1   **Worm Perturb-Seq: massively parallel whole-animal RNAi and RNA-seq**

2

3   Hefei Zhang[1$], Xuhang Li[1$], Dongyuan Song[2], Onur Yukselen[3], Shivani Nanda[1#], Alper

4   Kucukural[3,4], Jingyi Jessica Li[2,5], Manuel Garber[4*], Albertha J.M. Walhout[1,*]

5

6   [1]Department of Systems Biology, University of Massachusetts Chan Medical School,

7   Worcester, MA, USA

8   [2]Bioinformatics Interdepartmental Ph.D. Program, University of California, Los Angeles,

9   CA, USA

10   [3]Via Scientific Inc. Cambridge, MA, USA

11   [4]Department of Genomics and Computational Biology, University of Massachusetts Chan

12   Medical School, Worcester, MA, USA

13   [5]Department of Statistics and Data Science, Department of Biostatistics, Department of

14   Computational Medicine, and Department of Human Genetics, University of California,

15   Los Angeles, CA, USA

16   # Current address: Pathology, Beth Israel Deaconess Medical Center, Harvard Medical

17   School, Boston, MA.

18

19   [$]These authors contributed equally

20

21   * Correspondence to: marian.walhout@umassmed.edu;

22   manuel.garber@umassmed.edu

23

24 **The transcriptome provides a highly informative molecular phenotype to connect**
25 **genotype to phenotype and is most frequently measured by RNA-sequencing**
26 **(RNA-seq). Therefore, an ultimate goal is to perturb every gene and measure**
27 **changes in the transcriptome. However, this remains challenging, especially in**
28 **intact organisms due to different experimental and computational challenges.**
29 **Here, we present 'Worm Perturb-Seq (WPS)', which provides high-resolution RNA-**
30 **seq profiles for hundreds of replicate perturbations at a time in a living animal.**
31 **WPS introduces multiple experimental advances that combine strengths of bulk**
32 **and single cell RNA-seq, and that further provides an analytical framework,**
33 **EmpirDE, that leverages the unique power of the large WPS datasets. EmpirDE**
34 **identifies differentially expressed genes (DEGs) by using gene-specific empirical**
35 **null distributions, rather than control conditions alone, thereby systematically**
36 **removing technical biases and improving statistical rigor. We applied WPS to 103**
37 ***Caenhorhabditis elegans* nuclear hormone receptors (NHRs) to delineate a Gene**
38 **Regulatory Network (GRN) and found that this GRN presents a striking 'pairwise**
39 **modularity' where pairs of NHRs regulate shared target genes. We envision that**
40 **the experimental and analytical advances of WPS should be useful not only for *C.***
41 ***elegans*, but will be broadly applicable to other models, including human cells.**

42

43 Since the dawn of functional genomics, the transcriptome has proven to be one of the
44 most powerful molecular phenotypes to connect genotype to phenotype[1-3]. While early
45 work in yeast provided insights into the transcriptional responses to gene deletions[4,5],
46 similar large-scale and systematic studies in multicellular organisms have been lacking.
47 Moreover, statistical analyses of large-scale, high-throughput genomics data suffer from
48 technical biases and high false discovery rates (FDRs)[6], e.g., many false positives in the
49 identification of differentially expressed genes (DEGs). More recently, a method
50 commonly referred to as Perturb-seq has developed which uses pooled CRISPR-based
51 gene perturbation screens with single cell RNA-seq. This method has proven powerful in
52 in cell-based functional screens to annotate gene function, identify genetic interactions,
53 and to infer disease-related pathways [7-13]. However, this type of single-cell based

54 approach is limited by the low sensitivity for gene detection, lack of biological replicate

55 experiments and high cost, and challenges to perturb many genes *in vivo* [3,14-16].

56   Here, we present 'Worm Perturb-Seq' (WPS) in which individual genes are knocked

57 down in the nematode *C. elegans* by feeding bacteria expressing double-stranded RNA,

58 followed by RNA-seq using a strategy that adopts the high multiplexity of single-cell

59 sequencing but uses bulk samples to produce high-resolution RNA-seq profiles. WPS is

60 labor- and cost-efficient and enables replicate experiments. Using large-scale WPS data,

61 we found that subtle but systematic fluctuations in gene expression caused pervasive

62 false positive DEGs when analyzed by standard differential expression (DE) analysis

63 methods that compare experiment to control conditions. To circumvent this issue, WPS

64 includes a two-pronged data analysis framework, EmpirDE ('Empirical Differential

65 Expression'), that leverages the large WPS dataset to produce DEGs. EmpirDE

66 systematically removes technical biases and uses gene-specific models with empirical

67 null distributions to correct for anti-conservative *P* values (i.e., where the significance is

68 overestimated) obtained by standard DE analysis. We demonstrate the rigorous control

69 of FDR in WPS data and analysis by simulations and experimental benchmarking. We

70 applied WPS to the knockdown of 103 nuclear hormone receptors (NHRs) and discovered

71 that NHR pairs frequently share overlapping target genes, which cannot be explained by

72 protein similarity, but is more related to NHR coexpression. WPS will enable examining

73 different perturbations in addition to RNAi, including mutants, bacterial diets, and

74 exposure to drugs or toxins. Importantly, EmpirDE will also broadly enable statistically

75 rigorous analyses of large-scale transcriptomic data (i.e., > 100 conditions) in other

76 systems.

77

78 **Results**

79 In *C. elegans*, gene expression can be knocked down by feeding the animals bacteria

80 expressing double-stranded RNA for a target gene of interest[17,18]. This whole-animal RNA

81 interference (RNAi) is easy to perform for large sets of genes in parallel, and multiple

82 RNAi libraries have been developed[19-23]. We developed WPS, which is composed of two

83 major components: (1) an experimental approach to perform high-throughput whole-

84 animal RNAi experiments and generate multiplexed RNA-seq libraries, and (2) a

85 computational pipeline for quality control and rigorous statistical analysis of DEGs (**Fig.**
86 **1a**).

87

An overview of Worm Perturb-Seq (WPS)

89 WPS consists of several steps, many of which were optimized to enable high-throughput,
90 cost-effective experiments (**Fig. 1b**). Briefly, RNAi is started with animals at the first larval
91 stage (L1) and, when grown to the desired stage, animals are harvested, and total RNA
92 is extracted in a 96-well extraction plate. This streamlined workflow allows efficient
93 triplicate experiments for hundreds of knockdowns (**Fig. 1a**). Multiplex RNA-seq libraries
94 are constructed using an early barcoding step during reverse transcription in the 96-well
95 plates, with each barcode linked to a single perturbation, followed by pooling of ~50
96 samples and sequencing library construction. After sequencing, several quality control
97 steps are performed (see below) and DEGs are identified with EmpirDE, which uses
98 gene-specific models with empirical null distributions.

99

An experimental WPS platform

101 Several experimental steps of WPS were developed and optimized, including growing
102 animals, harvesting RNA, and generating, pooling, and sequencing of multiplexed
103 libraries (**Fig. 1b**, **Supplementary Protocols** and **Supplementary Note 1**). Notably,
104 WPS introduces a high-throughput worm lysis method for RNA extraction in 96-well
105 plates, which does not lyse eggs, making it suitable to use WPS for gravid adults
106 (**Extended Data Fig. 1a**). We optimized the 3' end barcoding method CEL-Seq2[24], which
107 was originally developed for single-cell RNA-seq[25], for multiplexing bulk RNA-seq
108 libraries, with significant reduction of costly reagents (**Fig. 1b** and **Supplementary**
109 **Protocol**).

110   The transcriptome of *C. elegans* changes greatly over the course of its lifetime; there
111 are oscillatory expression profiles during development, and gene expression continues to
112 change as the animals reproduce and age[26-28]. Therefore, if a knockdown has even a
113 small effect on development, it can result in many differentially expressed genes that are
114 secondary to the effect of the knockdown on development, rather than in response to the
115 perturbed gene. We therefore opted to use a period in the animal's lifetime in which the

116 transcriptome does not change to minimize developmental effects of knockdowns.
117 Animals develop from L1 to gravid adults in ~58 hours at 20°C and we found that the
118 gravid adult transcriptome was most stable between 60 and 68 hours post-L1-plating (**Fig.**
119 **2a, Extended Data Fig. 1b**). Therefore, we used a time of ~63 hours post-plating, which
120 is in between the first egg laid at 58 hours and the first egg hatched at 68 hours, allowing
121 enough time for sample collection and processing, and providing a buffer for perturbations
122 that elicit a mild developmental day (**Fig. 2a**).

123   In other systems, it has been shown that most genes can be quantified with a relatively
124 shallow read depth[29,30]. We performed down-sampling analysis of a reference dataset
125 with a depth of 53 million reads (**Extended Data Fig. 1c** and **Supplementary Table 1**).
126 We used an average of 6 million reads per sample, with which 90% of genes with >4
127 transcripts per million (TPM) and 80% of genes with > 2 TPM could be quantified (**Fig.**
128 **2b**, **Extended Data Fig. 1d**). For library multiplexing, we pooled ~54 samples, which
129 included 16 perturbations, each containing three biological replicates, together with six
130 negative controls (empty vector RNAi) into one sequencing library. This design was
131 intended to minimize batch effects by having all replicates of controls and perturbations
132 in the same library. Here, we combined WPS datasets for 103 NHRs (discussed below)
133 with those for ~900 metabolic genes [REWIRING][WIRING] for benchmarking analysis.

134

135 **WPS quality control**
136 Analysis of large-scale and high-throughput functional genomics experiments can be
137 complicated by batch effects and low-quality samples[6,31,32]. We followed standard
138 practices to ensure the quality of individual samples[33,34], and to identify and remove outlier
139 replicates (**Fig. 2c, Extended Data Fig. 1e, Supplementary Methods**). We next
140 developed two RNAi quality control (QC) analyses to verify the gene that was knocked
141 down. First, the reduction of targeted gene expression can be directly read out. For
142 instance, for 85% genes that are expressed at a high level (TPM ≥ 30), we found a > 2-
143 fold reduction in their mRNA levels when knocked down (**Extended Data Fig. 1f**).
144 Second, due to abundant reverse strand reads that map to the gene body of knocked
145 down genes, the identity of the perturbed gene can be directly identified from the WPS
146 data (**Fig. 2d,** red reads). These reads are likely derived from unspecific reverse

147 transcription of the anti-sense RNA generated during the RNA interference process[35].

148 This latter RNAi identity verification is particularly useful for genes that are expressed at

149 low levels and was able to verify the identity of almost all NHR RNAi clones that were

150 also confirmed by Sanger sequencing (**Fig. 2e**). We next performed WPS using RNAi

151 clones that were not confirmed a priori [REWIRING] and found three incorrect clones

152 (**Fig. 2f, g**, indicated in red). Importantly, we could identify the actual target by mapping

153 anti-sense sequences to the *C. elegans* genome (**Fig. 2f,** non-diagonal signals for the red

154 RNAi conditions). Two clones that returned a hit in the search were subsequently

155 confirmed by Sanger sequencing (**Extended Data Fig. 1g**). The other incorrect clone had

156 a partial insert that did not target a transcribed gene and was considered a non-targeting

157 perturbation (NTP). In the metabolic-gene WPS screen [REWIRING] we found that ~13%

158 samples had a wrong RNAi identity, including 67 NTPs (**Supplementary Table 2**),

159 showing the necessity of RNAi QC in large-scale screens. Taken together, WPS data can

160 be directly used to validate that the gene that has been knocked down and clones that

161 are incorrect can simply be removed from the dataset or analyzed with the corrected

162 target information.

163

164 **Standard DE analysis results in high false discoveries**

165 The 67 NTPs from the metabolic-gene WPS study [REWIRING] should have zero

166 differentially expressed genes and can therefore be used to evaluate the actual FDR in

167 WPS. We initially conducted DE analysis with DESeq2[36], by comparing each RNAi

168 perturbation to vector controls from the same sequencing library. Surprisingly, this

169 approach resulted in dozens of DEGs in both NTP and four randomly spike-in vector

170 control conditions (**Fig. 3a**, P_{adj} < 0.01, fold change (FC) > 2, collectively referred to as

171 NTPs hereafter). This suggests a high level of false discoveries despite stringent filtering

172 by estimated FDR and FC thresholds.

173 False positive DEG calls are common in RNA-seq studies[37-40] and are potentially

174 more profound when combined with large-scale screens because of systematic

175 variations[38][6]. By comparing mRNA levels among vector control, NTP, and RNAi samples

176 within the same sequencing library and across dozens of libraries, we discovered two

177 confounding issues. The first issue involved sequencing libraries in which a gene

178   consistently behaved differently in the vector control samples compared to the RNAi

179   samples in the same sequencing library (**Fig. 3b**, 'control-outlier gene', **Extended Data**

180   **Fig. 2a**). For instance, *swt-3* expression was lower in all RNAi samples when compared

181   to the vector control in the same batch (**Fig. 3b**), resulting in *swt-3* being identified as a

182   DEG in all these conditions, including in the NTP. The second issue involved genes with

183   highly variable mRNA levels across the WPS dataset (**Fig. 3c**, see **Extended Data Fig.**

184   **2b** for the entire dataset**,** 'noisy genes') that were frequently called as DEGs in both RNAi

185   perturbations and NTPs by DESeq2. This suggests that DESeq2 underestimates gene

186   expression variances, leading to anti-conservative *P* values.

187

188   **EmpirDE: an empirical null-based, gene-centered method to rigorously analyze**

189   **differential expression**

190   A major challenge in transcriptomics is that successful DE analysis hinges on reasonable

191   model assumptions and accurate parameter estimation. However, it is impossible to

192   accurately estimate parameters, such as gene-specific variance, or to identify model

193   misspecification, in small-sample-size experiments[41-44]. To systematically identify and

194   correct for false positive DEGs, we developed EmpirDE, which leverages the power of

195   having hundreds of conditions assessed in a uniform experimental setup, enabling the

196   rigorous identification of true DEGs that are elicited by the specific knockdown. EmpirDE

197   uses a two-pronged approach, first at the level of individual sequencing libraries and

198   second at the level of an entire dataset (**Fig. 3d**). The first step performs DE analysis

199   within each sequencing library (usually with ~16 conditions) using DESeq2. However,

200   instead of simply comparing an RNAi condition to the control, this step identifies control-

201   outlier genes and treats these differently using a control-independent DE analysis

202   procedure (**Supplementary Note 2**). Briefly, for each control-outlier gene, this step

203   empirically identifies a control-independent null population based on the distribution of the

204   gene's expression levels in the sequencing library and compares the level for the gene in

205   each RNAi condition to the newly defined null population. The second step of EmpirDE

206   combines the DE results from all sequencing libraries (here ~1,000 triplicate conditions)

207   to correct the anti-conservative *P* values based on the gene-specific empirical null

208   distributions of the DE test statistic (i.e., Wald statistic). Dozens to hundreds of conditions

209 uniformly collected in WPS provided a unique opportunity to directly estimate the

210 empirical null from the data in a gene-centered manner[45]. Assuming real effects are rare

211 in large-scale experiments, the gene expression in most conditions can be viewed as

212 'unchanged', thus defining an empirical null population. By fitting the central peak of the

213 distribution of the test statistic (i.e., the distribution of conditions that did not have a

214 significant effect)[45], this step estimates the empirical null and rescales the original test

215 statistic (i.e., Wald statistic) accordingly to obtain a corrected Wald statistic on a gene-by-

216 gene basis. This corrected statistic should follow a standard normal distribution and can

217 be converted to a *P* value (referred to as empirical *P* value) to identify DEGs (**Fig. 3d**).

218   EmpirDE identified fewer than 200 control-outlier genes in most sequencing libraries

219 (**Fig. 3e**), indicating a relatively low but significant number of confounded genes (1.4% of

220 14,000 detected genes). In the scenario of a well-fitted DE model, the empirical null

221 distribution of the Wald test statistic in DESeq2 analysis should adhere to its theoretical

222 null, a standard normal distribution[36]. Surprisingly, we observed a systematic difference

223 between the theoretical and empirical null over the ~14,000 detected genes in this dataset

224 (**Fig. 3f, g**, **Fig. 3d** shows an example gene *C06B3.7*). While the mean of the empirical

225 null distribution was symmetrically aligned around the mean of the theoretical Wald

226 distribution (0) (**Fig. 3f**), the empirical null had systematically larger standard deviations

227 than the theoretical expectation (1) (**Fig. 3g**). As a result, *P* values computed using the

228 theoretical Wald distribution were anti-conservative, but this can be corrected by EmpirDE

229 (**Fig. 3d**, empirical *P* value).

230   We next investigated the source of larger-than-expected standard deviations of

231 empirical null. By inspecting mRNA levels across all conditions for each gene, we found

232 a wide-spread random, yet consistent across samples, fluctuation of the mean (**Fig. 3h**,

233 **Extended Data Fig. 2b**). This fluctuation is typically mild in its effect size, thereby

234 distinguishing it from a specific change induced by RNAi (**Fig. 3h,** *acdh-1*). We

235 hypothesized that the broad Wald statistic distribution is caused by such random

236 fluctuations, and that the mean of observed gene expression (μ_{obs}) is the sum of the actual

237 biologically relevant expression change (μ_{bio}) and a random experimental or condition

238 specific fluctuation ($\Delta\mu$) (**Fig. 3h**). The random fluctuation can be driven by systematic

239 hidden covariates in the experiment, such as subtle difference in temperature associated

240   with position in the culture plate, a common occurrence in large-scale experiments[6]. As

241   such these technical fluctuations can be misinterpreted as biological signal from the RNAi

242   treatment in WPS.

243   To test the hypothesis that a random fluctuation of the mean can introduce the

244   observed test statistic inflation, we performed a simulation study using scDesign3[46] to

245   simulate the metabolic-gene WPS dataset [REWIRING]. To mirror the real data, we used

246   the DEGs identified in the WPS analysis as the ground truth, and synthesized a new

247   dataset in which the mean expression of each DEG was altered based on the DEG fold

248   change while remaining consistent otherwise. To test the role of $\Delta\mu$, we further introduced

249   a random fluctuation to each gene based on the level of fluctuation of a gene in real data

250   (**Extended Data Fig. 2c**). Consistent with our hypothesis, we found that the standard

251   deviations of the empirical null distributions in simulated data matched the inflated test

252   statistic of real data when a random $\Delta\mu$ was added, while being very close to the

253   theoretical distribution when removing $\Delta\mu$ (**Fig. 3i, Extended Data Fig. 2d-f**). Therefore,

254   the observed anti-conservative P values can be explained with random fluctuations of

255   mean mRNA levels.

256

257   **EmpirDE framework rigorously controls FDR**

258   To benchmark the performance of EmpirDE, we first used simulation data to assess both

259   FDR and power. As expected, both EmpirDE and conventional DESeq2 analysis had

260   FDR correctly controlled on the simulated data without $\Delta\mu$ (**Extended Data Fig. 3a, b**).

261   However, with $\Delta\mu$ added, only EmpirDE was able to rigorously control the observed false

262   discovery proportion (FDP) at the expected FDR (**Fig. 4a**). Importantly, the power of

263   EmpirDE is greater than that of DESeq2 at the same level of observed FDP (**Fig. 4b**),

264   indicating a true increase in performance instead of nominal rescaling of P values. Such

265   rigorous control of FDR depends not only on the optimized statistical modeling framework

266   but also on the proper adjustment for multiple testing. WPS experiments involve both

267   simultaneously testing the expression of thousands of genes in each perturbation

268   (column-wise multiple testing) and hundreds of gene perturbations (row-wise multiple

269   testing). Using the simulation data, we found that the worst-case adjusted P values in

270 both column-wise and row-wise adjustments of a DE test aligned best with the expected

271 FDR (**Fig. 4c**), while its loss of power was negligible (**Extended Data Fig. 3c**).

272   Next, we experimentally benchmarked the EmpirDE performance. We first used the

273 NTP experiments mentioned above (**Fig. 3a**) to compare the number of false positive

274 DEGs between DESeq2 and EmpirDE analysis with different thresholds for both FDR and

275 fold change (**Fig. 4d**). We found that the 90% quantile of the number of DEGs detected

276 in the NTPs was much lower with EmpirDE compared to DESeq2 (**Fig. 4d**). Specifically,

277 at a FC of 1.5 and FDR < 0.1 we detected 4 and 435 false positive DEGs in the EmpirDE

278 and DESeq2 analysis, respectively (**Fig. 4d,** white dashed line).

279   To further evaluate the performance of EmpirDE, we used the reproducibility of DEG

280 calls to empirically evaluate the power and error of the DE analysis. We randomly selected

281 and independently repeated, in triplicate, 36 RNAi experiments that yielded a broad range

282 of DEGs (**Fig. 4e**). DEGs that were identified in one experiment but had no significant

283 change in the other (FC < 1.1 or in reversed direction), are considered genuine false

284 discoveries. We used the rate of such irreproducible DEGs to estimate the true FDR and

285 found that EmpirDE showed a rate of irreproducible calls consistent with the FDR

286 threshold (10%), regardless of the effect size (number of DEGs) (**Fig. 4f**). In contrast,

287 DESeq2 analysis achieved the desired control of FDR only when the effect size was large.

288 The rigorous control of false positives of the two-pronged EmpirDE approach can be

289 further demonstrated by visually inspecting each of the 36 conditions (**Fig. 4g, h**,

290 **Extended Data Fig. 4**). For instance, many DEGs in the *metr-1* RNAi experiment were

291 control-outlier genes and, as expected, these did not replicate in the repeat experiment

292 (**Fig. 4g,** right side, dashed black circles). Although non-control-outlier DEGs were

293 generally reproduced with both DESeq2 and EmpirDE, the latter still eliminated a few

294 highly changed but unreproduced calls (**Fig. 4g,** left). Notably, the EmpirDE approach

295 was critical when true positives were sparse and false positives identified by DESeq2

296 analysis masked the retrieval of true positives (**Fig. 4h**). Finally, these analyses also

297 facilitated EmpirDE parameter optimization, for instance, selecting an optimal threshold

298 to determine control-outlier genes in the first step of the framework (**Supplementary**

299 **Methods**, **Extended Data Fig. 3d, e**).

300   Taken together, with the EmpirDE framework that uses gene-specific empirical null
301   models, WPS can robustly assign DEGs for large numbers of perturbations with high
302   signal-to-noise ratio and rigorously controlled FDR for perturbations eliciting from a few
303   to thousands of DEGs.

304

305   **A proof-of-principle of WPS with 103 NHRs**
306   NHRs are a family of transcription factors (TFs) that play important roles in various
307   physiological processes including metabolism, development, and homeostasis[47]. The *C.*
308   *elegans* genome is predicted to encode more than 250 NHRs, making it the largest TF
309   family. In contrast, the human genome encodes only 48[48-50]. Although many NHRs have
310   been studied in *C. elegans*[21,51-55], more than half remain completely uncharacterized[47,56].
311   We analyzed the expression levels and patterns of all 288 predicted NHRs encoded by
312   the *C. elegans* genome and selected a set of 103 that are expressed at relatively high
313   levels both in the whole body and in the intestine and/or hypodermis, tissues highly
314   suitable for RNAi[57], to test the WPS method (**Fig. 5a, b**).
315   WPS analysis of these NHRs yielded a gene regulatory network (GRN) comprising
316   6,778 interactions between 101 perturbations and 3,673 genes, with in- and out-degrees
317   following expected distributions[58,59] (**Fig. 5c, d, Supplementary Table 3**). We found that
318   ~80% of perturbed NHRs (81) were responsive (≥ 5 DEGs, a conservative threshold
319   compared with the false positives in the NTP analysis, **Fig. 4d**). This rate is much higher
320   than that reported in whole-genome single-cell Perturb-seq experiments (~30%)[13], and
321   almost double than that we found in metabolic gene screens (40%) [REWIRING],
322   indicating that the majority of the 103 NHRs tested are actively regulating gene
323   expression in adult animals. Most NHR knockdowns resulted in moderate number of
324   DEGs (5-100) (**Fig. 5d**) and the magnitude of gene expression changes were also modest
325   (**Extended Data Fig. 5a**). Notably, 70% of DEGs identified in more than one perturbation
326   changed in the same direction (up or down, **Extended Data Fig. 5b**). As expected, genes
327   that were mostly down-regulated are expressed at higher levels than those that are mostly
328   up-regulated in NHR perturbations (**Extended Data Fig. 5c**).
329   We used WormCat[60] to identify biological processes enriched in the DEGs for each of
330   the 54 NHRs that yielded more than 10 DEGs and found that many NHRs affected genes

331 involved in stress response and metabolism, specifically pathogen response and lipid

332 metabolism (**Fig. 5e, f** and **Extended Data Fig. 5d, e**). These observations indicate that

333 several NHRs may function to establish and/or maintain metabolic functions and to prime

334 the animal to respond to different stressors. The knockdown of individual NHRs was

335 associated with several other WormCat categories as well. Some of these were known,

336 including the association of *nhr-31* with the lysosomal vATPase, *nhr-49* with lipid

337 metabolism, and *nhr-10, 68, 114, 101* with the propionate shunt pathway[54,61,62]

338 (**Extended Data Fig. 5d, f**).

339   We wondered whether the regulation of stress response and metabolic genes was

340 due to a few hub genes that are affected by many NHR perturbations (**Extended Data**

341 **Fig. 5g**). We identified a set of hub genes (in-degree > 5) that are annotated in WormCat

342 as either stress response or metabolic genes and found that different NHRs influenced

343 the expression of distinct stress and metabolic genes (**Fig. 5g, h**). Together with the in-

344 degree distribution that shows that most genes are regulated only by few NHRs (**Fig. 5d**),

345 this analysis indicates that the enrichment for stress response and metabolic genes in the

346 NHR GRN is not driven by a few common genes. Interestingly, we found that the GRN

347 mostly consists of activating interactions, i.e., upon knockdown of an NHR, gene levels

348 tend to go down. However, on average, we found that stress response gene expression

349 was increased upon NHR knockdown (**Fig. 5i, j, Extended Data Fig. 6 a-f**). These results

350 indicate that NHRs ensure the correct active expression of metabolic genes, especially

351 those involved in lipid metabolism, while they repress stress response genes.

352   Of the 288 *C. elegans* NHRs, at least 269 are from a substantial expansion and

353 diversification of the HNF4 family[63]. This observation raises the question whether these

354 expanded HNF4 family members regulate similar targets, or whether they evolved distinct

355 and diverse functions. To answer this, we compared the target genes of the 81 NHRs in

356 the GRN in more detail using perturbation-perturbation similarity and made a remarkable

357 observation: these NHRs not only regulated various sets of targets, but also clustered

358 into modules consisting of distinct NHR pairs, which we named 'pairwise modularity'. In

359 fact, 52 of 81 NHRs (64%) shared a significant overlap only with one other NHR (**Fig. 6a,**

360 **b and Extended Data Fig. 7, 8a, b, Supplementary Table 4**), and such pairwise

361 modularity cannot be observed by random (**Fig. 6c, Extended Data Fig. 7**). Our analysis

362   identified numerous NHR pairs for which functional relationships were not yet known and

363   which provide hypotheses for further experimental evaluation (two examples in **Extended**

364   **Data Fig. 8c, d**). Notably, this observation was facilitated by the EmpirDE framework

365   because it increased the signal-to-noise ratio compared to DESeq2 (**Extended Data Fig.**

366   **7**). Thus, even with a relatively low number of perturbations (~100 RNAi conditions),

367   EmpirDE can effectively increase the interpretability of the data.

368   Interestingly, NHR sequence similarity only correlated with few pairs that shared target

369   genes. Overall, protein sequences of the NHR DNA binding domains showed relative low

370   similarity to each other (percent identity < 0.5), with only limited number of clusters (**Fig.**

371   **6d**). Although the few pairs with high protein sequence similarity were more likely to share

372   targets (e.g., *nhr-10, nhr-68, nhr-114,* and *nhr-101*) the protein similarity between most

373   NHR pairs was lower and did not correlate with perturbation-perturbation similarity (**Fig.**

374   **6e, f**). Remarkably, even some NHRs from different evolutionary origins significantly

375   shared targets (e.g., *nhr-107* and *nhr-41,* **Fig. 6a**). In comparison, we found that

376   perturbation-perturbation similarity correlated better with *nhr* coexpression levels (**Fig.**

377   **6g**). Therefore, the pairwise modularity unveiled by WPS may be affected more by

378   mechanisms involved in the regulation, and less by the biophysical properties (e.g., DNA

379   binding domains), of these NHRs.

380

381   **Discussion**

382   In this study, we provide a WPS platform that combines strengths of single-cell and bulk

383   RNA sequencing with whole animal gene perturbations by RNAi. WPS is both efficient

384   and cost-effective, *e.g.,* a 2-week timeframe for collecting 96 perturbations in triplicate

385   and more than 10-fold cost reduction compared with conventional methods

386   (**Supplementary Protocols**), which enables replicate screens with full transcriptome

387   readouts of hundreds of perturbations in a living animal. Future screens with additional

388   RNAi libraries, different bacterial diets, and supplementation of metabolites or drugs will

389   provide insights into how the animal responds to a variety of perturbations.

390   A key advantage of WPS is that it is based on whole-organism *in vivo*

391   perturbations. While this is not feasible in mammals, it should be applicable to organisms

392   amenable to large-scale RNAi screens, such as Drosophila. However, we do envision

393 that WPS-like screens will be feasible in bulk in tissue culture cells, especially when
394 smaller sub-libraries of genes (~100) are selected for perturbations. Another key feature
395 of WPS is the EmpirDE framework that uses an empirical null for each detected gene and
396 that can be applied due to the scale of WPS experiments, and which can alleviate
397 systematic errors such as batch effects. Our study also shows and corrects the anti-
398 conservative *P* values that result from incorrect theoretical assumptions of parametric
399 differential gene expression models, which is in line with recent studies[39,40]. Although the
400 concept of empirical null has been widely applied in genomics[64-66], to the best of our
401 knowledge, it has not been used to directly model the test statistic distribution at the level
402 of individual genes (features), possibly due to the lack of such massively parallel data
403 generated with WPS. EmpirDE exploits the unique power of having many conditions (>
404 100), each with three replicates, to achieve rigorous statistical analyses. Conventionally,
405 such level of statistical rigor is only achievable with a high number of replicates (e.g., 8-
406 12)[67], which is a rare occasion in biological research. Given the dramatic decrease in cost
407 and increase in throughput for RNA-seq in recent years [44,68,69], we envision the soon
408 explosion of large-scale RNA-seq datasets in both research and clinic settings, all of
409 which will benefit from using EmpirDE.

410 By applying WPS to more than 100 NHR perturbations, we discover a pairwise
411 modularity in which two or more NHRs regulate the expression of overlapping sets of
412 genes, which cannot be explained by protein (and presumably binding site) similarity.
413 Instead, this pairwise modularity suggests that 'AND-logic gates' are a common
414 mechanism of gene regulation in *C. elegans*. Future studies with other TFs will be
415 important to see if this is a general principle, or if it is a specific feature of NHRs.
416 Knockdown of many NHRs affected only few genes, suggesting that these TFs may either
417 not be active under the conditions tested, or are truly specialized in their regulatory
418 function. In two companion studies [WIRING][REWIRING], we further validated WPS by
419 perturbating ~900 metabolic genes. These studies generated high-quality, highly
420 interpretable datasets, providing tremendous insights into metabolic wiring and rewiring
421 at a systems level. Notably, WPS interrogates gene functions *in vivo*, thus linking genes
422 to their native physiological roles. For instance, using metabolic gene WPS data, we
423 identified an unconventional central carbon metabolism that consumes ribose, rather than

424 glucose, from dietary RNA and through the pentose phosphate pathway. Together, we

425 envision that WPS-style *in vivo* functional genomics will provide a powerful tool to uncover

426 gene functions in living organisms.

427

428 **REFERENCES**

429 1 Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109-
430 126 (2000).
431 2 Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-
432 391 (2002). https://doi.org/10.1038/nature00935
433 3 Przybyla, L. & Gilbert, L. A. A new era in functional genomics screens. *Nat Rev Genet* **23**, 89-103
434 (2022). https://doi.org/10.1038/s41576-021-00409-w
435 4 Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic
436 network. *Science* **292**, 929-934 (2001). https://doi.org/10.1126/science.292.5518.929
437 5 Kemmeren, P. *et al.* Large-scale genetic perturbations reveal regulatory networks and an
438 abundance of gene-specific repressors. *Cell* **157**, 740-752 (2014).
439 https://doi.org/10.1016/j.cell.2014.02.054
440 6 Baryshnikova, A. *et al.* Quantitative analysis of fitness and genetic interactions in yeast on a
441 genome scale. *Nature Methods* **7**, 1017-1024 (2010).
442 7 Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic
443 Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882 e1821 (2016).
444 https://doi.org/10.1016/j.cell.2016.11.048
445 8 Jaitin, D. A. *et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell
446 RNA-Seq. *Cell* **167**, 1883-1896 e1815 (2016). https://doi.org/10.1016/j.cell.2016.11.039
447 9 Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling
448 of Pooled Genetic Screens. *Cell* **167**, 1853-1866 e1817 (2016).
449 https://doi.org/10.1016/j.cell.2016.11.038
450 10 Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods*
451 **14**, 297-301 (2017). https://doi.org/10.1038/nmeth.4177
452 11 Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic
453 Screens. *Cell* **176**, 377-390 e319 (2019). https://doi.org/10.1016/j.cell.2018.11.029
454 12 Schraivogel, D. *et al.* Targeted Perturb-seq enables genome-scale genetic screens in single cells.
455 *Nat Methods* **17**, 629-635 (2020). https://doi.org/10.1038/s41592-020-0837-5
456 13 Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale
457 Perturb-seq. *Cell* **185**, 2559-2575 e2528 (2022). https://doi.org/10.1016/j.cell.2022.05.013
458 14 Jin, X. *et al.* In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism
459 risk genes. *Science* **370** (2020). https://doi.org/10.1126/science.aaz6063
460 15 Zheng, X. *et al.* Massively parallel in vivo Perturb-seq reveals cell type-specific transcriptional
461 networks in cortical development. *bioRxiv* (2023). https://doi.org/10.1101/2023.09.18.558077
462 16 Santinha, A. J. *et al.* Transcriptional linkage analysis with in vivo AAV-Perturb-seq. *Nature* **622**,
463 367-375 (2023). https://doi.org/10.1038/s41586-023-06570-y
464 17 Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis*
465 *elegans*. *Nature* **391**, 806-811 (1998).
466 18 Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
467 19 Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using
468 RNAi. *Nature* **421**, 231-237 (2003).
469 20 Rual, J.-F. *et al.* Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-
470 based RNAi library. *Genome Res* **14**, 2162-2168 (2004).
471 21 MacNeil, L. T. *et al.* Transcription factor activity mapping of a tissue-specific gene regulatory
472 network. *Cell Syst* **1**, 152-162 (2015). https://doi.org/10.1016/j.cels.2015.08.003
473 22 Horowitz, B. B., Nanda, S. & Walhout, A. J. M. A Transcriptional Cofactor Regulatory Network for
474 the *C. elegans* Intestine. *G3 (Bethesda)* (2023). https://doi.org/10.1093/g3journal/jkad096

475   23   Bhattacharya, S. *et al.* A metabolic regulatory network for the *Caenorhabditis elegans* intestine.
476   *iScience* **25**, 104688 (2022). https://doi.org/10.1016/j.isci.2022.104688
477   24   Yanai, I. & Hashimshony, T. CEL-Seq2-Single-Cell RNA Sequencing by Multiplexed Linear
478   Amplification. *Methods Mol Biol* **1979**, 45-56 (2019). https://doi.org/10.1007/978-1-4939-9240-9_4
479   25   Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol*
480   **17**, 77 (2016). https://doi.org/10.1186/s13059-016-0938-8
481   26   Spencer, W. C. *et al.* A spatial and temporal map of *C. elegans* gene expression. *Genome research*
482   **21**, 325-341 (2011). https://doi.org/10.1101/gr.114595.110
483   27   Hendriks, G. J., Gaidatzis, D., Aeschimann, F. & Grosshans, H. Extensive oscillatory gene
484   expression during *C. elegans* larval development. *Mol Cell* **53**, 380-392 (2014).
485   https://doi.org/10.1016/j.molcel.2013.12.013
486   28   Meeuse, M. W. *et al.* Developmental function and state transitions of a gene expression oscillator
487   in *Caenorhabditis elegans*. *Mol Syst Biol* **16**, e9498 (2020). https://doi.org/10.15252/msb.20209498
488   29   Heimberg, G., Bhatnagar, R., El-Samad, H. & Thomson, M. Low Dimensionality in Gene
489   Expression Data Enables the Accurate Extraction of Transcriptional Programs from Shallow
490   Sequencing. *Cell Syst* **2**, 239-250 (2016). https://doi.org/10.1016/j.cels.2016.04.001
491   30   Liu, Y., Zhou, J. & White, K. P. RNA-seq differential expression studies: more sequence or more
492   replication? *Bioinformatics* **30**, 301-304 (2014). https://doi.org/10.1093/bioinformatics/btt688
493   31   Taub, M. A., Corrada Bravo, H. & Irizarry, R. A. Overcoming bias and systematic errors in next
494   generation sequencing data. *Genome Med* **2**, 87 (2010). https://doi.org/10.1186/gm208
495   32   Tung, P. Y. *et al.* Batch effects and the effective design of single-cell gene expression studies. *Sci*
496   *Rep* **7**, 39921 (2017). https://doi.org/10.1038/srep39921
497   33   Love, M. I., Anders, S., Kim, V. & Huber, W. RNA-Seq workflow: gene-level exploratory analysis
498   and   differential   expression.   *F1000Res*   **4**,   1070   (2015).
499   https://doi.org/10.12688/f1000research.7035.1
500   34   Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol* **17**, 13 (2016).
501   https://doi.org/10.1186/s13059-016-0881-8
502   35   Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465-
503   476 (2001). https://doi.org/10.1016/s0092-8674(01)00576-1
504   36   Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-
505   seq data with DESeq2. *Genome Biol* **15**, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8
506   37   Consortium, S. M.-I. A comprehensive assessment of RNA-seq accuracy, reproducibility and
507   information content by the Sequencing Quality Control Consortium. *Nat Biotechnol* **32**, 903-914
508   (2014). https://doi.org/10.1038/nbt.2957
509   38   Li, S. *et al.* Detecting and correcting systematic variation in large-scale RNA sequencing data. *Nat*
510   *Biotechnol* **32**, 888-895 (2014). https://doi.org/10.1038/nbt.3000
511   39   Li, Y., Ge, X., Peng, F., Li, W. & Li, J. J. Exaggerated false positives by popular differential
512   expression methods when analyzing human population samples. *Genome Biol* **23**, 79 (2022).
513   https://doi.org/10.1186/s13059-022-02648-4
514   40   Pall, T., Luidalepp, H., Tenson, T. & Maivali, U. A field-wide assessment of differential expression
515   profiling by high-throughput sequencing reveals widespread bias. *PLoS Biol* **21**, e3002007 (2023).
516   https://doi.org/10.1371/journal.pbio.3002007
517   41   Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and
518   microarray studies. *Nucleic Acids Res* **43**, e47 (2015). https://doi.org/10.1093/nar/gkv007
519   42   Hart, S. N., Therneau, T. M., Zhang, Y., Poland, G. A. & Kocher, J. P. Calculating sample size
520   estimates   for   RNA   sequencing   data.   *J Comput Biol* **20**,   970-978   (2013).
521   https://doi.org/10.1089/cmb.2012.0283
522   43   Poplawski, A. & Binder, H. Feasibility of sample size calculation for RNA-seq studies. *Brief*
523   *Bioinform* **19**, 713-720 (2018). https://doi.org/10.1093/bib/bbw144
524   44   Schurch, N. J. *et al.* How many biological replicates are needed in an RNA-seq experiment and
525   which   differential   expression   tool   should   you   use?   *RNA* **22**,   839-851   (2016).
526   https://doi.org/10.1261/rna.053959.115
527   45   Efron, B. Large-Scale Simultaneous Hypothesis Testing - The Choice of a Null Hypothesis. *Journal*
528   *of the American Statistical Association* **99**, 96-104 (2004).
529   46   Song, D. *et al.* scDesign3 generates realistic in silico data for multimodal single-cell and spatial
530   omics. *Nat Biotechnol* (2023). https://doi.org/10.1038/s41587-023-01772-1

47   Tao, L. J., Seo, D. E., Jackson, B., Ivanova, N. B. & Santori, F. R. Nuclear Hormone Receptors and Their Ligands: Metabolites in Control of Transcription. *Cells* **9** (2020). https://doi.org/10.3390/cells9122606

48   Reece-Hoyes, J. S. *et al.* A compendium of *C. elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol* **6**, R110 (2005).

49   Antebi, A. in *WormBook, The C.elegans Research Community, ed. WormBook.* *http://www.wormbook.org (Nov 21, 2005).* (ed I. Greenwald) (2005).

50   Scholtes, C. & Giguere, V. Transcriptional control of energy metabolism by nuclear receptors. *Nat Rev Mol Cell Biol* **23**, 750-770 (2022). https://doi.org/10.1038/s41580-022-00486-7

51   Arda, H. E. *et al.* Functional modularity of nuclear hormone receptors in a *C. elegans* gene regulatory network. *Molecular Systems Biology* **6**, 367 (2010).

52   Reece-Hoyes, J. S. *et al.* Extensive rewiring and complex evolutionary dynamics in a *C. elegans* multiparameter transcription factor network. *Mol Cell* **51**, 116-127 (2013).

53   Fuxman Bass, J. I. *et al.* A gene-centered *C. elegans* protein-DNA interaction network provides a framework for functional predictions. *Mol Syst Biol* **12**, 884 (2016).

54   Van Gilst, M. R., Hajivassiliou, H., Jolly, A. & Yamamoto, K. R. Nuclear hormone receptor NHR-49 controls fat consumption and fatty acid composition in *C. elegans. PLoS Biol* **3**, e53 (2005).

55   Asahina, M., Valenta, T., Silhankova, M., Korinek, V. & Jindra, M. Crosstalk between a nuclear receptor and beta-catenin signaling decides cell fates in the *C. elegans* somatic gonad. *Dev Cell* **11**, 203-211 (2006). https://doi.org/10.1016/j.devcel.2006.06.003

56   Antebi, A. Nuclear receptor signal transduction in C. elegans. *WormBook*, 1-49 (2015). https://doi.org/10.1895/wormbook.1.64.2

57   Conte, D., Jr., MacNeil, L. T., Walhout, A. J. & Mello, C. C. RNA Interference in *Caenorhabditis elegans. Curr Protoc Mol Biol* **109**, 26 23 21-26 23 30 (2015). https://doi.org/10.1002/0471142727.mb2603s109

58   Albert, R. Scale-free networks in cell biology. *J Cell Sci* **118**, 4947-4957 (2005). https://doi.org/10.1242/jcs.02714

59   Deplancke, B. *et al.* A gene-centered *C. elegans* protein-DNA interaction network. *Cell* **125**, 1193-1205 (2006).

60   Holdorf, A. D. *et al.* WormCat: An Online Tool for Annotation and Visualization of *Caenorhabditis elegans* Genome-Scale Data. *Genetics* **214**, 279-294 (2020). https://doi.org/10.1534/genetics.119.302919

61   Hahn-Windgassen, A. & Van Gilst, M. R. The *Caenorhabditis elegans* HNF4alpha Homolog, NHR-31, mediates excretory tube growth and function through coordinate regulation of the vacuolar ATPase. *PLoS Genet* **5**, e1000553 (2009). https://doi.org/10.1371/journal.pgen.1000553

62   Bulcha, J. T. *et al.* A persistence detector for metabolic network rewiring in an animal. *Cell Rep* **26**, 460-468 (2019).

63   Robinson-Rechavi, M., Maina, C. V., Gissendanner, C. R., Laudet, V. & Sluder, A. Explosive lineage-specific expansion of the orphan nuclear receptor HNF4 in nematodes. *J Mol Evol* **60**, 577-586 (2005). https://doi.org/10.1007/s00239-004-0175-8

64   van Iterson, M., van Zwet, E. W., Consortium, B. & Heijmans, B. T. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* **18**, 19 (2017). https://doi.org/10.1186/s13059-016-1131-9

65   Cao, X., Wu, B. & Hertz, M. I. Empirical null distribution based modeling of multi-class differential gene expression detection. *J Appl Stat* **40**, 347-357 (2013). https://doi.org/10.1080/02664763.2012.743976

66   Schmitt, A. D., Hu, M. & Ren, B. Genome-wide mapping and analysis of chromosome architecture. *Nat Rev Mol Cell Biol* **17**, 743-755 (2016). https://doi.org/10.1038/nrm.2016.104

67   Halasz, G. *et al.* Optimizing murine sample sizes for RNA-seq studies revealed from large-scale comparative analysis. *BioRxiv* (2024). https://doi.org/https://doi.org/10.1101/2024.07.08.602525

68   Alpern, D. *et al.* BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. *Genome Biol* **20**, 71 (2019). https://doi.org/10.1186/s13059-019-1671-x

69   Janjic, A. *et al.* Prime-seq, efficient and powerful bulk RNA sequencing. *Genome Biol* **23**, 88 (2022). https://doi.org/10.1186/s13059-022-02660-8

586   70   Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011).
587   https://doi.org/10.1038/nbt.1754
588   71   Roux, A. E. *et al.* Individual cell types in *C. elegans age* differently and activate distinct cell-
589   protective responses. *Cell Rep* **42**, 112902 (2023). https://doi.org/10.1016/j.celrep.2023.112902
590   72   Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic*
591   *Acids Res* **50**, W276-W279 (2022). https://doi.org/10.1093/nar/gkac240
592   73   Nanda, S. *et al.* Systems-level transcriptional regulation of *Caenorhabditis elegans* metabolism.
593   *Mol Syst Biol*, e11443 (2023). https://doi.org/10.15252/msb.202211443
594

SUPPLEMENTARY TABLES LEGENDS

Supplementary Table 1: The reference TPM profile of adult *C. elegans* used in this study. This is obtained from the combined vector control data collected at 62 to 65 hours post L1 stage in one biological replicate (See **Supplementary Methods** for details).

Supplementary Table 2: List of 67 non-targeting perturbations (NTPs). The list does not include the 4 random spike-in vector controls.

Supplementary Table 3: Differentially expressed genes in NHR perturbations.

Supplementary Table 4: Perturbation-perturbation similarity matrix of NHR perturbations (only includes responsive conditions).

METHODS

See Supplementary Information.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

H.Z., X.L., and A.J.M.W. conceived the project and wrote the manuscript. H.Z. and X.L. jointly developed the WPS technology and analyzed the data. H.Z. conducted the experiments. X.L. wrote the codes. D.S. helped design the simulation study under the supervision of J.J.L. D.S. also provided critical discussions that led to the idea of modeling the empirical null. O.Y. and A.K. helped with the dolphinNext pipeline for WPS data processing. S.N. provided the NHR coexpression data. M.G. and A.J.M.W. supervised the study. The co-first authorship order was determined by a coin flip. XL and HZ contributed equally and reserve the right to list their name first in their resumes.

COMPETING INTEREST DECLARATION

M.G. and A.K. are co-founders of Via Scientific, Inc., a UMass Chan Medical School spin-off. A.K. is a board member of the company. M.G, A.K. and O.Y. have equity in the company. They ensure that steps have been taken to prevent these affiliations from affecting analysis integrity and are dedicated to upholding research transparency and integrity.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to A.J.M.W., or M.G. Supplementary Information is available for this paper.

DATA AVAILABILITY

The WPS data analysis pipeline is available at https://github.com/XuhangLi/WPS including detailed procedures of raw data processing, quality control and EmpirDE analysis. A walkthrough of the pipeline can be found in the repository. A standalone R package for the EmpirDE analysis can be found in https://github.com/XuhangLi/EmpirDE. All other source code to reproduce the figures and analyses in this study will be deposited to Zenodo before publication. Raw and processed data in this study are available in Gene Expression Omnibus (GEO) session GSE255865. Downloadable read count data are also available at the WPS portal hosted in our WormFlux website https://wormflux.umassmed.edu/WPS.

Figure 1

a
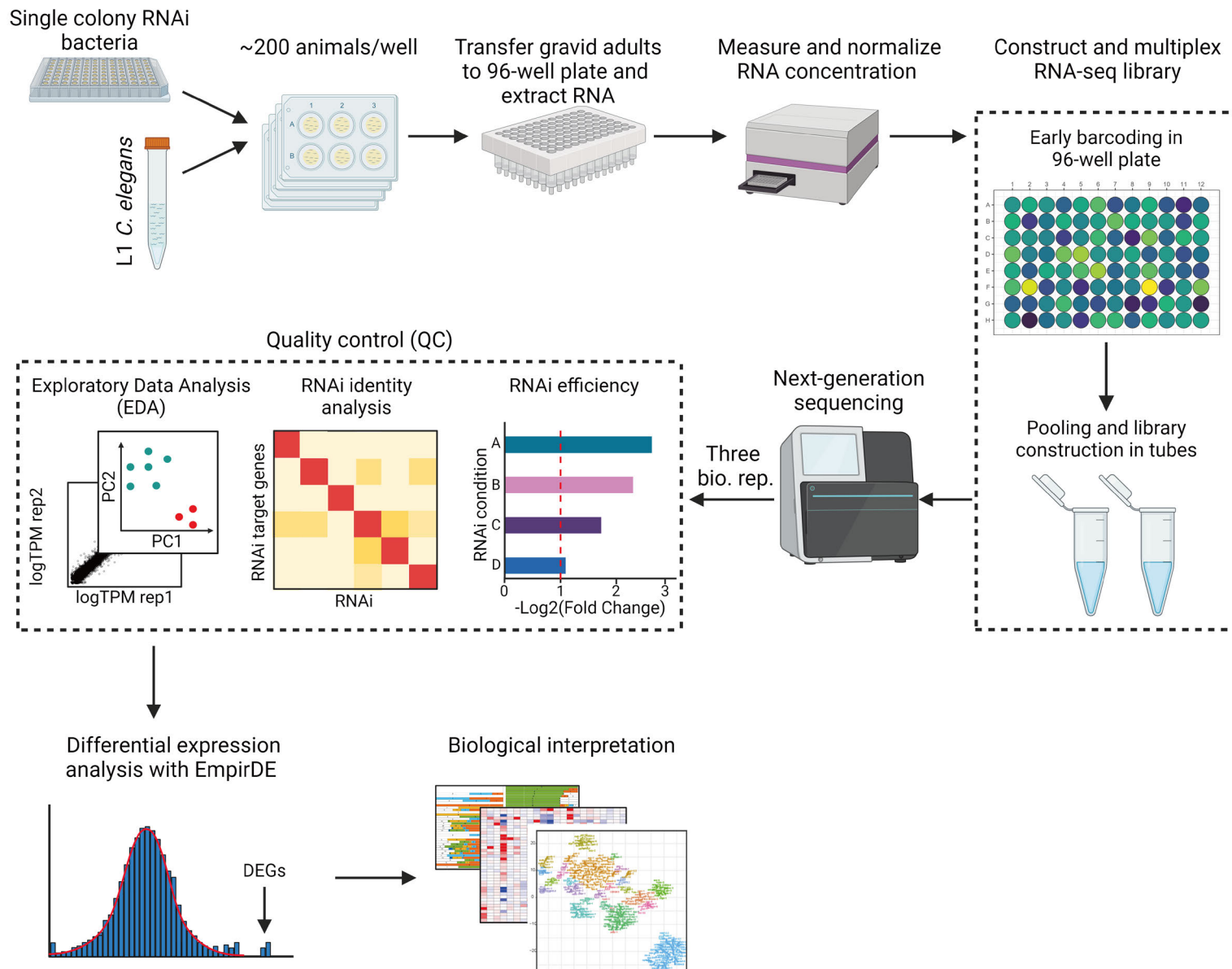
b

Optimization step	Highlights		
RNA extraction	**Scalable:** lyse animals by chemical method in 96-well plate	**Specific:** only break down animal bodies and not their eggs	**Robust:** identified a transcriptional steady state, 60-68 hours post L1, which allows for robust large-scale experimentation
RNA-seq library constructions	**Improved multiplexing:** pool individual samples after first-strand cDNA synthesis to improve robustness and cost efficiency	**Standardization:** optimize and standardize a protocol for bulk RNA-seq library construction based on CEL-Seq2 method	**Optimal read depth:** ~6M on average for optimal cost-benefit tradeoff
Quality control	**RNAi quality assurance:** identify reads from the anti-sense RNA generated during RNA interference to ensure the identity of RNAi target and quantify the efficiency of knockdown by the decrease of target gene expression		
Differential expression analysis	**Power:** three biological replicates for >95% RNAi conditions	**Rigorous false discovery control:** a two-pronged DE analysis framework, EmpirDE, that uses gene-specific models with empirical null distributions to rigorously identify DEGs at desired false discovery rate	

647

20

Figure 1

648   **Fig. 1: Worm Perturb-Seq (WPS) overview. a**, Overview of the WPS pipeline. This

649   figure was created with BioRender.com. **b**, WPS optimization highlights.

Figure 2
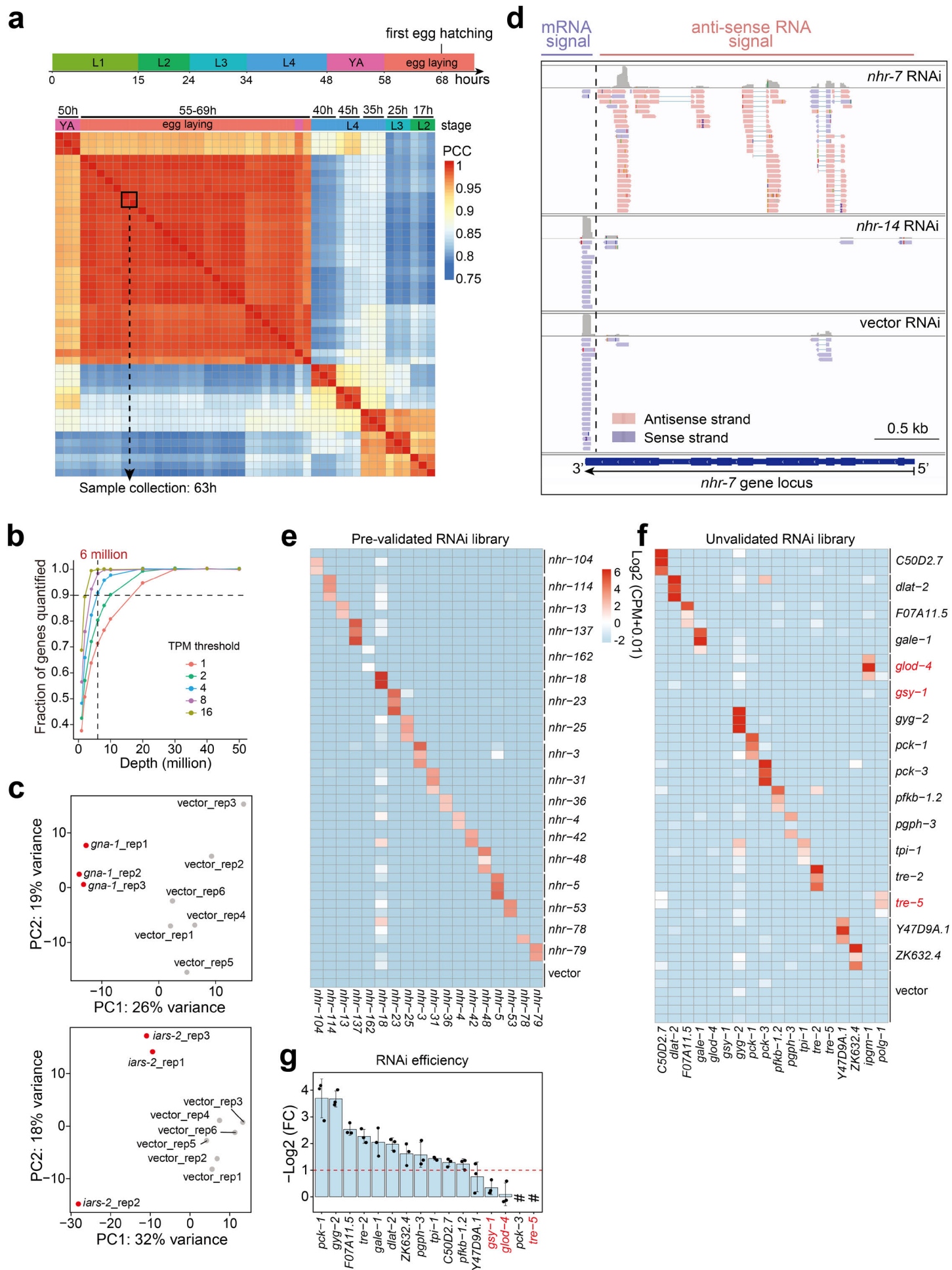
Figure 2

651 **Fig. 2: Development of WPS and data quality control. a**, Comparison of the *C. elegans*

652 transcriptome across developmental stages. The Pearson correlation coefficient (PCC)

653 was calculated by the WPS profiles of animals fed vector control bacteria and collected

654 at different time points post L1. **b**, Subsampling analysis of the reference WPS profile

655 showing the fraction of genes quantified versus sequencing depth (see **Supplementary**

656 **Methods**). **c**, Representative Principal Component Analysis (PCA) results for gene

657 expression profiles of perturbations without (*gna-1* on top panel) and with (*iars-2* in bottom

658 panel) a low-quality outlier replicate. Red and grey dots indicate RNAi and control

659 samples, respectively. **d,** An example showing reads mapped to the reverse strand of the

660 RNAi target gene (*nhr-7*) and a decrease in mRNA reads at the 3' end. The reads mapped

661 to *nhr-7* gene locus were visualized by Integrative Genomics Viewer (IGV [70]). *nhr-7* RNAi

662 was compared to vector control RNAi and another RNAi condition (*nhr-14*). **e, f**,

663 Quantification of anti-sense RNA reads in a Sanger-sequenced (**e**) and an unvalidated (**f**)

664 WPS sequencing library. Row names represent the intended RNAi gene (three replicates

665 each) and column names represent the actual knocked down genes. Row names in red

666 indicate wrong RNAi clones. Values are the log2(Count-Per-Million (CPM) + 0.01) of the

667 reads mapped to the reverse strand of each gene in the columns. **g,** Log2(Fold Change

668 (FC)) of the RNAi targeted gene expression for the WPS sequencing library shown in (**f**).

669 Pound key (#) indicates not detected.

23

Figure 3

670
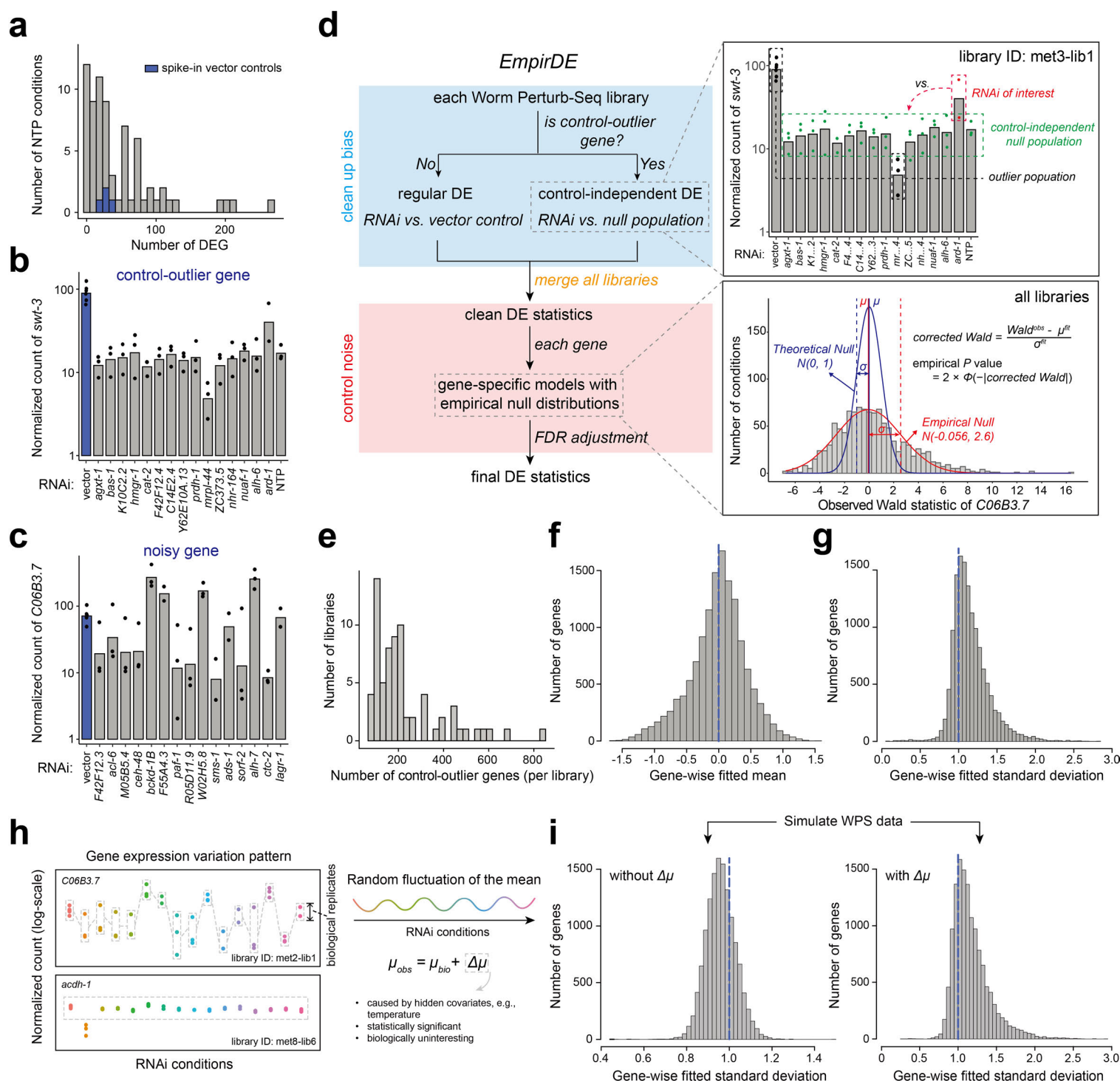
Fig. 3: **EmpirDE analysis framework reveals systematic anti-conservative *P* values caused by a deviation from the expected distribution of Wald test statistics. a**, Distribution of the number of DEGs in non-targeting perturbations (NTPs) identified by DESeq2 analysis (FC > 2, adjusted *P* value (P_{adj}) < 0.01). **b, c**, Examples of a control-outlier (**b**) and noisy (**c**) gene. Each bar plot shows the expression levels of the gene of

Figure 3

676   interest in a WPS sequencing library that includes RNAi perturbations and vector control

677   conditions. **d**, Schematic illustrating the EmpirDE framework. The zoom-in windows show

678   two example genes shown in (**b**) and (**c**). **e**, The number of control-outlier genes per WPS

679   sequencing library. **f**, **g**, Distribution of fitted means (**f**) and standard deviations (**g**) for all

680   genes in the empirical null modeling. The blue dashed line shows the values for

681   theoretical null. **h,** Examples showing the random fluctuation of the mean for two genes.

682   **i,** Distribution of fitted standard deviation using simulated WPS data with (right) and

683   without (left) adding a random fluctuation of the mean in the simulation.

Figure 4
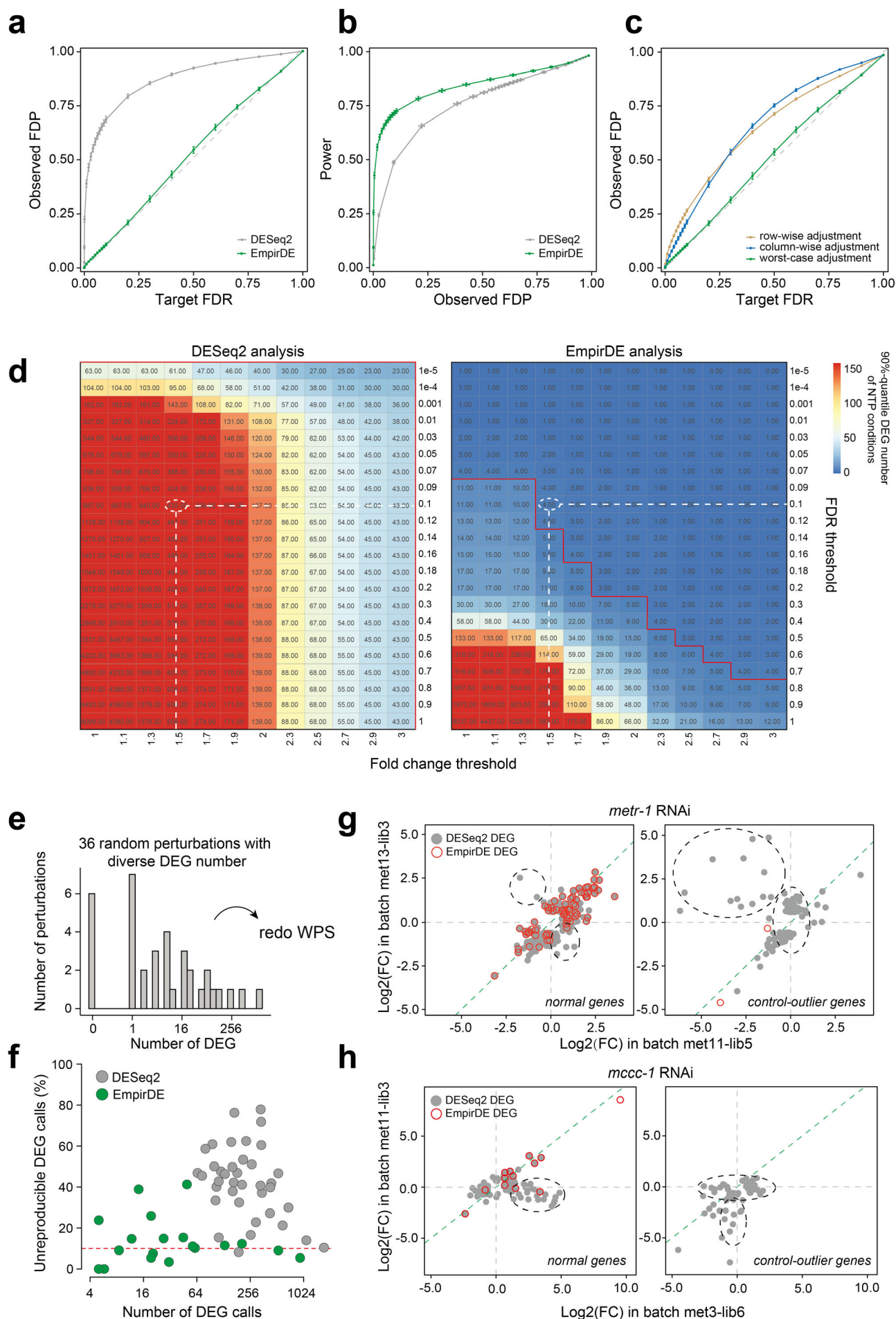
Figure 4

685 **Fig. 4: EmpirDE rigorously controls FDR. a, b,** Benchmarking the performance of
686 EmpirDE analysis framework. The observed False Discovery Proportion (FDP) is
687 compared to target FDR (**a**) and power (**b**). The full metabolic-gene WPS dataset (3,691
688 samples) was simulated 10 times with random mean fluctuation ($\Delta\mu$) to produce the error
689 bars of each metric. FDP and power were measured based on a pooled set of 117,096
690 simulated DE changes in all conditions (**Supplementary Methods**). **c**, Benchmarking
691 FDR control of different multiple testing adjustment strategies. **d,** Evaluating false
692 discoveries using NTP experiments. The color (and numbers) in the heatmap shows 90%-
693 quantile DEG numbers in the NTP experiments. The red lines show the threshold
694 boundary for five false positive DEG calls. **e,** Number of DEGs (defined by FDR < 0.1 and
695 FC > 1.5 using EmpirDE) for 36 perturbations that were repeated by a second WPS
696 experiment. **f,** Fraction of unreproducible DEGs for EmpirDE versus DESeq2 analysis.
697 Unreproducible DEGs were defined by genes that are called as DEG in one experiment
698 (FDR < 0.1, FC > 1.5) but confidently non-DEG in the other (FC < 1.1 or show a different
699 FC direction). The red dashed line shows the theoretical FDR (FDR = 0.1). **g**, **h**,
700 Comparison of log2(FC) measured in two independent experiments for representative
701 RNAi with either high (**g**) or moderate (**h**) number of DEGs. The green dashed line
702 indicates the diagonal (y=x) and black dashed circle indicates genes that are not
703 reproduced.

Figure 5
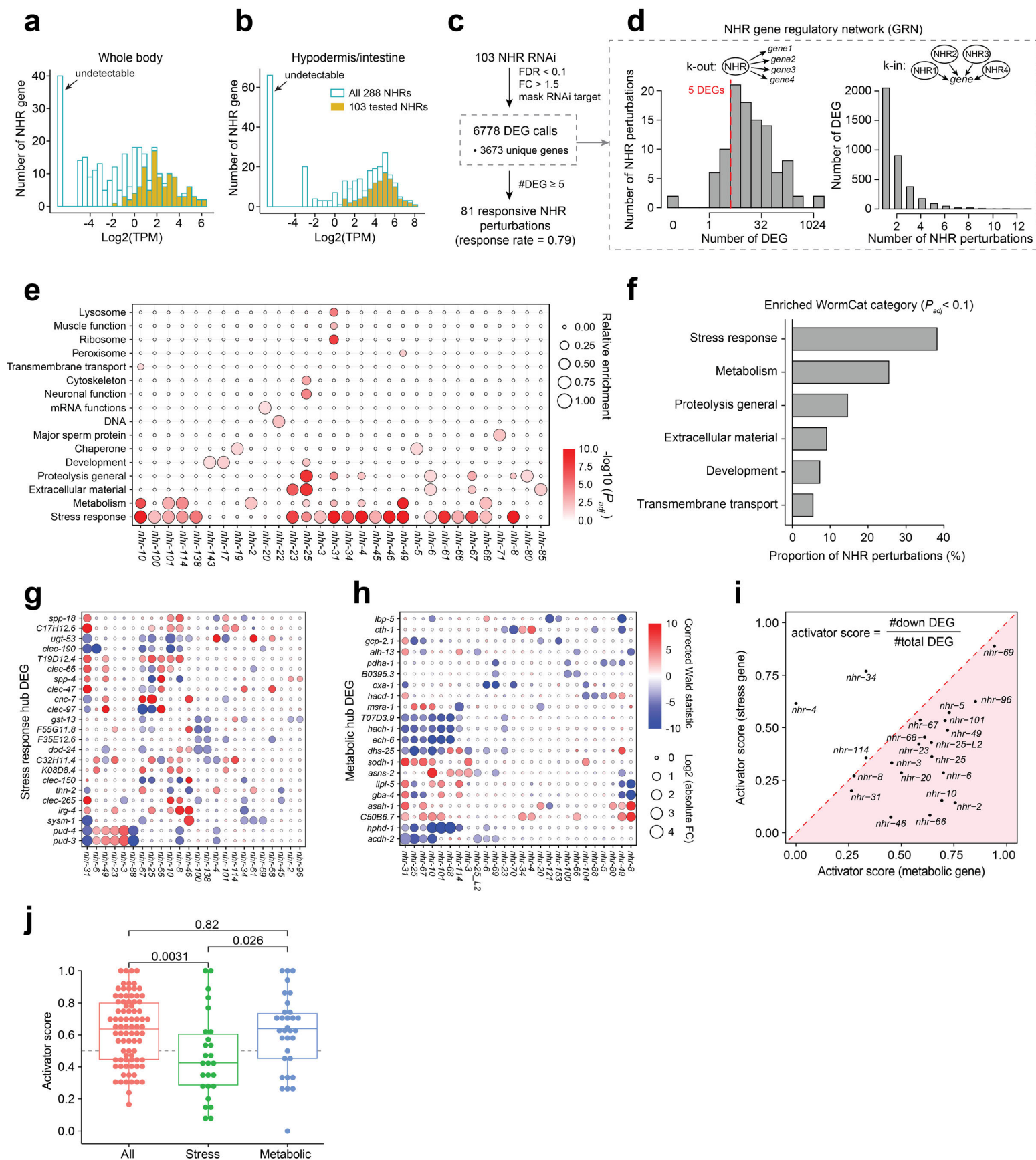
Figure 5

705   **Fig. 5: A NHR gene regulatory network. a**, Expression levels of *nhr* genes in adult

706   animals. The TPM was quantified by the reference profile used in **Extended Data Fig.**

707   **1c**. **b**, Tissue expression of *nhr* genes based on an adult stage single-cell RNA-seq

708   data[71]. The maximal TPM of hypodermis and intestine expression is shown in the plot. **c**,

709   Summary of NHR WPS experiments. **d**, Distribution of the number of DEGs in NHR

710   perturbations (i.e., k-out) and the number of NHRs regulating the same gene (i.e., k-in).

711   The red dashed line indicates the responsiveness threshold (≥ 5 DEGs). **e**, Functional

712   enrichment analysis of NHR RNAi conditions. Only the 54 NHRs (55 perturbations

713   because *nhr-25* was perturbed at both L1 and L2 stages) with more than 10 DEGs were

714   analyzed to ensure power. Conditions without significant (P_{adj} > 0.05) enrichment are not

715   shown in the plot. The relative enrichment is defined as the -log10(P_{adj}) normalized by its

716   maximum in a given RNAi condition. **f**. Proportion of testable RNAi (i.e. > 10 DEGs) that

717   showed significant enrichment (P_{adj} < 0.1) in WormCat Level 1. **g**, **h**, Visualization of hub

718   genes (i.e., regulated by >5 NHRs) of the stress response (**g**) and metabolic (**h**)

719   categories. **i**, Comparison of the activator score for metabolic (x-axis) and stress response

720   (y-axis) DEGs. NHR perturbations with ≥ 5 DEGs in the corresponding categories were

721   analyzed. **j**, Overall comparison of activator scores for different DEG categories. Two-

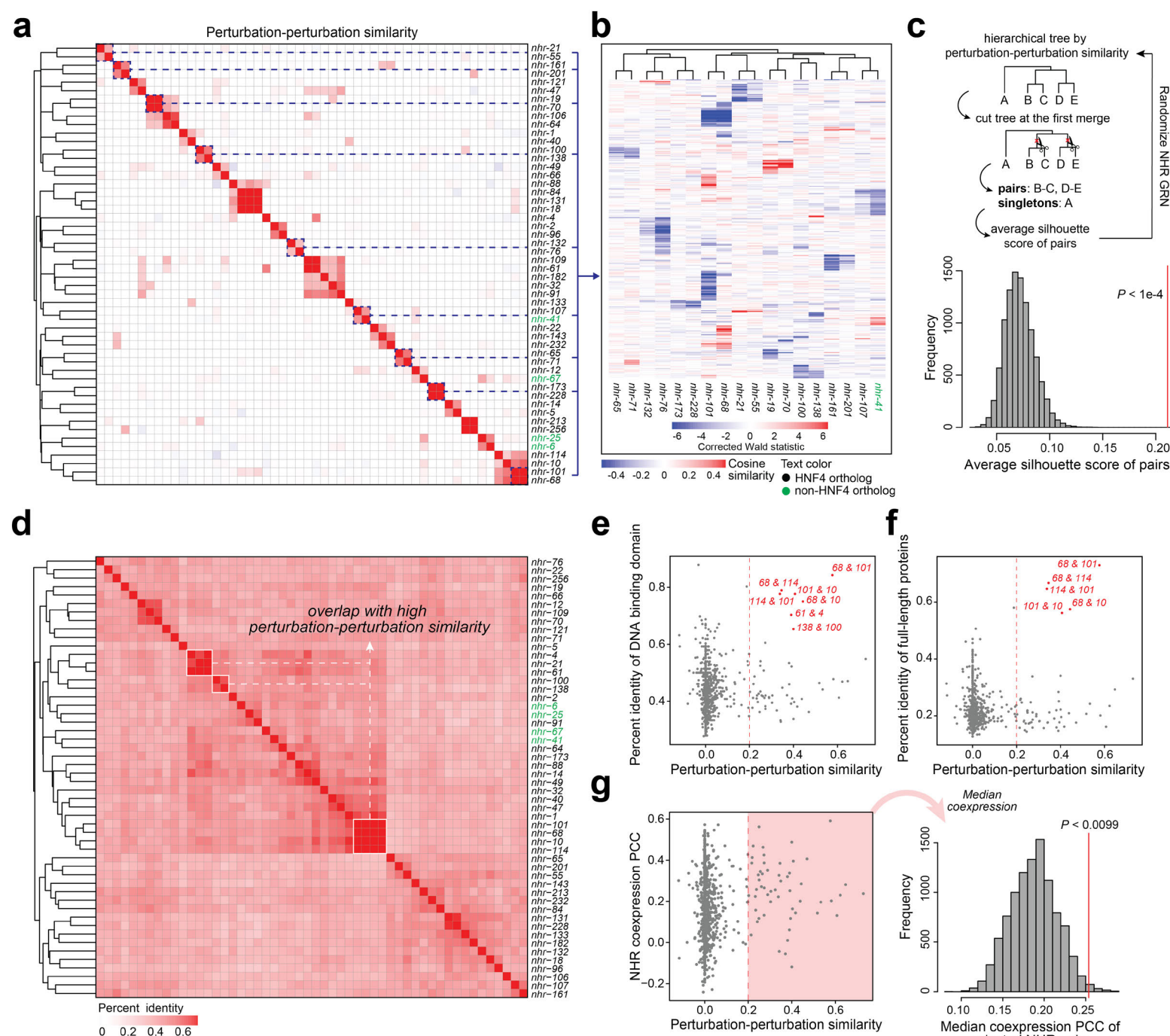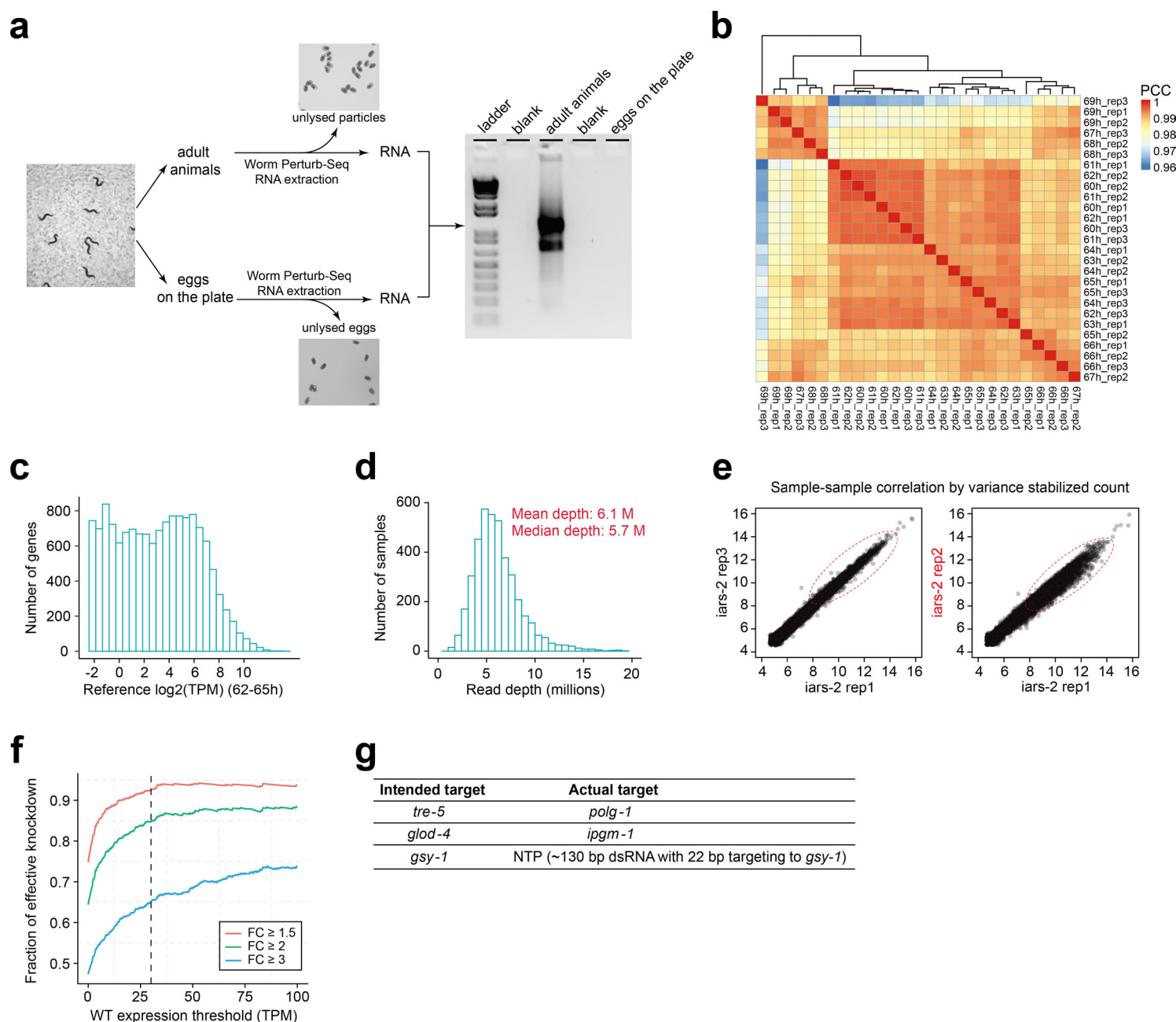722   sample Wilcoxon tests were performed to calculate the *P* values.

29

Figure 6

Fig. 6: The pairwise modularity of NHRs. a, Heatmap depicting perturbation-perturbation similarity of DEG profiles for the NHR perturbations. The perturbation-perturbation similarity was defined by cosine similarity of the filtered log2(FC) profile. The filtered log2(FC) was derived by masking the log2(FC) values of genes that are not called as DEGs (FDR < 0.1, FC > 1.5) to zero. **b**, Visualization of gene expression changes in selected NHR pairs. The gene expression change was measured by the corrected Wald

Figure 6

730 statistic. Rows are the union DEGs of these selected NHR perturbations. **c**,

731 Randomization test of the pairwise modularity of NHR gene family. The schematic shows

732 the design of the randomization test. Histogram shows the average silhouette score of

733 pairs in 10,000 randomizations. The red line indicates the observed score from real data.

734 **d**, Heatmap depicting protein sequence similarity (percent identity) for the DNA binding

735 domain (DBD) of NHRs. The heatmap was clustered using distance matrix generated by

736 Clustal Omega online tool from EMBL-EBI[72] (**Supplementary Methods**). **e, f**, Scatter

737 plots showing the comparison between perturbation-perturbation similarity and sequence

738 similarity of DBD (**e**) and full-length protein (**f**). Each data point indicates a pair of NHRs

739 and selected pairs are labeled. **g**, Scatter plot and randomization test for the associations

740 between perturbation-perturbation similarity and NHR coexpression. Coexpression was

741 measured based on the median Pearson Correlation Coefficient (PCC) of *nhr* gene

742 expression in a compendium of *C. elegans* gene expression data across various

743 conditions[73] (**Supplementary Methods**). The median coexpression level of pairs with a

744 cosine similarity greater than 0.2 (red region in (**g**)) is calculated and compared with that

745 from randomized data (**Supplementary Methods**). The histogram shows that the median

746 from real data (red line) is significantly greater than that from randomized data, indicating

747 a statistically significant association between NHR coexpression and perturbation-

748 perturbation similarity.

31

Extended Data Figure 1

a

b

c

d

e

Sample-sample correlation by variance stabilized count

f

g

| Intended target	Actual target
tre-5	*polg-1*
glod-4	*ipgm-1*
gsy-1	NTP (~130 bp dsRNA with 22 bp targeting to *gsy-1*)

749

Extended Data Fig. 1: Development of WPS. a, WPS RNA extraction method does not extract RNA from eggs. **b**, Zoom-in heatmap of **Fig. 2a**. **c**, Histogram of gene expression in the subsampling reference profile. The reference was generated by combining the data of a single replicate from 62 to 65 hours in (**b**). **d**, Read depth distribution of 4,083 WPS samples combining metabolic and NHR WPS datasets. **e**, Examples of gene expression correlation between replicates with good (left) and poor (right, labeled in red) quality. **f**, Fractions of perturbations with effective targeted gene knockdown versus targeted gene expression. Effective knockdown is defined by a minimum decrease of 1.5 (red), 2 (green)

32

758   and 3 (blue) folds in the targeted gene expression, based on differential expression

759   analysis. The fraction was calculated using 1,048 valid RNAi conditions from both

760   metabolic and NHR WPS datasets, titrated with a minimum targeted gene expression

761   threshold indicated in x-axis. The Wild-Type (WT) expression was based on the reference

762   profile in (**c**). **g**, Sanger sequencing results of the three fail-QC conditions in **Fig. 2f.**

Extended Data Figure 2

763
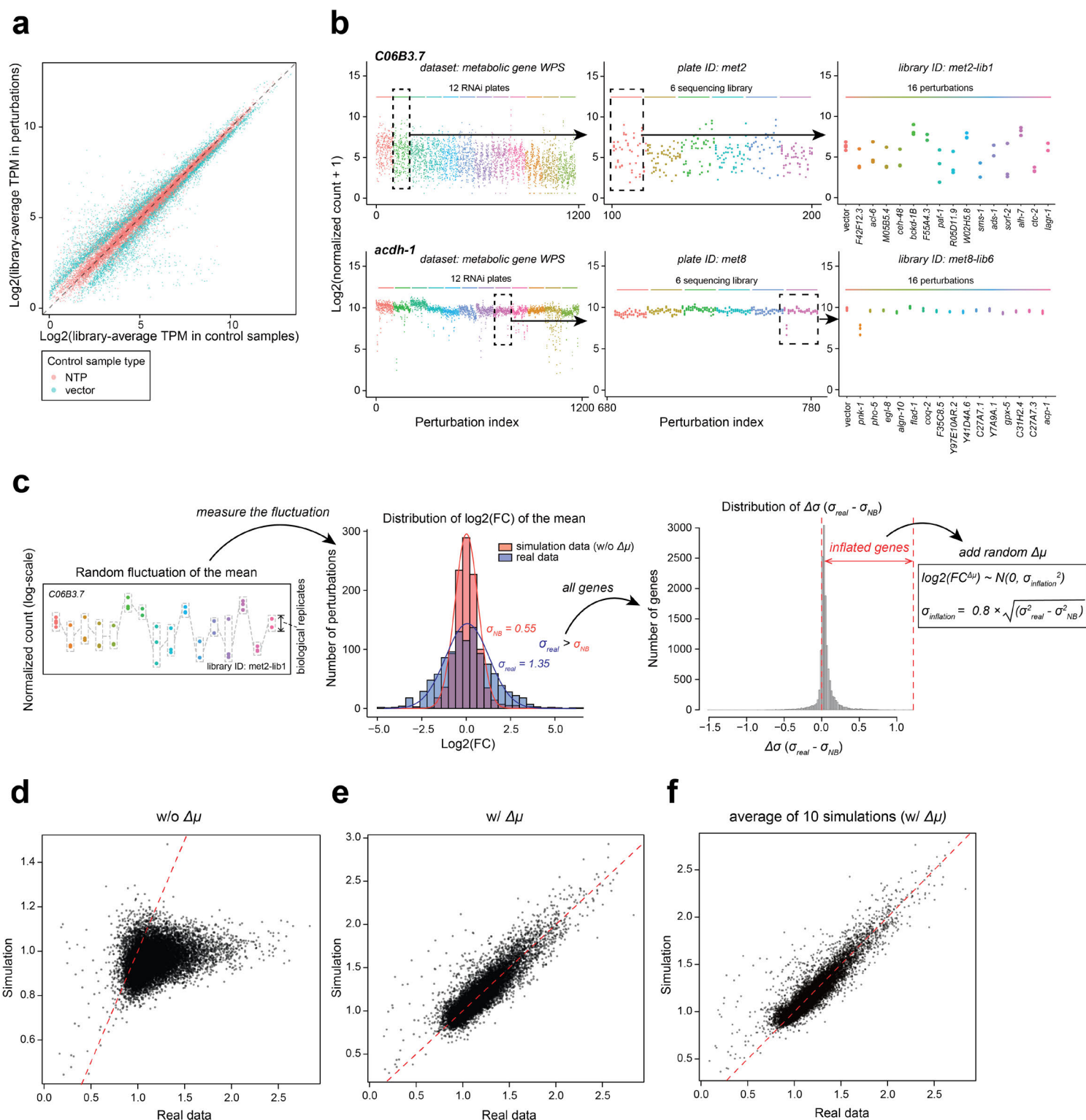
Extended Data Fig. 2: Investigating the source of false discovery in WPS analysis.

a, Scatter plot showing the expression level of control-outlier genes in different types of samples. The control-outlier genes in the 72 metabolic-gene WPS sequencing libraries are combined and displayed in the plot and each data point refers to a control-outlier gene

768 in a sequencing library. The y-axis represents the average TPM of all RNAi samples
769 (excluding NTPs and vector controls) in a sequencing library, which is compared to the
770 average TPM of either NTP (red) or vector control (blue) samples in the same sequencing
771 library on the x-axis. **b**, Gene expression levels of two example genes exhibiting high (top)
772 or moderate (down) fluctuation of the mean. The three panels from left to right present a
773 sequential zoom-in view of the expression pattern at the level of entire metabolic-gene
774 WPS dataset, an RNAi plate, and a sequencing library. The perturbation index is an
775 arbitrary index indicating the number of perturbations displayed in the plot. **c**, Illustration
776 of the quantification and simulation of random fluctuation of the mean. To quantify the
777 fluctuation level of each gene, a distribution of its log2(FC) across the thousand RNAi
778 conditions in the metabolic-gene WPS dataset was fitted by the same procedure used for
779 fitting the empirical null of the Wald statistic. The fitted sigma from real data was compared
780 to that from the simulation data without $\Delta\mu$. Inflation of the sigma level compared with
781 simulation fit was used as a proxy of mean fluctuation to create a heuristic (formula on
782 the right) for generating gene-specific random $\Delta\mu$ (**Supplementary Methods**). **d**, **e**,
783 Comparison of the fitted standard deviation of the Wald statistic between simulation and
784 real data for all genes. The red dashed line indicates the diagonal (y=x). The scatter plot
785 shows results from one simulation of the metabolic-gene WPS dataset. **f**, Similar to (**e**),
786 but showing the average of fitted standard deviations in 10 independent simulations.
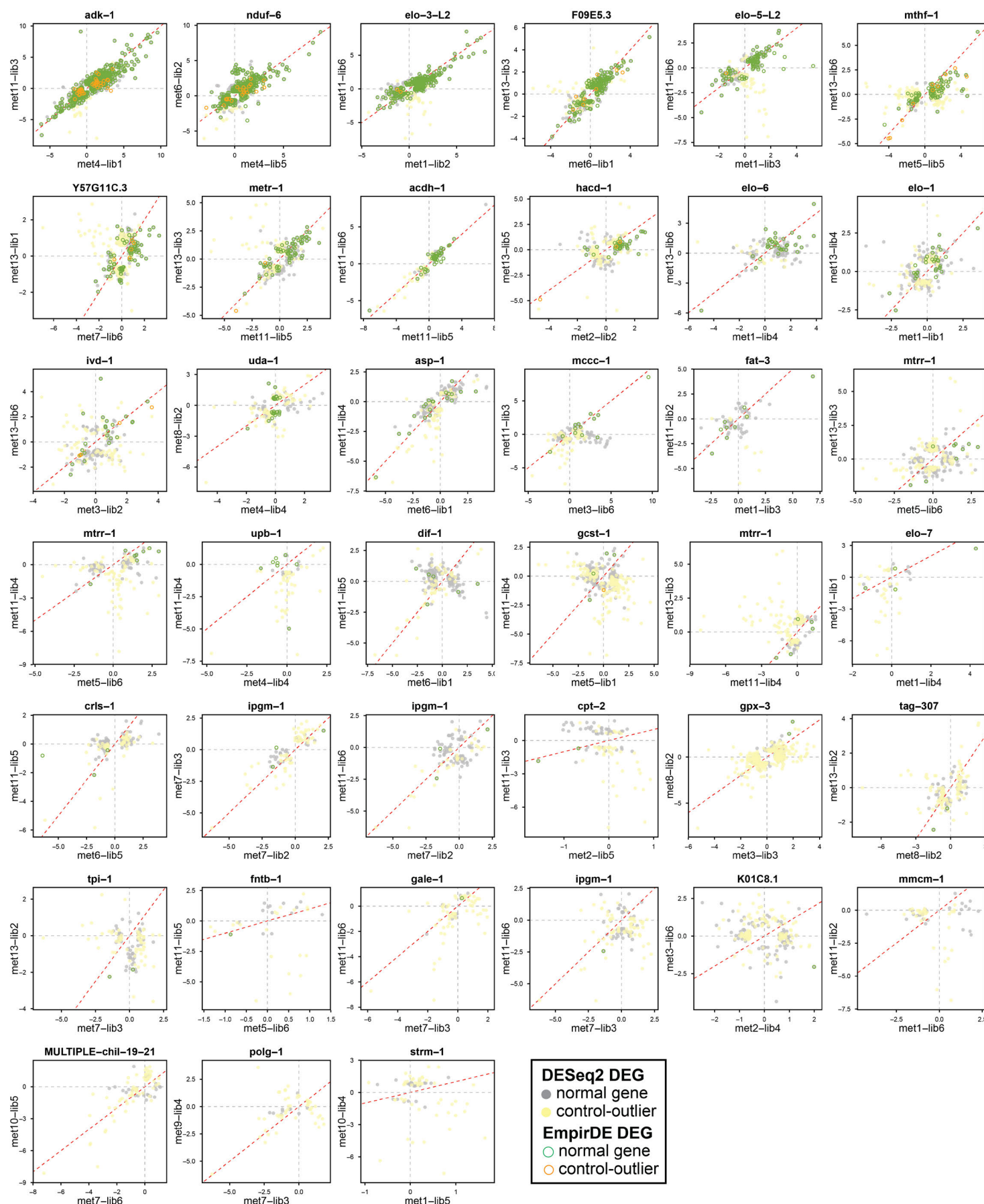
Extended Data Figure 3
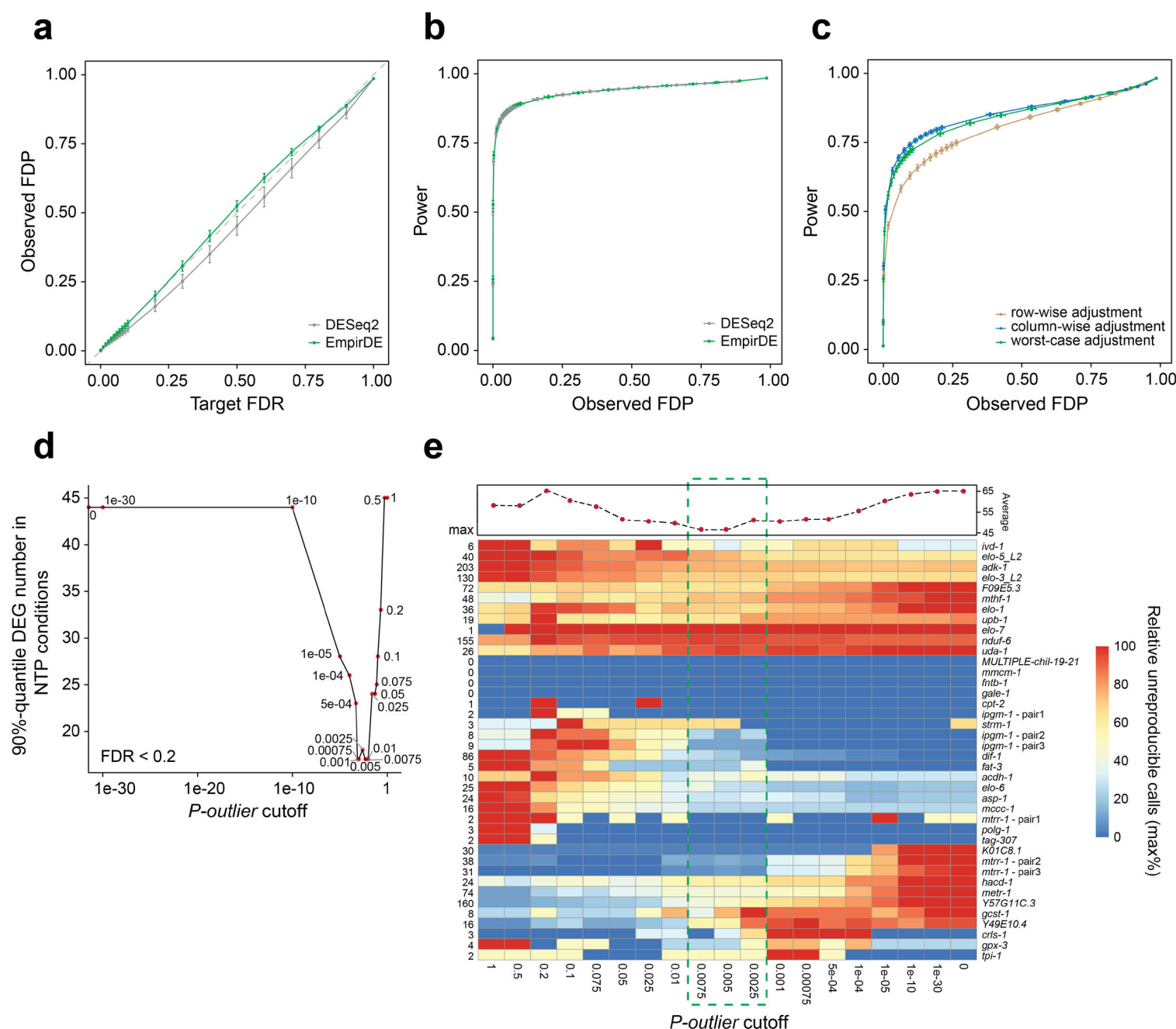
Extended Data Fig. 3: Benchmarking and parameter optimization of EmpirDE analysis framework. a, **b**, Performance benchmarking of the EmpirDE analysis using simulation data without $Δμ$. FDP, false discovery proportion. **c**, Benchmarking the FDR control of different multiple testing adjustment strategies. **d, e,** Two evaluations of false discovery to select the threshold for defining control-outlier genes in the first step of EmpirDE analysis. (**d**) shows the false discovery estimation by DEG calls in NTP samples across different P-outlier thresholds, and (**e**) shows the estimation by unreproducible calls

Extended Data Figure 3

795   in the 36 repeated conditions. To increase the sensitivity of the analysis, DEGs in (**d**) were

796   solely defined with FDR threshold of 0.2. Colors in (**e**) show the fraction of all

797   unreproducible DEGs with respect to the maximum level in each condition across a range

798   of *P-outlier* threshold. The maximum levels of unreproducible DEGs are shown on the left

799   of the heatmap. Scatter plot on the top shows the column-wise average (only conditions

800   with maximum level greater than 5 DEGs were used for better robustness). For definition

801   of unreproducible DEGs, please refer to **Supplementary Methods**.

Extended Data Figure 4

Log2(FC) in batch #1 (y-axis)

Log2(FC) in batch #2 (x-axis)

Legend:
DESeq2 DEG
- normal gene
- control-outlier

EmpirDE DEG
- normal gene
- control-outlier

803 **Extended Data Fig. 4: WPS analysis reproducibility.** Thirty-six WPS experiments were

804 repeated, and some were done more than twice (e.g., *mtrr-1*). Each experiment included

805 triplicate perturbations. Plots show comparison of the log2(FC) of genes called as DEGs

806 in either experiment. Red dashed lines indicate the diagonal (y=x).

Extended Data Figure 5
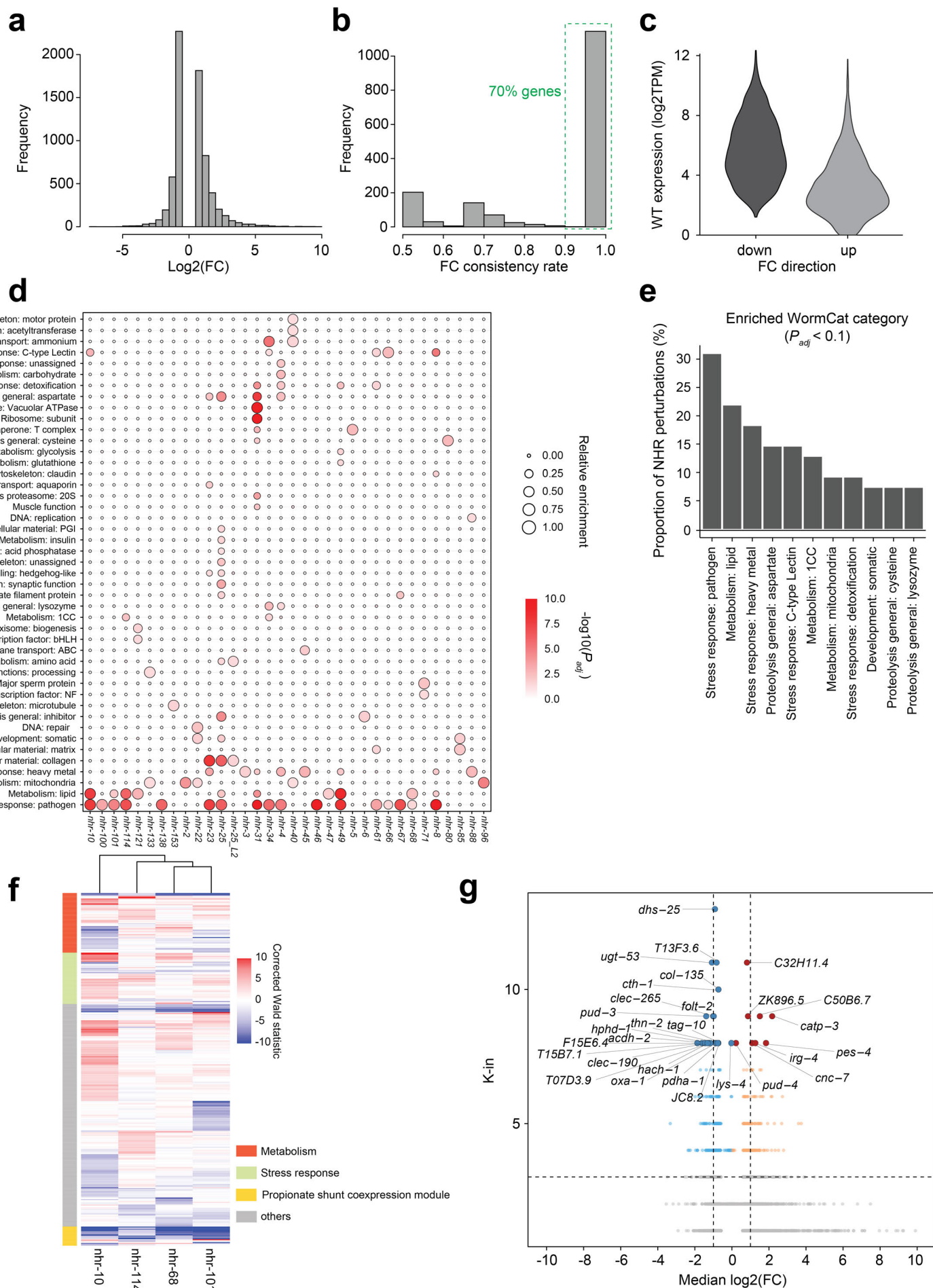
Extended Data Figure 5

Extended Data Fig. 5: Analysis of NHR WPS data. a, Log2(FC) distribution of all DEGs in the NHR WPS experiments. **b**, Distribution of the fold change consistency (up or down) for all DEGs with k-in ≥ 2, which was defined as the proportion of DEG calls that increase or decrease in the WPS experiments (whichever greater). **c**, Association between fold change direction and basal expression level of the DEGs. **d**, Gene set enrichment analysis of NHR perturbations using WormCat Level 2 categories[60]. Only NHR perturbations with more than 10 DEG calls were analyzed. Conditions without significant ($P_{adj} > 0.05$) enrichment are not shown in the plot. The relative enrichment is defined as the -log10(P_{adj}) normalized by its maximum in a given perturbation. **e,** Proportion of testable perturbations that showed significant enrichment ($P_{adj} < 0.1$) in WormCat Levels 2 categories. **f**, Heatmap displaying the corrected Wald statistics for the union of DEGs in *nhr-10*, *nhr-68*, *nhr-101*, and *nhr-114* perturbations. The propionate shunt coexpression module was defined based on a previous study[73]. **g,** Volcano plot showing the hub DEGs in the NHR GRN. The median log2(FC) is the median level of the log2(FC) values of all interactions for a corresponding gene in the GRN. The red and blue colors are used to indicate up and down regulated hub DEGs, respectively.

41

Extended Data Figure 6
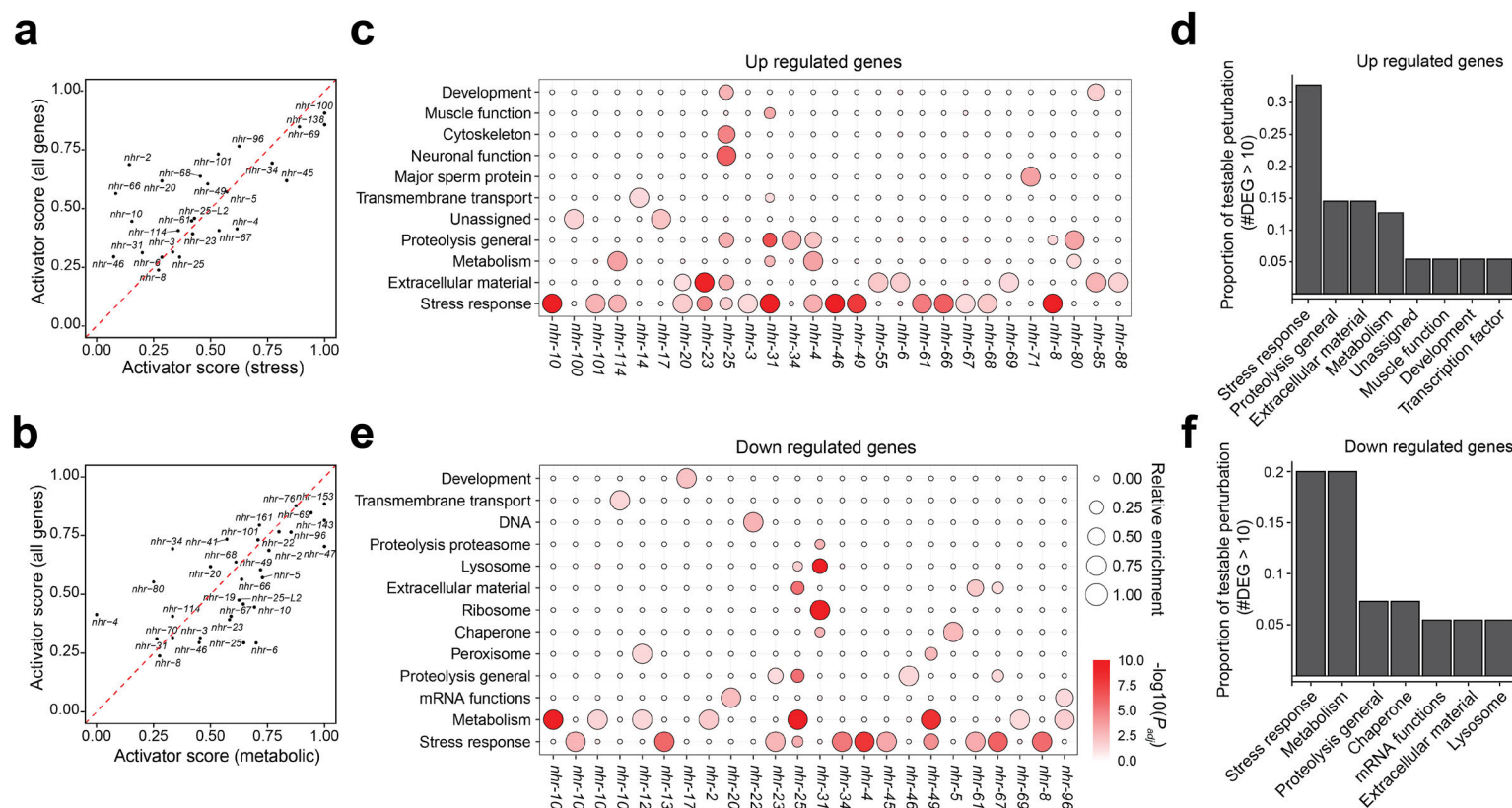
824

825   **Extended Data Fig. 6: Analysis of NHR GRN. a, b**, Correlation between the activator

826   score for all DEGs and that for stress response (**a**) and metabolic (**b**) genes. The stress

827   response and metabolic genes were defined based on WormCat Level 1 categories[60].

828   Only perturbations with ≥ 5 stress (**a**) or metabolic (**b**) DEGs are shown in the plot. **c, d**,
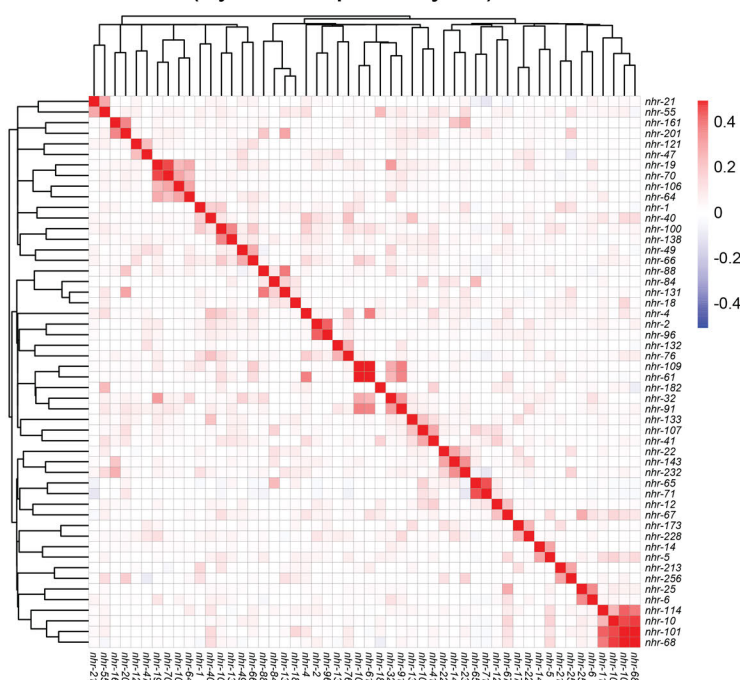
829   **e**, **f**, Gene set enrichment analysis of up-regulated (**c**,**d**) and down-regulated (**e**,**f**) genes.

830   WormCat Level 1 categories were used as the gene sets.

42

Extended Data Figure 7

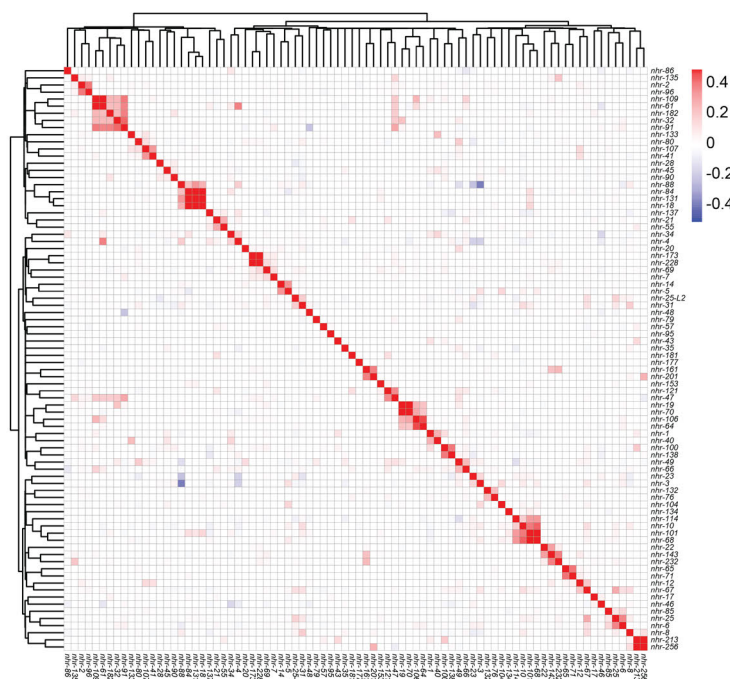**Supplementary heatmaps of perturbation-perturbation similarity**

52 selected NHR perturbations
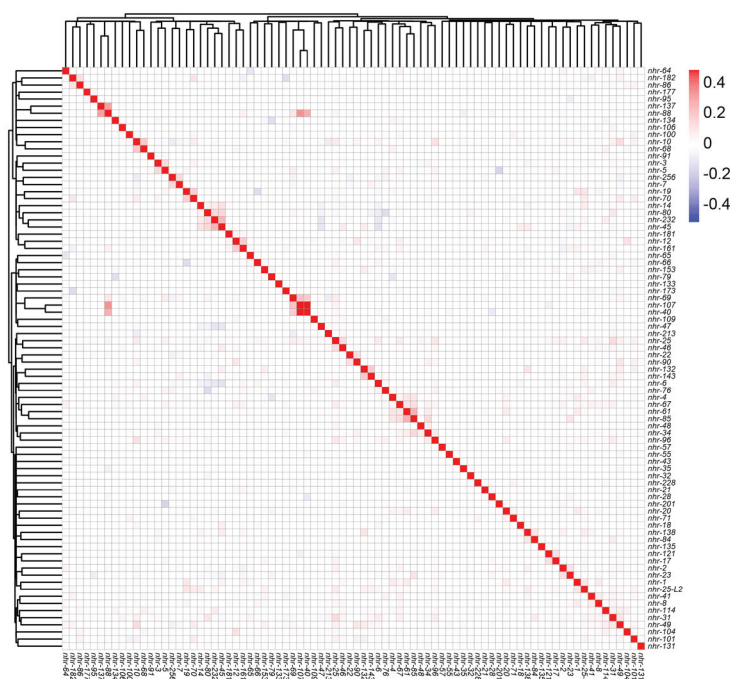(by DESeq2 analysis)

52 selected NHR perturbations
(by EmpirDE analysis, also in **Fig. 6a**)
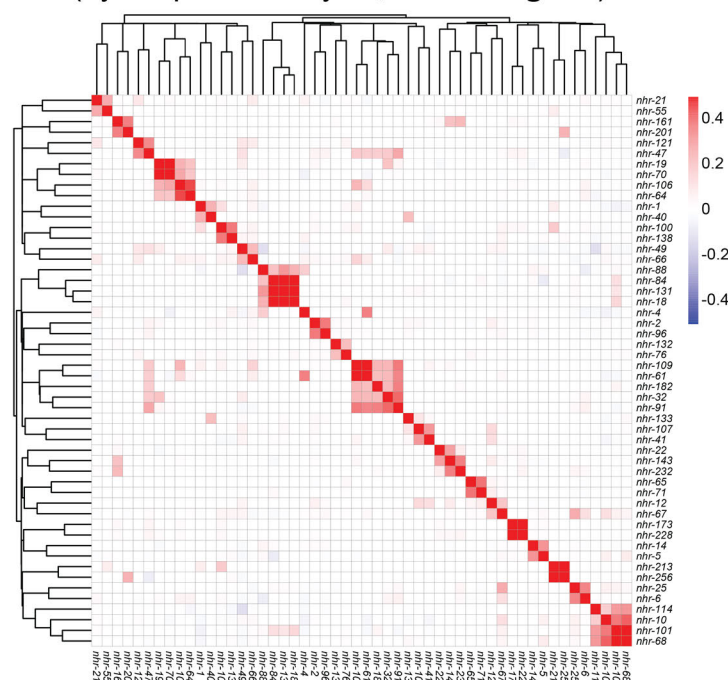
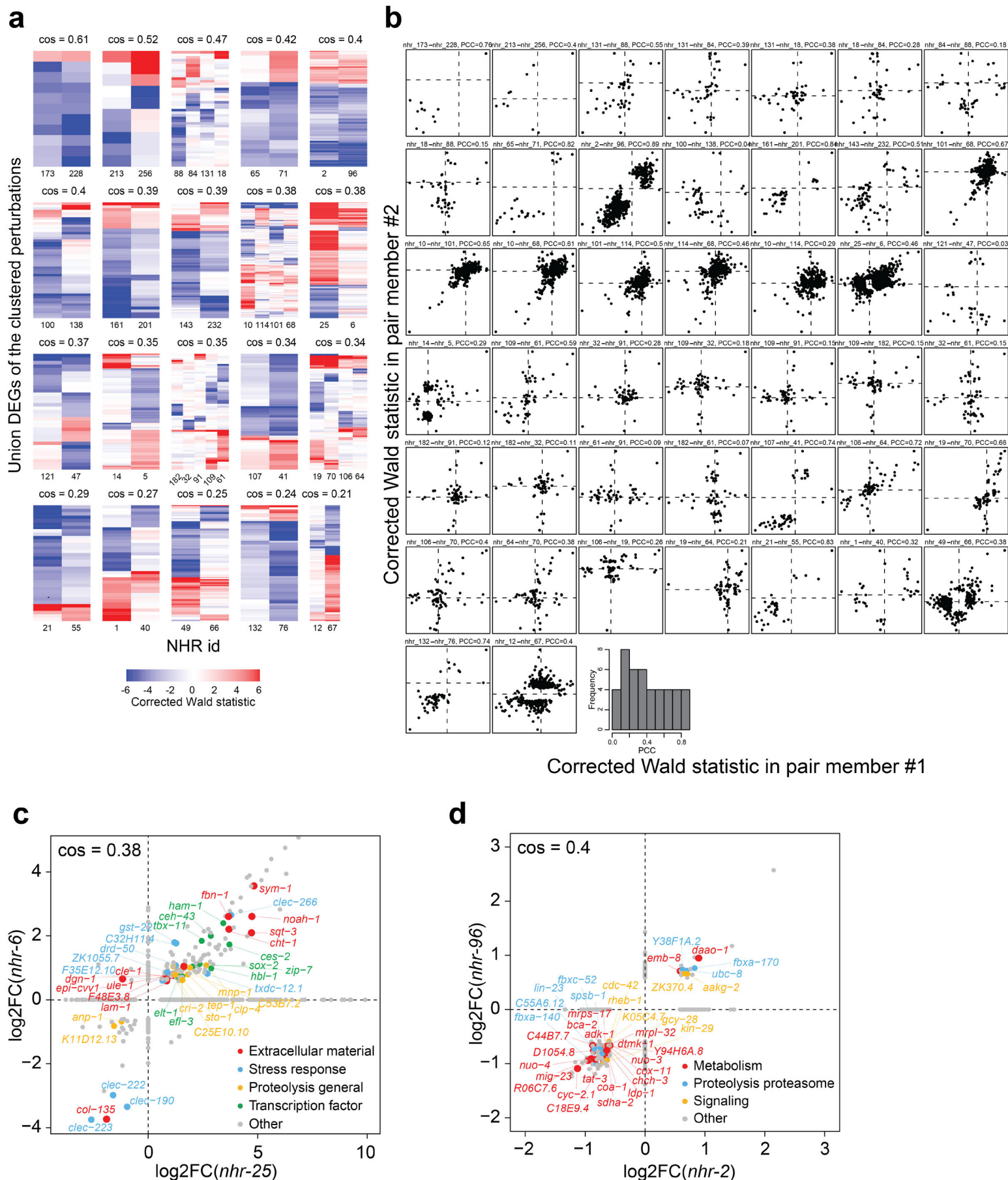All 82 NHR perturbations (81 NHR genes)
(by EmpirDE analysis)

All 82 NHR perturbations (81 NHR genes)
(randomized NHR GRN)

831

43

832 **Extended Data Fig. 7**: **Supplementary heatmaps of perturbation-perturbation**

833 **similarity.**

Extended Data Figure 8

835 **Extended Data Fig. 8: Visualization of the gene expression changes of selected**

836 **NHR perturbation clusters. a**, Heatmaps showing the corrected Wald statistic. The NHR

837 perturbations were grouped by clusters. These clusters were identified by cutting the

838 hierarchical tree in **Fig. 6a** using *cutree* function in R with h = 0.9. The union DEGs of

839 conditions in each cluster are shown in the y-axis. The numbers in x-axis indicate the

840 number in the corresponding *nhr* gene name. cos, cosine similarity. When there are more

841 than two conditions in a cluster, the average pairwise cosine similarity was labeled. **b**,

842 Scatter plots showing the correlation of expression changes in NHR pairs. The NHR pairs

843 were defined as the intra-cluster pairs in (**a**) and the corrected Wald statistics for the union

844 DEGs of the pair were shown in the plot. A histogram for the Pearson Correlation

845 Coefficient (PCC) of each pair is shown at the bottom. **c**, **d**, Comparison of filtered

846 log2(FC) profile for two example pairs of NHR perturbations. The union set of DEGs in

847 the two conditions are shown in the plot. cos, cosine similarity.

848