

28. Miles, R., Toth, K., Gulyas, A. I., Hajos, N. & Freund, T. F. Differences between somatic and dendritic inhibition in the hippocampus. *Neuron* **16**, 815–823 (1996).
29. Mel, B. W. Synaptic integration in an excitable dendritic tree. *J. Neurophysiol.* **70**, 1086–1101 (1993).
30. Poirazi, P. & Mel, B. W. Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron* **29**, 779–796 (2001).

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

**Acknowledgements**

This work was supported by the National Institutes of Health.

**Competing interests statement**

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.S. (e-mail: spruston@northwestern.edu).

**Identification of genes expressed in *C. elegans* touch receptor neurons**

Yun Zhang\*, Charles Ma\*, Thomas Delohery†‡, Brian Nasipak\*, Barrett C. Foat\*, Alexander Bounoutas\*, Harmen J. Bussemaker\*, Stuart K. Kim§ & Martin Chalfie\*

\* Department of Biological Sciences, Columbia University, New York, New York 10027, USA

† Memorial Sloan Kettering Institute, New York, New York 10021, USA

§ Department of Developmental Biology and Genetics, Stanford University Medical Center, Stanford, California 94305, USA

The extent of gene regulation in cell differentiation is poorly understood. We previously used saturation mutagenesis to identify 18 genes that are needed for the development and function of a single type of sensory neuron—the touch receptor neuron for gentle touch in *Caenorhabditis elegans*<sup>1,2</sup>. One of these genes, *mec-3*, encodes a transcription factor that controls touch receptor differentiation<sup>3,4</sup>. By culturing and isolating wild-type and *mec-3* mutant cells from embryos and applying their amplified RNA to DNA microarrays, here we have identified genes that are known to be expressed in touch receptors, a previously uncloned gene (*mec-17*) that is needed for maintaining touch receptor differentiation<sup>2,5</sup>, and more than 50 previously unknown *mec-3*-dependent genes. These genes are randomly distributed in the genome and under-represented both for genes that are co-expressed in operons and for multiple members of gene families. Using regions 5' of the start codon of the first 20 genes, we have also identified an over-represented heptanucleotide, AATGCAT, that is needed for the expression of touch receptor genes<sup>6</sup>.

Through mutagenesis screens for touch-insensitive mutants<sup>1,2</sup>, we previously identified *mec-3*, *mec-17* and eight *mec-3*-dependent genes that are needed for the function of the six touch receptor neurons (refs 7, 8; and G. Gu, L. Emtage and M.C., unpublished data). Those screens identified several alleles for each of these genes (except *mec-17*) and were therefore at or near saturation; however, they would not have identified genes whose activity is redundant, subtle or pleiotropic, or genes whose loss produces touch-super-sensitive animals.

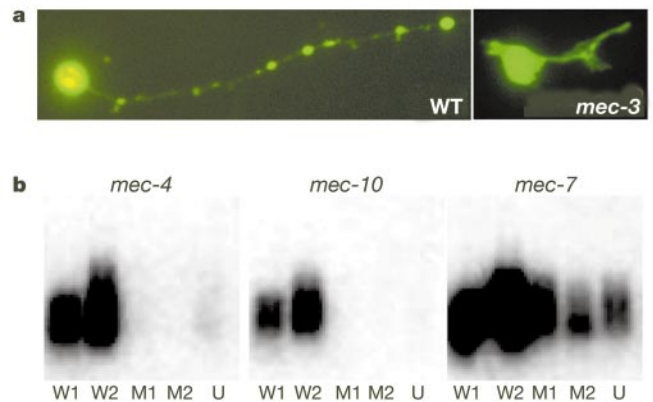
Because adult animals contain roughly 3,000 nuclei but only six touch receptor neurons, the identification of differences in RNA from animals with and without these cells on DNA microarrays has an inherent problem of sensitivity. Indeed, we could not identify any

touch receptor genes using RNA from wild-type and *mec-3* mutant animals on DNA microarrays (data not shown). We therefore obtained and cultured wild-type and *mec-3* mutant cells from embryos, isolated touch receptors by cell sorting, and amplified RNA from the isolated cells. We identified the cells by green fluorescent protein (GFP) fluorescence, which was expressed either from the promoter for *mec-18*, a touch-receptor-specific gene, in wild-type embryonic touch receptors (ALML/R and PLML/R), or from the *mec-3* promoter, which is specific to touch receptors and two other neurons (FLPL/R) in the embryo<sup>4</sup>, for *mec-3* mutant cells. Both promoters are expressed after the touch receptors are generated.

When initially dissociated, embryos produced a few faintly fluorescent cells. When cultured overnight, however, more cells (0.3–0.6%) elaborated processes and had intense GFP fluorescence. These GFP-positive cells, which were presumably generated and/or differentiated in culture, were similar in morphology (Fig. 1a) and gene expression (Fig. 1b and Table 1) to *in vivo* touch receptor neurons. The wild-type cultured cells were usually monopolar with a single long process or bipolar with a smaller second process at 180° from the first, in other words, they were similar to *in vivo* ALM or PLM cells, respectively. In contrast, the *mec-3* cells were bipolar or multipolar with smaller processes that branched more randomly.

To characterize gene expression in wild-type and mutant touch receptors, we isolated the cells by size, viability and GFP fluorescence intensity using fluorescence-activated cell sorting. A typical sorting produced 4 × 10<sup>6</sup> cells, 40–60% of which were GFP positive (~100-fold enrichment). From these cells, we obtained about 100 pg of poly(A)<sup>+</sup> RNA, which was amplified linearly by roughly one million times. All nine known *mec-3*-dependent *mec* genes were expressed at much higher levels in wild-type cells than in *mec-3* cells (Fig. 1b and Table 1). (*mec-3* is least differentially expressed probably because of its autoregulation; its expression declines in *mec-3* mutants as animals mature<sup>4</sup>; therefore, the reduction of *mec-3* messenger RNA in embryonic cells was probably not complete.) These results show that the differentiated touch receptors in culture express *mec-3*-dependent genes and that linear amplification can successfully amplify the corresponding RNAs.

To identify *mec-3*-dependent genes, we hybridized RNA from sorted wild-type cells and *mec-3* cells to genomic DNA microarrays containing DNA for 17,817 of the 18,967 known or predicted *C. elegans* genes<sup>9</sup>. Three sets of independently prepared RNA samples from mutant and wild-type touch receptors were each hybridized onto two separate arrays. We identified 71 *mec-3*-dependent candi-



**Figure 1** Morphology and gene expression of cultured *C. elegans* cells. **a**, A wild-type (WT) touch receptor neuron and a *mec-3* mutant cell after 12 h in culture. Both cells fluoresce because of expressed GFP. The wild-type cell has a single neurite, whereas the *mec-3* cell has several shorter neurites. **b**, Northern analyses of amplified RNA from sorted wild-type touch receptors (W1 and W2), sorted *mec-3* mutant cells (M1 and M2), and unsorted wild-type cells (U). Samples were probed with cDNAs for the indicated genes.

† Present address: Aventis Pharmaceuticals, 1041 Rt. 202-206, Bridgewater, New Jersey 08807-0800, USA.

date genes (Table 1) by choosing genes whose ratio of fluorescence intensity was  $\geq 2.0$  and was significant using Student's *t*-test ( $P < 0.05$ ). Using the same criteria, we identified 25 candidate genes that were overexpressed in the *mec-3* mutant cells (data not shown). These latter genes could be either *mec-3*-repressed genes or

genes that are expressed in a *mec-3*-independent manner in the FLP cells. Seven of the nine known *mec-3*-dependent genes and two other touch receptor-expressed genes, *mec-12* (ref. 10) and *unc-24* (T. Barnes and S. Hekimi, personal communication), were among the 71 *mec-3*-dependent candidate genes. Of these 11 genes, 7 were

**Table 1** *mec-3*-dependent genes

Gene name*	Description	Array ratio†	PCR ratio†	<i>P</i>	S1‡	S2‡
F57H12.7	MEC-17	62.10	2,928	0.002	•	•
<b><i>mec-9</i></b>	<b>EGF/Kunitz repeat protein</b>	<b>12.30</b>	<b>34</b>	<b>0.005</b>		
<b><i>mec-7</i></b>	<b><math>\beta</math>-Tubulin</b>	<b>11.91</b>	<b>9</b>	<b>0.023</b>	•	•
<b><i>mec-14</i></b>	<b>Aldo-keto reductase-like</b>	<b>9.34</b>	<b>20</b>	<b>0.011</b>	•	•
T07D1.3	Aldehyde dehydrogenase domain	8.75	69	0.004	•	•
<b><i>mec-18</i></b>	<b>coA synthetase-like</b>	<b>8.32</b>	<b>1,574</b>	<b>0.007</b>	•	•
C03A3.3	Serine protease/metallo- $\beta$ -lactamase-like	8.13	62	0.018		•
<b><i>mec-4</i></b>	<b>Degenerin</b>	<b>7.40</b>	<b>67</b>	<b>0.009</b>	•	•
<i>twk-28</i>	TASK potassium channel subunit	6.42	8	0.003		
<b><i>mec-12</i></b>	<b><math>\alpha</math>-Tubulin</b>	<b>5.91</b>	<b>16</b>	<b>0.001</b>	•	•
F42A9.9	Unknown	4.52	32	0.014	•	
ZC482.5	GABA receptor	4.24	17	0.014		•
F14E5.4	Acid phosphatase	3.91	280	0.005		•
<i>unc-24</i>	Stomatrin/lipid transfer protein-like	3.58		0.038		
F20D6.5	Tyrosine kinase catalytic domain	3.33	8	0.002		
C06A8.3	17k antigen ( <i>O. volvulus</i> )	3.29	7	0.024		
<i>far-3</i>	<i>O. volvulus</i> antigen peptide-like	3.17	8	0.014		
Y48G8AL.8	Contains ribosomal protein L22 signature	3.13		0.032		•
<b><i>mec-3</i></b>	<b>LIM homeodomain transcription factor</b>	<b>3.09</b>	<b>7</b>	<b>0.014</b>	•	•
<i>rpl-18</i>	Ribosomal protein L18	3.05		0.047		
Y54G11A.1	Unknown	2.92		0.029	•	
F01D5.8	Esterase/lipase/thioesterase active site	2.90	24	0.008	•	•
<i>cct-2</i>	T-complex chaperonin protein	2.89	3	0.017	•	
Y39A3CL.1	Unknown	2.81		0.027		•
F29B9.4	Jmjc and weak granin domains	2.79	2	0.004		
D2007.2	Major sperm protein (MSP) domain	2.75		0.026	•	
Y17D7B.4	Unknown	2.74		0.010		
<i>hsp-16.41</i>	Heat shock protein	2.74		0.040		•
F49E8.4	Cytidine deaminase	2.68	4	0.031		
W05H9.1	Unknown	2.50		0.032	•	
<i>rpl-20</i>	Ribosomal protein L20	2.49	4	0.027		
<i>rps-18</i>	Ribosomal protein S13	2.49	3	0.026		
<i>cct-4</i>	T-complex chaperonin protein	2.47	4	0.021		
W08E12.3	Member of uncharacterized protein family	2.46		0.050	•	•
Y38F2AR.4	Contains similarity to <i>C. elegans</i> MUT-2	2.44		0.046	•	
VW06B3R.1	Insulinase (proteinase M16)	2.42		0.027		
F54E2.2	Unknown	2.33		0.040	•	
W08E12.5	Member of uncharacterized protein family	2.30		0.045	•	
<i>rps-14</i>	Ribosomal protein S14	2.29		0.040	•	
F47B10.7	Acyl-coA binding protein	2.26		0.027	•	
<i>cyp-5</i>	Cyclophilin	2.25		0.047		
C05D11.2	Vacuolar protein sorting protein 16-like	2.21	2	0.001		
<i>cap-1</i>	F-actin capping protein $\alpha$ -subunit	2.21		0.024		•
EGAP7.1	Cuticular collagen	2.18		0.039		
C47D2.1	Unknown	2.17		0.027		
T04B2.3	Unknown; prenyl group binding site	2.16		0.006		
F52E4.1	Propionyl-CoA carboxylase $\beta$	2.16		0.029		•
<i>bag-1</i>	Contains ubiquitin domain	2.15		0.023		
F46E10.2	Contains similarity to elastin	2.15		0.003		
F54E12.4	Histone H2B	2.15		0.047		
<i>sra-6</i>	Sra family chemoreceptor	2.14		0.043		
K07C11.4	Similar to type-B carboxylesterase/lipase	2.13	4	0.024		
F19B2.5	Unknown; SNF2_N domain	2.12		0.006		•
<i>cct-1</i>	T-complex chaperonin protein	2.10	4	0.002		
C02F5.3	GTP-binding protein	2.10		0.047		
R05D7.5	Unknown	2.09		0.038		
M03F4.7	Calcium-binding protein	2.09		0.003		
C27B7.6	Serine/threonine protein phosphatase	2.09	4§	0.021		
B0336.11	MSP domain; like VAMP-associated proteins	2.08		0		
F39B2.11	Contains similarity to bacterial GntR	2.08		0.046		
F58E10.4	Unknown; AN1-like zinc finger	2.08		0.038		
F10A3.11	CW domain protein	2.08		0.009	•	•
F46G10.6	HLH DNA-binding domain	2.06		0.009	•	•
<i>srp-1</i>	Serpin, serine protease inhibitor	2.06		0.005	•	•
F42H10.3	Unknown; SH3 domain	2.05		0.004		
C55A1.6	Member of uncharacterized protein family	2.03		0.008		
<b><i>mec-1</i></b>	<b>EGF/Kunitz repeat protein</b>	<b>2.03</b>	<b>21</b>	<b>0.009</b>	•	•
K03H1.4	Transthyretin-like protein	2.03		0.003	•	•
F49E8.3	Puromycin-sensitive aminopeptidase-like	2.02	3§	0.031		
<i>arf-3</i>	GTP-binding protein	2.01		0.039		•
<i>acr-13</i>	Acetylcholine receptor	2.01		0.013		•

*mec* genes that have been cloned previously are in bold. *P* values are from the *t*-test.

\*Gene names (from the primers that generated the DNA microarray spots) and gene functions are from Wormbase (<http://www.wormbase.org>).

†The expression ratios from microarrays and from quantitative RT-PCR are given.

‡A dot denotes the presence of the MEC-3::UNC-86 consensus binding site (S1) or the heptanucleotide over-represented in the first 20 genes (S2). *mec-10* (PCR ratio 43) and *mec-2* (PCR ratio 70) have both sequences.

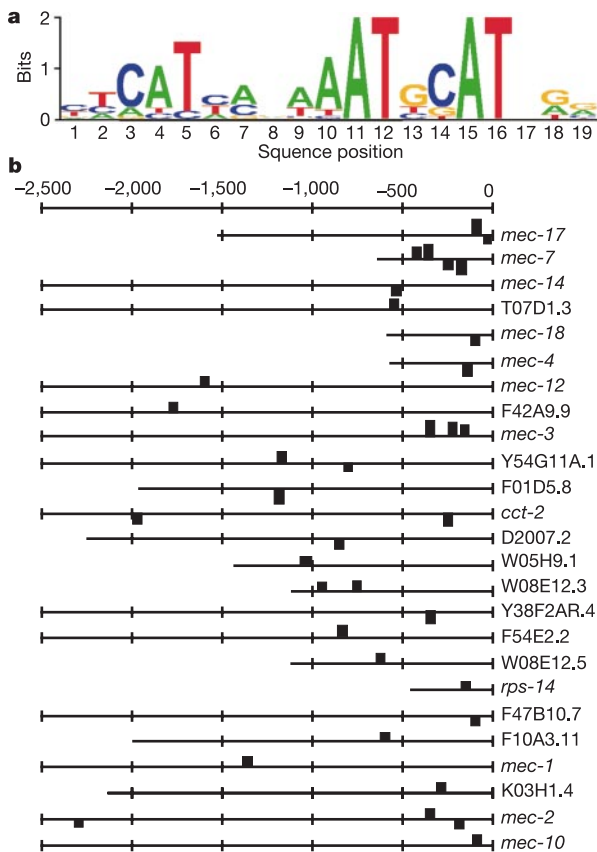
§One set of wild-type and mutant RNAs was used for the PCR ratio instead of two.

found among the first 14 genes in Table 1 (an additional *mec* gene, *mec-17*, is also in this group; see below). As described above, *mec-3*, one of the four remaining genes, is probably lower on the list because of the early age of the isolated cells. The remaining genes (*mec-1*, *mec-2* and *mec-10*) were probably lower or are missing from Table 1 because their RNAs hybridized poorly to the arrays. The arrays were spotted with coding-rich sequences from genomic DNA, and the DNAs for these three genes would not hybridize to at least 600 nucleotides at the 3' ends of the RNAs. Our amplification methods, however, preferentially produce sequences from the 3' ends of the RNAs. DNA microarrays that could detect the 3' ends of mRNAs would have more accurately identified the *mec-3*-dependent genes.

Quantitative polymerase chain reaction with reverse transcription (RT-PCR) indicated that the nine known *mec-3*-dependent genes were enriched from 7-fold to 1,574-fold in wild-type cells (Table 1). Eleven newly identified *mec-3*-dependent candidates gave similar ratios and were therefore likely to be expressed primarily in touch receptors and regulated directly by MEC-3. We confirmed this expression for three of these genes (Fig. 3): F57H12.7 and C03A3.3 were expressed in the six touch receptors, whereas T07D1.3 was expressed in touch receptors and the FLP cells. The expression of these genes in the touch receptors required *mec-3*.

Quantitative RT-PCR of 11 other genes in Table 1 gave modest differential ratios ( $\leq 4$ ). All of these genes ranked below *mec-3*. As described below for the *cct* genes, genes that give small ratios can be expressed differentially in the touch receptors. The modest

difference in expression might result from more general gene expression or an indirect effect on gene expression from the differentiation of the wild-type cells. If we assume that the genes above *mec-3* are true positives and use the fraction of previously uncharacterized genes below *mec-3* that have differential expression ratios of  $\leq 2$  (2/12) to estimate false positives, then 87%



**Figure 2** The MEC-3::UNC-86 binding site. **a**, Consensus derived from known MEC-3::UNC-86 binding sites from *mec-3*, *mec-4* and *mec-7* in *C. elegans*, *C. briggsae* and *C. vulgaris*. A perfect match would have a score of two bits. **b**, Positions of significant matches ( $P < 10^{-4}$ ) to the consensus in the regions 5' of the start codons of newly identified and known *mec-3*-dependent genes. The position of the start ATG is taken as 0; the heights of the bars are proportional to the extent that the identified sites match the consensus sequence. Bars shown underneath the lines represent the sequences in the other strand.

**a**

MEC-17	1	-MQVDADLRPLGPFQ---LIVRLD-----PMSVKQLQDPVIVYEAIIDNLA
W06B11.1	1	-MEIAFDLSTIFDNIQRFTTTL-----LLKYGPKRYAVVAQSIDCLG
CG3967	1	MVEFFFDIKPLFAQIIVKVTSMNLP-----NFFRGRDQCCLDATSMEFEIDQIG
CG17003	1	MVEFFFDIKHLFFQSIIRVQAHSLRPKVTQCRRYAQTREKSGKSTWTSCLRESLINMG
BAB29231		-----
BAB22194		-----
FLJ13158	1	MEFFFDVDALFPERITVLDQHLRFP-----ARRPQTMTTARV-DLQQQI-MTIIDELG
BAB14472	1	MEFFFDVDALFPERITVLDQHLRFP-----ARRPQTMTTARV-DLQQQI-MTIIDELG

	F	G
MEC-17	40	KLSAHLQLRTPLTCTCKLI--NSDSTLYLSKWKYDEEKVSRIMGFAKVGKRLFLY
W06B11.1	43	EMSKFHGWKRVITMYDKIVDHDDEQTYLIMWEKVNKSG-SILKGLLRVGYKTLTYT
CG3967	51	QLSATSQGLSKPVTTAQRLRMSDQNTTYLLAD--NEAGHNGAVLGLLKVGTKNLYLF
CG17003	58	KLSDAQGLCHAVTSADKLA--SDQVYVYVMA--DKAAGHWEITGLLKVGTKDLVVF
BAB29231	1	-----MQS-NRHVYIYIKDTSARPAGKGAIGFLKVGYKKLFLV
BAB22194	1	-----MQS-NRHVYIYIKDTSARPAGKGAIGFLKVGYKKLFLV
FLJ13158	52	KASAKAQNLSAPITTSASRMQS-NRHVYIYIKDSSARPAGKGAIGFLKVGYKKLFLV
BAB14472	52	KASAKAQNLSAPITTSASRMQS-NRHVYIYIKDSSARPAGKGAIGFLKVGYKKLFLV

MEC-17	95	DSQMOTVEGEI-LCLLDFVYHESVQRGCGVQGLLDMYFSEHTEPFYQALDNPSTVL
W06B11.1	99	DNE-QNYMEKMCILDFVVPVTEQRSGNCFKMFDFBMLKAEVTVDQCAFDPKSAAL
CG3967	106	DEAGKTRMVEQTPSLDFVYHESVQRGCGVQGLLDMYFSEHTEPFYQALDNPSTVL
CG17003	110	DQGGCYRRLNQTALDFVYHESVQRGCGVQGLLDMYFSEHTEPFYQALDNPSTVL
BAB29231	39	DDREAHNEVEP-LCLLDFVYHESVQRGCGVQGLLDMYFSEHTEPFYQALDNPSTVL
BAB22194	39	DDREAHNEVEP-LCLLDFVYHESVQRGCGVQGLLDMYFSEHTEPFYQALDNPSTVL
FLJ13158	108	DDREAHNEVEP-LCLLDFVYHESVQRGCGVQGLLDMYFSEHTEPFYQALDNPSTVL
BAB14472	108	DDREAHNEVEP-LCLLDFVYHESVQRGCGVQGLLDMYFSEHTEPFYQALDNPSTVL

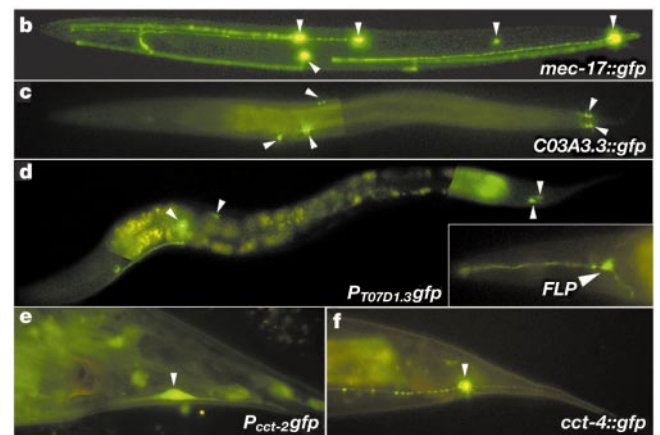
MEC-17	151	LGFMSSQYGLIKPWNQNTNFVFEELFLALSAEN-----
W06B11.1	155	QQFLEKYDRKDLVWQSNKYALCSNFFIGRHP-----
CG3967	163	LSFLSKHYGLKRIIPQANNFVLEGGFNDDGESGNGG-----G
CG17003	167	LAFMAKHVGLVRTIPQANNFVLEGGFNDDPDTT-----
BAB29231	95	LKFLNKHYNLETTVPQVNNFVIFEGFFAHQHRPPTSSLRATRHSRAAVADPTPAAPA
BAB22194	95	LKFLNKHYNLETTVPQVNNFVIFEGFFAHQHRPPTSSLRATRHSRAAVADPTPAAPA
FLJ13158	164	LKFLNKHYNLETTVPQVNNFVIFEGFFAHQHRPPTSSLRATRHSRAAVADPTPAAPA
BAB14472	164	LKFLNKHYNLETTVPQVNNFVIFEGFFAHQHRPPTSSLRATRHSRAAVADPTPAAPA

MEC-17	185	-----GIEKPPDQGWRFMTFPRRLCTGMDTFRWLQHAVSGHQSGKNAMAAPVDA
W06B11.1	187	-----TVPTFRQTKRASASSASVSHASRNSTSPFIGNRPRHDSVADLMRQD
CG3967	200	NGHANGTPNGLHITNSPNTLHFGATYLGEDSNRQSGSQQTTPNARLQQTIGTISFSG
CG17003	199	---TCKSASGLQATQSCCRSSRQCHYVROEQDAQIKHGQANRNTVQNDASGFFRQ
BAB29231	152	RKLPKRAEGDIKPYSSSDREFLKVAVEPPWPLNAPRRATPPAHPFRSSSLGNSP
BAB22194	149	GKACKDWELEKAVRPSGSGGSRVAVLSAGNSSEHPKATHFSSAVCPGAGSLHMGH
FLJ13158	198	RKLPKRAEGDIKPYSSSDREFLKVAVEPPWPLNAPRRATPPAHPFRSSSLGNSP
BAB14472	221	RKLPKRAEGDIKPYSSSDREFLKVAVEPPWPLNAPRRATPPAHPFRSSSLGNSP

MEC-17	234	DMTFQGLSNRAHQAKARKAHLSKPLW-----
W06B11.1	235	MLAGVRAEVDNSPTGLKNARDFGHRRIW-----
CG3967	257	RYGAKRPTCSMAEIIHAGNSKGGNGSGSAENRIK-----
CG17003	253	DQKIVVGTSIYRRRWKSPRTLARAGCRVSGGRRF-----
BAB29231	209	DRGLRPFVPEQELLSRLCLCPHPPTARLLATDPGGSQAQRRTSSSLPSDESRY
BAB22194	206	NQIPANAQAHTLTTLPYSKAHPALFSAVTLTKQALQELGWPFVSAEPLR-----
FLJ13158	255	ERGLRPFVPEQELLSRLCLCPHPPTARLLAADPGGSQAQRRTSSSLPSDESRY
BAB14472	278	ERGLRPFVPEQELLSRLCLCPHPPTARLLAADPGGSQAQRRTSSSLPSDESRY



**Figure 3** Newly identified *mec-3*-dependent genes. **a**, ClustalW (<http://clustalw.genome.ad.jp>) alignment of MEC-17 and similar proteins from *C. elegans* (W06B11.1), *Drosophila* (CG3967 and CG17003), mouse (BAB29231 and BAB22194) and human (FLJ13158 and BAB14472; BAB names are GenBank accession numbers, others are gene names). Amino acids that are identical (red) or similar (blue) to those in MEC-17 are indicated. The amino acid changes in MEC-17(U265) are indicated and correspond to nucleotide changes CTT (Leu) to TTT (Phe) and GAT (Asp) to GGT (Gly). **b-f**, GFP expression patterns (triangles indicate touch receptor cell bodies) of *mec-17::gfp* in a wild-type L3 larva (**b**), *C03A3.3::gfp* in a wild-type L4 larva (**c**), a *P<sub>T07D1.3</sub>gfp* promoter fusion in a wild-type adult (**d**), a *P<sub>cct-2</sub>gfp* promoter fusion in a wild-type adult PLM touch receptor (**e**) and *cct-4::gfp* in a wild-type adult PLM cell (**f**).

of the genes in Table 1 are likely to be differentially expressed in wild-type touch receptors.

Several factors may contribute to the differences between the array ratios and those obtained by quantitative RT-PCRs. As explained above for *mec-1*, *mec-2* and *mec-10*, some RNAs hybridize poorly to the arrays. In addition, the need to generate a ratio from the array data means that highly expressed RNAs in wild-type cells (that may saturate hybridization) and RNAs such as *mec-4* that are not expressed in *mec-3* mutant cells (which by default have a value that is equal to the standard deviation of the local background) will give lower ratios on the arrays.

MEC-3 maintains its own transcription and activates *mec-4* and *mec-7* by binding to DNA as part of a MEC-3::UNC-86 heterodimer<sup>6,11,12</sup>. The consensus derived from these binding sites (Fig. 2a) contains the binding site for UNC-86 (AAATT/GCAT) and an additional sequence (TCATCA) positioned 5' to this site. Seventeen newly identified *mec-3*-dependent genes in Table 1 and eight of the nine known *mec-3*-dependent genes contain the consensus sequence in the regions 5' of their start codons (up to 2.5 kb; Fig. 2b). They are presumably candidates for direct targets of *mec-3* regulation.

Analysis of the oligonucleotide frequencies<sup>13</sup> in the region lying 5' to the start codon of the top 20 genes in Table 1 showed that a heptanucleotide (AATGCAT) is significantly over-represented ( $P < 0.01$ ). This motif matches the UNC-86 binding core in the consensus sequence (Fig. 2a), and, thus provides a proof of principle that binding sites needed for transcription can be deduced from the array data. These results also provide a cautionary note, because this site is needed for UNC-86 binding, not MEC-3 binding. Of the 71 genes in Table 1, 27 contain this heptanucleotide. Consistent with the hypothesis that genes that show strong enrichment are direct targets of *mec-3* regulation, 12 of the 16 genes with an expression ratio of  $\geq 9$  from quantitative RT-PCR (including *mec-10*) contain this heptanucleotide, whereas none of the 15 genes with an expression ratio of  $< 9$  (excluding *mec-3*, which is a special case because of its autoregulation) has this heptanucleotide. Genes lacking these sites—for example, three of the four genes that encode ribosomal proteins—might be regulated indirectly by differentiation of the touch receptors.

We have noted several general properties of the *mec-3*-dependent genes in Table 1. First, the genes reflect the general distribution of genes on all six chromosomes. Second, this collection of genes is under-represented for genes that are expressed together in operons. Roughly 15% of *C. elegans* genes are initially expressed in multigene transcripts from operons<sup>14</sup>. Nine genes known to be members of operons in *C. elegans* are found in Table 1, but the other members of the same operons are not. Either these nine genes are expressed separately from the other members of their operons or they represent false positives. Two genes (F49E8.3 and F49E8.4) are sufficiently close enough to form an operon, but their expression as an operon has not been confirmed (they also have relatively low differential ratios by RT-PCR). Third, even though redundancy owing to gene families may have contributed to genes on this list being missed in our genetic screens, multiple members of only 4 out of 25 gene families are represented by the genes in Table 1. *mec-4* and *mec-10* both encode similar degenerin channel subunits, but these genes do not act redundantly<sup>15</sup>. Three genes (*cct-1*, *cct-2* and *cct-4*) encode similar, but probably not redundant components of the T-complex chaperonin (see below). In addition, Table 1 contains *unc-24*, whose product shares a stomatin-like region with MEC-2 (ref. 16), and *acr-13*, an acetylcholine receptor subunit (two other such subunits are expressed in the touch receptors<sup>17,18</sup> but are not included in Table 1). Thus, Table 1 may contain genes that provide functional redundancy, but redundancy through the use of multiple members of gene families does not seem to have a numerically important role in touch receptor differentiation. Last, the genes used by these cells are not more nematode-specific than are those of the whole *C. elegans* genome. Of the genes in Table 1, 38% (27/71)

show similarity to genes in other organisms, as compared with 36% of all *C. elegans* genes<sup>19</sup>.

One motivation for this project was to identify candidates for uncloned, but mutated, genes. The highest ranking gene in Table 1, F57H12.7, is such a gene. F57H12.7 maps to the position of the previously uncloned *mec-17* gene. *mec-17* seems to be needed for maintained differentiation of the touch receptors; *mec-17* mutants are touch sensitive at hatching, but become insensitive as they mature<sup>2</sup>. Expression of *mec-3* and *mec-7* is reduced in older *mec-17* animals<sup>4,5</sup>. F57H12.7 was identified as *mec-17* because the single mutant allele of *mec-17*, *u265*, contained two missense mutations (Fig. 3a) and microinjection of F57H12.7 genomic DNA into *mec-17(u265)* animals rescued the mutant phenotype. F57H12.7 was expressed solely and strongly in the touch receptors from late embryos to adults in wild-type animals but not in *mec-3* mutants (Fig. 3b and data not shown). *mec-17* encodes a predicted polypeptide of 262 amino acids. Although this protein has not been described previously and has no readily identifiable motifs, MEC-17, another predicted *C. elegans* gene, W06B11.1, and two predicted genes each from *Drosophila*, mouse and humans share a domain of 158 amino acids that has 29–41% identity (Fig. 3a).

Although many of the genes encode proteins of unknown function, others suggest specific roles in the touch receptors. For example, one of the most striking features of the *C. elegans* touch receptors is that their processes are packed with unusual, 15-prot filament microtubules<sup>20</sup> that are formed from  $\beta$ -tubulin and  $\alpha$ -tubulin encoded by *mec-7* (ref. 7) and *mec-12* (ref. 10), respectively. Thus, the presence of three genes encoding components of the T-complex chaperonin (*cct-1*, *cct-2* and *cct-4*) in Table 1 is apt, because this chaperonin has been implicated in the correct folding of  $\alpha$ - and  $\beta$ -tubulin and in microtubule assembly<sup>21</sup>. *cct-2* and *cct-4* were expressed in touch receptors (Fig. 3e, f) in a *mec-3*-dependent fashion (data not shown). These genes were expressed widely in *C. elegans*, but their expression in other cells was not affected by mutation of *mec-3* (data not shown).

The role of many of the genes, especially those with strong *mec-3*-dependent expression, in the function of touch receptors is intriguing. Such genes include *twk-28*, encoding a TASK potassium channel protein; ZC482.5, encoding a GABA ( $\gamma$ -aminobutyric acid) receptor subunit; four genes that encode predicted enzymes or proteins with catalytic domains, C03A3.3 (metallo- $\beta$ -lactamase), F14E5.4 (acid phosphatase), F20D6.5 (tyrosine kinase) and F01D5.8 (esterase); and two genes (C06A8.3 and *far-3*) that encode proteins that are similar to antigens from the parasitic nematode *Onchocerca volvulus*.

In multicellular organisms, tissue and cell heterogeneity has restricted the usefulness of DNA microarrays<sup>22</sup>. In *C. elegans*, for example, RNA from whole animals has been applied to DNA microarrays to see gross differences in expression between male and hermaphrodite animals<sup>9</sup> or germline cells<sup>23</sup>, which make up about two-thirds of the cells in adults, or between animals at different developmental stages<sup>9,24</sup>. But genes that are expressed in a few cells, such as the six touch receptors, cannot be identified systematically using whole-worm RNA. Thus, a main advantage of our technique, which uses isolated cells, linearly amplified RNA, and DNA microarrays, is its ability to analyse global gene expression for individual cell types. Although the cell culturing needed in this method may prevent the detection of differences resulting from cell-cell interactions, these techniques should be applicable towards the study of cellular development and function under many different experimental conditions. □

## Methods

### Cell culture, FACS and RNA amplification

Strain TU2583, containing a *P<sub>mec-18</sub>gfp* integrated array, and strain TU2767, containing a *P<sub>mec-3</sub>gfp* integrated array with *mec-3(e1338)*, were used, respectively, to generate wild-type and *mec-3* mutant touch receptors. Cells were obtained from embryos by a

modification of the method of L. Bloom (<http://cobweb.dartmouth.edu/cgi-bin/cgiwrap/~ambros/protocol.cgi?id=16>). Details of the modified protocol and other methods are given in the Supplementary Information. We carried out all flow cytometric analyses and cell sorting on a FACS VantageSE equipped with CellQuest software (Becton Dickinson). RNA from the sorted cells was amplified through two rounds by a modification of the method of ref. 25.

**Real-time RT-PCR**

Roughly 3–5 µg of second-round amplified RNA from purified wild-type or *mec-3(e1338)* touch receptors was converted to complementary DNA using random primers and PowerScript reverse transcriptase (Clontech) according to the manufacturer's protocol. We used 30 ng of cDNA from each (two wild-type and two *mec-3(e1338)* RNAs) sample in a 20-µl RT-PCR reaction using the FastStart DNA Master SYBR Green I kit (Roche) supplemented with 3 mM MgCl<sub>2</sub> and Platinum Taq Polymerase (Invitrogen). Quantitative RT-PCR was done using the Roche LightCycler system. The efficiency of amplification for each pair of primers was determined by a standard curve that was generated using serially diluted genomic DNA. The ratio of expression level of each investigated gene between wild-type and *mec-3(e1338)* samples was calculated as described<sup>26</sup> using *unc-86* as the reference gene. For at least eight genes, one intron was included in the PCR product from genomic DNA whose length and melting temperature would be different from that of the PCR product from RNA samples. As we did not see the melting curve for the PCR product generated from genomic DNA, we could be certain that there was no detectable genomic DNA contamination in the generated RNA samples.

**Microarray analysis**

Twenty-microgram samples of amplified RNA were reverse transcribed into cDNAs using random hexamers, labelled with Cy3-dUTP or Cy5-dUTP (Amersham), and applied to microarrays as described<sup>25</sup>. We scanned hybridized microarrays with an Axon scanner and determined the expression levels with GenePix software. The fluorescence intensities for Cy3-labelled RNA and Cy5-labelled RNA were normalized by GenePix software and the background standard deviations were normalized accordingly. Three amplified RNA samples were generated from three independent primary cell cultures and cell sorting for wild-type touch receptors (w1, w2 and w3) and for *mec-3(e1338)* touch receptors (m1, m2 and m3). Each hybridization (w1–m1, w2–m2, w3–m3) was done twice. The data for these experiments are stored in the Stanford Microarray Database in the *C. elegans* section, under Chalfie (<http://genome-www5.stanford.edu/cgi-bin/SMD/cluster/QuerySetup.pl>). All non-flagged spots for which the net fluorescence intensity at least in one channel was three background standard deviations above background were considered to be well hybridized and measured. If the net fluorescence intensity for a non-flagged spot was within 1 s.d. of the background, the net fluorescence intensity was set as 1 s.d. of the background. We averaged the natural logs of the ratios for the duplicate hybridizations. Student's *t*-test *P* values were determined with the averages from independent experiments.

**Gene analysis**

We used the motif-finding program MEME<sup>27</sup> (<http://meme.sdsc.edu/meme/website/meme.html>) to derive a 19-nucleotide consensus sequence from 19 known MEC-3/UNC-86 binding sites in the promoter region of *mec-3*, *mec-4* and *mec-7* in *C. elegans*, *Caenorhabditis briggsae* and *Caenorhabditis vulgaris*. A weight matrix for the defined consensus site was built on the basis of the multiple sequence alignment and was used to search the regions 5' of the start codon (up to 2.5 kb) of the genes in Table 1 for the significant matches (*P* < 10<sup>-4</sup>) using a pattern search tool (<http://www.ucmb.ub.ac.be/bioinformatics/rsa-tools>). A graphical representation of the weight matrix was then generated (<http://ep.ebi.ac.uk/EP/SEQLOGO>).

We used the Blast protein-protein tool (BlastP; <http://www.ncbi.nlm.nih.gov/BLAST>) with the non-redundant Protein Data Base at the National Center for Biotechnology Information and with *C. elegans* proteins at Wormbase (<http://www.wormbase.org>) to identify similar genes, that is, those with an expectation value (*E*) < 10<sup>-10</sup> and that were aligned for at least 80% of their lengths<sup>19</sup>.

We amplified the *mec-17(u265)* genomic fragment by PCR from worm lysates and the PCR product was sequenced by GeneWiz Inc. A 5.5-kb fragment of *mec-17* genomic DNA was amplified by PCR using the Clontech polymerase mixture and subcloned using the TA cloning kit (Invitrogen). The plasmid (0.75 ng µl<sup>-1</sup>) was injected together with the *rol-6(su1006)* marker plasmid pRF4 (75 ng µl<sup>-1</sup>; ref. 28) and pBR322 (75 ng µl<sup>-1</sup>) into *mec-17(u265)* animals. The rescue of the touch-insensitive phenotype was determined in transgenic adults.

We generated *mec-17::gfp* by inserting a 5-kb genomic DNA fragment (produced by PCR) including 1.9 kb of upstream sequence and all of the predicted genomic sequence except that for 42 amino acids at the carboxy terminus of MEC-17 into pPD95.77 (a gift from A. Fire), which contains the *gfp* gene with added introns. The construct was transformed with DA#735 (ref. 29), a plasmid containing *lin-15* + DNA, into MT1642 (*lin-15(n765)*) to produce three stable lines that were examined using an Axiophot microscope with a FITC filter set. We also transformed TU353 (*mec-3(e1338)*; *dpy-20(e1282)*) with *mec-17::gfp* DNA and pMH86 (ref. 30), a plasmid containing *dpy-20* + DNA, and obtained three stable lines. Animals from these lines were mated with *dpy-20(e1282)* males to produce *mec-3/+* animals.

The *cct-4::gfp* construct was generated by inserting 2,389 bases of *cct-4* genomic DNA, including 585 bases of upstream sequence, in-frame into pPD95.77. *P<sub>cct-2</sub>gfp* was generated by inserting 2,436 bases of upstream sequence of *cct-2* and 42 bases of coding sequence in-frame into pPD95.77. Each construct was injected with pMH86 into TU353 and produced three stable lines. Animals were mated with *dpy-20(e1282)* males to obtain *mec-3/+* heterozygotes.

Similar constructs were formed for T07D1.3 by inserting a 2.2-kb DNA fragment including the upstream sequence and the starting codon of the gene into pPD95.75 (a gift from A. Fire), and for C03A3.3 by inserting a 3.6-kb genomic DNA fragment including the upstream sequence and all of the predicted genomic sequence for the gene in-frame into pPD95.77.

Received 5 February; accepted 20 May 2002; doi:10.1038/nature00891.

- Chalfie, M. & Sulston, J. Developmental genetics of the mechanosensory neurons of *Caenorhabditis elegans*. *Dev. Biol.* **82**, 358–370 (1981).
- Chalfie, M. & Au, M. Genetic control of differentiation of the *Caenorhabditis elegans* touch receptor neurons. *Science* **243**, 1027–1033 (1989).
- Way, J. C. & Chalfie, M. *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell* **54**, 5–16 (1988).
- Way, J. C. & Chalfie, M. The *mec-3* gene of *Caenorhabditis elegans* requires its own product for maintained expression and is expressed in three neuronal cell types. *Genes Dev.* **3**, 1823–1833 (1989).
- Mitani, S., Du, H., Hall, D. H., Driscoll, M. & Chalfie, M. Combinatorial control of touch receptor neuron expression in *Caenorhabditis elegans*. *Development* **119**, 773–783 (1993).
- Xue, D., Finney, M., Ruvkun, G. & Chalfie, M. Regulation of the *mec-3* gene by the *C. elegans* homeoproteins UNC-86 and MEC-3. *EMBO J.* **11**, 4969–4979 (1992).
- Savage, C. *et al.* *mec-7* is a  $\beta$ -tubulin gene required for the production of 15- protofilament microtubules in *Caenorhabditis elegans*. *Genes Dev.* **3**, 870–881 (1989).
- Gu, G., Caldwell, G. A. & Chalfie, M. Genetic interactions affecting touch sensitivity in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **93**, 6577–6582 (1996).
- 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).
- Fukushige, T. *et al.* MEC-12, an  $\alpha$ -tubulin required for touch sensitivity in *C. elegans*. *J. Cell Sci.* **112**, 395–403 (1999).
- Xue, D., Tu, Y. & Chalfie, M. Cooperative interactions between the *Caenorhabditis elegans* homeoproteins UNC-86 and MEC-3. *Science* **261**, 1324–1328 (1993).
- Duggan, A., Ma, C. & Chalfie, M. Regulation of touch receptor differentiation by the *Caenorhabditis elegans mec-3* and *unc-86* genes. *Development* **125**, 4107–4119 (1998).
- Van Helden, J., Andre, B. & Collado-Vides, J. A website for the computational analysis of yeast regulatory sequences. *Yeast* **16**, 177–187 (2000).
- Blumenthal, T. *et al.* A global analysis of *Caenorhabditis elegans* operons. *Nature* **417**, 851–854 (2002).
- Huang, M. & Chalfie, M. Gene interactions affecting mechanosensory transduction in *Caenorhabditis elegans*. *Nature* **367**, 467–470 (1994).
- Barnes, T. M., Jin, Y., Horvitz, H. R., Ruvkun, G. & Hekimi, S. The *Caenorhabditis elegans* behavioural gene *unc-24* encodes a novel bipartite protein similar to both erythrocyte band 7.2 (stomatins) and nonspecific lipid transfer protein. *J. Neurochem.* **67**, 46–57 (1996).
- Treinin, M. & Chalfie, M. A mutated acetylcholine receptor subunit causes neuronal degeneration in *C. elegans*. *Neuron* **14**, 871–877 (1995).
- Treinin, M., Gillo, B., Liebman, L. & Chalfie, M. Two functionally dependent acetylcholine subunits are encoded in a single *Caenorhabditis elegans* operon. *Proc. Natl Acad. Sci. USA* **95**, 15492–15495 (1998).
- Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215 (2000).
- Chalfie, M. & Thomson, J. N. Structural and functional diversity in the neuronal microtubules of *Caenorhabditis elegans*. *J. Cell Biol.* **93**, 15–23 (1982).
- Liang, P. & MacRae, T. H. Molecular chaperones and the cytoskeleton. *J. Cell Sci.* **110**, 1431–1440 (1997).
- Lockhart, D. J. & Winzler, E. A. Genomics, gene expression and DNA arrays. *Nature* **405**, 827–836 (2000).
- Reinke, V. *et al.* A global profile of germline gene expression in *C. elegans*. *Mol. Cell.* **6**, 605–616 (2000).
- Hill, A. A., Hunter, C. P., Tsung, B. T., Tucker-Kellogg, G. & Brown, E. L. Genomic analysis of gene expression in *C. elegans*. *Science* **290**, 809–812 (2000).
- Eberwine, J. *et al.* Analysis of gene expression in single live neurons. *Proc. Natl Acad. Sci. USA* **89**, 3010–3014 (1992).
- Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, E45–E45 (2001).
- Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
- 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).
- Huang, L. S., Tzou, P. & Sternberg, P. W. The *lin-15* locus encodes two negative regulators of *Caenorhabditis elegans* vulval development. *Mol. Biol. Cell* **5**, 395–411 (1994).
- Han, M. & Sternberg, P. W. Analysis of dominant-negative mutations of the *Caenorhabditis elegans let-60 ras* gene. *Genes Dev.* **5**, 2188–2198 (1991).

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

**Acknowledgements**

We thank X. Chen for help with analysis of the microarray data; M. Goodman for constructing the integrated *mec-18::gfp* strain; B. Tycko for the LightCycler; and E. Schwarz, J. Wang and J. Eberwine for discussion. This work was supported by a grant from the National Center for Research Resources to S.K.K. and a grant from the National Institutes of Health to M.C.

**Competing interests statement**

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.C. (e-mail: mc21@columbia.edu).