The concept of germ layers has been one of the foremost organizing principles in developmental biology, classification, systematics and evolution for 150 years (refs 1–3). Of the three germ layers, the mesoderm is found in bilaterian animals but is absent in species in the phyla Cnidaria and Ctenophora, which has been taken as evidence that the mesoderm was the final germ layer to evolve4,6–9. The origin of the ectoderm and endoderm germ layers, however, remains unclear, with models supporting the antecedence of each as well as a simultaneous origin6–9. Here we determine the temporal and spatial components of gene expression spanning embryonic development for all Caenorhabditis elegans genes and use it to determine the evolutionary ages of the germ layers. The gene expression program of the mesoderm is induced after those of the ectoderm and endoderm, thus making it the last germ layer both to evolve and to develop. Strikingly, the C. elegans endoderm and ectoderm expression programs do not co-induce; rather the endoderm activates earlier, and this is also observed in the expression of endoderm orthologues during the embryology of the frog Xenopus tropicalis, the sea anemone Nematostella vectensis and the sponge Amphimedon queenslandica. Querying the phylogenetic ages of specifically expressed genes reveals that the endoderm comprises older genes. Taken together, we propose that the endoderm program dates back to the origin of multicellularity, whereas the ectoderm originated as a secondary germ layer freed from ancestral feeding functions.

Embryonic development in C. elegans begins with a series of asymmetric cell divisions producing five somatic founder cells (AB, MS, E, C, D), each giving rise to a limited number of tissue types, and a single germline founder cell (P4) (Fig. 1a).10 To determine globally the spatial-temporal gene expression in the C. elegans embryo, we isolated five blastomeres (AB, MS, E, C and P3) that collectively amount to the entire embryo and cultured them in vitro,30 min (Fig. 1a). We found a general coherence between the time courses: 82% of the genes are within one log2 unit difference (Extended Data Fig. 3b). Genes with ‘missing’ expression tend to be expressed late in development (Extended Data Fig. 1d). For 380 genes expressed in the whole-embryo time course, we detected no expression at all in the blastomere time courses (Supplementary Table 2; see, for example, C55B7.3 in Fig. 1d). Genes with ‘missing’ expression are maintained (for example, wrn-1, a β-catenin-like protein, pol-1/caudal and pie-1, a zinc-finger protein; Fig. 1d), for some genes expression is higher than in the whole embryo (flp-15; Fig. 1d) and for others expression is lower (ceh-27, a homeodomain protein and Y41D4B.26; Fig. 1d). We found a general coherence between the time courses: 82% of the genes are within one log2 unit difference (Extended Data Fig. 3b). Of the genes that do differ, we found a strong bias for genes with lower expression in the blastomere time course as opposed to higher expression. For 380 genes expressed in the whole-embryo time course, we detected no expression at all in the blastomere time courses (Supplementary Table 2; see, for example, C55B7.3 in Fig. 1d). Genes with ‘missing’ expression tend to be expressed late in development (Extended Data Fig. 1d), indicating that, although in earlier development very few genes are unaccounted for in the data set, by the end of the time course noticeable deviations from standard development are apparent.

Performing principal component analysis on the blastomere transcriptomes distinguished the three germ layers (Fig. 2a). The three principal components collectively explained 41% of the variation in gene expression across the five lineages. The first principal component (PC1) correlated with development time, reflecting the expression of genes with non-specific expression (Extended Data Fig. 4). In general, PC2 distinguished the endoderm while PC3 distinguished ectoderm from mesoderm (Fig. 2a). The C lineage clusters with the other mesodermal lineages, although it produces both muscle and epidermis, probably because it contains twice as many muscle cells as epidermal cells. The overall distribution of the time courses into germ layers provides evidence for their distinction at the transcriptomic level.

To identify the specific genes uniquely expressed in each germ layer, we computed the correlation of the expression profile of each of the dynamically expressed genes to all others, and clustered them using hierarchical clustering (Fig. 2b). We detected 25 clusters, each comprising at least 10 genes. Gene members in a given cluster tended to have the same timing and location of expression (Fig. 2b, see right-hand bars). Fifty-four per cent of dynamically expressed genes are not specific to particular lineages (Fig. 2b), with nearly half deriving from the maternal transcriptome. The dynamically expressed genes with lineage specificity were divided according to their germ layer of expression (Extended Data Fig. 5), while further requiring each germ-layer annotated gene to have at least two-thirds of its expression in that germ layer (Supplementary Table 1). Our data reveal the spatial and temporal expression profile for each gene (Fig. 1d). For example, unc-120/ SRF has expression in MS, C and P3, as expected from its known role as a myogenic master regulator.

Since the five lineages each develop in isolation from one another, their context in the embryo is lost and, consequently, absence of signalling between cell lineages must affect some gene regulation. Most noticeably, the specification of the pharynx in the AB lineage is dependent upon two Notch signalling events and indeed we do not see expression of pharyngeal specification genes in the AB lineage (Extended Data Fig. 3a). Thus, although we found that for some genes expected levels are maintained (for example, wrn-1, a β-catenin-like protein, pol-1/caudal and pie-1, a zinc-finger protein; Fig. 1d), for some genes expression is higher than in the whole embryo (flp-15; Fig. 1d) and for others expression is lower (ceh-27, a homeodomain protein and Y41D4B.26; Fig. 1d).
The dynamics of the germ-layer expression programs may be unique to *C. elegans* or a general property of animal development. To test this, we analysed the previously characterized transcriptomes of the distantly related species *X. tropicalis*, *N. vectensis* and *A. queenslandica*. For each species, we mapped the orthologues of the *C. elegans* germ-layer genes in the respective genome and computed their average developmental expression profiles. We found a general recapitulation of
the order found in *C. elegans* (Fig. 3). The onset of the endodermal program in *Xenopus* occurs during gastrulation, well before that of the ectodermal and mesodermal programs (*P* < 0.01, Kolmogorov–Smirnov test). In *Nematostella*, we also detected a major rise in the expression of endodermal orthologues during gastrulation (*P* < 10^{-5}). The observation that mesoderm orthologues in *Nematostella* are expressed in the planula is consistent with the notion that the bilaterian mesoderm was co-opted from late-expressed genes. In *Amphimedon*, endodermal orthologues are enriched for expression during the ‘brown’ stage, in which two layers first become visible. Expression of the orthologues of the ectoderm and mesoderm germ-layer genes, in contrast, is seen only in the early stages (*P* < 10^{-14}), reflecting that they are solely deposited as maternal transcripts.

The distinct and conserved temporal inductions of germ-layer-specific expression (Fig. 3), with the mesoderm both appearing last in evolutionary timescales and developing last in the embryo, support accretion of processes as a mechanism in the evolution of development in *Amphimedon* 

Examining the evolutionary age of the individual germ layers, we found that genes specifically expressed in the endoderm have a significantly higher fraction of older genes (*P* < 10^{-5}, *χ^2* test). In contrast, the ectoderm and mesoderm genes are significantly younger (*P* < 10^{-3}, *χ^2* test). Since the phylogenetic analysis revealed that endodermal gene clusters are of older origin, we enquired into their functional properties. We found that endodermal-specific genes are enriched for energy production, metabolism and transport functions (Fig. 4b and Extended Data Fig. 7). The observation that the endoderm is enriched in general feeding functions suggests that it is closer, relative to the ectoderm, in its characteristics to the choanoflagellate-like ancestor. To test this, we

**Figure 3** | The endodermal expression program precedes the ectodermal program in diverse species. Expression of germ-layer genes in *C. elegans*, and their orthologues in *X. tropicalis*, *N. vectensis* and *A. queenslandica*. The average is computed on the maximum-normalized gene profiles.
examined the level of orthology with the choanoflagellate *Monosiga brevicollis* for each of the functional classes. Indeed we found a higher fraction of *M. brevicollis* orthologues in endoderm–enriched functional classes, such as transport and metabolism (Fig. 4c), suggesting that the endoderm is most closely aligned with the feeding capabilities of the free-living choanoflagellates. Moreover, while transport and metabolism appear to be related to ‘housekeeping’ functions, we observe, in contrast, that they are induced early on in embryogenesis in the endoderm–germ-layer program.

Our results shed light on the evolutionary history of the endoderm germ layer (Fig. 4d). At the dawn of the metazoans, choanoflagellate–like colonial organisms comprised individual cells that probably all retained feeding functions. However, with the evolution of epithelial cells, the possibility of distinct cell–types emerged, as cells could communicate by strong membrane connections. Our analysis of the composition and dynamics of the germ-layer transcriptomes leads us to propose that the endoderm program has retained the feeding functions of its choanoflagellate–like ancestor. Expression in the Amphimedon sponge is informative since physical layers of epithelia exist in this organism. The expression of sponge orthologues of the endoderm gene set suggests that *Amphimedon* only has a functional ‘propto-endoderm’ germ layer. This is also supported by recent evidence that the GATA gene in *Amphimedon* is expressed in the internal layer in the sponge.

In the lineage leading to the eumetazoans, the transport and metabolic functions performed by internal cells may have allowed the external cells to specialize into an ectodermal germ layer (Fig. 4d). In this model, the ancestry of the endoderm follows from its role in feeding, whereas only later in evolution was it coupled with its current function as the gastrulating internal layer. This scenario is in line with Haeckel’s gastrea hypothesis which posits a layered spherical organism as the urmetazooan. However, our model of feeding processes driving selection of the endodermal identity is also consistent with an ancestral flattened placula, as proposed by Bütschli, that subsequently evolved into a two-layered stage where the lower epithelia specialized in digestion.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 2 July; accepted 23 October 2014.**

**Published online 10 December 2014.**

**1. Hall, B. K. Evolutionary Developmental Biology 2nd edn (Chapman & Hall, 1998).**

**2. Wolpert, L. Principles of Development 4th edn (Oxford Univ. Press, 2011).**

**3. Technau, U. & Scholz, C. B. Origin and evolution of endoderm and mesoderm. Int. J. Dev. Biol. 47, 531–539 (2003).**

**4. Ryan, J. F. et al. The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution. Science 342, 1242592 (2013).**

**5. Martindale, M. Q., Pang, K. & Finnerty, J. R. Investigating the origins of triploblasty: ‘mesodermal’ gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa). Development 131, 2463–2474 (2004).**

**6. Buss, L. W. The Evolution of Individuality (Princeton Univ. Press, 1987).**

**7. Gould, S. J. Ontogeny and Phylogeny (Belknap Press of Harvard Univ. Press, 1977).**

**8. Nielsen, C. Animal Evolution: Interrelationships of the Living Phyla 3rd edn (Oxford Univ. Press, 2012).**

**9. Valentine, J. W. On the Origin of Phyla (Univ. Chicago Press, 2004).**

**10. 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).**

**11. Edgar, L. G. & Goldstein, B. Culture and manipulation of embryonic cells. Methods Cell Biol. 107, 151–175 (2012).**

**12. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-seq: single-cell RNA-seq by multiplexed linear amplification. Cell Rep. 2, 666–673 (2012).**

**13. Shahak, A. et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 498, 236–240 (2013).**

**14. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nature Rev. Genet. 14, 618–630 (2013).**

**15. Murray, J. J. et al. Multidimensional regulation of gene expression in the C. elegans embryo. Genome Res. 22, 1282–1294 (2012).**

**16. Fukusugi, T., Brodigan, T. M., Schriefer, L. A., Waterston, R. H. & Krause, M. Defining the transcriptional redundancy of early bodywall muscle development in *C. elegans*: evidence for a unified theory of animal muscle development. Genes Dev. 20, 3395–3406 (2006).**

**17. Neves, A. & Priess, J. R. THE REF-1 family of bHLH transcription factors pattern C. elegans embryos through Notch-dependent and Notch-independent pathways. Dev. Cell 8, 867–879 (2005).**

**18. Yanai, I., Peshkin, L., Jorgensen, P. & Kirschner, M. W. Mapping gene expression in two Xenopus species: evolutionary constraints and developmental flexibility. Dev. Cell 20, 483–496 (2011).**

**19. Helm, R. R., Siebert, S., Tulin, S., Smith, J. & Dunn, C. W. Characterization of differential transcript abundance through time during Nematostella vectensis development. BMC Genomics 14, 266 (2013).**

**20. Anany, L. et al. BLIND ordering of large-scale transcriptomic developmental timecourses. Development 141, 1161–1166 (2014).**

**21. Domazet-Loso, T. & Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature 468, 815–818 (2010).**

**22. Levin, M., Hashimshony, T., Wagner, F. & Yanai, I. Developmental milestones punctuate gene expression in the Caenorhabditis embryo. Dev. Cell 22, 1101–1108 (2012).**

**23. Kalinka, A. T. et al. Gene expression divergence recapitulates the developmental hourglass model. Nature 468, 811–814 (2010).**

**24. King, N. et al. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature 451, 783–788 (2008).**

**25. Leys, S. P. & Riesgo, A. Epithelia, an evolutionary novelty of metazoans. J. Exp. Zool. B 318, 438–447 (2012).**

**26. Nakashibi, N., Sogabe, S. & Degnan, B. M. Evolutionary origin of gastrulation: insights from sponge development. BMC Biol. 12, 26 (2014).**

**27. Haeckel, E. Die Gastraea-Theorie, die phylogenetische Classification des Thierreichs und die Homologie der Keimblätter. Jenaerische Nachrichten, 8, 1–55 (1874).**

**28. Leininger, S. et al. Developmental gene expression provides clues to relationships between sponge and eumetazoan body plans. Nature Commun. 5, 3905 (2014).**

**29. Schierwater, B. et al. Concatenated analysis sheds light on early metazoan evolution and fuels a modern “urmetazoon” hypothesis. PLoS Biol. 7, e20 (2009).**

**30. Bütschli, O. Bemerkungen zur Gastraea-Theorie. Morph. Jahrb. 18, 415–427 (1884).**

**Supplementary information** is available in the online version of the paper.
METHODS

Blastomere isolation and culturing. Egg shells were removed from C. elegans embryos and the resulting blastomeres cultured as previously described. The egg shell and vitelline membrane were removed at the two-cell stage, and the embryo separated to the AB and P1 blastomeres by pipetting. P1 was allowed to undergo one cell division and separated to EMS and P2, or two cell divisions before being separated to the MS, E, C and P3 blastomeres, to allow the Wnt signalling from P2 to EMS (Extended Data Fig. 1b). The five lineages were cultured in a humid chamber in EGM, and division of the E blastomere was used as a clock (Extended Data Table 2). All lineages from a single embryo were frozen at the same time. Individual samples were transferred with a micro-pipette into a 0.5 µl drop of egg salts placed on the cap of a 0.5 ml LoBind Eppendorf tube, excess liquid was aspirated off, and the samples frozen in liquid nitrogen. Samples were stored at −80 °C. Samples were collected in triplicates; correlations between replicates are shown in Extended Data Fig. 2a. Throughout this work, ‘correlation’ denotes Pearson’s correlation coefficient.

Whole-embryo time course. Precisely staged single embryos were collected at the one-, two- and four-cell stages, and 10 min intervals thereafter up to muscle movement, then roughly every 30 min; 50 embryos were used in total. RNA from each embryo was prepared using TRIzol as previously described, with one modification: 1 µl of the ERCC spike-in kit (5:1,000,000 dilution) was added to the TRIzol to each sample.

Single cell and whole-embryo transcriptomics. CEL-seq was used to amplify and sequence both RNA from the whole embryos and the cultured blastomeres. For the whole embryos, RNA was re-suspended in 5 µl water and 1 µl primer added; 1.2 µl were taken for the amplification. For the blastomeres, 1 µl of a 1:50,000 dilution of the ERCC spike-in kit and 0.2 µl of the primer were mixed (a total of 1.2 µl) and added directly to the lid of the Eppendorf tube where the cell was frozen. Linear amplification and library preparation were as previously described. Libraries were sequenced on an Illumina HiSeq2000 according to standard protocols. Paired-end sequencing was performed, reading at least 11 bases for read 1, 35 bases for read 2, and the Illumina barcode when needed. The complete data set has been deposited in the Gene Expression Omnibus database under accession number GSE50548.

Expression analysis pipeline. Transcript abundances were obtained from the sequencing data as previously described. Briefly, libraries were sequenced on an Illumina HiSeq2000 according to standard protocols, paired-end sequencing, using the CEL-seq protocol. Mapping of the reads used BWA, version 0.6.1, against the C. elegans WBCel215 genome (bwa aln -n 0.04 -o 1 -e -1 -d 16 -i 5 -k 2 -M 3 -O 11 -E 4). Read counting used hts整齐 version 0.3.5p1 default, against WS230 annotation exons. The counts were normalized by dividing by the total number of mapped reads for each gene and multiplying by 10^6, yielding the estimated gene expression levels in transcripts per million (t.p.m.).

Warped whole-embryo time course. The whole-embryo time course (Extended Data Fig. 2c) was compared with the blastomere time courses (Fig. 1b) using a restricted set of 4,527 genes with a log fold-change of at least 5 across the 50-embryo time course, greater than 100 t.p.m. maximum expression, and less than 101 t.p.m. minimum expression. These cutoffs were used to limit analysis to only the most dynamically expressed genes given the distinct dynamics of the whole-embryo time course. The minimum expression threshold further selected for temporally restricted expression. For each blastomere time point, the five lineages were summed up to represent the whole embryo, taking into account the fraction of the whole embryo represented by the specific lineage (half for AB, one eighth each for E, MS, C and P3). An eleven-stage warped whole-embryo time course was generated by taking for each stage a weighted average across the 50 embryos based upon the correlations of this set resulted in very similar warped profiles.

Spatial and temporal gene expression profiles. In the profiles shown in Fig. 1d, the log expression is split among the lineages according to the fraction in the natural scale expression. The black line indicates the expression of the whole-embryo time course.

Definition of gene sets for dynamically expressed and differentiation genes. The 3,910 dynamically expressed genes were defined based upon the warped whole-embryo time course with >3 log, fold-change, >10 t.p.m. maximum expression and <100 t.p.m. minimum expression (Extended Data Fig. 2b). These parameters were adapted to the warped time course, which is less dynamic owing to averaging effects. ‘Constitutively expressed’ genes (Extended Data Fig. 3b) were defined as highly expressed genes (>500 t.p.m. maximum expression) but not members of the dynamically expressed genes. ‘Expressed genes’ (Extended Data Fig. 3b) were defined as those genes with >10 t.p.m. maximum expression. The differentiation gene sets (Fig. 1c and Extended Data Fig. 2d) were generated for each group—the neurons (AB), muscle (MS, C and P3), endoderm (E), epidermis (AB and C), pharynx (MS) and germline (P3)—by examining terminal expression in the time courses. Genes were assigned to one of the seven sets if they exhibited expression ≥50 t.p.m. in that group and a correlation coefficient greater than 0.7 of expression across the lineages with the expected expression pattern, as highlighted in red on the lineage trees. The parameters were set according to their definition of similarly sized sets.

Clusters of temporal gene expression patterns. A correlation coefficient was computed for each gene’s temporal warped whole-embryo time course against each of 17 idealized expression profiles (Extended Data Fig. 3c). The idealized profiles were constructed based upon average expression of clusters using the k-means algorithm and represent the general patterns of the transcriptome. The idealized profiles are vectors of the same length (11) as the warped time-course profile but with digital expression of three possible values: 0, 1 and 2. Each dynamically expressed gene was then assigned to the idealized profile to which it best correlated. Seven of the 17 idealized profiles correspond to ‘maternal’ profiles (Extended Data Fig. 3c) in which expression is initially high and then drops. We collapsed these seven profiles to one profile and denoted it as the ‘0’ cluster in Fig. 2b.

Hierarchical clustering and definition of germ-layer genes. Hierarchical clustering used the ‘linkage’ function in MATLAB using the unweighted centre of mass distance (UPGMC) algorithm. The top 20 clusters with at least ten genes were examined (Fig. 2b). Clusters with at least 65% of the genes of the same germ layer contributed their genes with the dominant germ layer. Germ layers were assigned by correlating the average expression with germ-layer-specific patterns with a cutoff of 0.6 correlation with the following ID: ectoderm = 0101011; mesoderm = 01011; where the order is AB, MS, E, C and P3. Germ-layer genes were defined according to the sum of the genes identified by the clusters and are indicated in Fig. 2b. We further filtered the germ-layer gene sets by keeping only those genes whose expression was partitioned across the germ layers such that at least two-thirds of the expression was in that germ layer.

Gene age. Orthologues were retrieved from the MetaPhoS project using the 2010 release. Taxonomies were retrieved from the NCBI Taxonomy. For each C. elegans gene, if the gene was also present in at least 25% of the examined non-metazoan opisthokont eukaryotes, it was annotated as ‘old’. Similar results were also observed for the definition of ‘old’ genes at the level of eukaryotic and cellular life (Extended Data Fig. 6). MetaPhoRs were also used to delineate the orthologues shown in Fig. 4c for M. brevicollis.

Orthologous gene expression profiles. The developmental time courses of A. queenslandica, X. tropicalis and N. vectensis have been previously described. For these species, the latest protein annotations were used to detect orthologs as follows: A. queenslandica, AqH2; X. tropicalis, JGI_4.2; N. vectensis, GCA_000209225. A. queenslandica orthologues were delineated using OrthoMCL, and those of X. tropicalis and N. vectensis were retrieved from Biomart which contained the annotations on the noted versions. We included in the analysis genes whose maximum expression was greater than the data-set-specific threshold; this was computed as the average expression across all genes. Expression pm = [passing this threshold were each normalized to their own maximum expression. A Kolmogorov–Smirnov test was used to test for significantly different temporal dynamics between endoderm and ectoderm expression. For this analysis the timing of expression for each gene was computed as the stage at which half of the sum expression had occurred.

Functional categories analysis. COG functional category annotations were retrieved from WormMart. For simplicity, annotations ‘of general function prediction only’ and ‘function unknown’ were ignored, as well as those categories capturing fewer than 3% of the genes. Enrichments were computed using the hypergeometric distribution.

31. Goldstein, B. Induction of gut in Caenorhabditis elegans embryos. Nature 357, 255–257 (1992).
32. Fischer, S. et al. The External RNA Controls Consortium: a progress report. Nature Methods 7, 731–734 (2005).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
34. Przybucz, L. P., Huerta-Cepas, J. & Gabaldon, T. MetaPhoRs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. Nucleic Acids Res. 39, e32 (2011).
35. Fischer, S. et al. in Current Protocols in Bioinformatics Ch. 6, Unit 6.12, 11–19 (2011).
36. Dohm, J. C. et al. BioMart Central Portal: an open database network for the biological community. Database 2011, bar041 (2011).
37. Tatsuo, R. L., Kooin, E. V. & Lipman, D. J. A genomic perspective on protein families. Science 278, 631–637 (1997).
38. Szewczak, J. E. et al. WormBase: better software, richer content. Nucleic Acids Res. 34, D475–D478 (2006).
39. Murray, J. I. Cell 475–478 (2006).
40. Cowan, A. E. & McIntosh, J. R. Mapping the distribution of differentiation potential for intestine, muscle, and hypodermis during early development in Caenorhabditis elegans. Cell 41, 923–932 (1985).
41. Good, K. et al. The T-box transcription factors TBX-37 and TBX-38 link GLP-1/Notch signaling to mesoderm induction in C. elegans embryos. *Development* **131**, 1967–1978 (2004).
42. Goldstein, B. An analysis of the response to gut induction in the *C. elegans* embryo. *Development* **121**, 1227–1236 (1995).
43. Laufer, J. S., Bazzicalupo, P. & Wood, W. B. Segregation of developmental potential in early embryos of *Caenorhabditis elegans*. *Cell* **19**, 569–577 (1980).
44. Goldstein, B. Establishment of gut fate in the E lineage of *C. elegans*: the roles of lineage-dependent mechanisms and cell interactions. *Development* **118**, 1267–1277 (1993).
45. Fox, R. M. et al. The embryonic muscle transcriptome of *Caenorhabditis elegans*. *Genome Biol.* **8**, R188 (2007).
46. McGhee, J. D. et al. ELT-2 is the predominant transcription factor controlling differentiation and function of the *C. elegans* intestine, from embryo to adult. *Dev. Biol.* **327**, 551–565 (2009).
Extended Data Figure 1 | *In vitro* culturing of the *C. elegans* embryonic founder blastomeres. The cells are separated as shown in the left schematic and then cultured in embryonic growth medium as shown in the micrographs on the right. The numbers indicate the stages at which the cells were collected for transcriptome analysis. Six of the 11 stages are shown in the micrographs.
Extended Data Figure 2 | A transcriptomic survey of *C. elegans* embryonic founder cell lineages. a, Replicates of the embryonic blastomere time courses. The heat maps show the correlations among the replicates for each blastomere lineage at each of the eleven examined stages. For three blastomere stages there were no replicates. The median correlation coefficient is 0.9. Samples were collected in triplicates. Only samples with at least 750,000 reads were used, which has been previously shown to be of sufficient sequencing depth for CEL-seq12. Supplementary Table 3 provides the sequencing statistics for each sample. b, Expression profiles of the 3,910 dynamic genes across the blastomere lineage time courses. See Methods for definition of dynamic genes. c, Correlation coefficients between samples of the whole-embryo time course. Each of the 50 samples comprises a single embryo, collected at the indicated minutes past the four-cell stage. Again, only samples with at least 750,000 reads were used and Supplementary Table 3 provides the sequencing statistics for each sample. d, The expression profiles of the 1,664 genes with differentiated expression analysed in Fig. 1c. Each profile was ‘standardized’ by subtracting its mean and dividing by its standard deviation. e, Comparison of the blastomere time courses to the EPIC data set15. For 115 genes, we could compare gene expression to previously published embryonic expression profiles generated by microscopic lineaging until the ~300-cell stage15,39. Of these, 75% of our profiles had consistent localized expression (Supplementary Table 1). Of those, 54% matched completely, and 21% of the genes expressed in all of the lineages in our data set had some missing expression in the EPIC data set because the lineaging was not performed until the end of the developmental process. The remaining genes have some overlap in expression. Such differences in expression could be caused by the transgene in the EPIC data set not recapitulating the profile of the endogenous gene, or missing signals between cells in the blastomere data set, as is seen from the whole-embryo/blastomere expression level ratio (see Supplementary Table 1, ratios defined as equal, slightly higher/lower or much higher/lower). Expression profile compared with the EPIC data set deviates more when expression in the blastomeres is low compared with the whole embryo, but the blastomere data set has the advantage that all genes are assayed simultaneously, no transgenes are used, maternal transcripts are seen and downregulation of genes is observable.
Extended Data Figure 3 | Lineage-restricted gene expression identifies genes dependent upon coherence of the lineages and tissue specificity.

a, Expression profiles of genes involved in pharynx specification. The left and right panels correspond to the two Notch signalling events. The top and bottom images correspond to the expected regulatory patterns in the whole embryo and isolated blastomeres, respectively. The *tbx-37* gene is not shown since it is identical to *tbx-38* in expression profile. 

b, Comparison of the overall sum of expression between the two time courses, plotted on a log2 scale (black). Genes ‘missing’ in the separated lineage time course were manually added to the graph at −3. The additional plots indicate the same measure for dynamically expressed genes (blue) and constitutive genes (red). 

c, Idealized expression profiles used to identify gene expression clusters. 

d, The gene expression profiles for the temporally restricted gene expression profiles. Each profile was ‘standardized’ by subtracting its mean and dividing by its standard deviation. 

e, Average expression profiles of ten clusters of dynamically expressed genes determined on the basis of the whole-embryo expression data (see Methods). 

f, The number of dynamic genes in each temporal period. In each group, the genes not expressed in the lineage time course (b) are marked in red.
Extended Data Figure 4 | The first principal component correlates with developmental time. Principal component analysis as described in Fig. 2a. Colour codes are the same as in Fig. 1. PC1, PC2 and PC3 capture 18%, 12% and 11%, respectively, of the variation in the expression, in the 1,320 dynamically expressed genes with no expression in the first stage (to exclude genes with maternal expression).
Extended Data Figure 5 | Germ-layer-specific expression. Expression profiles of the germ-layer-specific genes in each of the lineages. The x and y axes are the 11 examined temporal stages and individual genes, respectively. Germ-layer-specific genes were identified by hierarchical clustering based upon correlation among dynamically expressed genes (see Methods).
Extended Data Figure 6 | Robustness of gene age analysis. a, Same format as Fig. 4a but with the definition of old genes as those present in at least 25% of the examined eukaryotes (see Methods) that are not ophistokonts. b, Same as Fig. 4a with a definition of ‘old’ as those present in 25% of the examined organisms that are not eukaryotes (Eubacteria and Archaea).
Extended Data Figure 7  |  Truncated endoderm gene set control. To exclude the possibility that general genes were included as ‘endoderm-specific’ because the endoderm program is induced earlier, we excluded temporal clusters 8, 9 and 10 from the endoderm genes and repeated the relevant analyses. We found that there was no marked change in the results. The results are shown in the same format as Figs 3 and 4b, c.
Extended Data Table 1 | The fates of the progeny of each blastomere \textit{in vivo} and in isolated cultured blastomeres

|       | Fates in whole embryo | Expected \textit{in vitro} | References |
|-------|-----------------------|-----------------------------|------------|
| AB    | Neurons               | Unknown                     | 10         |
|       | Epidermis             | Yes                         | 40         |
|       | Pharynx               | No                          | 41         |
|       | 1 muscle cell         | Unknown                     |            |
| MS    | Muscle                | Yes                         | 42         |
|       | Pharynx               | Yes                         | 42         |
| E     | Endoderm              | Yes                         | 43,44      |
| C     | Muscle                | Yes                         | 40         |
|       | Epidermis             | Yes                         | 40         |
| P3    | D                     | Muscle                      | Yes        | 40         |
|       | P4                    | Germ line                   | Unknown    |            |

Data are from refs 40–44.
Extended Data Table 2 | Description of the developmental stages queried in this study

| Stage number | Stage name | Description | Time* |
|--------------|------------|-------------|-------|
| 1            | 2-cell     | 2-cell embryo | 0     |
| 2            | 4-cell     | 4-cell embryo | 20    |
| 3            | E          | After division of EMS to E and MS | 40    |
| 4            | 2E         | After division of E to Ea and Ep | 60    |
| 5            | 2E+        | After division of MSa and MSP to MSaa, MSap, MSPa and MSPp | 90    |
| 6            | 4E         | After division of Ea and Ep to Eal, Ear, Epl and Epr | 110   |
| 7            | 4E+        | 60 minutes after division of Ea and Ep to Eal, Ear, Epl and Epr | 140   |
| 8            | 8E         | After division of Eal, Ear, Epl and Epr to Eala, Ealp, Eara, Earp, Epla, Eplp, Epra and Eprp | 180   |
| 9            | 8E+        | 90 minutes after division of Eal, Ear, Epl and Epr to Eala, Ealp, Eara, Earp, Epla, Eplp, Epra and Eprp | na    |
| 10           | 8E++       | 180 minutes after division of Eal, Ear, Epl and Epr to Eala, Ealp, Eara, Earp, Epla, Eplp, Epra and Eprp | na    |
| 11           | o.n.       | After an over-night incubation – more than 8 E cells are visible. | na    |

* Timing of the stage in the Sulston lineage. Timing is indicated as minutes from the 2-cell stage.
## Extended Data Table 3 | Tissue-specific gene sets

| Tissue       | Gene sets                                                                                                                                 |
|--------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| Neuronal     | Genes with the following GO terms:                                                                                                        |
|              | GO:0001764 neuron migration                                                                                                                |
|              | GO:0004983 neuropeptide Y receptor activity                                                                                               |
|              | GO:0005328 neurotransmitter:sodium symporter activity                                                                                     |
|              | GO:0006836 neurotransmitter transport                                                                                                     |
|              | GO:0007218 neuropeptide signaling pathway                                                                                                 |
|              | GO:0007268 synaptic transmission                                                                                                           |
|              | GO:0007411 axon guidance                                                                                                                  |
|              | GO:0008021 synaptic vesicle                                                                                                                |
|              | GO:0030424 axon                                                                                                                              |
|              | GO:0030425 dendrite                                                                                                                        |
|              | GO:0030594 neurotransmitter receptor activity                                                                                                |
|              | GO:0043005 neuron projection                                                                                                               |
|              | GO:0045202 synapse                                                                                                                          |
|              | GO:0045211 postsynaptic membrane                                                                                                           |
|              | GO:0048489 synaptic vesicle transport                                                                                                      |
|              | GO:0048666 neuron development                                                                                                             |
| Muscle       | Genes identified by Fox et al.                                                                                                             |
| Endoderm     | Genes identified by McGhee et al.                                                                                                          |
| Epidermis    | Genes with the following GO term:                                                                                                          |
|              | GO:0018996 molting cycle, collagen and cuticulin-based cuticle                                                                            |
| Pharynx      | Genes with the following GO term:                                                                                                          |
|              | GO:0007631 feeding behavior                                                                                                               |
| Germline     | Genes with the following GO terms:                                                                                                         |
|              | GO:0051729 germline cell cycle switching, mitotic to meiotic cell cycle                                                                     |
|              | GO:0048477 oogenesis                                                                                                                       |
|              | GO:0045132 meiotic chromosome segregation                                                                                                  |
|              | GO:0043186 P granule                                                                                                                        |
|              | GO:0007276 gamete generation                                                                                                               |
|              | GO:0007281 germ cell development                                                                                                           |
|              | GO:0007126 meiosis                                                                                                                          |
|              | GO:0001556 oocyte maturation                                                                                                               |
|              | GO:0000003 reproduction                                                                                                                    |

Data for muscle and endoderm are from refs 45 and 46, respectively.