Data analysis

One- or two-way analyses of variance (ANOVAs) with genotype as between-subject factor and session, block, day or quadrant as within-subject factor, followed by Fisher post-hoc tests (when necessary) were performed. Data are mean ± s.e.m.

Received 8 May; accepted 10 June 2002; doi:10.1038/nature00928.

1. Spear, N. E. *The Processing of Memories: Forgetting and Retention* (Erlbaum, Hillsdale, New Jersey, 1978).
2. Spear, N. E. & Riccio, D. C. *Memory: Phenomena and Principles* (Allyn and Bacon, Needham Heights, Massachusetts, 1994).
3. Sara, S. J. Retrieval and reconsolidation: toward a neurobiology of remembering. *Learn. Mem.* **7,** 73–84 (2000).
4. Abel, T., Martin, K. C., Bartsch, D. & Kandel, E. R. Memory suppressor genes: inhibitory constraints on the storage of long-term memory. *Science* **279,** 338–341 (1998).
5. Sweatt, J. D. Memory mechanisms: the yin and yang of protein phosphorylation. *Curr. Biol.* **11,** 391–394 (2001).
6. Mansuy, I. M., Mayford, M., Jacob, B., Kandel, E. R. & Bach, M. E. Restricted and regulated overexpression reveals calcineurin as a key component in the transition from short-term to long-term memory. *Cell* **92,** 39–49 (1998).
7. Mansuy, I. M. *et al.* Inducible and reversible gene expression with the rtTA system for the study of memory. *Neuron* **21,** 257–265 (1998).
8. Malleret, G. *et al.* Inducible and reversible enhancement of learning, memory, and long-term potentiation by genetic inhibition of calcineurin. *Cell* **104,** 675–686 (2001).
9. Lisman, J. E. A mechanism for the Hebb and anti-Hebb processes underlying learning and memory. *Proc. Natl Acad. Sci. USA* **86,** 9574–9578 (1989).
10. Morishita, W. *et al.* Regulation of synaptic strength by protein phosphatase 1. *Neuron* **32,** 1133–1148 (2001).
11. Save, E., Poucet, B., Foreman, N. & Buhot, M. C. Object exploration and reactions to spatial and nonspatial changes in hooded rats following damage to parietal cortex or hippocampal formation. *Behav. Neurosci.* **106,** 447–456 (1992).
12. Alberts, A. S., Montminy, M., Shenolikar, S. & Feramisco, J. R. Expression of a peptide inhibitor of protein phosphatase 1 increases phosphorylation and activity of CREB in NIH 3T3 fibroblasts. *Mol. Cell Biol.* **14,** 4398–4407 (1994).
13. Huang, F. L. & Glinsmann, W. H. Separation and characterization of two phosphorylase phosphatase inhibitors from rabbit skeletal muscle. *Eur. J. Biochem.* **70,** 419–426 (1976).
14. Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268,** 1766–1769 (1995).
15. Davis, H. P. & Squire, L. R. Protein synthesis and memory: a review. *Psychol. Bull.* **96,** 518–559 (1984).
16. Silva, A. J., Kogan, J. H., Frankland, P. W. & Kida, S. CREB and memory. *Annu. Rev. Neurosci.* **21,** 127–148 (1998).
17. Hagiwara, M. *et al.* Transcriptional attenuation following cAMP induction requires PP1-mediated dephosphorylation of CREB. *Cell* **70,** 105–113 (1992).
18. Impey, S. *et al.* Induction of CRE-mediated gene expression by stimuli that generate long-lasting LTP in area CA1 of the hippocampus. *Neuron* **16,** 973–982 (1996).
19. Gonzalez, G. A. & Montminy, M. R. Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. *Cell* **59,** 675–680 (1989).
20. Morris, R. G., Garrud, P., Rawlins, J. N. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297,** 681–683 (1982).
21. Strack, S., Barban, M. A., Wadzinski, B. E. & Colbran, R. J. Differential inactivation of postsynaptic density-associated and soluble Ca^{2+}/calmodulin-dependent protein kinase II by protein phosphatases 1 and 2A. *J. Neurochem.* **68,** 2119–2128 (1997).
22. Lisman, J., Malenka, R. C., Nicoll, R. A. & Malinow, R. Learning mechanisms: the case for CaMKII. *Science* **276,** 2001–2002 (1997).
23. Lisman, J. E. & Zhabotinsky, A. M. A model of synaptic memory: a CaMKII/PP1 switch that potentiates transmission by organizing an AMPA receptor anchoring assembly. *Neuron* **31,** 191–201 (2001).
24. Blitzer, R. D. *et al.* Gating of CaMKII by cAMP-regulated protein phosphatase activity during LTP. *Science* **280,** 1940–1942 (1998).
25. Gallagher, M. & Rapp, P. R. The use of animal models to study the effects of aging on cognition. *Annu. Rev. Psychol.* **48,** 339–370 (1997).
26. Foster, T. C., Sharrow, K. M., Masse, J. R., Norris, C. M. & Kumar, A. Calcineurin links Ca^{2+} dysregulation with brain aging. *J. Neurosci.* **21,** 4066–4073 (2001).
27. Norris, C. M., Halpain, S. & Foster, T. C. Alterations in the balance of protein kinase/phosphatase activities parallel reduced synaptic strength during aging. *J. Neurophysiol.* **80,** 1567–1570 (1998).
28. Villarreal, D. M., Do, V., Haddad, E. & Derrick, B. E. NMDA receptor antagonists sustain LTP and spatial memory: active processes mediate LTP decay. *Nature Neurosci.* **5,** 48–52 (2002).
29. Connor, J. H., Quan, H., Oliver, C. & Shenolikar, S. Inhibitor-1, a regulator of protein phosphatase 1 function. *Methods Mol. Biol.* **93,** 41–58 (1998).
30. Mayford, M. *et al.* Control of memory formation through regulated expression of a CaMKII transgene. *Science* **274,** 1678–1683 (1996).

Acknowledgements

We thank S. Shenolikar for the *I-1* reagents; I. Weiss, G. Hédou, F. Dey, A. Hirschy and M. Nemir for technical help; V. Taylor for assistance with animals; A. Jongen-Relo for help with stereology; D. Benke for help with membrane-enriched preparations; and T. Bliss for reading the manuscript. This work was supported by the Swiss Federal Institute of Technology, the Swiss National Science Foundation and the National Center of Competence in Research.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.M.M. (e-mail: isabelle.mansuy@cell.biol.ethz.ch).

Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*

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

** Departments of Developmental Biology and Genetics, Stanford University Medical Center, Stanford, California 94305, USA*
† Stanford Medical Informatics, 251 Campus Drive, MSOB X-215, Stanford, California 94305, USA

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[1]. Classic examples of open chromatin domains include 'puffs' on polytene chromosomes in *Drosophila* and extended loops from lampbrush chromosomes[2,3]. 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 whether co-expressed genes are clustered, we examined the chromosomal positions of the genes expressed in muscle of *Caenorhabditis elegans* at the first larval stage. Here we show 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 messenger RNA tagging to isolate muscle mRNA, because 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[4], mRNAs from specific tissues can be enriched by cross-linking them to the tagged PAB-1, and co-immunoprecipitating the complex of mRNA and tagged PAB-1 using an anti-epitope monoclonal antibody[20]. 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[5] (*myo-3*p; 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). Experiments using dot blot techniques and polymerase chain reaction with reverse transcription (RT–PCR) show that *unc-54*, which is specifically expressed in muscle[5], was co-immunoprecipitated with the muscle-expressed Flag::PAB-1 but that *gld-1*, which is specifically expressed in the germ line[6], was not (Fig. 2c and data not shown).

Next, we used DNA microarrays to analyse 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 labelled probes (see Methods) were then hybridized to DNA microarrays[7] containing 90% of the 19,733 genes currently estimated in the *C. elegans* genome[8]. We repeated the mRNA-tagging experiment six times to assess statistically which genes are 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 rank of about 50%, whereas genes expressed in muscle should have a rank significantly higher. A Student's *t*-test identified 1,364 genes that are significantly enriched in the muscle mRNA-tagging experiments ($P < 0.001$) (Fig. 3; see Supplementary Information Table 1). In control experiments using worms that do not express Flag::PAB-1 (see Methods), only 85 genes are enriched ($P < 0.001$; Supplementary Information Table 1).

We verified the mRNA-tagging approach by first showing that the list of 1,364 genes contains 16 of 18 muscle positive controls (89%) (Supplementary Information 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://nematode.lab.nig.ac.jp; Supplementary Information Table 2). Second, the expression pattern for 47 of the 1,364 genes was known from prior work, and 44 are expressed in body wall muscle (94%) (Supplementary Information Table 3). In contrast, only 13 of 51 genes selected at random are expressed in muscle (25.5%; http://nematode.lab.nig.ac.jp; Supplementary Information Table 4). Third, we compared the list of 1,364 genes with two groups of negative controls. The first group is 559 genes that are known not to be expressed in larval stage 1 (L1) muscle. Of these, 13 are in the list of 1,364 genes (2.3%) (Supplementary Information Table 5). The second group is 7,681 genes that show undetectable hybridization signals in DNA microarray experiments using RNA from L1 hermaphrodites. Of these, only 132 are in the list of 1,364 muscle-expressed genes (1.7%) (Supplementary Information Table 6). Finally, the mRNA-tagging approach is not strongly biased against rare mRNAs. The distribution of the signal intensities of the 1,364 L1 muscle genes was found to be nearly identical to that of all genes present on the microarray (Supplementary Information 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 for profiling gene expression in specific tissues in whole-genome expression studies. There are many previously characterized promoters that can direct the expression of epitope-tagged PAB-1 in many cell types in *C. elegans*, so it is now feasible to identify mRNAs expressed in any tissue at different developmental times, under different growth conditions or in different genetic backgrounds. Because poly(A)-binding protein is highly conserved from yeast to humans[9], mRNA tagging is applicable to other model organisms to isolate mRNAs expressed in cells or tissues that were previously inaccessible.

The list of 1,364 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. Five genes encode transcription factors that may specify muscle cell function, and fourteen genes 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 more than 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[8] within 10 kilobases (kb) of the start position of another muscle gene (see Methods). This number of clustered muscle genes was then compared with 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

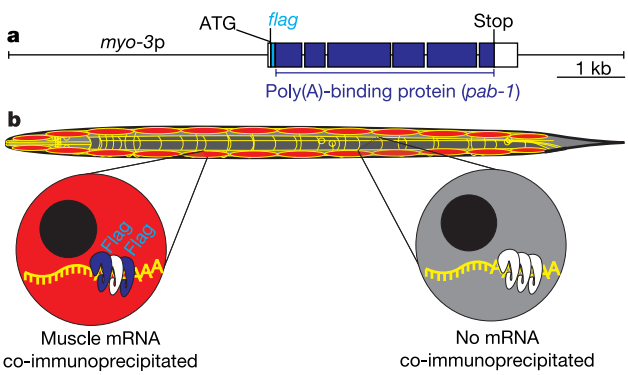

Figure 1 Outline of the mRNA-tagging technique. **a**, Schematic of *myo-3*p::*flag*::*pab-1*. The initiator methionine and stop codon are shown. Dark blue boxes show the predicted *pab-1* exons, white boxes show untranslated regions, and the light blue box represents the Flag tag. **b**, Schematic of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

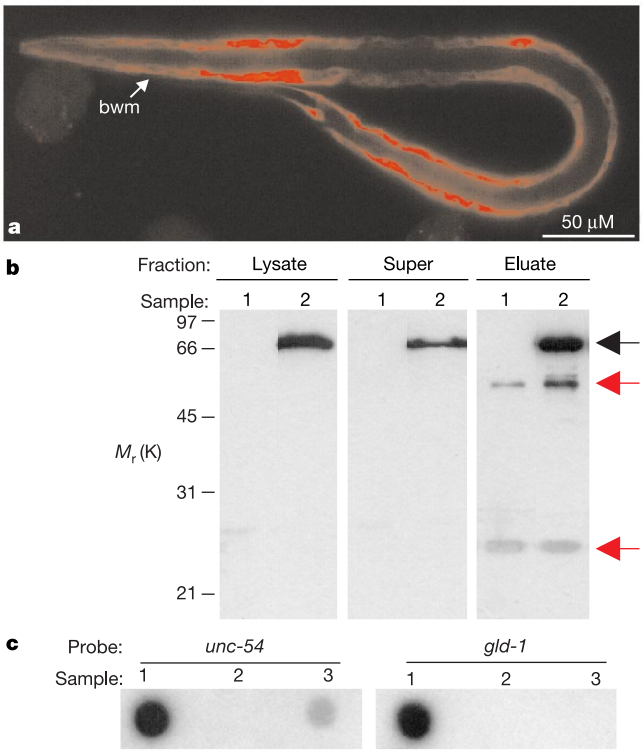

Figure 2 Muscle mRNA tagging. **a**, Anti-Flag antibody staining of a worm expressing *myo-3*p::Flag::PAB-1 showing specific expression in striated muscle. bwm, body wall muscle. **b**, Western blot analysis of immunoprecipitation of Flag::PAB-1 (black arrow) from cell-free extracts. Western blots were 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 1.6 × more sample than the lysate and supernatant lanes. M_r, relative molecular mass. **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-3*p::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.

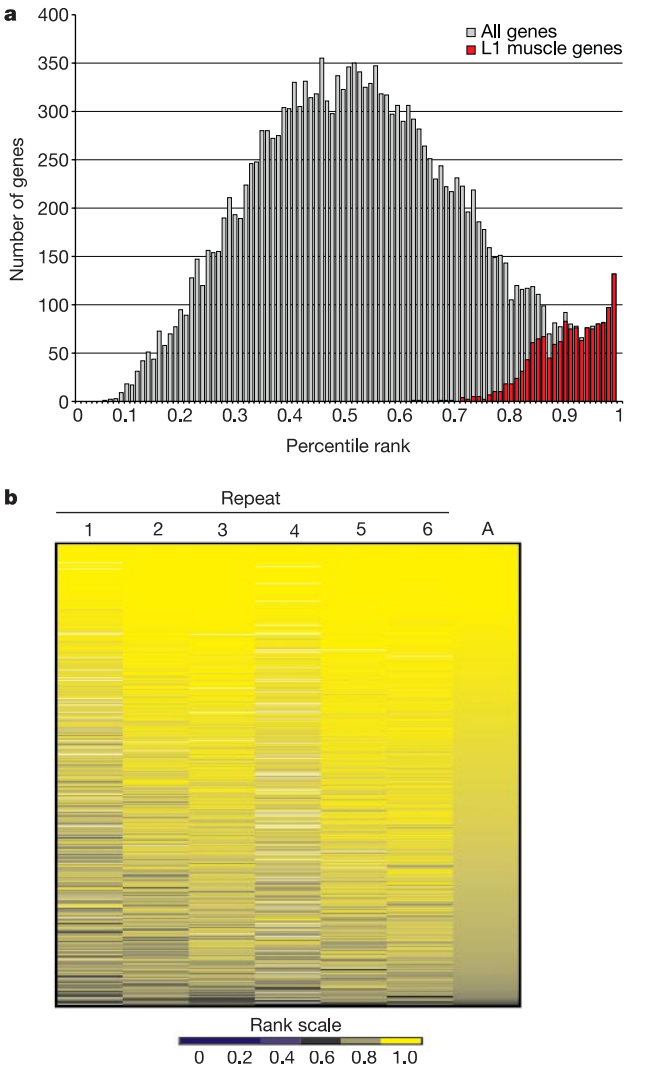

Figure 3 Genes enriched from muscle by mRNA tagging. **a**, Histogram of the average percentile rank of enrichment after mRNA tagging. The *x* axis shows the average percentile rank of enrichment, and the *y* axis shows the number of genes. Open bars represent the distribution for all genes; red bars represent the distribution of the 1,364 genes that are significantly enriched (Student's *t*-test, $P < 0.001$). **b**, Graphical representation of the percentile ranks of the 1,364 muscle-enriched genes. Each row represents a single gene. Colour represents the percentile rank of enrichment. Columns 1–6 are the six repeats and column A is the average percentile rank. Genes are listed in descending order of their average percentile rank. See Supplementary Information for how the graphical representation was generated. **c**, Pie chart showing the types of genes enriched in L1 muscle. The numbers in parentheses show the number of genes in the list that are expressed in muscle, out of the total number of genes in that biological group.

of clustered genes was calculated, however, special consideration was given to neighbouring 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[10,11]. Thus, clustering of different genes from the same operon on the chromosome would not be evidence for chromatin domains, and so we used only 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 Supplementary Information Figs 3 and 4), 1,304 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 for a random distribution (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 neighbours 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 2–5 genes each (Fig. 4b, c). Eighty-five 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 (Supplementary Information Table 7), are enriched to about the same extent by mRNA tagging, and are expressed at about the same level as the rest of the muscle-expressed genes (Supplementary Information 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 whether 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 germline-intrinsic genes[12]. All three germline groups showed sig-

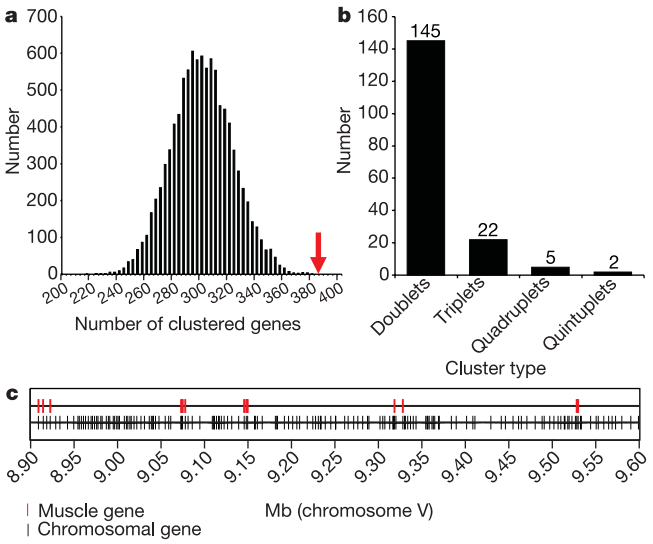

Figure 4 Muscle genes are clustered in small groups of 2–5 genes along the chromosomes. **a**, Histogram comparing 10,000 random samples with the observed muscle gene clusters (red arrow). **b**, Distribution of the number of muscle-expressed genes in each cluster. **c**, Schematic of 0.7 Mb of chromosome 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	Total genes	Genes considered*	Observed clustered†	Expected clustered‡	$P^{§}$
L1 muscle-enriched genes	1,364	1,304	386	310.1	7.2×10^{-5}
Sperm-enriched genes	650	616	198	112	$<10^{-15}$
Oocyte-enriched genes	258	242	26	14.4	0.017
Germline-intrinsic genes	508	475	136	67.9	$<10^{-15}$
Mountain 00	2,703	2,377	962	868.9	5.8×10^{-4}
Mountain 01	1,818	1,662	435	465.9	0.89
Mountain 02	1,465	1,100	360	270.6	$<10^{-15}$
Mountain 03	1,363	1,166	316	271.8	0.0017
Mountain 04	1,195	1,113	466	268.2	$<10^{-15}$
Mountain 05	1,012	944	248	160.7	$<10^{-15}$
Mountain 06	978	915	214	159.8	1.6×10^{-4}
Mountain 07	909	840	133	130.7	0.32
Mountain 08	810	761	225	119.8	$<10^{-15}$
Mountain 09	786	744	172	102.8	$<10^{-15}$
Mountain 10	635	555	97	64.8	0.0013
Mountain 11	587	546	133	71.5	$<10^{-15}$
Mountain 12	462	435	86	35.9	$<10^{-15}$
Mountain 13	396	368	49	27.3	1.3×10^{-5}
Mountain 14	353	331	46	22.1	4.4×10^{-5}
Oxidoreductases	323	317	21	16.7	0.23
Transcription factors	255	214	8	8.6	0.56
Lipid, fatty acid and sterol metabolism genes	281	280	12	12.4	0.53

The genes in each group, except for the mountains, are given in Supplementary Information Table 1. The genes in each mountain can be accessed elsewhere (http://cmgm.stanford.edu/~kimlab/topomap/c._elegans_topomap.htm).
*The number of genes considered, after removing operons, putative tandem duplicates, and genes with uncertain genomic positions.
†The number of genes in the gene group that are within 10 kb of another gene from the same group.
‡The number of clusters expected for a random distribution. 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.
§The probability that the observed number of genes that cluster could be matched or exceeded by chance.

nificant levels of gene clustering ($P < 0.05$) (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[13]. In all, 13 of the 15 largest gene clusters showed significant levels of gene clustering ($P < 0.05$) (Table 1).

As a negative control, we analysed 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 is 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. 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[14,15]. 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. Because opening chromatin for one gene can result in the opening of neighbouring chromatin[16,17], transcriptional machinery may access two co-expressed genes more efficiently if they are neighbours than if they are far apart. In this model, the fortuitous juxtaposition of co-expressed genes would be selected over an evolutionary timescale. An alternative interpretation of clustering is that genes expressed in the same tissue appear in clusters because neighbouring genes share a single DNA enhancer element. In this model, neighbouring 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 techniques of molecular biology and worm culture were used to respectively generate the *myo-3*p::*flag*::*pab-1* construct (*pPRSK9*) and the resulting SD1075 strain (see Supplementary Information for details).

Messenger RNA tagging

The mRNA-tagging protocol is a modification of a formerly published protocol[18]. For each sample, 1 ml of packed worms was resuspended in M9 buffer, and then fixed in 0.5% formaldehyde in M9 for 1 h at 4 °C. Animals were rinsed in homogenization buffer (HB), and then resuspended in 3 ml of HB. HB is prepared by diethyl pyrocarbonate (DEPC) treatment of 300 mM NaCl, 50 mM Hepes buffer at pH 7.6, 10 mM $MgCl_2$, 1 mM EGTA, 30 mM EDTA, 0.2 mg ml^{-1} heparin (sodium salt) and 10% glycerol, and then adding dithiothreitol (DTT) to 1 mM, vanadyl ribonucleoside complex to 8 mM (Sigma), 50 U ml^{-1} RNasin, and half a protease inhibitor cocktail tablet (Roche). Animals were lysed with two or three passes on a French press at 8,000 p.s.i. (~55 MPa), and then 25 passes on a Wheaten homogenizer. Large debris was sedimented by centrifugation at 4,300*g* for 6 min, and then smaller debris was separated from the supernatant by centrifugation at 48,000*g* for 20 min. At this point, the lysate could be flash frozen and stored for future use.

RNA bound to Flag::PAB-1 was enriched with anti-Flag-M2 affinity gel beads (Sigma). The affinity beads were prepared by rinsing twice 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. We mixed together 1 ml of lysate, 50 U of RNasin, 8 μl of 200 mM vanadyl ribonucleoside, and 75 μl of pelleted anti-Flag-M2 affinity gel beads for 2–3 h at 4 °C. The affinity beads were washed four times with cold HB, and then the RNA–protein crosslinks were reversed by incubating the beads in 125 μl of elution buffer (50 mM Tris/HCl at pH 8.0, 10 mM EDTA, 1.3% SDS; 20 U RNasin) at 65 °C for 30 min. We collected the supernatant containing the RNA, repeated the elution once more, and then combined the two supernatants.

RNA was isolated by mixing 250 μl of the immunoprecipitation supernatant (or cell extract) with 1 ml of trizol, and then mixing 250 μl of chloroform. After letting the mixture stand for 10 min, it was spun in a microfuge at 12,000 *g* for 15 min at 4 °C, and then the supernatant was extracted with chloroform. RNA was precipitated by adding 600 μl of isopropanol, incubating at room temperature for 10 min and then centrifuging in a microfuge for 20 min at 12,000 *g* at 4 °C. The RNA pellet was rinsed twice with 1 ml of 70% ethanol.

RNA was linearly amplified as previously described[19]. The DNA microarrays, probe preparation, and microarray hybridization are described in Supplementary Information.

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 genes in each hybridization. The mRNA-tagging experiment was repeated six times, and the mean percentile rank from all repeats was determined. A Student's *t*-test was used to determine which genes showed a mean enrichment significantly greater than the mean 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 given in Supplementary Information. Briefly, the calculations included only those genes that were present on the microarray and for which we could determine a chromosomal position[8]. The calculations used only the first gene in an operon, and only one gene of tandem repeats. The number of clusters expected for a random distribution was calculated separately for each chromosome, and then summed to give the total number. This was done so that the number expected for a random distribution reflected any bias in the experimental list. Finally, because the germline data were obtained from experiments using microarrays containing 11,917 genes[12], lists of genes were randomly selected from only these genes to avoid bias.

Received 23 January; accepted 16 July 2002; doi:10.1038/nature01012.

1. Weintraub, H. Tissue-specific gene expression and chromatin structure. *Harvey Lect.* **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 Uridine incorporation in lampbrush chromosomes. *Proc. Natl Acad. Sci. USA* **48,** 562–570 (1962).
4. Gorlach, M., Burd, C. G. & Dreyfuss, G. The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. *Exp. Cell Res.* **211,** 400–407 (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,** 165–183 (1996).
7. 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).
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,** 82–86 (2001).
9. Fukami-Kobayashi, K., Tomoda, S. & Go, M. Evolutionary clustering and functional similarity of RNA-binding proteins. *FEBS Lett.* **335,** 289–293 (1993).
10. Blumenthal, T. Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet.* **11,** 132–136 (1995).
11. Blumenthal, T. Gene clusters and polycistronic transcription in eukaryotes. *BioEssays* **20,** 480–487 (1998).
12. Reinke, V. *et al.* A global profile of germline gene expression in *C. elegans*. *Mol. Cell* **6,** 605–616 (2000).
13. Kim, S. K. *et al.* A gene expression map for *Caenorhabditis elegans*. *Science* **293,** 2087–2092 (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. *Nature Genet.* **26,** 183–186 (2000).
15. Caron, H. *et al.* The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291,** 1289–1292 (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 β-globin chromosomal domain. *EMBO J.* **13,** 1823–1830 (1994).
17. Stalder, J. *et al.* Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell* **20,** 451–460 (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,** 329–339 (1993).
19. Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T. & Marincola, F. M. High-fidelity mRNA amplification for gene profiling. *Nature Biotechnol.* **18,** 457–459 (2000).
20. Tenenbaum, S. A., Carson, C. C., Lager, P. J. & Keene, J. D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl Acad. Sci. USA* **97,** 14085–14090 (2000).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com/nature).

Acknowledgements

We thank J. Shaw, K. Mach, M. Laub, J. Lieb, M. Hiller, E. Harmon and J. Wang for 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 Supplementary Information. 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 National Institutes of Health (S.K.K.).

Competing interests statement

The authors declare that they have no competing financial interests.

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

Shh and *Gli3* are dispensable for limb skeleton formation but regulate digit number and identity

Ying Litingtung*†, Randall D. Dahn†‡, Yina Li*, John F. Fallon‡ & Chin Chiang*

* *Vanderbilt University Medical Center, Department of Cell and Developmental Biology, Nashville, Tennessee 37232, USA*
‡ *Department of Anatomy, University of Wisconsin, 1300 University Avenue, Madison, Wisconsin 53706, USA*
† *These authors contributed equally to this work*

Most current models propose *Sonic hedgehog* (*Shh*) as the primary determinant of anteroposterior development of amniote limbs[1]. Shh protein is said to be required to direct the formation of skeletal elements and to specify digit identity through dose-dependent activation of target gene expression. However, the identity of genes targeted by Shh, and the regulatory mechanisms controlling their expression, remain poorly understood. *Gli3* (the gene implicated in human Greig cephalopolysyndactyly syndrome) is proposed to negatively regulate *Shh* by restricting its expression and influence to the posterior mesoderm[2–4]. Here we report genetic analyses in mice showing that *Shh* and *Gli3* are dispensable for formation of limb skeletal elements: $Shh^{-/-}$ $Gli3^{-/-}$ limbs are distally complete and polydactylous, but completely lack wild-type digit identities. We show that the effects of Shh signalling on skeletal patterning and ridge maintenance are necessarily mediated through Gli3. We propose that the function of *Shh* and *Gli3* in limb skeletal patterning is limited to refining autopodial morphology, imposing pentadactyl constraint on the limb's polydactyl potential, and organizing digit identity specification, by regulating the relative balance of Gli3 transcriptional activator and repressor activities.

In *Drosophila*, signalling by Hedgehog (Hh) is necessarily mediated through the regulation of opposing activator and repressor activities of the bifunctional transcription factor cubitus interruptus (Ci)[5,6]. Ci undergoes default proteolysis to truncated forms (CiR) that repress expression of *hh* target genes[7]; stimulation of Hh prevents CiR formation and generates a full-length transcriptional activator (CiA) that upregulates expression of *hh* target genes[5]. Shh signalling in vertebrates is mediated by the GLI family (*Gli1–3*) of Ci orthologues, and evidence suggests that the activator and repressor functions of Ci have been uncoupled and partitioned among distinct GLI proteins[1].

Anteroposterior (A/P) patterning in the limbs of amniotes is controlled by posterior limb bud mesoderm, called the zone of polarizing activity (ZPA), through secretion of Shh protein; ectopic Shh or ZPA tissue grafted to the anterior border induces mirror-image digit duplications[8]. Gain- and loss-of-function studies suggest that Shh signalling progressively specifies increasing numbers of digits, with more-posterior identities, through dose-dependent activation of putative *Shh* target genes[9–13], but the identity and function of *Shh* effector genes remain unclear[1]. *Gli1* has been proposed to positively mediate Shh signalling (reviewed in ref. 1), but analysis of single- and double-mutant combinations of *Gli1–3* demonstrates that *Gli1* and *Gli2* perform no significant limb A/P patterning function[1]. Studies *in vivo* and *in vitro* suggest that *Gli3*, normally expressed in an anterior domain complementary to *Shh*[3], negatively regulates the expression of both *Shh* and its target genes through generating a repressor form (Gli3R) in a manner analogous to that in *Drosophila*[2,4,14]. *Xenopus*[15,16] and cell culture studies provide evidence that Gli3 may positively mediate Shh signalling in the limb; Shh stimulation prevents Gli3R formation[4,14], and Gli3