Title

Whole-organism eQTL mapping at cellular resolution with single-cell sequencing

Authors

Eyal Ben-David[1,2*], James Boocock[1], Longhua Guo[1], Stefan Zdraljevic[1], Joshua S. Bloom[1*], Leonid Kruglyak[1*]

(*) To whom correspondence should be addressed: eyal.bendavid@mail.huji.ac.il, jbloom@mednet.ucla.edu, lkruglyak@mednet.ucla.edu

Affiliations

1. Department of Human Genetics, Department of Biological Chemistry, and Howard Hughes Medical Institute, University of California, Los Angeles, CA 90095, USA.

2. Department of Biochemistry and Molecular Biology, Institute for Medical Research Israel-Canada, The Hebrew University School of Medicine, Jerusalem, Israel

Abstract

Genetic regulation of gene expression underlies variation in disease risk and other complex traits. The effect of expression quantitative trait loci (eQTLs) varies across cell types; however, the complexity of mammalian tissues makes studying cell-type eQTLs highly challenging. We developed a novel approach in the model nematode *Caenorhabditis elegans* that uses single cell RNA sequencing to map eQTLs at cellular resolution in a single one-pot experiment. We mapped eQTLs across cell types in an extremely large population of genetically distinct *C. elegnas* individuals. We found cell-type-specific *trans*-eQTL hotspots that affect the expression of core pathways in the relevant cell types. Finally, we found single-cell-specific eQTL effects in the nervous system, including an eQTL with opposite effects in two individual neurons. Our results show that eQTL effects can be specific down to the level of single cells.

Introduction

Gene expression differences have a strong genetic basis, and genome-wide studies have identified thousands of regions affecting gene expression, termed expression quantitative trait loci (eQTLs) [1]. eQTLs have been proposed to underlie genetic associations with complex traits and diseases [2,3], and genome-wide eQTL mapping holds great promise for uncovering the molecular underpinnings of phenotypic variation.

Studies in purified blood cell populations [4-6] and computational analyses in human tissues [7,8] indicate that many eQTLs are cell-type specific. Studying eQTLs in relevant tissues and cell types is thus important for understanding their role in trait variation [9]. However, major challenges remain for cell-type-specific eQTL mapping. First, studying multiple tissues or cell types is laborious and expensive, and has mostly been limited to large consortia [10]. Additionally, diseases can be associated with specific subsets of cells that may not be readily separable from the rest of the tissue, limiting the insight that can be gleaned from studying bulk tissue samples [11].

The recently developed technology of single-cell RNA sequencing (scRNA-seq) has the potential to significantly mitigate both of the above issues. In scRNA-seq, gene expression is profiled at the level of individual cells, and different cell populations are identified based on their gene expression profiles. This approach allows simultaneous measurement of expression in multiple tissues and cell types in a “one-pot” experiment, reducing the number of samples that need to be processed and profiled separately. Furthermore, distinct cell types can be interrogated directly, facilitating eQTL mapping in disease-relevant but rare cell types.

The size and complexity of mammalian tissues has so far limited the ability to use scRNA-seq for eQTL mapping, and studies have focused on purified cell populations and cell lines [12,13].

55 To pilot a whole-organism approach, we therefore turned to the model nematode *Caenorhabditis*
56 *elegans*. One of the mainstays of modern genetic research, *C. elegans* has an invariant cell
57 lineage that leads to each individual having the same number and identity of cells [14]. Decades
58 of research have uncovered the functions of many of those cells, as well as expression markers
59 that uniquely identify them [15,16]. These features and resources make *C. elegans* exceptionally
60 well-suited for using scRNA-seq to map eQTLs across cell types in the natural physiological
61 context of a whole animal, down to cellular resolution.
62
63 **Results**
64 **Cell-type-specific eQTL mapping in a single one-pot experiment**
65 Genome-wide eQTL mapping involves acquiring genotypes and gene expression profiles for a
66 genetically diverse cohort. We recently developed a method, *C. elegans* extreme quantitative
67 trait locus (ceX-QTL) mapping, for genetic analysis of complex traits in extremely large
68 populations of segregants [17]. The method takes advantage of a mutation in the gene *fog-2* that
69 forces the normally hermaphroditic *C. elegans* to reproduce via obligate outcrossing, allowing
70 us to propagate a large crossing experiment for multiple generations.
71 Here we build on ceX-QTL by combining it with single-cell RNA sequencing (scRNA-seq)
72 to carry out eQTL mapping at cellular resolution in a single one-pot experiment (Fig. 1A). In this
73 approach, a large heterogeneous pool of cells from thousands of genetically distinct individuals
74 is profiled using scRNA-seq, cell types are inferred by clustering scRNA-seq profiles and
75 studying known cell-type markers, and genotype information is reconstructed using expressed
76 genetic variants, enabling eQTL mapping in multiple cell types simultaneously.

77 We propagated a cross between the laboratory strain N2 and a highly divergent isolate
78 from Hawaii, CB4856, for four generations, generating a pool of 200,000 genetically distinct F4
79 segregants. We dissociated the segregant pool to single cells at the L2 larval stage and profiled
80 the cells with scRNA-seq. We identified clusters in a Uniform Manifold Approximation and
81 Projection (UMAP) of the dataset [18–20], and determined their cell-type identities using known
82 markers [21,22]. Our final dataset comprises 55,508 cells classified into 19 different cell types (Fig.
83 1B) (Table S1). The observed number of cells of each type was strongly correlated with the
84 known cell-type abundance in L2 larvae (Spearman's $\rho = 0.87$, $p = 2.2 \times 10^{-6}$, Supplementary
85 Fig. 1).

86 Most of the cells in our sample were expected to carry unique genotypes (Methods). This
87 design is advantageous for eQTL mapping because it maximizes the sample size [23], but requires
88 *de novo* genotype calling, since the genotype of each cell is unknown beforehand. Rather than
89 assign deterministic genotype calls based on sparse scRNA-seq data, we derived genotype
90 probabilities for each cell using a Hidden Markov Model (HMM) (Supplementary Fig. 2). We then
91 performed eQTL mapping with these genotype probabilities in a negative binomial modeling
92 framework (Methods).

93

94 **eQTL mapping in multiple cell types**

95 We mapped 1,718 *cis* eQTLs in 1,294 genes, and 451 *trans* eQTLs in 390 genes, at a false
96 discovery rate (FDR) of 10% across the different cell types (Fig. 2A-B, Table S2). The number
97 of eQTLs detected in each cell type was strongly correlated with the number of cells of that type
98 (Spearman's $\rho = 0.91$, $p < 2.2 \times 10^{-16}$). In cell types with >1000 cells, we mapped between 52
99 and 415 eQTLs (Table S1). For 1,071 of the 1,294 genes with a *cis* eQTL (83%), the eQTL was

100 detected in only one cell type. For 208 of the remaining 223 genes (93%), the direction of the
101 eQTL effect was the same in all cell types in which it was detected.

102 We studied to what degree our *cis* eQTL results were concordant with gene expression
103 differences between the parents. We generated a scRNA-seq dataset from 6,721 N2 and 3,104
104 CB4856 cells, and used a classifier trained on the segregant dataset to identify cell types in the
105 parental scRNA-seq dataset. We then carried out a differential expression analysis in each cell
106 type. We found 870 differentially expressed genes (at a >2 fold change and FDR of 10%), of
107 which 201 (23%) had a *cis* eQTL in the same tissue (OR = 18.8, $p < 2.2 \times 10^{-16}$, Fisher's Exact
108 Test). 191 of these *cis* eQTL (95%) showed the same direction of effect as the parental
109 difference. Further, the effect sizes of the significant *cis* eQTLs were strongly correlated with the
110 sizes of the parental differences (Spearman's $\rho = 0.66$, $p < 2.2 \times 10^{-16}$) (Supplementary Fig. 3-
111 4). These results provide independent support for our *cis*-eQTL mapping, and show that for a
112 sizable fraction of the genes, those *cis* eQTLs are a major cause of differential gene expression
113 between the strains.

114

115 **Comparison between bulk and single-cell *cis*-eQTL mapping**

116 To investigate the relationship between single-cell and bulk eQTL mapping, we compared our
117 single-cell eQTLs to those previously identified in a panel of 200 recombinant inbred lines (RILs)
118 generated from crossing N2 and CB4856 [24]. In the bulk study, a large population of whole worms
119 from each RIL was recovered at a late larval stage, L4, and profiled on expression microarrays.
120 We reanalyzed data for 11,535 genes expressed in both datasets and identified 981 *cis* eQTLs
121 in the bulk dataset (at an FDR cutoff of 10%). Despite major differences in experimental design,
122 including the developmental stage of the worms, the overlap with the single-cell *cis* eQTLs was

123 highly significant, with 335 *cis* eQTLs shared between the studies (Odds Ratio = 7.2, $p < 2.2 \times$
124 10^{-16}, Fisher's exact test) (Fig. 2C). These shared loci represented 34% of the bulk *cis* eQTLs
125 and 32% of the single-cell *cis* eQTLs. Furthermore, the bulk and single-cell eQTL effect sizes
126 were highly correlated (Spearman's $\rho = 0.64$, $p < 2.2 \times 10^{-16}$) (Fig. 2D). Lastly, single-cell eQTLs
127 detected in multiple cell types were more likely to also be seen in the bulk study: 50% of the
128 genes with *cis* eQTLs detected in multiple cell types were also identified in bulk, compared to
129 28% of the eQTLs detected in only one cell type (OR = 2.58, $p = 2.1 \times 10^{-8}$) (Fig. 2C). This
130 observation suggests that the single-cell eQTL mapping approach improves the power to detect
131 cell-type specific effects.
132

133 **Shared and cell-type specific *trans*-eQTL hotspots**

134 We observed that 90 of the 451 *trans* eQTLs clustered at 5 hotspots, each containing 12-31
135 eQTLs (Supplementary Fig. 5, Table S3). A hotspot on Chr I was identified independently in
136 both neurons and seam cells; the top associated variant (Chr. I:10890182) was the same for
137 both cell types (Supplementary Fig. 5A-B). The other hotspots were identified in the body wall
138 muscle (on Chr. I) (Supplementary Fig. 5C), the intestine (on Chr. V) (Supplementary Fig 5D),
139 and neurons (two distinct hotspots on Chr. III) (Supplementary Fig. 5B).

140 To test whether the target genes of these 5 hotspots are involved in coherent biological
141 processes, we relaxed the FDR threshold to 20%, which increased the number of genes linked
142 to each hotspot to 21-42, and performed Gene Ontology (GO) enrichment analysis (Table S4).
143 For three of the hotspots, we found significant enrichments that were consistent with the cell-
144 type specificity of the hotspot. The targets of the hotspot detected in intestinal cells were weakly
145 enriched for genes involved in the innate immune response (FDR-corrected p = 0.042), a major

146 role of that tissue [25]. The targets of the hotspot detected in the body wall muscle were enriched
147 for genes associated with the term *myofilament* (FDR corrected p = 6.4 x 10^{-8}), *actin cytoskeleton*
148 (FDR-corrected p = 4.2 x 10^{-6}), and related terms. The enrichment was driven by the genes *mup-*
149 *2, tni-1, tnt-2, mlc-2, mlc-3, lev-11* and *act-4*. *mlc-2* and *mlc-3* encode a myosin light chain, and
150 *act-4* encodes an actin protein. *lev-11* encodes a tropomyosin, and *mup-2*, *tni-1*, and *tnt-2*
151 encode 3 of the 4 proteins in *C. elegans* that are expressed in the body-wall muscle and form
152 troponin complexes, highly conserved regulators of muscle contraction [26] (Supplementary Fig.
153 6).

154 The targets of the neuronal hotspot on the right arm of Chr III were enriched for genes
155 involved in *vesicle localization* (FDR-corrected p = 7.5 x 10^{-3}), as well as for BMP receptor
156 binding genes (FDR-corrected p = 2.8 x 10^{-3}). The latter enrichment was driven by *dbl-1* and *tig-*
157 *2*, orthologs of human bone morphogenetic protein (BMP) genes BMP5 and BMP8 and ligands
158 of the transforming growth factor beta (TGF-β) pathway [27]. Notably, *dbl-1* was discovered as a
159 gene that regulates body size in *C. elegans* [28], the hotspot peak marker is located <300 kb from
160 the peak of a QTL we previously identified for body size [29], and the corresponding confidence
161 intervals overlap (Table S3), suggesting that differential regulation of the TGF-β pathway is
162 involved in variation in body size between N2 and CB4856.

163

164 **Cell-specific eQTL effects in the *C. elegans* nervous system**

165 *C. elegans* is a premier model for studying neurobiology at the cellular level, which is aided by
166 its invariant cell lineage and the diverse functions associated with specific individual neurons.
167 Importantly, many of the neurons are highly variable in their gene expression, and express
168 specific gene markers [16]. To identify specific subtypes of neuronal cells, we separately clustered

169 the 12,467 cells identified as neurons and compared the clusters to previous *C. elegans* scRNA-
170 seq datasets, including the recently published *C. elegans* Neuronal Gene Expression Map &
171 Network (CeNGEN) [21,22,30,31] (Table S5). The neurons fell into 81 distinct clusters, ranging from
172 17 to 872 cells. We mapped these clusters onto 100 (83%) of the 120 neuronal clusters identified
173 in CeNGEN (Supplementary Fig. 7). We also identified CEM neurons, which are male specific
174 and absent from CeNGEN, based on the expression of the marker *cwp-1* [32].

175 We mapped *cis* eQTLs in each of the single neuronal subtypes (sn-eQTLs) and identified
176 a total of 163 sn-eQTLs in 132 genes at an FDR of 10% (Fig. 3A, Table S6). Of these, 117 (88%)
177 were identified in only a single neuronal subtype. Functional annotation of sn-eQTLs identified
178 25 genes involved in signaling (FDR-corrected p = 0.047), including 12 genes involved in G-
179 protein Coupled Receptor (GPCR) signaling (FDR-corrected p = 0.047) and 8 genes involved in
180 neuropeptide signaling (FDR-corrected $p = 9.9 \times 10^{-3}$).

181 We compared the sn-eQTLs to those identified when all neurons were analyzed jointly
182 ("pan-neuronal mapping"), and found that a sizable fraction of the sn-eQTLs did not have
183 evidence for a pan-neuronal signal: 92 were not identified pan-neuronally at an FDR of 10%,
184 and 69 were not identified even at a highly permissive FDR of 50%, suggesting that they exert
185 their effects only in specific neuronal subtypes (Fig. 3A). Regardless of statistical significance,
186 pan-neuronal eQTLs should have consistent effect directions across neuronal subtypes, while
187 subtype-specific eQTLs should not. We therefore compared the direction of effect of each sn-
188 eQTL in the subtype in which it was detected with its direction of effect in the set of all neurons
189 excluding that subtype. Among the 69 sn-eQTLs with no signal in the pan-neuronal mapping
190 even at the permissive FDR, the direction of the effect was concordant for 33 and discordant for
191 36, not significantly different from chance ($p > 0.5$; binomial test), as would be expected if these

192 effects are truly subtype-specific (Fig. 3B). In contrast, among the 94 that had a pan-neuronal
193 signal at an FDR of 50%, the direction of the effect was concordant for 88 and discordant for
194 only 6 ($p < 0.000001$; binomial test), consistent with differences in detection arising from limited
195 statistical power.

196 In a striking case, we observed an sn-eQTL in the neuropeptide gene *nlp-21* that showed
197 significant and opposing effects in two neurons (Fig. 3C-D). In the RIC neuron, higher *nlp-21*
198 expression was associated with the CB4856 allele ($\beta = 4.4$, FDR-corrected $p = 0.03$), while in
199 the RIM neuron, higher *nlp-21* expression was associated with the N2 allele ($\beta = -5.4$, FDR-
200 corrected $p = 9.8 \times 10^{-7}$). In the pan-neuronal mapping, no significant effect is observed for this
201 gene. We identified the RIC and RIM neurons in the parental dataset, and although the small
202 number of cells in each group (35 and 27, respectively, with only 9 and 5 of them from CB4856)
203 was insufficient for statistical testing, the directions of the differences agreed with the eQTL
204 effects (Fig. 3E). These results provide direct evidence that eQTLs can be specific down to the
205 cellular level.

206

207 **Discussion**

208 We used scRNA-seq to map eQTLs in *C. elegans* across cell types in a single one-pot
209 experiment. Earlier scRNA-seq eQTL mapping studies were limited in sample size to at most
210 ~100 individuals, but nevertheless highlighted the potential of this approach to identify cell-type
211 [13] and developmental [12] eQTLs, as well as loci affecting expression variance [33]. Our novel
212 approach allowed us to map eQTLs in tens of thousands of genotypically distinct individuals and
213 enabled detection of both *cis* and *trans* eQTLs, as well as resolution of their effects down to the
214 level of specific cells.

215 One of the major factors affecting gene expression studies is variation resulting from
216 uncontrolled environmental differences between individuals that are grown or processed
217 separately. By using scRNA-seq, we were able to process all individuals jointly. After the initial
218 parental cross, all subsequent steps carried out over the course of five *C. elegans* life-cycles
219 (three weeks) were performed in bulk, limiting any confounding environmental factors. To
220 minimize the influence of genotype on development, we synchronized the worms at the first
221 larval stage, L1, and collected samples at the L2 stage, limiting the time for differences to
222 accumulate post-synchronization. Even careful synchronization is not expected to completely
223 remove the effects of genetic variation on developmental timing, and such variation can be
224 combined with gene expression time course data collected during development to increase the
225 power of eQTL mapping and to study the developmental dynamics of eQTLs [34]. This raises the
226 possibility that future scRNA-seq studies of *C. elegans* across developmental stages would open
227 the door to a similar analysis of our single-cell eQTL dataset.

228 Previous work suggested the existence of cell-type specific eQTL hotspots in *C. elegans*
229 based on the expression patterns of hotspot targets [34]. We discovered three hotspots that are
230 cell-type specific, with targets that are involved in core functions performed by these cell types.
231 Recently, eQTL hotspots have been identified in human blood cells [35,36], as well as in cell lines
232 [37]. These results suggest hotspot and *trans* eQTL discovery is facilitated by expression studies
233 that can distinguish cell types and point to a larger role of hotspots in the genetics of gene
234 expression in animals.

235 A comparison of the single-cell *cis* eQTLs to those mapped in a previous whole-worm
236 eQTL study from our laboratory showed a highly significant overlap despite major differences in
237 experimental design. These results join accumulating evidence that a sizable fraction of *cis*

238 eQTLs have robust, consistent effects [10,38], and show that many of the effects are conserved
239 across worm development. The strong overlap of *cis* eQTLs mapped by scRNA-seq and by
240 whole-worm analysis also suggests that the effect of many *cis* eQTLs is conserved across cell
241 types. By generating an extremely large number of unique segregants, our method enables
242 scaling up the number of studied individuals simply by sequencing a larger number of cells.
243 Thus, the increasing throughput of single-cell technologies and sequencing platforms will enable
244 future work to study cell-type specificity of *cis* and *trans* eQTLs in greater detail.

245 Lastly, we discovered *cis* eQTLs that act in single subtypes of *C. elegans* neurons,
246 including many that were not found when all neurons were analyzed jointly. Importantly, we also
247 discovered an eQTL that influences expression of the gene *nlp-21* in opposing directions in two
248 different neurons. Such antagonism could result from context-dependent effects of the same
249 variant, or from different variants acting in each cell. Our study joins a growing body of work that
250 identified eQTLs that are specific to cell types [5–8,13,39,40], as well as developmental [12,34] and
251 environmental [41–43] contexts. Distinct genetic effects are found across cell types and conditions,
252 and large-scale study of such effects is therefore crucial for gaining a comprehensive
253 understanding of regulatory variation.

254

255

Figures

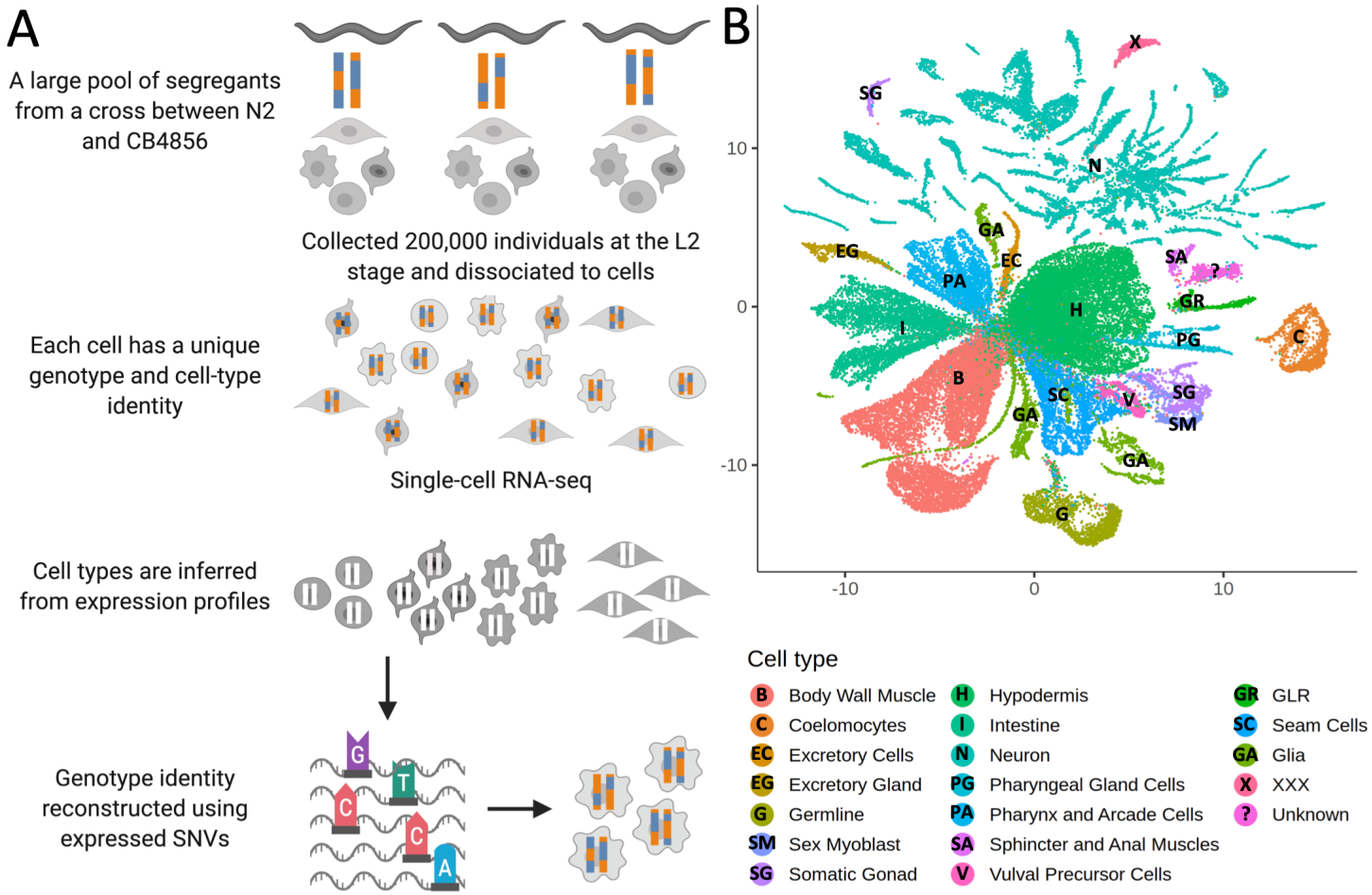


Figure 1. Whole-organism eQTL mapping with single-cell RNA-sequencing (scRNA-seq). (A) A large population of segregants is dissociated to single cells. Each cell in the suspension has an unknown genotype and cell-type identity. The suspension is profiled using scRNA-seq. Cell-type identity is inferred by clustering cells and comparing the expression of known marker genes. Genotypes are reconstructed from expressed single-nucleotide variants (SNVs). (B) The UMAP projection of 55,508 scRNA-seq expression profiles from approximately 200,000 *C. elegans* F4 segregants collected at the L2 larval stage is shown. Cells are colored based on the inferred cell type.

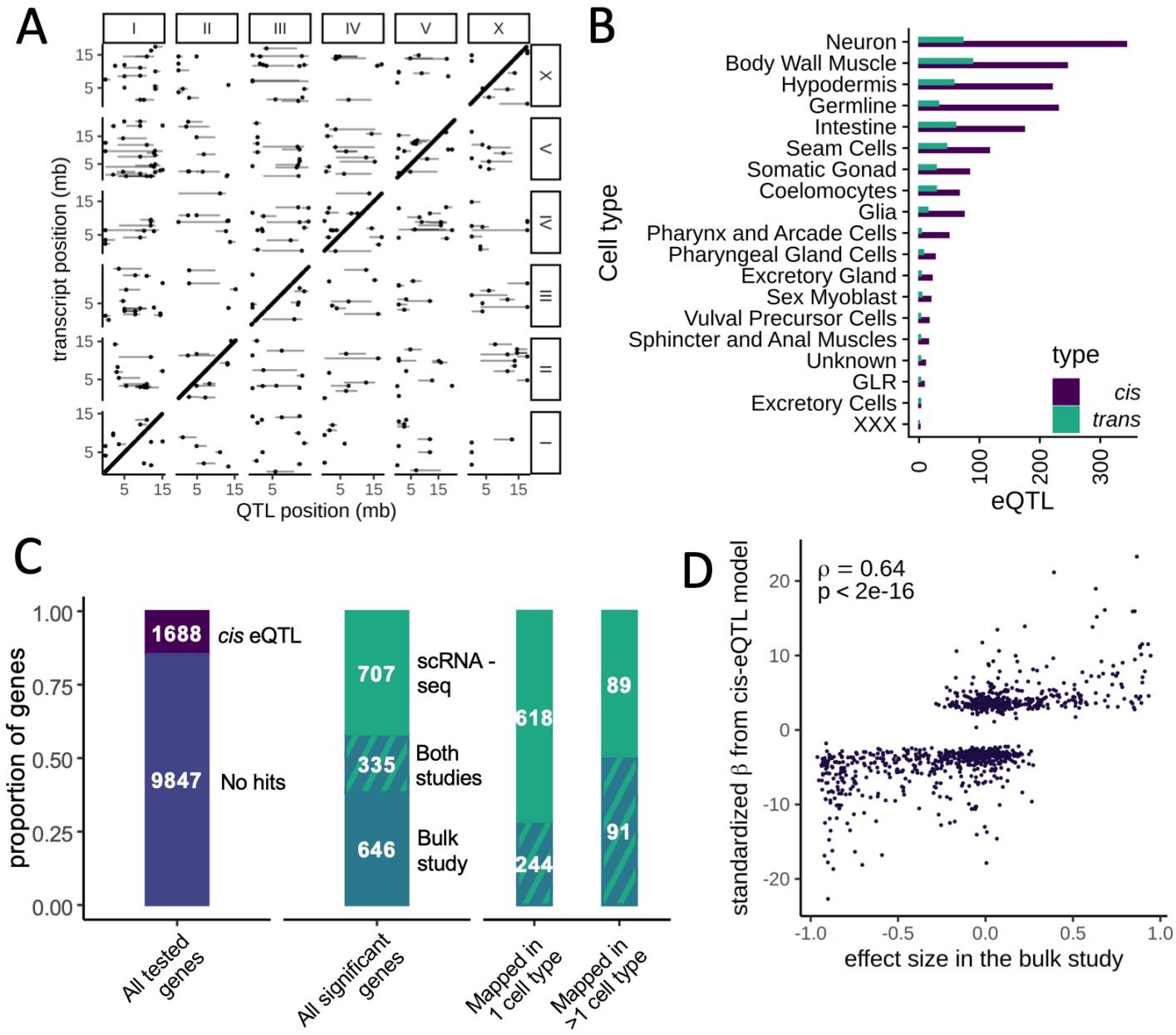


Figure 2. eQTL mapping in cell types. (A) A genome-wide map of eQTLs across all cell types is shown. The position of the eQTLs is shown on the x-axis, while the y-axis shows the position of the associated transcripts. Points along the diagonal are *cis* eQTLs (those mapping to nearby genes). (B) The number of *cis* and *trans* eQTLs mapped in each cell type. (C) The overlap between a previous study that mapped eQTLs in whole worms in a panel of recombinant inbred lines (RIL) and our dataset.

274 (Left) The proportion of genes with a *cis* eQTL in at least one dataset, out of all genes tested. (Middle)
275 Of the 1,688 significant *cis*-eQTL genes, 355 had a *cis* eQTL in both datasets, representing a highly
276 significant enrichment. (Right) Hits mapped in more than one cell type were more likely to also be
277 found in the whole-worm ("bulk") dataset. (D) Quantitative comparison between normalized effect
278 sizes in our dataset and in the whole-worm dataset.
279

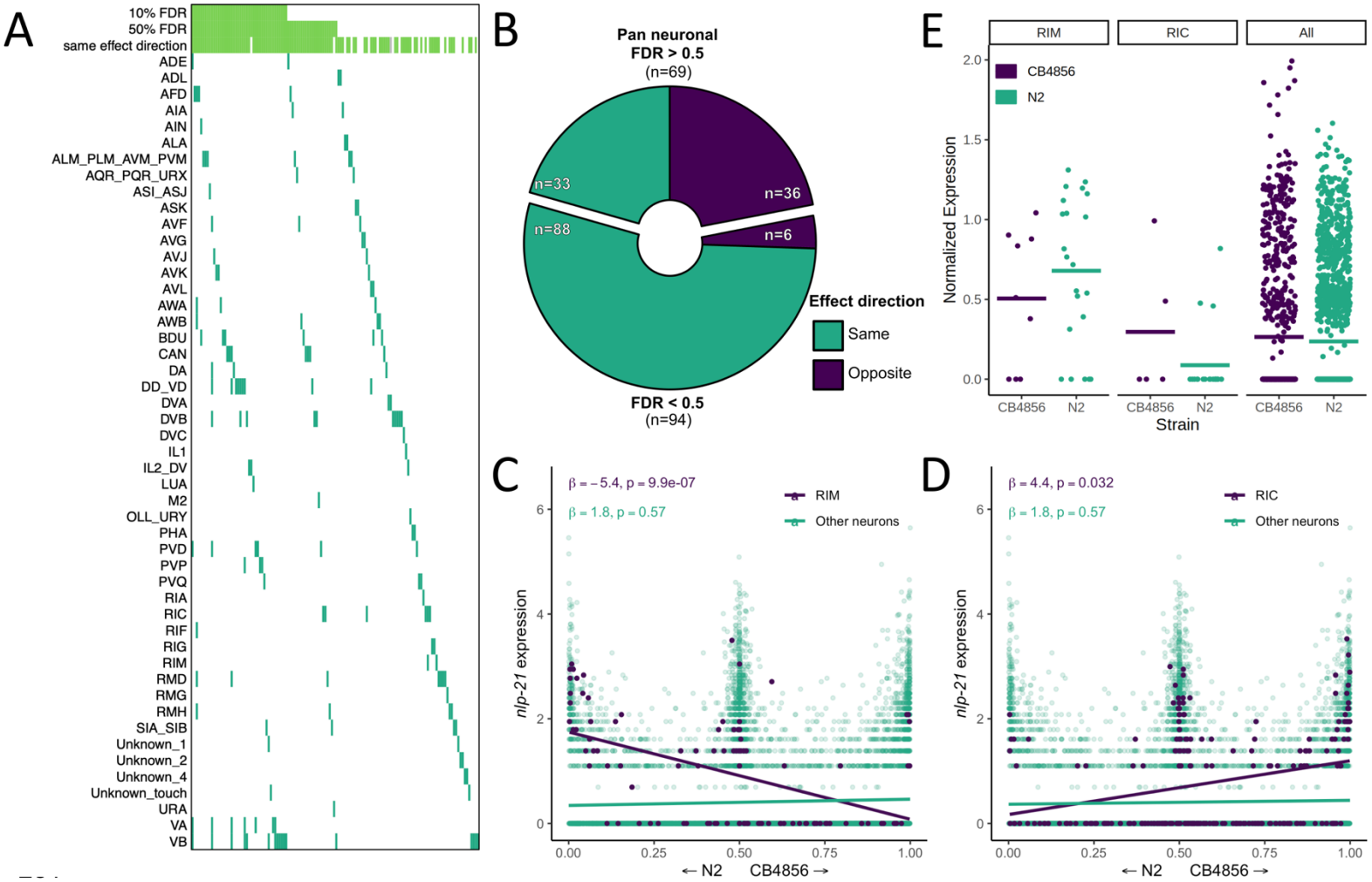


Figure 3. Neuron-specific eQTL mapping. (A) *cis*-eQTLs mapped in single neuronal subtypes (sn-eQTLs) are shown. The top three rows indicate whether the eQTL was mapped pan-neuronally at a 10% FDR threshold (row 1), at a 50% FDR threshold (row 2), and whether the sign of the effect estimate ("effect direction") was the same in the pan-neuronal and single cell mapping (row 3). (B) Comparing the effect direction between the sn-eQTL mapping and mapping in a set of neurons excluding the sn-eQTL neuron shows evidence for subtype-specific effects. The number of genes showing the same ("purple") or opposite ("turquoise") effect directions is shown for genes with pan-neuronal FDR > 50% (top) and < 50% (bottom). (C-D) An eQTL with antagonistic effects in two

290 neurons. Higher expression of the gene *nlp-21* in the RIM neuron is associated with the N2 allele (C),
291 while higher expression in the RIC neuron is associated with the CB4856 allele (D). In C and D, a
292 linear fit is shown for illustration. All p values are FDR-corrected. Read counts were normalized to the
293 number of UMIs in each cell and log-transformed. (E) Expression of *nlp-21* in the parental dataset.
294 The direction of effect is concordant between the left panel and (C) (RIM neuron) and between the
295 middle panel and (D) (RIC neuron). Horizontal lines are averages.
296

297 **Methods**

298 ***C. elegans* culturing**

299 *C. elegans* strains were cultured at 20°C using standard conditions with the exception that the
300 agar in the nematode growth media (NGM) was replaced with a 4:6 mixture of agarose and agar
301 (NGM+agarose), to prevent burrowing of the CB4856 strain. Parental strains used were QX2314
302 (N2 *fog-2*(q71) V; *hsp-90p*::GFP II) and PTM299 (CB4856 *fog-2*(kah89)). Large segregant
303 panels were generated as before [17]. Briefly, 500 L4 males from PTM299 and 500 L4
304 hermaphrodites from QX2314 were seeded on a plate for 30 hours, and gravid worms and eggs
305 were collected and bleached. Eggs were synchronized to L1 larvae for 24h, and seeded on 10cm
306 NGM+agarose plates. In each generation, gravid worms were bleached, their progeny
307 synchronized for 24h and seeded. The entire process was repeated up to F4, with 3-4 days per
308 generation.

309

310 **Cell extraction and sequencing**

311 192,000 F4 were seeded on four 10cm NGM+agarose plates seeded with OP50. L2 were
312 recovered after 24 hours, and staging was validated under a stereomicroscope. L2 cell
313 dissociation was carried out as previously described [44], implementing modifications from a later
314 study [45], as well as our own. Worms were recovered off the plates and washed three times in
315 M9. Lysis was then done with an SDS-DTT solution (200 mM DTT, 0.25% SDS, 20 mM HEPES,
316 pH 8.0, 3% sucrose) in a hula mixer set on low speed to prevent worms from settling. The lysate
317 was observed under the stereoscope every two minutes, and lysis was stopped when a blunted
318 head shape appeared in the majority of worms [45], after ~4 minutes. Worms were then washed
319 quickly three times in 1ml of M9, and two additional times in 1ml of egg buffer (118 mM NaCl,

320 48 mM KCl, 2 mM CaCl2, 2 mM MgCl2, 25 mM HEPES, pH 7.3, osmolarity adjusted to 340
321 mOsm with sucrose). Worms were then resuspended in 0.5ml of 20mg/ml Pronase E that was
322 freshly prepared in L15 media supplanted with 2% fetal bovine serum (L15-FBS) and adjusted
323 to 340 mOsm with sucrose. Worm dissociation was done by continuous pipetting on the side of
324 the tube, and monitored every 2-3 minutes on a microscope equipped with a 40x phase contrast
325 objective lens. Dissociation was stopped when few intact worms remained and a high density of
326 cells was visible. 0.5ml of L15-FBS was added to stop the reaction, and the lysate was spun for
327 6 minutes at 500g at 4°C. The cell pellet was resuspended in PBS (PBS was adjusted to 340
328 mOsm with sucrose). Cell suspension was spun for 1 minute in 100g at 4°C to remove remaining
329 undigested worms, counted and diluted to 1M cells/ml in osmolarity adjusted PBS, and loaded
330 directly onto 5 lanes of 3' Chromium single-cell RNA-sequencing flow cells (10x Genomics),
331 targeting 10,000 cells on each lane. Library prep was carried out according to manufacturer's
332 protocol. Prepared libraries were sequenced together on an S4 lane of Novaseq 6000. A paired-
333 end 2x150 run was done, to maximize the recovery of single-nucleotide variants. In all
334 downstream processing, each of the five 3' Chromium lanes processed concurrently was treated
335 as a separate "batch", and lane identity corresponds to the "batch" identity for the rest of the
336 methods.
337

338 **Single-cell RNA-sequencing data processing**

339 Raw sequencing reads were analyzed using *CellRanger* (Ver 3.0.2). We used a gene transfer
340 format (GTF) file that was corrected for misannotation of 3' untranslated regions (3'UTR) that
341 was generated in a previous study [22]. *C. elegans* cell-types differ widely in the number of UMIs
342 that are recovered using scRNA-seq. Therefore, a simple UMI cutoff, as is commonly used, may

343 be biased for cell-types with more UMIs. We therefore implemented an iterative pipeline to
344 recover clusters of *bona fide* cells and remove cell doublets as well as degraded cells. We took
345 20,000 cells with the most UMIs in each cluster (twice the targeted number of cells, 100,000
346 overall), and processed them in *Monocle* (Ver 3) [46]. Default parameters were used, with the
347 exception that 100 dimensions were used for reduction, and batch was added as a covariate.
348 Leiden clustering identified a total of 154 clusters, and we used the *top_markers* function in
349 *Monocle* to identify the genes upregulated in each. We then removed clusters whose top genes
350 included any ribosomal genes or the mitochondrial genes *ndfl-4*, *nduo-6*, *atp-6*, *ctc-2*, *ctc-3*, *ctc-*
351 *1*, which we noticed were usually found together as the most upregulated genes in clusters that
352 did not specifically express any known markers for *C. elegans* cell-types. This removed a total
353 of 30,980 cells (31%). For the remaining 69,020 cells, raw counts were processed using the R
354 package *SoupX* to reduce ambient RNA contamination [47]. We then normalized, reduced
355 dimensions and clustered the background corrected cell profiles in *Monocle* using the same
356 parameters as above.

357 To annotate cell-types, we used the markers described in a previous study (Table S12 in
358 Ref [22]) that reanalyzed a previous L2 single-cell dataset [21]. Our cell-type annotation corresponds
359 to the "UMAP" column in that table, with the following exceptions: (a) we separated hypodermis
360 from seam cells, somatic gonad from sex muscle cells, and glia from excretory cells, since those
361 groups were not clustering together in our data. (b) cells identified as "Miscellaneous" in that
362 table were annotated as individual cell-type identifications in our data, with the exception of the
363 sphincter and anal muscles which were not differentiated from each other in our data. Finally,
364 we re-evaluated our cell type identifications, and filtered cell doublets as well as dead cell or
365 debris that may still contaminate *bona fide* cell-type clusters. We trained a classifier using our

366 manually curated cell-type classifications with a L2-penalized multinomial logistic regression
367 framework, as implemented in the *Scikit-learn* Python package (v0.22) [48]. We read the raw gene
368 expression matrices into python using scanpy (v1.4.2). We removed 2,582 genes that were
369 expressed in less than 10 cells. The gene expression levels of each cell were corrected so that
370 the total gene expression counts added up to 10,000. Per gene, these corrected counts were
371 normalized using a log(1+x) transformation. To speed up the computation of the multinomial
372 logistic regression, we only used the 2,037 genes with a mean expression between 0.0125 and
373 3, and a minimum dispersion of 0.5. We scaled the gene expression matrix so that the
374 expression level of each gene across cells had a mean of 0 and a variance of 1, after scaling
375 expression values over 20 were set to 20. We fit a multinomial logistic regression model using
376 the scaled gene expression values for the 2,037 highly variable genes from the complete set of
377 69,020 cells to obtain an estimate for the inverse regularization strength (C). Using the estimated
378 C of 7.74×10^{-04}, we performed 5 fold cross validation to estimate the probability that each cell
379 belongs to one of the manually curated cell-type classifications. Any cell with a probability higher
380 than 0.2 of belonging to 2 or more cell types (9,198) was classified as a doublet. Any cell which
381 did not belong to a cell-type with probability >= 0.4 (5,547) was classified as low quality. In total,
382 we removed 11,398 cells that were classified as a doublet or low quality. We removed an
383 additional 2,114 cells classified as Neurons as described in the section “Neuronal cell-type
384 classification”. For the remaining final list of 55,508 cells, we used the output of the classifier as
385 the final cell-type classification. The final classification is shown in Figure 1. For display purposes
386 the plot in Figure 1 was generated by rerunning umap on the finalized dataset with *euclidean*
387 distance metric and *umap.min_dist*=0.5, resulting in a more compressed visualization of the
388 dataset.

389

390 **Estimating the number of unique genotypes**

391 We note that calculating the number of expected unique genotypes is akin to the well-
392 known "Birthday problem" in statistics. Given C cells sampled from I individuals, the expected
393 number of cells with a unique genotype is $C(1-1/I)^{C-1}$. Assuming 50%-90% of worms were
394 successfully dissociated (a conservative range), we expect 31,134-40,257 unique genotypes.

395

396 **Single-nucleotide variant counting**

397 We used a list of single-nucleotide variants (SNVs) we previously curated for CB4856 compared
398 to the N2 reference [17]. We derived genotype informative UMI counts for N2 and CB4856 variants
399 using *Vartrix version 1.0* (https://github.com/10XGenomics/vartrix) directly on the output of
400 *CellRanger*. To reduce SNV counts that result from SNVs in the ambient RNA background, we
401 only kept SNVs that resided in genes with positive counts in the background corrected matrix.

402

403 **Genotype inference using a hidden Markov model**

404 We set up a hidden Markov model (HMM) to infer the genotypes of the recombinant progeny
405 [49,50]. The HMM is used to calculate the probability of underlying genotypes for each individual
406 and requires three components: (1) prior probabilities for each of the possible genotypes, (2)
407 emission probabilities for observing variant informative reads given each of the possible
408 genotypes, (3) and transition probabilities - the probabilities of recombination occurring between
409 adjacent genotype informative sites.

410 For the autosomal chromosomes we defined prior genotype probabilities as 0.25 for
411 homozygote N2, 0.5 for heterozygote CB4856/N2, and 0.25 for homozygote CB4856. For the

412 sex chromosome we defined prior genotype probabilities as 0.44 for homozygotes N2, 0.44 for
413 the heterozygote CB4856/N2, and 0.11 for homozygote CB4856. These values were chosen
414 because to generate the segregant population, N2 hermaphrodites were crossed to CB4856
415 males, and thus contributed twice as many X chromosomes to the progeny as CB4856.

416 Emission probabilities were calculated as previously described for low coverage
417 sequencing data [51,52] under the assumption that the observed counts of reads for both possible
418 variants (Y) at a genotype informative site (g) arise from a random binomial sampling of the
419 alleles present at that site and that sequencing errors (e) occur independently between reads at
420 a rate of 0.002:

421

422 $p(Y|g = NN) = \binom{D}{r}(1-e)^r(1-(1-e))^{D-r}$

423 $p(Y|g = NC) = \binom{D}{r}\frac{1}{2}^D$

424 $p(Y|g = CC) = \binom{D}{r}(e)^r(1-e)^{D-r}$

425

426 Where (D) is the total read depth at a genotype informative site for a given individual, (r) is the
427 total read depth for the N2 variant at that site, and N represents the N2 variant and C represents
428 the CB4856 variant.

429 Transition probabilities were derived from an existing N2 x CB4856 genetic map [53]. We
430 linearly interpolated genetic map distances from the existing map to all genotype informative
431 sites in our cross progeny. We scaled these genetic map distances, multiplying them by a factor
432 of 0.4, to account for the fact that the previous genetic map was built using ten generations of
433 intercrossing whereas progeny from our cross are derived from four generations of intercrossing
434 [54].

For QTL mapping we used an additive coding, summing the probability that the genotype was homozygote CB4856 with one half the probability that the genotype was heterozygote N2/CB4856.

eQTL mapping

Genotype probabilities were standardized, and markers in very high LD (r>.9999) were pruned. This LD pruning is approximately equivalent to using markers spaced 5 centimorgans (cm) apart. For each transcript, we counted the number of cells for which at least one UMI count was detected in each cell type. Transcripts with non-zero counts in at least 20 cells in a cell type were considered expressed in that cell type and used for downstream analyses.

As has been previously described for droplet scRNA-seq, counts of UMIs can be adequately parameterized by a gamma-poisson distribution, which is also known as the negative binomial distribution [55]. Thus we used a negative binomial regression framework for eQTL mapping here. We also note that simpler approaches using log(counts+1) with ordinary least squares behave pathologically, especially in regard to behavior with multiple partially correlated covariates, and simulations (not shown) showed such models lead to inflated false positive rates.

For each expressed transcript in each cell type we first fit the negative binomial generalized linear model:

$$E[Y] = \mu \quad (1)$$

$$Var(Y) = \mu + \frac{1}{\theta}\mu^2 \quad (2)$$

$$\mu = exp(\beta_i + X_t\beta_t + \mathbf{X_b}\beta_b + X_c\beta_c) \quad (3)$$

Which has the following log-likelihood:

$$\ell(\beta,\theta) = -\sum_{n=1}^{N}[(y_n+\theta)log(\mu_n+\theta) - y_n log(\mu_n) + log(|\ \Gamma(y_n+1)\ |) - \theta log(\theta) + log(|\ \Gamma(\theta)\ |) - log(|\ \Gamma(\theta+y_n)\ |)] \quad (4)$$

And where Y is a vector of UMI counts per cell, X_t is a vector of the log(total UMIs per cell) and controls for compositional effects, $\mathbf{X_b}$ is an indicator matrix assigning cells to batches, and X_c is the vector of standardized genotype probabilities across cells for the closest genotypic marker to each transcript from the pruned marker set. In addition, β is a vector of estimated coefficients from the model, μ_n is the expected value of Y for a given cell n, N is the total number of cells in the given cell type, and θ is a negative binomial overdispersion parameter. Model parameters were estimated using iteratively-reweighted least squares as implemented in the *negbin.reg* function in the *Rfast2* R package. If the model did not converge, model parameters were estimated with the *gam* function in the *mgcv* package [56], which opts for certainty of convergence over speed. We note that due to the computational burden of fitting so many GLMs in the context of sc-eQTL mapping, we chose to estimate θ once for each transcript in each cell type for this model, and use that estimate of θ in the additional models for that transcript within the cell type, as described below. This approach is conservative, as the effects of unmodeled factors (for example *trans* eQTLs) will be absorbed into the estimate of overdispersion, resulting in larger estimated overdispersion $(\frac{1}{\theta})$ and lower model likelihoods. Computational approaches that re-estimate θ for each model, that jointly model all additive genetic effects, or that regularize θ across models and transcripts [57], may further increase statistical power to identify linkages.

479 To evaluate the statistical significance of *cis* eQTLs, a likelihood ratio statistic
480 $-2(\ell_{nc} - \ell_{fc})$, was calculated comparing the log-likelihood of this model described above (ℓ_{fc})
481 to the log-likelihood of the model where β is re-estimated while leaving out the covariate X_c for
482 the *cis* eQTL marker (ℓ_{nc}). A p-value was derived under the assumption that this statistic is χ^2
483 distributed with one degree of freedom. This p-value was used for the evaluation of significance
484 of *cis* eQTLs for the neuronal subtypes. Within each neuronal subtype, FDR adjusted p-values
485 were calculated using the method of Benjamini and Hochberg [58]. For the other cell types (with
486 typically much larger cell numbers) and for the genome-wide scans for eQTLs, a permutation
487 procedure was used to calculate FDR adjusted p-values, and is described further below.
488 For each expressed transcript in each cell type we also scanned the entire genome for
489 eQTLs, enabling detection of *trans* eQTLs. A similar procedure was used as for *cis* eQTL except
490 that equation (3) was replaced with:
491
492 $$\mu = exp(\beta_i + X_t\beta_t + \mathbf{X_b}\beta_b + X_g\beta_g) \quad (5)$$
493
494 where X_g is a vector of the scaled genotype probabilities at the gth genotypic marker, and the
495 model is fit separately, one at a time, for each marker across the genome for each transcript. A
496 likelihood ratio statistic for each transcript, within each cell type, for each genotypic marker is
497 calculated by comparing this model to the model where β is re-estimated while leaving out the
498 covariate X_g. The likelihood ratio statistic was transformed into a LOD score, by dividing it by
499 $2log_e(10)$. We also used functions in the *fastglm* R package for this scan, again re-using
500 estimates of θ obtained as described above for each transcript for each cell type. For each
501 transcript and each chromosome, QTL peak markers were identified as the marker with the

502 highest LOD score. The 1.5 LOD-drop procedure was used to define approximate 95%
503 confidence intervals for QTL peaks [59].

504 FDR-adjusted p-values were calculated for QTL peaks. They were calculated as the ratio
505 of the number of transcripts expected by chance to show a maximum LOD score greater than a
506 particular LOD threshold vs. the number of transcripts observed in the real data with a maximum
507 LOD score greater than that threshold, for a series of LOD thresholds ranging from 0.1 to 0.1+the
508 maximum observed LOD for all transcripts within a cell type, with equal-sized steps of 0.01. Per
509 chromosome, the number of transcripts expected by chance at a given threshold was calculated
510 by permuting the assignments of segregant identity within each batch relative to segregant
511 genotypes, calculating LOD scores for all transcripts across the chromosome as described
512 above, and recording the maximum LOD score for each transcript. In each permutation instance,
513 the permutation ordering was the same across all transcripts. We repeated this permutation
514 procedure 10 times. Then, for each of the LOD thresholds, we calculated the average number
515 of transcripts with maximum LOD greater than the given threshold across the 10 permutations.
516 We used the *approxfun* function in R to interpolate the mapping between LOD thresholds and
517 FDR and estimate an FDR-adjusted p-value for each QTL peak [60].

518 The same procedure was performed for *cis* eQTL analysis, with the difference being that
519 the expected and observed number of transcripts at a given LOD threshold were calculated only
520 at the marker closest to the transcript. We note that, as expected, Benjamini and Hochberg
521 adjusted p-values, and FDR adjusted p-values from this permutation procedure for *cis* eQTLs
522 were nearly identical.

523

524 **scRNA-seq in the parental strains**

525 The parental QX2314 and PTM299 strains were grown separately for 4 generations on 10cm
526 plates, with recurrent cycles of bleaching and synchronization as was done for the segregant
527 population. For single cell preparation, synchronized L1 from both strains were seeded together
528 in equal numbers on 10cm plates, and they were processed together from that point onwards,
529 to limit any environmental effects. We believe that differences in efficiency of the cell preparation
530 procedure between N2 and CB4856 could explain the imbalanced representation in the final
531 dataset (6,721 N2 and 3,104 CB4856 cells). We took advantage of the different parental
532 genotypes when processing the cells, and called cells as those with at least 50 SNV counts
533 supporting one genotype, and less than 50 supporting the other. Cell-type identification was
534 automated by using the logistic regression model trained on the segregants which is discussed
535 above. Differential expression analysis was carried out using the *DEsingle* R package in each
536 cell type, as well as globally in all cells combined [61].

537 To compare differential expression results with our *cis* eQTL results, we first normalized
538 the effect size of each *cis* eQTL by its standard error. Those were used directly in comparisons
539 done within each cell type. To compare with global differential expression, those standardized
540 effects were combined across all cell types in which an eQTL was identified using Stouffer's
541 weighted-Z method [62].

542

543 **Processing whole-worm eQTL data**

544 Microarray genotype and gene expression data for our published expression QTL data were
545 acquired from the gene expression omnibus (GEO) [24]. Probe sequences were realigned to the
546 WBcel235 transcriptome using BWA, and uniquely mapping probes were used. Expression
547 probes that were present in less than 2/3 of the sample were removed. The genotype and

548 expression matrices were standardized. To map eQTLs, we calculated the Pearson correlation
549 between each probe and every genotype. Correlation coefficients were transformed to LOD
550 scores using $-n \cdot \frac{ln(1-R^2)}{2ln(10)}$. To assess significance and account for multiple testing, we
551 permuted the sample identities 100 times and calculated the average number of transcripts with
552 an identified eQTL at different LOD scores. We compared these results to the unpermuted LOD
553 scores to estimate the false-discovery rate (FDR), and selected a cutoff corresponding to a rate
554 of 10% (LOD = 4.2), equivalent to the single-cell mapping. *cis* eQTLs were derived by calculating
555 the Pearson correlation between transcript expression and the normalized genotypes in the
556 variant nearest to a given transcript, transforming to LOD score and comparing against the global
557 threshold.
558
559 **Hotspot analysis**
560 To discover hotspots, we split the genome into 130 bins of 5 centimorgans each. We then
561 counted the number of eQTLs in each bin identified in each cell type (applying a 10% FDR
562 significance threshold), after removing all *cis* linkages. *Cis* linkages were defined here as those
563 where the transcribed gene falls within the 95% eQTL confidence-interval range extended by
564 1MB on both sides. A bin was considered to have an excess of linkages if the number of linkages
565 exceeded the number expected by chance from a Poisson distribution, given the average
566 number of linkages per bin for that cell type and a Bonferroni correction for the total number of
567 bins ($p<3.8e-4$) [63]. The *findpeaks* function in the *pracma* R package was used to identify peak
568 hotspot bins and prevent identifying sets of adjacent bins as hotspots.

569 For gene ontology (GO) analysis, we identified hotspot targets using the same procedure
570 above, but relaxed the significance threshold to 20% FDR. We then used the R package *topGO*
571 to identify enriched terms, with the genes expressed in the cell type used as background.
572

573 **Neuronal cell-type classification**

574 Neuronal classification was carried out using a combination of available *C. elegans* scRNA-seq
575 datasets, including a published L2 dataset [21,22], and the *C. elegans* Neuronal Gene Expression
576 Map & Network (CeNGEN) project [31]. Neuronal cells were processed separately using *monocle3*
577 with default parameters, with the exception that 100 dimensions were specified for the
578 *preprocess_cds* step. The analysis was carried out in two passes. In the first pass, we processed
579 all cells identified by our classifier as neurons. Following Leiden clustering, we removed 2,114
580 cells that were in clusters whose top genes were mostly mitochondrial and ribosomal genes,
581 similar to the analysis described above for the global dataset. We then processed the pruned
582 dataset in *monocle3* as described above. To annotate the final neuronal clusters, we first used
583 the list of marker genes from two previous publications [22,30], to derive candidate clusters that
584 uniquely express marker genes. We next used the *top_markers* function in *monocle3* to identify
585 upregulated genes in each cluster compared to the rest. These were compared with the data
586 available in the online SCeNGEN *Shiny* application (https://cengen.shinyapps.io/SCeNGEA) for
587 the candidate cluster. The full list of genes used for classification is found in Table S5. In the
588 final dataset, clusters Unknown_1 - Unknown_4 are of unknown identity and do not correspond
589 to the clusters of the same name in CeNGEN, while the clusters Unknown_touch and
590 Unknown_glut_2 do correspond to cell clusters of the same names in CeNGEN.
591

592 **Single neuronal subtype eQTL (sn-eQTL) analysis**

593 sn-eQTL mapping is described above (section "eQTL mapping"). GO annotation of genes with
594 sn-eQTL was done in *topGO*, with the genes expressed in neurons (determined using the criteria
595 for inclusion in eQTL mapping) used as background. A heatmap was plotted using the
596 *ComplexHeatmap* package [64]. To determine the consistency in effect direction between the sn-
597 eQTL neuron and the rest of the neurons, we repeated the eQTL mapping, aggregating cells
598 from all neuronal cell-types but omitting the neuron with the sn-eQTL. The RIC and RIM neurons
599 in the parental datasets were identified using the same gene markers used in the segregant
600 eQTL dataset, as described in Table S5.

601

602 **Acknowledgments**

603 The work was supported by the Howard Hughes Medical Institute and NIH grants R01
604 HG004321 (L.K.) and K99 HG010369 (E.B-D). Figure 1A was created with Biorender.com. We
605 thank Michael Mashock, Marco De Simone, and other members of the Technology Center for
606 Genomics & Bioinformatics (TCGB) at UCLA for preparing and sequencing 10x 3' chromium
607 single-cell libraries.

608

609 **Author contributions**

610 Conceptualization: E.B-D, J.B., J.S.B, and L.K.; Methodology: E.B-D, J.B. and J.S.B.;
611 Investigation: E.B-D, J.B., L.G., S.Z. and J.S.B.; Formal analysis: E.B-D, J.B., and J.S.B; Writing
612 – original draft: E.B-D and L.K.; Supervision: J.S.B and L.K; Funding acquisition: E.B-D. and L.K.
613 All authors reviewed the manuscript.

614

615 **Data availability**

616 Raw sequencing data are available under NCBI bioproject PRJNA658829.

617

618 **Code availability**

619 Code for the HMM and eQTL mapping can be found at
620 https://github.com/joshsbloom/single_cell_eQTL. Data and code to recreate the figures in the
621 manuscript can be found at https://github.com/eyalbenda/worm_sceQTL.

622

623 **Competing interests**

624 The authors declare no competing interests

625

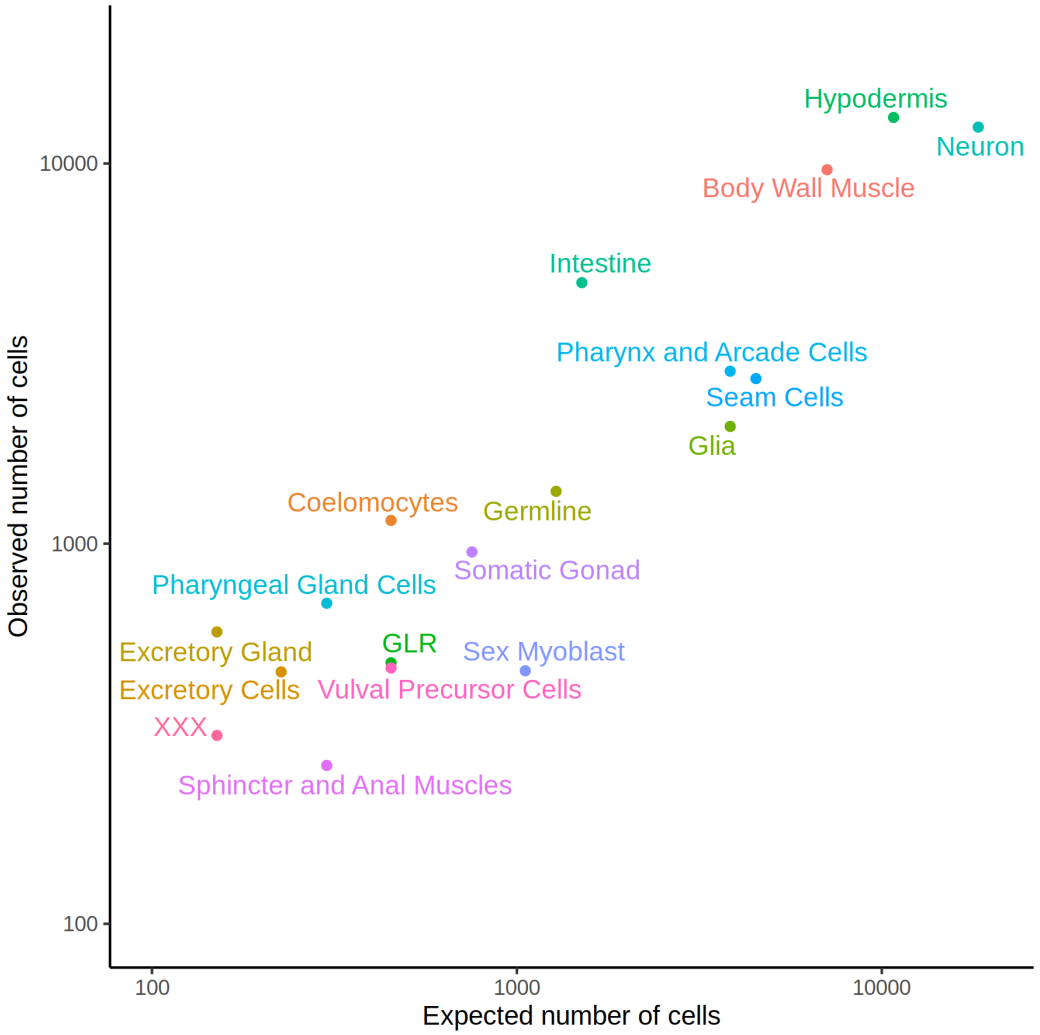
626 **Materials & Correspondence**

627 Should be addressed to E.B-D, JS.B or L.K

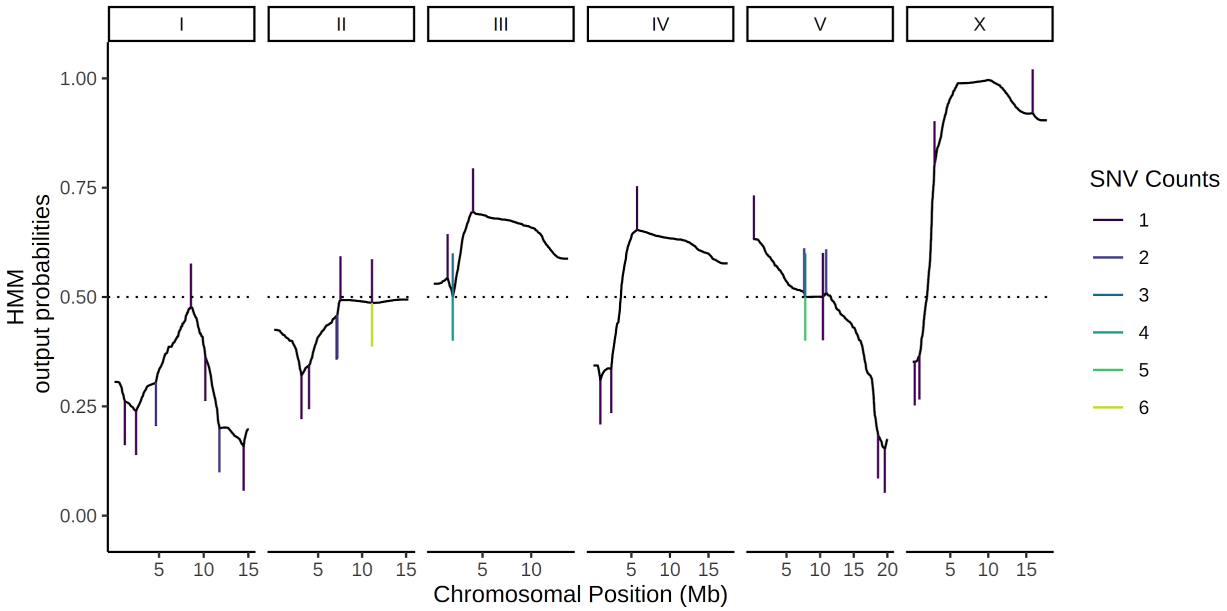
628

629

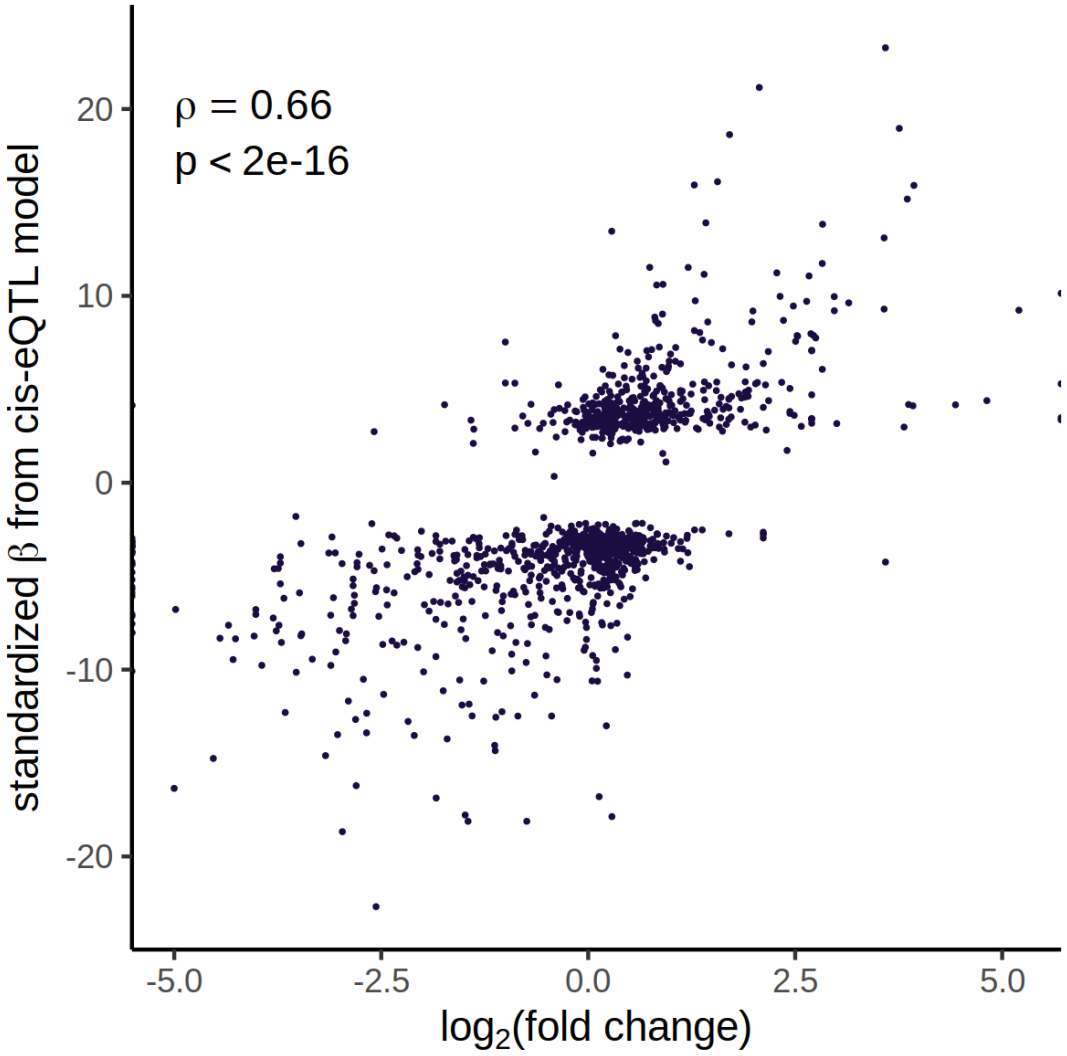
Supplementary Figures



Supplementary Figure 1. Observed representation of cell-types in our dataset compared to expected. The expected number of cells was calculated by manually curating the cellular lineage information available at https://www.wormatlas.org/ for the L2 stage.



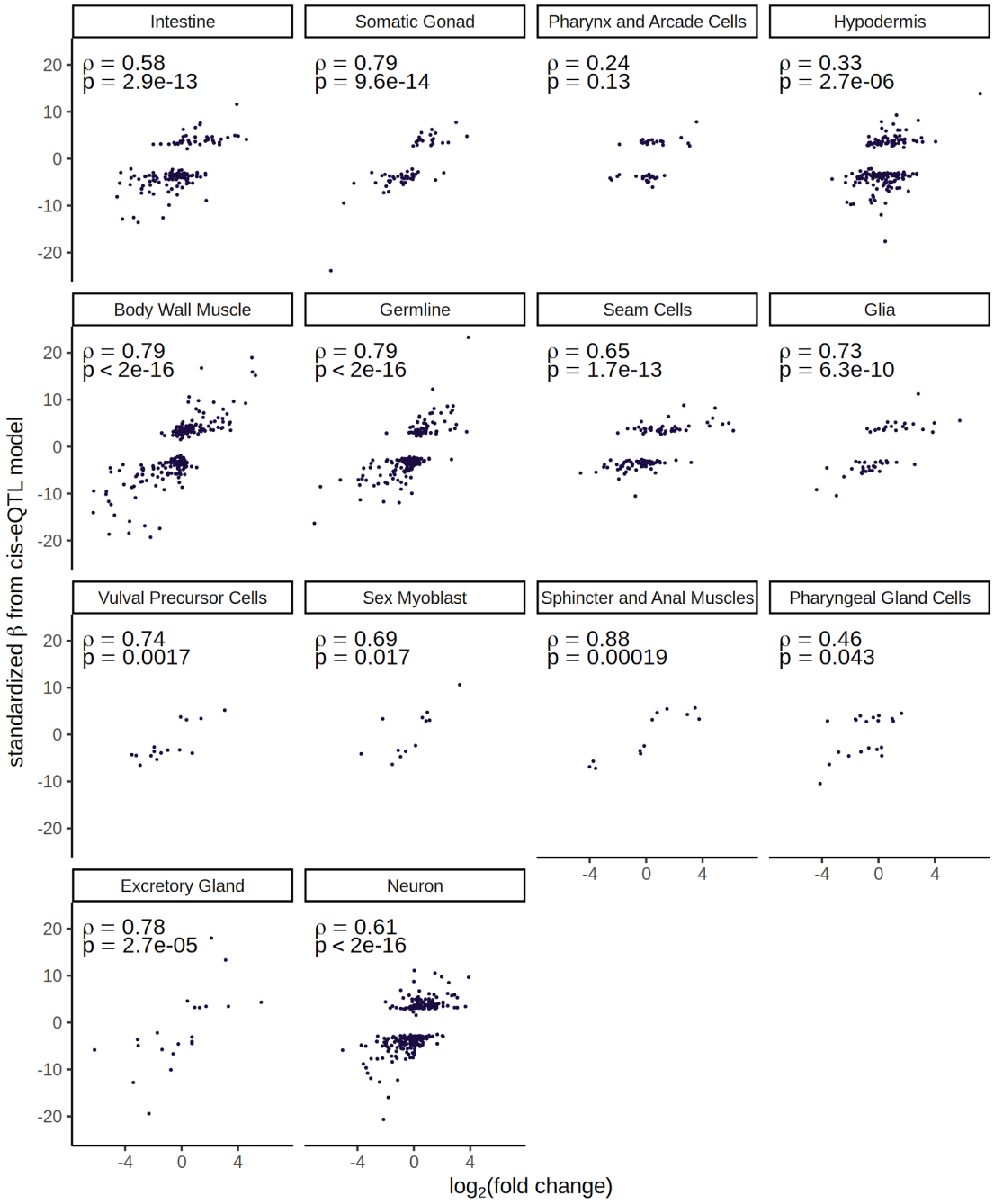
Supplementary Figure 2. Probabilistic genotyping using a hidden markov model (HMM). A cell with the median (69) number of unique genotype-informative SNV UMI counts is shown for illustration. The trace is a summation of the probability of a CB4856 homozygous genotype and half the probability of CB4856 heterozygous genotype at each position. Each vertical line is a count for an SNV, and colors correspond to the count depth. Vertical lines pointing upwards denote counts supporting the CB4856 variant, while lines pointing downwards are counts supporting the N2 variant.



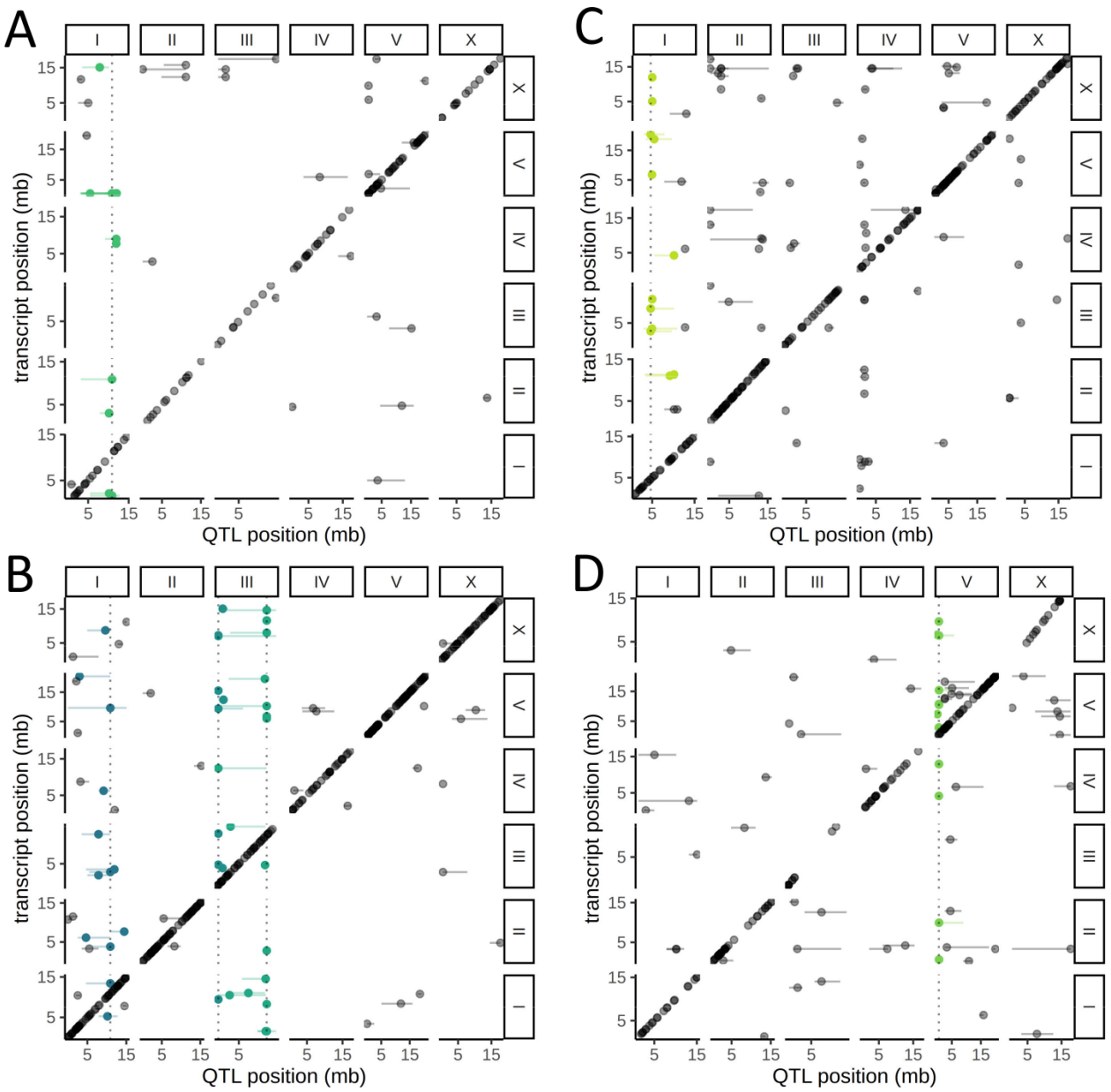
653

654 **Supplementary Figure 3. *cis* eQTLs reflect gene expression differences in the parent**
655 **strains.** Comparison between *cis* eQTLs mapped across all cell types and gene expression
656 differences in a dataset of 6,721 N2 and 3,104 CB4856 cells. For *cis* eQTLs mapped in multiple
657 cell-types, effect sizes were combined using Stouffer's weighted-Z method [62]. Differential gene
658 expression between N2 and CB4856 was calculated using R package DEsingle .

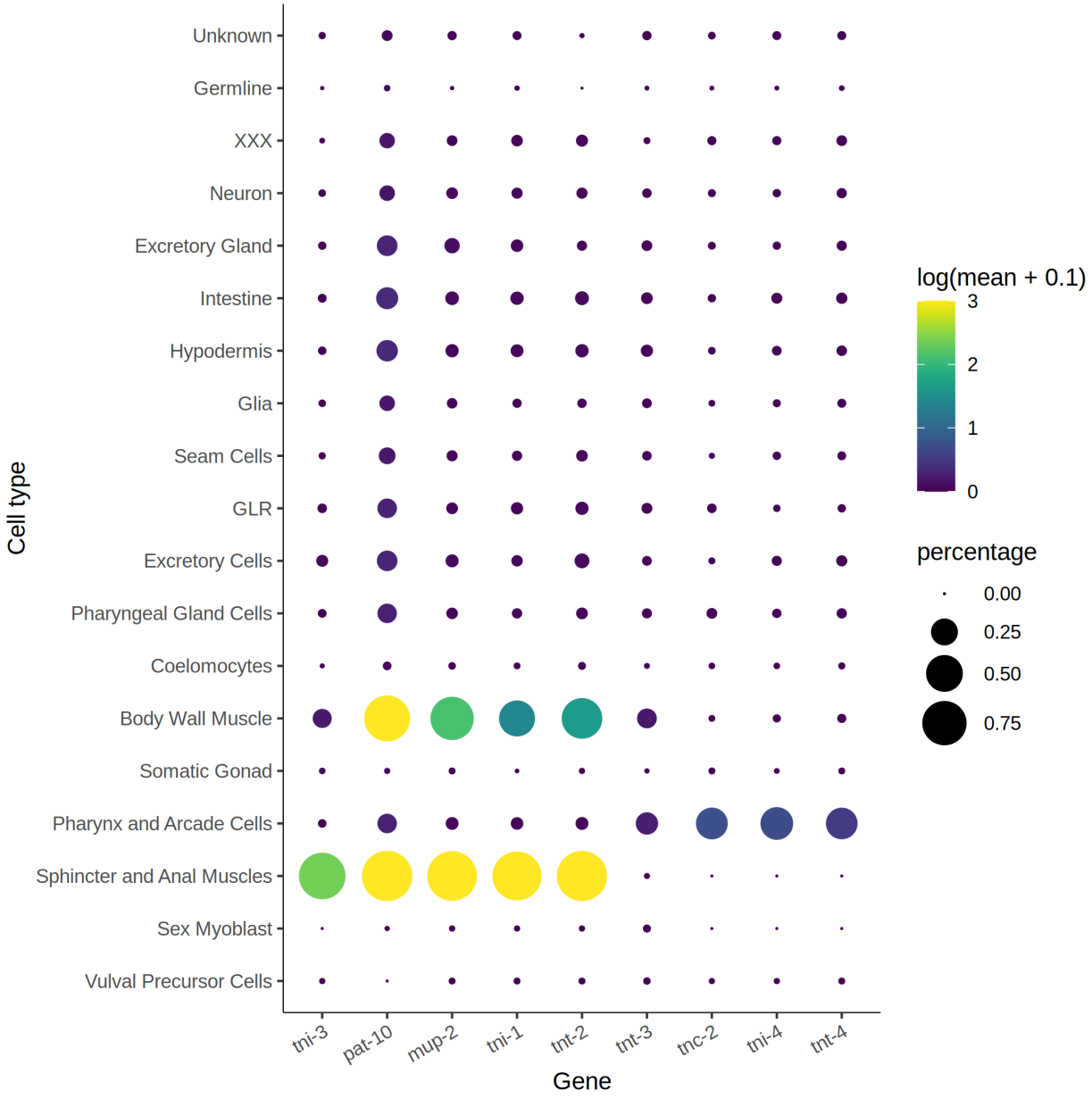
659



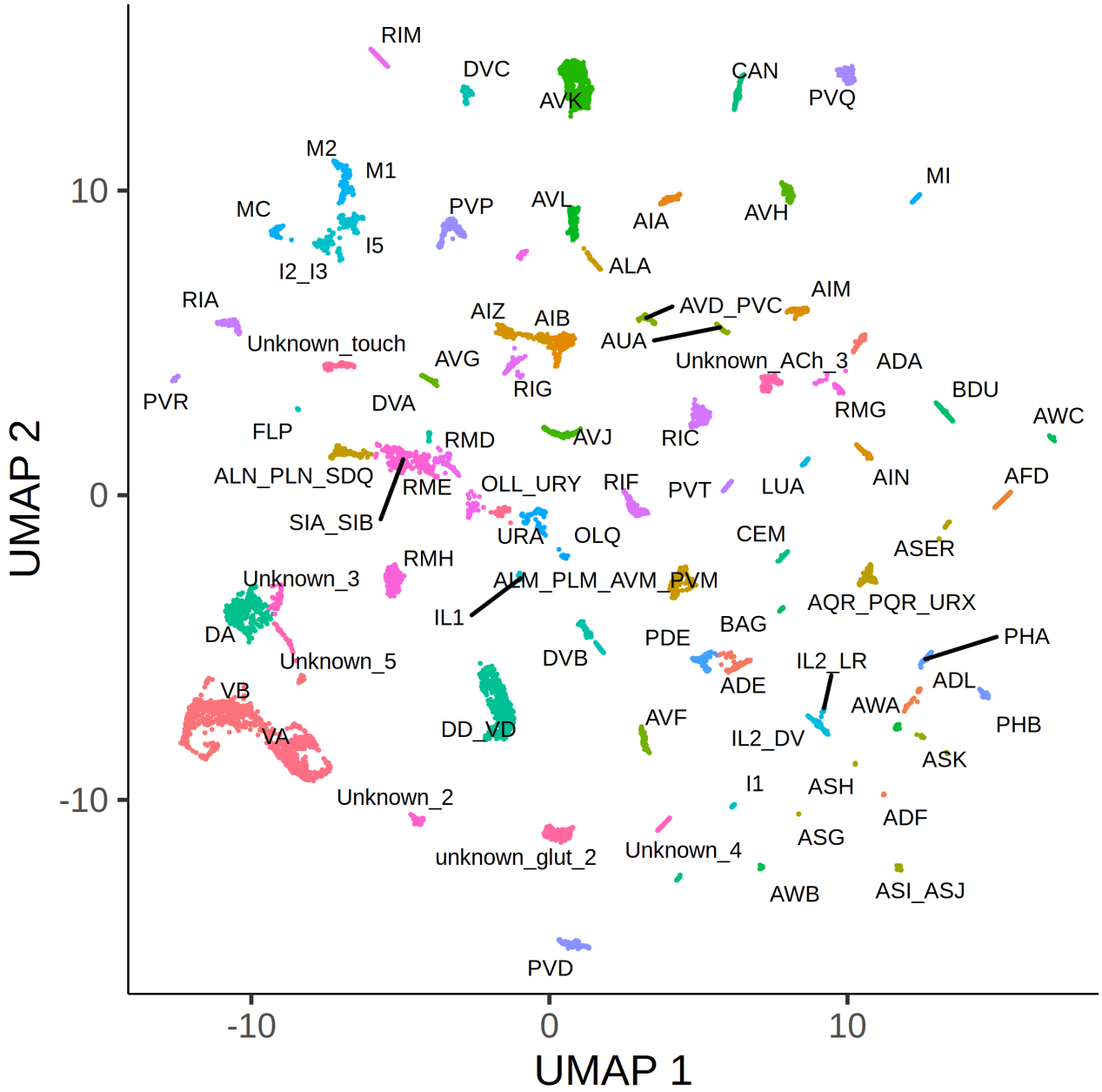
Supplementary Figure 4. Correlation between cell-type *cis*-eQTL signal and differential gene expression between the parental strains. The data corresponds to Figure S3, split out by each cell type separately.



Supplementary Figure 5. cell-type-specific *trans*-eQTL Hotspots. A genome-wide map of eQTLs in seam cells (A), neurons (B), body-wall muscle cells (C) and intestinal cells (D) is shown. The position of the eQTLs is shown on the x-axis, while the y-axis shows the position of the associated transcripts. The dotted line marks the peak position of the hotspot, while targets of each hotspot are colored.



Supplementary Figure 6. Cell-type expression of genes that form troponin complexes. Size of circles corresponds to the percentage of cells expressing each gene in each cell-type, and the color corresponds to average expression. Of the four troponin genes strongly expressed in the body wall muscle, three are affected by a *trans*-eQTL hotspot on Chr. I.



Supplementary Figure 7. UMAP projection of 12,468 neurons. Each cluster is labeled based on the neuronal identity. Clusters represent either single neurons, or few neurons with shared function.

684 **References**

685 1. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev.*
686 *Genet.* **16**, 197–212 (2015).
687 2. Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11
688 common diseases. *Am. J. Hum. Genet.* **95**, 535–552 (2014).
689 3. Hormozdiari, F. *et al.* Leveraging molecular quantitative trait loci to understand the genetic
690 architecture of diseases and complex traits. *Nat. Genet.* **50**, 1041–1047 (2018).
691 4. Raj, T. *et al.* Polarization of the Effects of Autoimmune and Neurodegenerative Risk Alleles in
692 Leukocytes. *Science* **344**, 519–523 (2014).
693 5. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type–specific
694 master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012).
695 6. Ishigaki, K. *et al.* Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid
696 arthritis. *Nat. Genet.* **49**, 1120–1125 (2017).
697 7. Donovan, M. K. R., D'Antonio-Chronowska, A., D'Antonio, M. & Frazer, K. A. Cellular deconvolution
698 of GTEx tissues powers discovery of disease and cell-type associated regulatory variants. *Nat.*
699 *Commun.* **11**, 955 (2020).
700 8. Kim-Hellmuth, S. *et al.* Cell type–specific genetic regulation of gene expression across human
701 tissues. *Science* **369**, (2020).
702 9. Yao, D. W., O'Connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease
703 mediated by assayed gene expression levels. *Nat. Genet.* **52**, 626–633 (2020).
704 10. Consortium, T. Gte. The GTEx Consortium atlas of genetic regulatory effects across human
705 tissues. *Science* **369**, 1318–1330 (2020).
706 11. van der Wijst, M. *et al.* The single-cell eQTLGen consortium. *eLife* **9**, e52155 (2020).
707 12. Cuomo, A. S. E. *et al.* Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic
708 genetic effects on gene expression. *Nat. Commun.* **11**, 1–14 (2020).

709 13. van der Wijst, M. G. P. *et al.* Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and
710 co-expression QTLs. *Nat. Genet.* **50**, 493–497 (2018).
711 14. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, Caenorhabditis
712 elegans. *Dev. Biol.* **56**, 110–156 (1977).
713 15. Hall, D. H. & Altun, Z. F. *C. elegans atlas*. (Cold Spring Harbor Laboratory Press, 2007).
714 16. Hobert, O., Glenwinkel, L. & White, J. Revisiting Neuronal Cell Type Classification in
715 Caenorhabditis elegans. *Curr. Biol. CB* **26**, R1197–R1203 (2016).
716 17. Burga, A., Ben-David, E., Vergara, T. L., Boocock, J. & Kruglyak, L. Fast genetic mapping of
717 complex traits in C. elegans using millions of individuals in bulk. *Nat. Commun.* **10**, 2680 (2019).
718 18. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for
719 Dimension Reduction. *ArXiv180203426 Cs Stat* (2018).
720 19. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected
721 communities. *Sci. Rep.* **9**, 1–12 (2019).
722 20. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by
723 pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
724 21. Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*
725 **357**, 661–667 (2017).
726 22. Packer, J. S. *et al.* A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell
727 resolution. *Science* **365**, (2019).
728 23. Mandric, I. *et al.* Optimized design of single-cell RNA sequencing experiments for cell-type-specific
729 eQTL analysis. *Nat. Commun.* **11**, 5504 (2020).
730 24. Rockman, M. V., Skrovanek, S. S. & Kruglyak, L. Selection at linked sites shapes heritable
731 phenotypic variation in C. elegans. *Science* **330**, 372–376 (2010).
732 25. Pukkila-Worley, R. & Ausubel, F. M. Immune defense mechanisms in the Caenorhabditis elegans
733 intestinal epithelium. *Curr. Opin. Immunol.* **24**, 3–9 (2012).

734 26. Ono, K. & Ono, S. Tropomyosin and Troponin Are Required for Ovarian Contraction in the
735 Caenorhabditis elegans Reproductive System. *Mol. Biol. Cell* **15**, 2782–2793 (2004).
736 27. Gumienny, T. L. TGF-β signaling in C. elegans. *WormBook* 1–34 (2013)
737 doi:10.1895/wormbook.1.22.2.
738 28. Suzuki, Y. *et al.* A BMP homolog acts as a dose-dependent regulator of body size and male tail
739 patterning in Caenorhabditis elegans. *Dev. Camb. Engl.* **126**, 241–250 (1999).
740 29. Andersen, E. C. *et al.* A Powerful New Quantitative Genetics Platform, Combining Caenorhabditis
741 elegans High-Throughput Fitness Assays with a Large Collection of Recombinant Strains. *G3*
742 *Genes Genomes Genet.* **5**, 911–920 (2015).
743 30. Hammarlund, M., Hobert, O., Miller, D. M. & Sestan, N. The CeNGEN Project: The Complete Gene
744 Expression Map of an Entire Nervous System. *Neuron* **99**, 430–433 (2018).
745 31. Taylor, S. R. *et al.* Expression profiling of the mature C. elegans nervous system by single-cell
746 RNA-Sequencing. *bioRxiv* 737577 (2019) doi:10.1101/737577.
747 32. Portman, D. S. & Emmons, S. W. Identification of C. elegans sensory ray genes using whole-
748 genome expression profiling. *Dev. Biol.* **270**, 499–512 (2004).
749 33. Sarkar, A. K. *et al.* Discovery and characterization of variance QTLs in human induced pluripotent
750 stem cells. *PLOS Genet.* **15**, e1008045 (2019).
751 34. Francesconi, M. & Lehner, B. The effects of genetic variation on gene expression dynamics during
752 development. *Nature* **505**, 208–211 (2014).
753 35. Yao, C. *et al.* Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits.
754 *Am. J. Hum. Genet.* **100**, 571–580 (2017).
755 36. Kolberg, L., Kerimov, N., Peterson, H. & Alasoo, K. Co-expression analysis reveals interpretable
756 gene modules controlled by trans-acting genetic variants. *eLife* **9**, e58705 (2020).
757 37. Brynedal, B. *et al.* Large-Scale trans-eQTLs Affect Hundreds of Transcripts and Mediate Patterns of
758 Transcriptional Co-regulation. *Am. J. Hum. Genet.* **100**, 581–591 (2017).

759 38. Smith, E. N. & Kruglyak, L. Gene–Environment Interaction in Yeast Gene Expression. *PLOS Biol.* **6**,
760 e83 (2008).
761 39. Zhang, T. *et al.* Cell-type–specific eQTL of primary melanocytes facilitates identification of
762 melanoma susceptibility genes. *Genome Res.* **28**, 1621–1635 (2018).
763 40. Westra, H.-J. *et al.* Cell Specific eQTL Analysis without Sorting Cells. *PLOS Genet.* **11**, e1005223
764 (2015).
765 41. Orozco, L. D. *et al.* Unraveling Inflammatory Responses using Systems Genetics and Gene-
766 Environment Interactions in Macrophages. *Cell* **151**, 658–670 (2012).
767 42. Fairfax, B. P. & Knight, J. C. Genetics of gene expression in immunity to infection. *Curr. Opin.*
768 *Immunol.* **30**, 63–71 (2014).
769 43. Kim-Hellmuth, S. *et al.* Genetic regulatory effects modified by immune activation contribute to
770 autoimmune disease associations. *Nat. Commun.* **8**, 266 (2017).
771 44. Zhang, S., Banerjee, D. & Kuhn, J. R. Isolation and Culture of Larval Cells from C. elegans. *PLOS*
772 *ONE* **6**, e19505 (2011).
773 45. Kaletsky, R. *et al.* The C. elegans adult neuronal IIS/FOXO transcriptome reveals adult phenotype
774 regulators. *Nature* **529**, 92–96 (2016).
775 46. Qiu, X. *et al.* Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods*
776 **14**, 309–315 (2017).
777 47. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet based single-
778 cell RNA sequencing data. *bioRxiv* 303727 (2020) doi:10.1101/303727.
779 48. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830
780 (2011).
781 49. Broman, K. W. The Genomes of Recombinant Inbred Lines. *Genetics* **169**, 1133–1146 (2005).
782 50. Arends, D., Prins, P., Jansen, R. C. & Broman, K. W. R/qtl: high-throughput multiple QTL mapping.
783 *Bioinforma. Oxf. Engl.* **26**, 2990–2992 (2010).

784 51. Dodds, K. G. *et al.* Construction of relatedness matrices using genotyping-by-sequencing data.
785 *BMC Genomics* **16**, 1047 (2015).
786 52. Bilton, T. P. *et al.* Accounting for Errors in Low Coverage High-Throughput Sequencing Data When
787 Constructing Genetic Maps Using Biparental Outcrossed Populations. *Genetics* **209**, 65–76 (2018).
788 53. Rockman, M. V. & Kruglyak, L. Recombinational Landscape and Population Genomics of
789 Caenorhabditis elegans. *PLOS Genet.* **5**, e1000419 (2009).
790 54. Rockman, M. V. & Kruglyak, L. Breeding Designs for Recombinant Inbred Advanced Intercross
791 Lines. *Genetics* **179**, 1069–1078 (2008).
792 55. Svensson, V. Droplet scRNA-seq is not zero-inflated. *Nat Biotechnol* **38**, 147–150 (2020).
793 56. Wood, S. N. *Generalized Additive Models: An Introduction with R*. (Chapman and Hall/CRC, 2017).
794 57. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq
795 experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288–4297 (2012).
796 58. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful
797 approach to multiple testing. *J R Stat Soc* (1995).
798 59. Dupuis, J. & Siegmund, D. Statistical methods for mapping quantitative trait loci from a dense set of
799 markers. *Genetics* **151**, 373–386 (1999).
800 60. Albert, F. W., Bloom, J. S., Siegel, J., Day, L. & Kruglyak, L. Genetics of trans-regulatory variation
801 in gene expression. *eLife* **7**, e35471 (2018).
802 61. Miao, Z., Deng, K., Wang, X. & Zhang, X. DEsingle for detecting three types of differential
803 expression in single-cell RNA-seq data. *Bioinformatics* **34**, 3223–3224 (2018).
804 62. Whitlock, M. C. Combining probability from independent tests: the weighted Z-method is superior to
805 Fisher's approach. *J. Evol. Biol.* **18**, 1368–1373 (2005).
806 63. Smith, E. N. & Kruglyak, L. Gene–Environment Interaction in Yeast Gene Expression. *PLOS Biol.* **6**,
807 e83 (2008).
808 64. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in
809 multidimensional genomic data. *Bioinforma. Oxf. Engl.* **32**, 2847–2849 (2016).

810