiles were acquired until all animal sizes were represented.

Chronogram generation. As the animals pass through the profiler either tail or head first, we oriented all profiles relative to one another by using an automated method based on the best fit with their neighbors as determined by Pearson's correlation coefficient (PCC).

Clustering. Before calculating the correlation between chronograms we reduced their complexity by applying a linear filter to smooth out the signal and eliminate high-frequency noise. Image compression and PCC calculation were performed in Matlab. Images displaying no detectable variation from background were excluded from the clustering analysis (details in **Supplementary Methods** section "Hierarchical clustering of the chronograms"). The neighbor-joining tree was calculated using average-linkage hierarchical clustering[22]. Tree illustrations were performed with the tree editor TreeDyn (http://www.treedyn.org/)[33].

Cluster enrichments. The significance of enrichment for gene attributes within a list of genes induced by each given subtree was calculated using the cumulative hypergeometric distribution[34]. False discovery rates were associated with each observed nominal *P*-value according to an empirical null distribution of nominal *P*-values calculated similarly from 100 random permutations of all genes within the same tree structure (full details in **Supplementary Methods** in the section "Extraction of functionally enriched subtrees").

Strains origin. Most transgenic lines analyzed here were generated either by a modification of the microinjection method described by Mello *et al.* (1991) where 5′ regulatory DNA::GFP constructs and *dpy-5*(+) plasmid (pCeh-361) and selection for rescue of the Dpy-5 mutant phenotype[35], or by microprojectile bombardment[14]. A small fraction was provided by the Caenorhabditis Genetics Center. Individual strain origins are available for each expression patterns on the localizome webpage (see below) and at Wormbase (http://www.wormbase.org/).

Data availability. Data collected in the course of this project are available on Wormbase for batch download of both raw data and processed chronograms. We also created a searchable Localizome database that is freely accessible through the Internet and that provides the user with the anatomic annotation defined by microscopic observation and the associated chronogram. For any query gene, the web interface also displays the best chronogram matches, providing a convenient way to identify genes with similar expression patterns (http://vidal.dfci.harvard.edu/localizome/).

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

This work was funded by the National Cancer Institute (NCI 4 R33 CA097516-02)(M.V.), Genome British Columbia, the Canadian Institute of Health Research and Genome Canada (D.B. and D.G.M.), the NetWork Bench National Science Foundation (IIS-0513650), the Muscular Dystrophy Association, the National Institutes of Health (NIH) (1 P20 CA11300-01, A.L.B. and HD43156, W.A.M.) and the Helen Kellogg Institute for International Studies (C.A.H.R.). F.P.R. was supported in part by NIH grant HG003224. M.T. was supported by NIH NRSA Fellowship HG004098. Some nematode strains used in this work were provided by the *Caenorhabditis Genetics* Center, which is funded by the NIH National Center for Research Resources. Thanks to Tracey Clingingsmith and Abigail Bird for their indispensable assistance and to Michael Cusick and Mike Boxem for careful proofreading of the manuscript.

AUTHOR CONTRIBUTIONS

Transgenic animals were generated by ballistic transformation by J.S. and J.R.-H. under the supervision of I.A.H., and by microinjection by D.T. and D.L. under the supervision of R.J. and D.B., anatomic annotations of the strains were performed by R.H.-N. and R.V. under the supervision of D.G.M. The chronogram concept was conceived by W.A.M. and implemented by D.D., N.B. and C.A.H. D.D. generated the Gateway *promoter::GFP* constructs with the help of J.R., and performed the profiling experiments with the technical support of R.P. Computational analyses were performed by N.B., C.A.H., K.V., A.-R.C., A.C. and M.T. under the supervision of D.D., F.P.R., C.L., A.-L.B. and M.V. Lab support was provided by J.R., N.S. and A.B. The manuscript was written by D.D., N.B., C.A.H., A.-L.B. and M.V. The project was conceived and codirected by D.B. and M.V.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Davidson, E.H. *et al.* A genomic regulatory network for development. *Science* **295**, 1669–1678 (2002).
2. Inoue, T., Wang, M., Ririe, T.O., Fernandes, J.S. & Sternberg, P.W. Transcriptional network underlying *Caenorhabditis elegans* vulval development. *Proc. Natl. Acad. Sci. USA* **102**, 4972–4977 (2005).
3. Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L. & Myers, R.M. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* **16**, 1–10 (2006).
4. Sulston, J.E., Schierenberg, E., White, J.G. & Thomson, J.N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
5. Sulston, J.E. & Horvitz, H.R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977).
6. Kimble, J. & Hirsh, D. The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. *Dev. Biol.* **70**, 396–417 (1979).
7. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. & Prasher, D.C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).
8. Bao, Z. *et al.* Automated cell lineage tracing in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **103**, 2707–2712 (2006).
9. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
10. Vidal, M. A biological atlas of functional maps. *Cell* **104**, 333–339 (2001).
11. Dupuy, D. *et al.* A first version of the *Caenorhabditis elegans* Promoterome. *Genome Res.* **14**, 2169–2175 (2004).
12. Hobert, O. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *Biotechniques* **32**, 728–730 (2002).
13. McKay, S.J. *et al.* Gene expression profiling of cells, tissues, and developmental stages of the nematode *C. elegans*. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 159–169 (2003).
14. Reece-Hoyes, J.S. *et al.* Insight into transcription factor gene duplication from *Caenorhabditis elegans* Promoterome-driven expression patterns. *BMC Genomics* **8**, 27 (2007).
15. Mello, C.C., Kramer, J.M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
16. Toms, N., Cooper, J., Patchen, B. & Aamodt, E. High copy arrays containing a sequence upstream of *mec-3* alter cell migration and axonal morphology in *C. elegans*. *BMC Dev. Biol.* **1**, 2 (2001).
17. Kim, S.K. *et al.* A gene expression map for *Caenorhabditis elegans*. *Science* **293**, 2087–2092 (2001).
18. Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
19. Gunsalus, K.C. *et al.* Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* **436**, 861–865 (2005).
20. Ge, H., Liu, Z., Church, G.M. & Vidal, M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat. Genet.* **29**, 482–486 (2001).
21. Rual, J.F. *et al.* Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res.* **14**, 2162–2168 (2004).
22. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).
23. Jiang, M. *et al.* Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **98**, 218–223 (2001).
24. Kamath, R.S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
25. Simmer, F. *et al.* Genome-wide RNAi of *C. elegans* using the hypersensitive *rrf-3* strain reveals novel gene functions. *PLoS Biol.* **1**, E12 (2003).
26. Fernandez, A.G. *et al.* New genes with roles in the *C. elegans* embryo revealed using RNAi of ovary-enriched ORFeome clones. *Genome Res.* **15**, 250–259 (2005).
27. Lundquist, E.A. *et al.* The *mec-8* gene of *C. elegans* encodes a protein with two RNA recognition motifs and regulates alternative splicing of *unc-52* transcripts. *Development* **122**, 1601–1610 (1996).
28. Spike, C.A., Davies, A.G., Shaw, J.E. & Herman, R.K. MEC-8 regulates alternative splicing of *unc-52* transcripts in *C. elegans* hypodermal cells. *Development* **129**, 4999–5008 (2002).
29. Anyanful, A. *et al.* The RNA-binding protein SUP-12 controls muscle-specific splicing of the ADF/cofilin pre-mRNA in *C. elegans*. *J. Cell Biol.* **167**, 639–647 (2004).
30. Loria, P.M., Duke, A., Rand, J.B. & Hobert, O. Two neuronal, nuclear-localized RNA binding proteins involved in synaptic transmission. *Curr. Biol.* **13**, 1317–1323 (2003).
31. Fujita, M. *et al.* The role of the ELAV homologue EXC-7 in the development of the *Caenorhabditis elegans* excretory canals. *Dev. Biol.* **256**, 290–301 (2003).
32. Han, J.D. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004).
33. Chevenet, F., Brun, C., Banuls, A.L., Jacq, B. & Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**, 439 (2006).
34. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J. & Church, G.M. Systematic determination of genetic network architecture. *Nat. Genet.* **22**, 281–285 (1999).
35. Thacker, C., Sheps, J.A. & Rose, A.M. *Caenorhabditis elegans dpy-5* is a cuticle procollagen processed by a proprotein convertase. *Cell. Mol. Life Sci.* **63**, 1193–1204 (2006).
36. Kim, S. & Wadsworth, W.G. Positioning of longitudinal nerves in *C. elegans* by nidogen. *Science* **288**, 150–154 (2000).
37. Sagasti, A., Hobert, O., Troemel, E.R., Ruvkun, G. & Bargmann, C.I. Alternative olfactory neuron fates are specified by the LIM homeobox gene *lim-4*. *Genes Dev.* **13**, 1794–1806 (1999).
38. Kostic, I. & Roy, R. Organ-specific cell division abnormalities caused by mutation in a general cell cycle regulator in *C. elegans*. *Development* **129**, 2155–2165 (2002).
39. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
40. Harfe, B.D. *et al.* Analysis of a *Caenorhabditis elegans* Twist homolog identifies conserved and divergent aspects of mesodermal patterning. *Genes Dev.* **12**, 2623–2635 (1998).
41. Hall, D.H. *et al.* Ultrastructural features of the adult hermaphrodite gonad of *Caenorhabditis elegans*: relations between the germ line and soma. *Dev. Biol.* **212**, 101–123 (1999).
42. Bulow, H.E. & Hobert, O. Differential sulfations and epimerization define heparan sulfate specificity in nervous system development. *Neuron* **41**, 723–736 (2004).