Control of neuronal terminal differentiation through cell context-dependent CFI-1/ARID3 functions

Yinan Li [1,2], Jayson J. Smith[1,4], Filipe Marques[1,4], Anthony Osuma[1,2], Hsin-Chiao Huang[3], Paschalis Kratsios[1,2,3,4]

[1] Department of Neurobiology, University of Chicago, Chicago, IL, USA

[2] Committee on Neurobiology, University of Chicago, Chicago, IL, USA

[3] Committee on Development, Regeneration and Stem Cell Biology, University of Chicago, Chicago, IL, USA

[4] University of Chicago Neuroscience Institute, Chicago, IL, USA

Correspondence: pkratsios@uchicago.edu

1

19
20
21 **SUMMARY**
22

23 ARID3 transcription factors are expressed in the nervous system, but their functions and

24 mechanisms of action are largely unknown. Here, we generated *in vivo* a genome-wide binding

25 map for CFI-1, the sole *C. elegans* ARID3 ortholog. We identified 6,396 protein-coding genes as

26 putative direct targets of CFI-1, most of which (77%) are expressed in post-mitotic neurons and

27 encode terminal differentiation markers (e.g., neurotransmitter receptors, ion channels,

28 neuropeptides). To gain mechanistic insights, we focused on two neuron types. In sensory

29 neurons (IL2 class), CFI-1 exerts a dual role: it acts directly to activate, and indirectly to repress,

30 distinct terminal differentiation genes. In motor neurons, however, CFI-1 acts directly as a

31 repressor, continuously antagonizing three transcriptional activators (UNC-3/Ebf, LIN-39/Hox4-

32 5, MAB-5/Hox6-8). By focusing on a glutamate receptor gene (*glr-4/GRIK1*), we found CFI-1

33 exerts its repressive activity through proximal binding to the *glr-4* locus. Further, the core DNA

34 binding domain of CFI-1 is partially required for *glr-4* repression in motor neurons. Altogether,

35 this study uncovers cell context-dependent mechanisms through which a single ARID3 protein

36 controls the terminal differentiation of distinct neuron types.

37
38
39

INTRODUCTION

Members of the ARID family of proteins are found in plants, yeast, fungi, and invertebrate and vertebrate animals (Kortschak et al., 2000, Patsialou et al., 2005, Wilsker et al., 2002, Wilsker et al., 2005). ARID family proteins are expressed either ubiquitously or in a tissue-specific fashion and control various biological processes, such as cell proliferation, differentiation, and embryonic patterning (Wilsker et al., 2002, Wilsker et al., 2005). Additionally, mutations in ARID family proteins are associated with cancer and several neurodevelopmental disorders (Shang et al., 2015, Bramswig et al., 2017, Kosho et al., 2014, Miyake et al., 2014, Smith et al., 2016, Lin et al., 2014).

Humans possess fifteen ARID family proteins, divided into seven subfamilies (ARID1-5, JARID1-2) based on the degree of sequence similarity (Wilsker et al., 2005). The **A**T-**R**ich **I**nteraction **D**omain (ARID), after which the family is named, was first identified in ARID3 proteins. These bind DNA in a sequence-specific manner, prefer AT-rich sequences, and are known to function as transcription factors (Wilsker et al., 2005). The ARID5 subfamily also encodes transcription factors (Patsialou et al., 2005, Wilsker et al., 2002), but the remaining five subfamilies (ARID1-2, ARID4, JARID1-2) encode proteins that bind DNA in a non-sequence-specific manner (Patsialou et al., 2005). For example, ARID1A, ARID1B, and ARID2 constitute subunits of the SWI/SNF (BAF/PBAF) chromatin-remodeling complex that can move and/or eject nucleosomes. Although the precise functions of all ARID proteins are not known, accumulating evidence suggests they can act either as positive or negative regulators of gene transcription, or as components of chromatin-remodeling complexes (Kortschak et al., 2000, Patsialou et al., 2005, Wilsker et al., 2002, Wilsker et al., 2005).

Mouse *Bright* (*Arid3a*) and *Drosophila dead ringer* (*retained*) are the founding members of the ARID family, and belong to the ARID3 subfamily, which is specific to metazoans. Single orthologs exist in *C. elegans* (CFI-1) and *Drosophila (dead ringer),* whereas mammals contain three orthologs (*ARID3A-C*) (**Fig. 1A**) (Kortschak et al., 2000, Wilsker et al., 2005, Gregory et al., 1996, Shaham and Bargmann, 2002). A defining feature of ARID3 proteins is the extended ARID (eARID) domain, a ~40 residue-long region next to the core ARID domain (Wilsker et al., 2002, Wilsker et al., 2005) (Kortschak et al., 2000). Structural studies showed both ARID and eARID domains contact DNA (Iwahara and Clubb, 1999, Iwahara et al., 2002).

ARID3 proteins have several early developmental roles, as identified by genetic studies. Mice lacking *Bright/Arid3a* display early embryonic lethality due to defects in hematopoiesis (Webb et al., 2011). *Bright/Arid3a* is best studied in B cell lineages, where it acts as an activator

74 and increases immunoglobulin transcription(Herrscher et al., 1995, Ratliff et al., 2014, Webb et

75 al., 2011, Webb et al., 1998). However, *Bright/Arid3a* is also critical for embryonic stem cell

76 differentiation (An et al., 2010, Popowski et al., 2014, Rhee et al., 2014). In this context, it can

77 act either as an activator or repressor of gene expression (Rhee et al., 2014). Similar to mice

78 lacking *Bright/Arid3a,* null mutants for *dead ringer* in *Drosophila* display early lethality (Shandala

79 et al., 1999, Shandala et al., 2002)*. Dead ringer* is essential for anterior-posterior patterning and

80 muscle development in the fly embryo, and can act either as an activator or repressor of gene

81 transcription(Hader et al., 2000, Shandala et al., 1999, Valentine et al., 1998). Lastly, *Arid3a* and

82 *Arid3b* have been associated with tumorigenesis by acting as direct inducers of cell cycle

83 regulators (Lestari et al., 2012, Saadat et al., 2021).

84 　　　　Genetic studies in *Drosophila* and *C. elegans* have also identified late developmental

85 roles of ARID3 proteins in the nervous system. In *Drosophila* larvae, *dead ringer* is expressed in

86 distinct neuron types and controls axonal pathfinding (Ditch et al., 2005, Shandala et al., 2003,

87 Sibbons, 2004), though its downstream targets in neurons - and thus, whether *dead ringer*

88 behaves as an activator or repressor – remain unknown. In *C. elegans*, *cfi-1* is selectively

89 expressed in head muscle and several neuron types: the IL2 sensory neurons, the AVD, PVC,

90 and LUA interneurons, and various classes of motor neurons (Shaham and Bargmann, 2002, Li

91 et al., 2020). Because *C. elegans* animals lacking *cfi-1* are viable (Shaham and Bargmann,

92 2002), these mutant strains provided a glance into the potential functions of ARID3 proteins

93 during post-embryonic life. Candidate approaches that examined a handful of effector genes

94 encoding neurotransmitter (NT) biosynthesis proteins and receptors suggested that CFI-1 acts

95 as an activator of gene expression both in sensory neurons (IL2) and interneurons (AVD, PVC)

96 (Shaham and Bargmann, 2002, Zhang et al., 2014, Ahn et al., 2022). However, in ventral nerve

97 cord motor neurons, CFI-1 is thought to act as a repressor of the glutamate receptor gene *glr-*

98 *4/GRIK1* (Kerk et al., 2017). The molecular mechanisms underlying the differential activities of

99 CFI-1 (activator versus repressor) in these distinct neuron types remain unknown. Elucidating

100 such mechanisms in *C. elegans* may provide clues as to how CFI-1 orthologs in other species

101 control cell differentiation. Lastly, unbiased approaches to identify the *in vivo* targets of CFI-1

102 (and any other ARID3 protein) in the nervous system are currently lacking, preventing a

103 comprehensive understanding of the neuronal functions controlled by ARID3 transcription

104 factors.

105 　　　　Here, we performed chromatin immunoprecipitation for CFI-1 followed by sequencing

106 (ChIP-Seq). By generating an *in vivo* binding map on the *C. elegans* genome, we identified

4

107   6,396 protein-coding genes as putative direct targets of CFI-1, the majority of which (77%) are

108   expressed in post-mitotic neurons. Gene ontology analysis suggests CFI-1 is primarily involved

109   in the process of neuronal terminal differentiation. To gain mechanistic insight into how CFI-1

110   controls the terminal differentiation of different neuron types, we focused on head sensory

111   neurons (IL2 class) and nerve cord motor neurons (DA, DB, VA, and VB classes). In sensory

112   IL2 neurons, CFI-1 exerts a dual role: it acts directly to activate and indirectly to repress distinct

113   terminal differentiation genes (e.g., NT receptors, ion channels). In nerve cord motor neurons,

114   however, CFI-1 acts directly to repress expression of the glutamate receptor gene *glr-4/GRIK1*.

115   CRISPR/Cas9-mediated mutagenesis of endogenous CFI-1 binding sites suggests proximal

116   binding to the *glr-4* locus is necessary for repression, advancing our understanding of ARID3-

117   mediated gene repression. Importantly, the core DNA binding domain of CFI-1 is partially

118   required for *glr-4* repression in motor neurons. Altogether, this study offers mechanistic insights

119   into cell context-dependent functions of CFI-1 (ARID3), a critical regulator of the terminal

120   differentiation program of distinct neuron types.

121

122

123 **RESULTS**
124
125 **A map of CFI-1/ARID3 binding on the *C. elegans* genome**
126 To identify CFI-1 binding events, we first generated an endogenous reporter strain through in-
127 frame insertion of the *flag* epitope sequence (*3xflag*) immediately after the *cfi-1* start codon (**Fig.**
128 **1B**). Immunostaining against FLAG on adult *3xflag::cfi-1* animals showed nuclear expression in
129 head muscle cells, as well as in neurons of the head, ventral nerve cord, and tail regions (**Fig.**
130 **1C-E**), indicating this reporter allele faithfully recapitulates the known expression pattern of *cfi-1*
131 (Kerk et al., 2017, Shaham and Bargmann, 2002). Unlike *cfi-1* null animals (Shaham and
132 Bargmann, 2002), homozygous *3xflag::cfi-1* animals do not display any defects in posterior
133 touch response (**Supplementary Fig. 1**). This suggests that insertion of the *3xflag* sequence
134 does not alter *cfi-1* gene function. We therefore conducted ChIP-Seq using a FLAG antibody on
135 homozygous *3xflag::cfi-1* animals at the third larval stage (L3), as all *cfi-1*-expressing cells are
136 generated by this stage.
137 　　　Our ChIP-Seq experiment revealed strong enrichment of CFI-1 binding in the *C. elegans*
138 genome, identifying 14,806 unique binding peaks (q-value cutoff: 0.05) (**Fig. 1F-G**). The CFI-1
139 peaks are predominantly located between 0 and 3kb upstream of transcription start sites (**Fig.**
140 **1H**), suggesting CFI-1 acts at promoter and enhancer regions to regulate gene expression.
141 Altogether, ChIP-Seq for CFI-1 generated the first *in vivo* binding map of an endogenously
142 tagged ARID3 protein, offering an opportunity to comprehensively identify the biological
143 processes controlled by CFI-1.
144
145 **The majority of CFI-1/ARID3 target genes encode neuronal terminal differentiation**
146 **markers**
147 Subsequent bioinformatic analysis of the 14,806 CFI-1 binding peaks revealed 6,396 protein-
148 coding genes as putative CFI-1 targets (see Materials and Methods). Because the majority of
149 *cfi-1*-expressing cells are neurons (**Fig. 1C-E**) (Kerk et al., 2017, Shaham and Bargmann,
150 2002), we reasoned that a significant portion of the 6,396 protein-coding genes may be
151 expressed in the nervous system. To test this, we used available single-cell expression profiles
152 (CeNGEN project: www.cengen.org) for all known *cfi-1*-expressing neurons (IL2, URA, AVD,
153 PVC, LUA, DA, DB, VA, VB, DD, VD), and indeed found that 77.1% of the global CFI-1 targets
154 (4,931 out of 6,396) are expressed in these neurons (**Fig. 1J, Supplementary file 1**). To gain
155 insights into the biological functions of CFI-1, we conducted gene ontology (GO) analysis with
156 PANTHER (Mi et al., 2013). Strikingly, the majority of CFI-1 target genes (~70%) encodes

157 proteins essential for neuronal terminal differentiation (e.g., NT receptors, transporters, ion

158 channels, transmembrane receptors, cell adhesion molecules) (**Fig. 1I-K, Supplementary file**

159 **2**). The second largest category (23% of CFI-1 targets) contains transcription factors, chromatin

160 factors, as well as proteins involved in DNA/RNA metabolism (**Fig. 1I-K**), suggesting CFI-1 can

161 affect gene expression indirectly through these factors. Altogether, the downstream targets

162 identified via our unbiased approach suggest that CFI-1 plays a prominent role in neuronal

163 terminal differentiation.

164

165 **CFI-1/ARID3 acts directly to activate terminal differentiation genes in IL2 sensory**

166 **neurons**

167 Although *cfi-1* is expressed in several neuron types, a handful of CFI-1 target genes have only

168 been identified in head sensory neurons of the IL2 class (**Fig. 2A-B**) (Shaham and Bargmann,

169 2002, Zhang et al., 2014). In these neurons, genetic experiments suggested CFI-1 influences

170 gene expression both positively and negatively (Shaham and Bargmann, 2002, Zhang et al.,

171 2014). CFI-1 activates various terminal differentiation genes (e.g., *cho-1/ChT, unc-17/VAChT,*

172 *gcy-19* [receptor-type guanylate cyclase]*, klp-6* [kinesin-like protein], *unc-5* [netrin receptor]),

173 and represses expression of two ion channel-encoding genes (*pkd-2/Polycystin-2 like 1*

174 [PKD2L1] and *lov-1/Polycystin-1 like 3* [PKD1L3]) associated with polycystic kidney disease

175 (Zheng et al., 2018b). It remained unknown, however, whether CFI-1 acts directly or indirectly to

176 control these genes. Leveraging our ChIP-Seq dataset, we found that CFI-1 binds directly to all

177 known terminal differentiation genes (e.g., *cho-1/ChT, unc-17/VAChT)* that require *cfi-1* gene

178 activity for their activation in IL2 neurons (**Fig. 2C**). However, we did not detect any binding in

179 the *cis*-regulatory regions of genes repressed by *cfi-1* (*pkd-2, lov-1*) (**Fig. 2D**). Altogether,

180 biochemical evidence (ChIP-Seq) combined with genetic studies (Shaham and Bargmann,

181 2002, Zhang et al., 2014) strongly suggest that in IL2 sensory neurons, CFI-1 exerts a dual role:

182 it directly activates a set of terminal differentiation genes, but indirectly (via intermediary factors)

183 represses the expression of ion channel-encoding genes (*pkd-2/Polycystin-2* and *lov-*

184 *1/Polycystin-1 like)* (**Fig. 2B**).

185   We next undertook an unbiased approach to investigate whether genes expressed in

186 mature IL2 neurons are also bound by CFI-1. To this end, we used available single-cell

187 expression profiles (CeNGEN project: www.cengen.org) and selected the most highly expressed

188 genes (top 1,000) in IL2 neurons. Strikingly, most of these genes (~70%) are bound by CFI-1

189 (**Fig. 2E**). Among them, GO analysis revealed an overrepresentation of terminal differentiation

7

190   genes (**Fig. 2E, Supplementary file 3**). This analysis provides biochemical evidence to support

191   the idea that CFI-1 directly activates scores of terminal differentiation genes in IL2 neurons.

192

193   **CFI-1/ARID3 acts directly to repress endogenous *glr-4/GRIK1* expression in cholinergic**

194   **motor neurons**

195   Our findings in IL2 sensory neurons suggest CFI-1 is a direct activator and indirect repressor of

196   gene expression (**Fig. 2B-D**). Next, we interrogated the function of CFI-1 in cholinergic nerve

197   cord motor neurons that control locomotion (**Fig. 3A**). Using an endogenous reporter allele (Li et

198   al., 2020), we found that *cfi-1* is selectively expressed in 29 cholinergic motor neurons (of the

199   DA, DB, VA, VB classes) located in the mid-body region (**Fig. 3A, Supplementary Fig. 2**). Next,

200   we asked whether in these neurons CFI-1 functions as an activator of gene expression. To this

201   end, we examined for *cfi-1* dependency five available motor neuron-specific terminal

202   differentiation markers (*twk-40, twk-43* [TWiK potassium channels]*; ncs-2* [neuronal calcium

203   sensor]*; npr-29* [neuropeptide]*; dbl-1* [Bmp-like]) (Li and Kratsios, 2021). These showed no

204   difference in expression in motor neurons of *cfi-1(-)* mutants (**Supplementary Fig. 3**).

205   Consistently, a previous study also found that three other terminal differentiation markers (*acr-5*

206   [acetylcholine receptor], *del-1* [SCNN1 sodium channel], *inx-12* [gap junction protein]) are not

207   affected in motor neurons of *cfi-1(-)* mutants (Kerk et al., 2017). Interestingly, our ChIP-seq data

208   indicate that these eight genes are all bound by CFI-1 (**Supplementary Fig. 3**), raising the

209   possibility of CFI-1 operating redundantly with other transcription factors to activate expression

210   of terminal differentiation genes in motor neurons.

211   The terminal differentiation gene *glr-4* (ortholog of human GRIK1 [glutamate inotropic

212   receptor kainite type subunit 1]) is the only known CFI-1 target in motor neurons, where it is

213   negatively regulated by CFI-1 (Kerk et al., 2017), providing an opportunity to obtain mechanistic

214   insights into how ARID3 proteins mediate gene repression in the nervous system. We therefore

215   carried out an in-depth investigation focused on *glr-4*, as detailed below.

216   Because previous studies employed transgenic reporters (Brockie et al., 2001, Kerk et

217   al., 2017), the endogenous expression pattern of *glr-4* in *C. elegans* neurons remained unclear.

218   We therefore generated an endogenous *glr-4* reporter allele by inserting the

219   *2xNLS::mScarlet::SL2* cassette immediately after the start codon (see Materials and Methods)

220   and established the *glr-4* expression pattern in motor neurons with single-cell resolution (**Fig.**

221   **3B**). Consistent with previous studies (Brockie et al., 2001, Kerk et al., 2017), we observed high

222   levels of *glr-4* (*mScarlet*) expression in head neurons, as well as in SAB motor neurons that

223 innervate head muscle (**Fig. 3A, C**). This endogenous reporter also revealed new sites of

224 expression. In the nerve cord of WT animals, we detected low levels of *glr-4* (*mScarlet*)

225 expression in 14 of the 29 *cfi-1*-expressing motor neurons, as well as in AS motor neurons,

226 which do not express *cfi-1* (**Fig. 3A-B,D, Supplementary Fig. 2**). Further, the observed levels

227 of *glr-4* expression in head (SAB) and ventral cord motor neurons were independently confirmed

228 by available scRNA-Seq data (CeNGEN project: www.cengen.org) (**Fig. 3E**).

229   Next, we tested whether endogenous *glr-4* expression in motor neurons depends on *cfi-1*

230 gene activity. Indeed, expression of the *glr-4* (*mScarlet*) reporter allele is increased in motor

231 neurons of *cfi-1* loss-of-function mutants, as we observed a higher number of cells expressing

232 *glr-4,* and at higher levels, compared to WT motor neurons (**Fig. 3D**). No effects on *glr-4* were

233 observed in SAB neurons, as they do not express *cfi-1* (arrowheads in **Fig. 3C**). Moreover, we

234 performed single-molecule mRNA fluorescent *in situ* hybridization (sm mRNA FISH) in WT and

235 *cfi-1* mutant animals. Loss of *cfi-1* led to increased levels of *glr-4* mRNA in nerve cord motor

236 neurons (**Fig. 3F**). Lastly, these results were corroborated by RT-PCR in WT and *cfi-1* mutant

237 animals (**Fig. 3G**). Altogether, we conclude CFI-1 limits the endogenous expression of *glr-*

238 *4/GRIK1* in nerve cord motor neurons. In WT animals, we can either detect low or no *glr-4*

239 expression in motor neurons, whereas loss of *cfi-1* results in robust *glr-4* expression in these

240 cells (**Fig. 3H**). Because ChIP-Seq revealed extensive CFI-1 binding immediately upstream of

241 the *glr-4* locus (**Fig. 3B**), we propose CFI-1 acts as a direct repressor of *glr-4/GRIK1*.

242

243 ***cfi-1/ARID3* is required to maintain *glr-4* repression in nerve cord motor neurons**

244 The *cfi-1*-expressing motor neurons (DA, DB, VA, VB) are generated in two waves (DA/DB are

245 born during embryogenesis; VA and VB at larval stage 1 [L1]) (**Fig. 4A-B**). Although all *cfi-1*-

246 expressing motor neurons have been generated by L2, we found no *glr-4* (*mScarlet)* expression

247 in WT animals at this stage. However, we observed a progressive increase in the number of WT

248 motor neurons expressing low levels of *glr-4* (*mScarlet)* at subsequent stages (L3, L4, Day 2

249 [D2] adult), indicating a correlation between *glr-4* expression and motor neuron maturation (**Fig.**

250 **4A**).

251   To test when *cfi-1* gene activity is required for repression, we monitored endogenous *glr-*

252 *4* (*mScarlet)* expression in motor neurons of WT and *cfi-1* mutant animals at larval (L2, L3, L4)

253 and adult (day 2) stages. Compared to controls, we identified a statistically significant increase

254 in the number of *glr-4* (*mScarlet)*-expressing motor neurons at L3, L4 and adult (day 2) stages

255 (**Fig. 4A**). Next, we used a conditional *cfi-1* allele (*mNG::3xFLAG::AID::cfi-1*) that enables

256   temporally controlled CFI-1 protein depletion upon administration of the plant hormone auxin (Li

257   et al., 2020, Zhang et al., 2015). Depletion of CFI-1 during the first 2 days of adulthood led to a

258   significant increase in the number of motor neurons expressing *glr-4* (*mScarlet*) (**Fig. 4B-C**),

259   suggesting *cfi-1* is continuously required to maintain *glr-4/GRIK1* repression in the adult (see

260   Discussion).

261

262   **cfi-1/ARID3 is sufficient to repress glr-4 expression**

263   To test whether *cfi-1* is sufficient to repress *glr-4*, we ectopically expressed *cfi-1* in the SAB

264   neurons. Using an SAB-specific-promoter (*unc-4*) to drive *cfi-1*, we observed a significant

265   decrease in the number of SAB neurons expressing a *glr-4* reporter gene (**Fig. 4D-E**), indicating

266   *cfi-1* is not only necessary (**Fig. 3**), but also sufficient to repress *glr-4* expression.

267

268   **CFI-1 binding sites proximal to glr-4 are necessary for repression in motor neurons**

269   CFI-1 binds to both proximal and distal *cis*-regulatory elements upstream of *glr-4* (**Fig. 5A**). To

270   precisely identify the elements through which CFI-1 mediates repression, we conducted *cis*-

271   regulatory analysis in the context of transgenic reporter animals. When *tagRFP* was driven by

272   distal regulatory elements (2.23kb or 938bp), we did not observe differences in the number of

273   *tagRFP* expressing motor neurons between WT and *cfi-1* mutants (**Fig. 5B**). However, we found

274   an increase in the number of *tagRFP*-expressing motor neurons in *cfi-1* mutants when *tagRFP*

275   was driven by a 3.7kb element (**Fig. 5B**), suggesting this element contains sequences

276   necessary for CFI-1 repression. A translational reporter (GLR-4::GFP) driven by a 4.9kb

277   element yielded similar results (**Fig. 5B**). Interestingly, a shorter *tagRFP* reporter (3.14kb) that

278   specifically lacks the most proximal CFI-1 binding peak did not show an increase in the number

279   of *tagRFP* expressing motor neurons in *cfi-1* mutants (**Fig. 5B**), suggesting proximal CFI-1

280   binding to the *glr-4* locus is needed for repression.

281   We next sought to determine whether proximal CFI-1 binding sites are required for *glr-4*

282   repression. The CFI-1 site (NNATHDNN) has been previously determined *in vitro* through

283   protein binding microarrays (**Fig. 5A**) (Weirauch et al., 2014). Within the most proximal region of

284   *glr-4*, we identified eleven predicted CFI-1 binding sites (**Fig. 5A**). To test their functionality, we

285   introduced nucleotide substitutions to all eleven sites in the context of the endogenous *glr-4*

286   (*mScarlet)* reporter through CRISPR/Cas9 genome editing (**Fig. 5C**). This manipulation nearly

287   phenocopied the *cfi-1* null mutant phenotype, as it led to a dramatic increase in the number of

288   *mScarlet*-expressing motor neurons (**Fig. 5C-D**). We conclude that CFI-1 binding sites located

289   in the proximal region of *glr-4* are necessary for its repression in motor neurons.

290

291   **The transcription factor UNC-3 (Collier/Ebf) and two Hox proteins (LIN-39, MAB-5)**

292   **activate basal levels of *glr-4/GRIK1* expression in motor neurons**

293   Because *glr-4* is expressed at basal levels in cholinergic motor neurons of WT animals (**Fig. 3**),

294   we reasoned this occurs due to CFI-1 antagonizing the function of *glr-4* activators in these

295   neurons. The transcription factor UNC-3 (Collier/Ebf) and the Hox proteins LIN-39

296   (Scr/Dfd/Hox4-5) and MAB-5 (Antp/Hox6-8) are known to act as transcriptional activators in

297   cholinergic motor neurons (Feng et al., 2020, Kerk et al., 2017) (**Fig.6A**), leading us to

298   hypothesize that they can also activate *glr-4* expression. Indeed, we found that expression of

299   the endogenous *glr-4* (*mScarlet*) reporter is reduced in *unc-3* loss-of-function mutant animals

300   (**Fig. 6B**). Importantly, the decrease of *glr-4* expression was significantly exacerbated in *unc-3;*

301   *lin-39; mab-5* triple mutants compared to *unc-3* single mutants, indicating that these three

302   factors cooperate to activate *glr-4* in cholinergic motor neurons (**Fig. 6B**). Interestingly, the Hox

303   requirement is only revealed in the absence of *unc-3* gene activity, as *glr-4* expression appears

304   normal in *lin-39; mab-5* double mutants.

305

306   **The most proximal UNC-3 biding site is necessary for *glr-4* expression in motor neurons**

307   Interrogation of available ChIP-Seq data for UNC-3, LIN-39 and MAB-5 showed overlapping

308   binding upstream of *glr-4* (**Fig. 6C**), suggesting a direct mode of activation by these factors. To

309   functionally test this notion, we focused on UNC-3 because its binding site (termed COE motif)

310   is well-defined in the *C. elegans* genome (Kratsios et al., 2012, Li et al., 2020). Through a

311   bioinformatic search (Materials and Methods), we found two COE motifs, one proximal and one

312   distal to the *glr-4* locus (**Fig. 6C**). Using CRISPR/Cas9 genome editing, we introduced

313   nucleotide substitutions to the proximal COE motif (COE1) in the context of the endogenous *glr-*

314   *4* (*mScarlet*) reporter(**Fig. 6C**). Animals carrying this *glr-4* (*mScarlet*) [COE1 MUT] reporter allele

315   showed a significant reduction in the number of *mScarlet*-expressing motor neurons,

316   reminiscent of the effect seen in *unc-3 (-)* null mutants (**Fig. 6D**). These data indicate that, in WT

317   animals, the most proximal COE motif is necessary for basal *glr-4* expression in motor neuros.

318

319

320

11

CFI-1 antagonizes the ability of UNC-3 to activate *glr-4* expression in motor neurons

Because UNC-3 activates basal *glr-4* expression in motor neurons of WT animals, we wondered whether it also controls the increased levels of *glr-4* expression observed in *cfi-1* mutants. We found this to be the case through double mutant analysis. The number of *mScarlet*-expressing cells in *cfi-1; unc-3* mutants is dramatically decreased compared to *cfi-1* single mutants (**Fig. 6E**). These data indicate that CFI-1 antagonizes the ability of UNC-3 to activate *glr-4* (*mScarlet*) expression in motor neurons.

Because the proximal UNC-3 binding site (COE motif) is required to activate *glr-4* expression in WT motor neurons (**Fig. 6D**), this site may also be necessary for the increased expression of *glr-4* in motor neurons of *cfi-1* mutants. Indeed, we observed a significant reduction in the number of *mScarlet*-expressing cells in *cfi-1* mutants carrying the *glr-4* (*mScarlet*) $^{COE1\ MUT}$ reporters compared to the intact version of the *glr-4* (*mScarlet*) reporter (**Fig. 6E**). Lastly, we hypothesized that the extensive binding of CFI-1 (four binding peaks identified by ChIP-seq) immediately upstream of the *glr-4* locus may limit the ability of UNC-3 to access the locus, resulting in basal levels of *glr-4* expression in WT motor neurons. To test this, we performed ChIP-Seq for UNC-3 in WT and *cfi-1* null mutant animals (at the L3 stage). We found that UNC-3 binding on the *glr-4* locus remains largely unaltered upon *cfi-1* loss (**Fig. 6C**), suggesting UNC-3 can access the locus independently of the presence of CFI-1.

The core ARID domain of CFI-1 is partially required for *glr-4* repression

To gain molecular insights into ARID3-mediated gene repression, we deleted portions of the CFI-1 DNA-binding domain and then conducted rescue assays to assess *glr-4* expression. ARID3 proteins are defined by the eARID domain, a ~40 residue-long highly conserved domain immediately following the core ARID domain (**Fig. 7A**). Because structural studies on Dead ringer showed that eARID contacts DNA (Iwahara and Clubb, 1999, Iwahara et al., 2002, Patsialou et al., 2005), we tested whether the CFI-1 eARID domain is required for *glr-4* repression. Transgenic expression of either WT CFI-1 or CFI-1 lacking the eARID domain (ΔeARID) in motor neurons of *cfi-1* null mutant animals led to complete rescue. That is, *glr-4* expression was no longer observed in motor neurons when either WT or ΔeARID CFI-1 were provided (p = 0.37) (**Fig. 7B**), indicating eARID is dispensable for CFI-1-mediated gene repression. Next, we mutated the helix-turn-helix (HTH) domain within the core ARID region, as the HTH domain of Dead ringer contacts the major groove of DNA (Iwahara and Clubb, 1999, Iwahara et al., 2002, Patsialou et al., 2005). Again, transgenic expression of CFI-1 lacking the

354   HTH domain (ΔHTH) in motor neurons of *cfi-1* mutants led to significant repression of *glr-4*

355   expression - the effect is comparable to WT CFI-1 (p = 0.05) (**Fig. 7B**). However, transgenic

356   expression of CFI-1 lacking the entire core ARID (ΔARID) domain (including the HTH domain) in

357   motor neurons of *cfi-1* mutants led to partial rescue, i.e., ΔARID CFI-1 did not completely

358   repress *glr-4* expression compared to WT CFI-1 (p = 0.0011) (**Fig. 7B**). Altogether, our analysis

359   suggests that the eARID and HTH domains of CFI-1 are dispensable, but the core ARID domain

360   is partially required for *glr-4* repression in motor neurons.

361

362
363

DISCUSSION

ARID3 proteins are the founding members of the ARID family, but their functions in the nervous system are poorly understood. Here, we performed ChIP-Seq for CFI-1, the sole ARID3 ortholog in *C. elegans*, and established its genome-wide binding map *in vivo*. We found that most CFI-1 target genes (77%) are expressed in post-mitotic neurons and encode proteins critical for terminal differentiation and neuronal function (e.g., neurotransmitter receptors, ion channels, neuropeptides). Further, our study offers mechanistic insights into how a single ARID3 protein controls the terminal differentiation program of distinct neuron types, uncovering cell context-depending CFI-1 functions.

ARID3 proteins bind to proximal and distal *cis*-regulatory regions
In vitro assays like PBM (protein binding microarrays) and SELEX (systematic evolution of ligands by exponential enrichment) have defined the binding site of CFI-1, Dead ringer and Arid3a (Mathelier et al., 2014, Nitta et al., 2015, Weirauch et al., 2014). To assess their mechanism of action, however, unbiased assays to monitor genome-wide binding of ARID3 proteins remain necessary. ChIP-Seq in *C. elegans* for endogenously tagged CFI-1 revealed binding events predominantly located between 0 and 3kb upstream of transcription start sites (**Fig. 1H**). Because the size of intergenic regions for most *C. elegans* genes is less than 3kb (Dupuy et al., 2004, Nelson et al., 2004), our *in vivo* binding map strongly suggests that CFI-1 can act both at proximal (e.g., promoters) and distal (e.g., enhancers) regions to regulate gene expression. Consistently, a previous *in vitro* study found that overexpressed Arid3a in mouse embryonic stem cells also binds to both proximal and distal *cis*-regulatory elements (Rhee et al., 2014).

The genome-wide binding map suggests a prominent role for CFI-1 in neuronal terminal differentiation
Our current knowledge of ARID3 functions in the nervous system remains rudimentary. In vertebrate nervous systems, the role of ARID3 proteins is completely unknown. In *Drosophila*, *dead ringer* has been implicated in the control of axonal pathfinding (Ditch et al., 2005, Shandala et al., 2003, Sibbons, 2004). In *C. elegans*, a handful of effector genes have been identified as CFI-1 targets in distinct neuron types (Glenwinkel et al., 2021, Kerk et al., 2017, Shaham and Bargmann, 2002, Zhang et al., 2014). Hence, a comprehensive understanding of ARID3-mediated biological processes in the nervous system is lacking.

14

398 To address this knowledge gap, we globally identified 6,396 protein-coding genes as

399 putative direct targets of CFI-1, most of which (77%) are expressed in post-mitotic *C. elegans*

400 neurons. GO analysis suggested a prominent role for CFI-1 in neuronal terminal differentiation,

401 as ~70% of its target genes encode NT receptors, transporters, ion channels, transmembrane

402 receptors, cell adhesion molecules, etc. Moreover, 23% of CFI-1 targets encode transcription

403 factors, chromatin factors, and proteins involved in DNA/RNA metabolism, suggesting CFI-1 can

404 affect gene expression indirectly through these factors. Altogether, our ChIP-Seq dataset

405 illuminates the biological processes under the control of an ARID3 protein in *C. elegans*.

406

407 **CFI-1 acts as a terminal selector in IL2 sensory neurons**

408 Transcription factors that bind directly to the *cis*-regulatory region of terminal differentiation

409 genes (e.g., NT biosynthesis components, NT receptors, ion channels, neuropeptides,

410 membrane proteins) and activate their expression have been termed "terminal selectors"

411 (Hobert, 2008, Hobert, 2016b, Hobert and Kratsios, 2019). Terminal selectors are continuously

412 required in individual neuron types to initiate and maintain expression of terminal differentiation

413 genes, thereby safeguarding neuronal functionality throughout life. To date, terminal selectors

414 have been described in *C. elegans*, *Drosophila*, simple chordates and mice (Hobert and

415 Kratsios, 2019), suggesting an evolutionarily conserved role for these critical regulators of

416 neuronal differentiation. For most terminal selectors, however, biochemical evidence for direct

417 binding to their target genes is currently lacking. For example, CFI-1 is a candidate terminal

418 selector in IL2 sensory neurons because five terminal differentiation genes (*cho-1/ChT, unc-*

419 *17/VAChT, gcy-19* [receptor-type guanylate cyclase], *klp-6* [kinesin-like protein], *unc-5* [netrin

420 receptor]) fail to be properly expressed in *cfi-1* mutants (Zhang et al., 2014). Here, we provide

421 biochemical evidence that CFI-1 binds directly to all these genes, indicating CFI-1 acts as a

422 direct activator. Further, our analysis of the top 1,000 highly expressed genes in IL2 neurons

423 revealed that the majority of CFI-1 binding (~70%) occurs at terminal differentiation genes,

424 consistent with an *in silico* prediction study for CFI-1 binding (Glenwinkel et al., 2021).

425 Altogether, a synthesis of our findings with the aforementioned studies indicate that CFI-1

426 functions as a bona fide terminal selector and directly activates IL2 terminal differentiation

427 genes. This mode of action is reminiscent of mouse Arid3a, which is known to also act as a

428 direct activator of gene expression in cells outside the nervous system (Herrscher et al., 1995,

429 Ratliff et al., 2014, Rhee et al., 2014).

430 Accumulating evidence suggests that terminal selectors act in combination with other

431 transcription factors to determine the differentiation of individual neuron types (Glenwinkel et al.,

15

432 2021, Lloret-Fernandez et al., 2018). Supporting this notion, genetic and *in silico* prediction

433 studies found that CFI-1 collaborates with the POU homeodomain transcription factor UNC-86

434 to activate IL2-specific terminal differentiation genes (Glenwinkel et al., 2021, Zhang et al.,

435 2014). Similarly, it was proposed that CFI-1 collaborates with two different homeodomain

436 proteins, UNC-42/Prop-1 like and CEH-14/LIM, to respectively control the terminal differentiation

437 of AVD and PVC interneurons (Berghoff et al., 2021, Glenwinkel et al., 2021), albeit the

438 underlying mechanisms remain unclear.
439
440 **Insights into ARID3-mediated gene repression**

441 In IL2 sensory neurons, CFI-1 exerts a dual role (**Fig. 2**). It functions as a direct activator of IL2-

442 specific terminal differentiation genes and indirect repressor of ion channel-encoding genes

443 (*pkd-2, lov-1*), which are normally expressed in CEM neurons (cells responsible for pheromone

444 detection) (Chasnov et al., 2007). Hence, it promotes IL2 terminal differentiation and inhibits an

445 alternative neuronal identity (CEM). However, in nerve cord motor neurons, CFI-1 functions as a

446 direct repressor of the glutamate receptor-encoding gene *glr-4/GRIK1*. Altogether, these

447 observations indicate that a single ARID3 protein can control the terminal differentiation program

448 of distinct neuron types through mechanisms that depend on cell context, possibly due to CFI-1

449 participating in distinct neuron type-specific regulatory complexes that function either as

450 dedicated activators or repressors.

451   Through an in-depth regulatory analysis of *glr-4/GRIK1*, our findings advance our

452 understanding of how ARID3 proteins mediate gene repression in several aspects. First, *cis*-

453 regulatory analysis combined with mutagenesis of endogenous CFI-1 binding sites strongly

454 suggests that proximal CFI-1 binding to the *glr-4* locus is required for repression (**Fig. 5**).

455 Second, CFI-1 antagonizes three conserved transcription factors (UNC-3/Ebf, LIN-39/Hox4-5,

456 MAB-5/Hox6-8) that directly activate basal levels of *glr-4/GRIK1* expression in nerve cord motor

457 neurons (**Fig. 6**). Third, ChIP-Seq for UNC-3 in WT and *cfi-1(-)* mutants did not reveal any

458 changes in the UNC-3 binding pattern to *glr-4*, arguing against a model in which *glr-4* is

459 repressed because the repressor (CFI-1) competes with the activator (UNC-3) for binding to the

460 same *cis*-regulatory elements. Lastly, we found that deletion of highly conserved protein

461 domains (eARID, HTH) predicted to bind DNA (based on Dead ringer structural studies) did not

462 affect the ability of CFI-1 to repress *glr-4* expression (**Fig. 7**). This suggests that either eARID

463 and HTH act redundantly, or other CFI-1 domains are responsible for *glr-4* repression.

464 Supporting the latter, deletion of the entire core ARID domain (including the HTH domain) of

465 CFI-1 led to a partial failure to repress *glr-4* in motor neurons.

16

466   One unresolved question is how exactly CFI-1 represses gene expression in *C. elegans*

467 motor neurons. Because proximal binding is required for *glr-4* repression, it is possible that CFI-

468 1 interferes with the function of the basal transcription complex (Baumann et al., 2010). Another

469 possibility stems from patterning studies in the *Drosophila* embryo, where *Dead ringer* binds to

470 Groucho, a well-characterized transcriptional corepressor (Hader et al., 2000, Valentine et al.,

471 1998). In mouse stem cells, Arid3a directly represses pluripotency genes by recruiting histone

472 deacetylases (HDACs)(Rhee et al., 2014). Future studies are needed to determine whether CFI-

473 1 acts through any of these repressive mechanisms in motor neurons.

474

475 **CFI-1 continuously antagonizes the activator function of terminal selectors**

476 The *glr-4* gene receives positive regulatory input from three conserved transcription factors

477 (UNC-3/Ebf, LIN-39/Hox4-5, MAB-5/Hox6-8) and negative input from CFI-1. UNC-3 and the Hox

478 proteins LIN-39 and MAB-5 are known terminal selectors in *C. elegans* nerve cord motor

479 neurons (Feng et al., 2020, Kratsios et al., 2017, Kratsios et al., 2012). As such, they are

480 continuously required from embryonic to adult stages to activate expression of multiple terminal

481 differentiation genes (Feng et al., 2020, Li et al., 2020). On the other hand, our constitutive (null

482 alleles) and conditional (temporally controlled protein depletion) approaches also revealed a

483 continuous requirement for CFI-1 in motor neurons (**Fig. 4**). This argues against a transient role

484 where, for example, CFI-1 establishes a repressive chromatin environment during early

485 development and its activity becomes unnecessary at later stages of life. Instead, we favor a

486 model where CFI-1 is required continuously to prevent high levels of *glr-4* expression driven by

487 terminal selectors (UNC-3, Hox) (**Fig. 7C**). Such model is also supported by the continuous

488 requirement of two other repressor proteins (BNC-1/Bnc1, MAB-9/Tbx20) in *C. elegans* motor

489 neurons (Kerk et al., 2017). Antagonism between repressor proteins and terminal selectors has

490 also been reported in *C. elegans* touch receptor neurons with EGL-44/Tead3 and EGL-46/Insm2

491 repressing terminal selector target genes (Zheng et al., 2018a). Further, studies in mice

492 indicated that the activity of two terminal selectors, Nurr1 and Crx, which respectively control

493 dopamine neuron and photoreceptor identities, is counteracted by repressor proteins (Otx2,

494 Nr2e3)(Di Salvio et al., 2010, Peng et al., 2005). Additional work is needed, however, to

495 determine whether all these repressor proteins (e.g., EGL-44, Otx2) act directly and are

496 continuously required.

497

498 **Evolutionary implications**

17

499   The terminal selector UNC-3 is required to maintain *cfi-1* expression in nerve cord motor

500   neurons (Li et al., 2020). Hence, the repressor protein (CFI-1) and *glr-4* are both targets of

501   UNC-3, thereby generating an incoherent feedforward loop (FFL) (**Fig. 7C**). In SAB motor

502   neurons, however, CFI-1 is not expressed, and UNC-3 is able to drive high levels of *glr-4*

503   expression (**Fig. 7C**)(Kratsios et al., 2015). From an evolutionary perspective, incoherent FFLs

504   have been proposed to diversify a ground state into various substates (Hobert, 2016a). One can

505   envision an ancestral state where an UNC-3 ortholog is present in a relatively homogeneous

506   population of motor neurons, but the recruitment of a repressor (CFI-1) enabled their

507   diversification. Hence, the UNC-3 → CFI-1 –I *glr-4* incoherent FFL may help distinguish, at the

508   molecular level, nerve cord motor neurons that control locomotion from SAB motor neurons that

509   control head movement. In agreement with this idea, incoherent FFLs are known to diversify

510   gustatory neurons in *C. elegans* and photoreceptor cells in *Drosophila* (Etchberger et al., 2009,

511   Johnston, 2013).

512

513

514

18

515   **MATERIALS AND METHODS**

516

517   ***C. elegans* strain culture**

518   Worms were grown at 20˚C or 25˚C on nematode growth media (NGM) plates supplied with *E.*

519   *coli* OP50 as food source (Brenner, 1974). All strains used or generated for this study are listed

520   in **Supplementary File 4**.

521

522   **Generation of transgenic animals carrying transcriptional fusion reporters and**

523   **overexpression or rescue constructs**

524   Reporter gene fusions for *cis*-regulatory analyses of *glr-4* were made with PCR fusion(Hobert,

525   2002). Genomic regions were amplified and fused to the coding sequence of *tagrfp* followed by

526   the *unc-54* 3' UTR. PCR fusion DNA fragments were injected into young adult *pha-1(e2123)*

527   hermaphrodites at 50 ng/μl together with *pha-1* (pBX plasmid) as co-injection marker (50 ng/μl).

528   To generate animals with *cfi-1* overexpression in the SAB neurons, the *unc-4* promoter was

529   fused to the cDNA sequence of *cfi-1* followed by the *unc-54* 3' UTR. The fluorescent co-injection

530   marker *myo-2::gfp* was used (2 ng/ul) and the PCR fusion DNA fragments were injected into

531   young adult animals carrying the *glr-4::tagrfp* reporter at 50 ng/ul. To generate transgenic

532   animals carrying different versions of the *cfi-1* cDNA rescue constructs (WT, ΔARID, ΔeARID,

533   ΔHTH), the *cfi-1* enhancer driving expression in motor neurons was fused to the corresponding

534   version of *cfi-1* cDNA followed by the *unc-54* 3' UTR. The fluorescent co-injection marker *myo-*

535   *2::gfp* was used (2 ng/ul) and the PCR fusion DNA fragments were injected into young adults of

536   *cfi-1(-)* mutants carrying the *glr-4::tagrfp* reporter at 50 ng/ul.

537

538   **Targeted genome editing**

539   The endogenous *glr-4* reporter allele *syb3680 [2xNLS::mScarlet::glr-4]* was generated by

540   SunyBiotech via CRISPR/Cas9 genome editing by inserting the *2xNLS::mScarlet* cassette

541   immediately after the ATG of *glr-4*. Moreover, the endogenous *glr-4* reporter allele *syb5348*

542   [*2xNLS::mScarlet::SL2::glr-4*$^{11\ CFI-1\ sites\ MUT}$] that carries nucleotide substitutions in eleven CFI-1

543   binding sites was also generated by SunyBiotech. The endogenous *glr-4* reporter allele *kas29*

544   [*2xNLS::mScarlet::SL2::glr-4*$^{COE1\ MUT}$] that carries nucleotide substitutions in a single

545   UNC-3 binding site (COE1 motif) was generated in the Kratsios lab by using homology

546   dependent repair and inserting a synthesized DNA fragment that carries the desired mutations.

547

Microscopy

Imaging slides were prepared by anesthetizing worms with sodium azide (NaN$_3$, 100 mM) and mounting them on a 4% agarose pad on glass slides. Images were taken with an automated fluorescence microscope (Zeiss, Axio Imager Z2). Images containing several z stacks (0.50 µm intervals between stacks) were taken with Zeiss Axiocam 503 mono using the ZEN software (Version 2.3.69.1000, Blue edition). Representative images are shown following max-projection of 2-5 µm Z-stacks using the maximum intensity projection type. Image reconstruction was performed with Image J (Schindelin et al., 2012).

Motor neuron subtype identification

Motor neuron subtypes were identified based on combinations of the following factors: [1] co-localization with or exclusion from additional reporter transgene with known expression patterns; [2] Invariant position of neuronal cell bodies along the ventral nerve cord, [3] Birth order of specific motor neuron subtypes (e.g., during embryonic or post-embryonic stages); [4] Total cell numbers in each motor neuron subtype.

Bioinformatic prediction of binding motifs

Information of the CFI-1 binding motif is curated in the Catalog of Inferred Sequence Binding Preferences database (http://cisbp.ccbr.utoronto.ca). To predict and identify CFI-1 binding motifs in the *glr-4* promoter, we utilized tools provided by MEME (Multiple Expectation maximization for Motif Elicitation) bioinformatics suite (http://meme-suite.org/), and performed FIMO (Find Individual Motif Occurrences) motif scanning analysis.

Chromatin Immunoprecipitation (ChIP)

ChIP assay was performed as previously described, with the following modifications (Yu et al. 2017; Zhong et al. 2010). Synchronized L1 *cfi-1(syb1778[3xFLAG::cfi-1])* worms and N2 worms were cultured on 10 cm plates seeded with OP50 at 20°C overnight. Late L2 worms were cross-linked and resuspended in FA buffer supplemented with protease inhibitors (150 mM NaCl, 10 µl 0.1 M PMSF, 100 µl 10% SDS, 500 µl 20% N-Lauroyl sarsosine sodium, 2 tablets of cOmplete ULTRA Protease Inhibitor Cocktail [Roche Cat.# 05892970001] in 10ml FA buffer). For each IP experiment, 200 µl worm pellet was collected. The sample was then sonicated using a Covaris S220 with the following settings: 200 W Peak Incident Power, 20% Duty Factor, 200 Cycles per Burst for 1 min. Samples were transferred to centrifuge tubes and spun at the

581 highest speed for 15 min. The supernatant was transferred to a new tube, and 5% of the

582 material was saved as input and stored at -20°C. The remainder was incubated with FLAG

583 antibody at 4°C overnight. Wild-type (N2) worms do not carry the 3xFLAG tag and serve as

584 negative control. The *cfi-1(syb1778[3xFLAG::cfi-1])* CRIPSR generated allele was used in order

585 to immunoprecipitate the endogenous CFI-1 protein. On the next day, 20 µl Dynabeads Protein

586 G (1004D) was added to the immunocomplex which was then incubated for 2 hr at 4°C. The

587 beads were then washed at 4°C twice with 150 mM NaCl FA buffer (5 min each), once with 1M

588 NaCl FA buffer (5 min). The beads were transferred to a new centrifuge tube and washed twice

589 with 500 mM NaCl FA buffer (10 min each), once with TEL buffer (0.25 M LiCl, 1% NP-40, 1%

590 sodium deoxycholate, 1mM EDTA, 10 mM Tris-HCl, pH 8.0) for 10 min, twice with TE buffer (5

591 min each). The immunocomplex was then eluted in 200 µl elution buffer (1% SDS in TE with

592 250 mM NaCl) by incubating at 65°C for 20 min. The saved input samples were thawed and

593 treated with the ChIP samples as follows. One (1) µl of 20 mg/ml proteinase K was added to

594 each sample and the samples were incubated at 55°C for 2 hours then 65°C overnight (12-20

595 hours) to reverse cross-link. The immonuprecipitated DNA was purified with Ampure XP beads

596 (A63881) according to manufacturer's instructions.

597

598 **ChIP-seq data analysis**

599 Unique reads were mapped to the *C. elegans* genome (ce10) with bowtie2 (Langmead and

600 Salzberg 2012). Peak calling was then performed with MACS2 (minimum q-value cutoff for peak

601 detection: 0.005) (Zhang et al. 2008). For visualization purposes, the sequencing depth was

602 normalized to 1x genome coverage using bamCoverage provided by deepTools (Ramírez et al.

603 2016) and peak signals were shown in Integrated Genome Viewer (Siponen et al.). Heatmap of

604 peak coverage in regard to CFI-1 enrichment center was generated with NGSplot (Shen et al.

605 2014). The average profile of peaks binding to TSS region was generated with ChIPseeker (Yu

606 et al. 2015).

607

608 **Enrichment of CFI-1 targets in IL2 expressed genes**

609 The top 1,000 highest expressed genes in IL2 (transcripts per million, tpm) were mined from

610 available single-cell RNA-sequencing data (CenGEN). This dataset was computationally

611 compared to a dataset of CFI-1 ChIP-Seq targets using the 'semi_join' function in R (package

612 Dplyr 1.0.7). This generated a new data frame containing genes in the scRNA-seq dataset that

613 are also putatively bound by CFI-1. Similarly, the 'set_diff' function (Dplyr 1.0.7) was used to

614   generate a new data frame containing genes that are expressed in IL2 based on scRNA-seq but

615   which are not found in the CFI-1 ChIP-seq dataset. Gene list analysis (PANTHER 17.0) was

616   performed on both data frames to functionally classify all genes based on protein class ontology.

617

618   **Temporally controlled protein degradation**

619   Temporally controlled protein degradation was achieved with the auxin-inducible degradation

620   system (Zhang et al., 2015). TIR1 expression was driven by the pan-neuronal promoter in the

621   transgene *otTi28[unc-11prom8+ehs-1prom7+rgef-1prom2::TIR1::mTurquoise2::unc-54 3'UTR]*.

622   To induce degradation of CFI-1 proteins, we used the allele *kas16[cfi-1::mNG::AID]*. Worms at

623   the L4 stage were grown at 20 ˚C on NGM plates coated with 4 nM auxin (indole-3-acetic acid

624   [IAA] dissolved in ethanol) or ethanol (negative control) for 2 days before testing (see figure

625   legends for exact time in specific experiments). All plates were shielded from light.

626

627   **Statistical analysis**

628   For data quantification, graphs show values expressed as mean ± standard error mean of

629   animals. The statistical analyses were performed using the unpaired t-test (two-tailed).

630   Calculations were performed using the GraphPad QuickCalcs online software

631   (http://www.graphpad.com/quickcalcs/). Differences with $p<0.05$ were considered significant.

632

633   **Single molecule RNA fluorescent *in situ* hybridization (sm RNA FISH)**

634   Synchronized L1 worms were collected from the plates and washed with M9 buffer 3 times.

635   Worms were incubated in the fixation buffer (3.7% formaldehyde in 1x PBS) for 45 minutes at

636   room temperature. Worms were then washed twice with 1x PBS, resuspended in 70% ethanol

637   and left at 4˚C for two nights. After removing the ethanol, worms were incubated in the wash

638   buffer (10% formamide in 2x SSC buffer) for 5 minutes and the wash buffer was removed

639   afterwards. A *glr-4* probe was designed using the Stellaris Probe Designer website (Biosearch

640   Technologies). The probe was mixed in hybridization buffer (0.1 g/ml dextran sulfate [Sigma

641   D8906-50G], 1 mg/ml *Escherichia coli* tRNA [ROCHE 10109541001], 2 mM vanadyl

642   ribonucleotide complex [New England Biolabs S1402s], 0.2 mg/ml RNase-free BSA [Ambion

643   AM2618], 10% formamide) and added to the worms. The hybridization buffer was removed and

644   worms were washed twice in wash buffer (DAPI was added during the second wash and

645   incubated for 30 minutes in the dark for nuclear counterstaining). Worms were washed once in

646   2x SSC, incubated in GLOX buffer (0.4% glucose, 0.1 M Tris-HCl, 2x SSC) for 2 minutes for

22

647 equilibration, and the resuspended in GLOX buffer with glucose oxidase and catalase added.

648 The samples were then examined under the fluorescent microscope.

649

650 **Real-time PCR assay for *glr-4* expression level analysis**

651 Synchronized L4 stage wildtype and *cfi-1(-)* worms were collected, and mRNA was extracted.

652 cDNA library was prepared using the Superscript first strand cDNA synthesis kit (Invitrogen

653 #11904-018). RT-PCR TaqMan assays for the genes *glr-4* (assay ID: Ce02435302_g1) and

654 *pmp-3* (Ce02485188_m1) were performed, and the expression level of *glr-4* was determined in

655 each genotype after normalizing to the expression of the housekeeping gene *pmp-3*.

656

657 **Harsh touch behavioral assay**

658 Harsh touch was delivered with a platinum wire pick as previously described (Marques et al.,

659 2019). The stimulus was applied from above the animals by pressing down with the edge of the

660 pick on the tail of non-moving adult animals. Each animal was tested only once. Worms that

661 moved forward in response to harsh touch were scored as normal response. Results are

662 presented as fractions of animals that responded normally.

663

664
665
666

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Y. L., Conceptualization, Data curation, Investigation, Visualization, Methodology, Writing—review and editing; J.J.S. Formal analysis, Validation, Investigation, Writing—review and editing; F.M., A.O., H.C.H., Formal analysis, Validation, Investigation; P. K., Conceptualization, Supervision, Investigation, Funding acquisition, Project administration, Writing— original draft, review and editing.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY

Sequencing data have been deposited in GEO under accession code GSE 205628. Moreover, all data generated or analyzed for this study are included in the manuscript and supporting files.

FIGURE LEGENDS

Figure 1. Mapping genome-wide CFI-1 binding with ChIP-seq.

A: Schematics of the coding sequences of CFI-1 and its *Drosophila* homolog *Dead ringer (Dri)* and Mouse homologs Arid3a-c. The ARID (yellow) and REKLES (blue) domains are highlighted.

B: (Doitsidou et al.) Diagram of the *3xflag::cfi-1* allele. The endogenous CFI-1 proteins are tagged with 3xFLAG which are inserted immediately after the ATG. (Bottom) Schematic of *C. elegans* with dashed line boxes highlighting regions where immunostaining images are shown in C-E.

C-E: The expression of 3xFLAG::CFI-1 fusion protein is confirmed by immunostaining (DAPI, blue; anti-FLAG, red) in the ventral cord motor neurons (C), muscles (indicated by arrowheads) and neurons in the head (D), and neurons in the tail (E).

F: Fingerprint plot indicating localized, strong enrichment of CFI-1 binding events in the genome. Specifically, when counting the reads contained in 86% of all genomic bins, only 50% of the maximum number of reads are reached, which indicates 14% of the genome contain half of total reads.

G: Heatmap of CFI-1 ChIP-seq signal around 1.0 kb of the center of the binding peak.

H: Summary plot of CFI-1 ChIP-seq signal with a 95% confidence interval (grey area) around 3.0 kb of the transcription start site (Kadkhodaei et al.). The average signal peak is detected at ~140 bp upstream of the TSS.

I: Graph summarizing protein class ontology analysis of global CFI-1 target genes identified by ChIP-seq. A total number of 7,995 genes are analyzed, 4,984 of which hit known protein class terms.

J: Pie chart summarizing three main categories of genes bound by CFI-1.

K: Venn diagram showing that 77.1% of the protein-coding genes (4,931 out of 6,396) that are bound by CFI-1 are expressed in the nervous system based on RNA-Seq data from the CenGEN project.

Figure 2. CFI-1 directly activates terminal differentiation genes in the IL2 sensory neurons.

A: Diagram showing the location of the IL2 sensory neurons.

B: Model summarizing known targets of CFI-1 in the IL2 sensory neurons.

C: CFI-1 binding signals at the loci of genes activated by CFI-1 in the IL2 neurons.

D: CFI-1 binding signals at the loci of genes repressed by CFI-1 in the IL2 neurons.

721 **E:** Protein class ontology analysis of genes expressed in IL2 sensory neurons binned by CFI-1

722 binding based on ChIP-seq data. Bar colors correspond to colors in Fig 1I: Blue, genes involved

723 in terminal differentiation; yellow, genes involved in regulating gene expression; green,

724 miscellaneous genes.

725

726 **Figure 3. CFI-1/Arid3a represses *glr-4*/GRIK1 in the ventral cord motor neurons**

727 **A:** Summary of the endogenous expression patterns of cfi-1 and glr-4 in the five cholinergic

728 nerve cord motor neuron subtypes and the SAB neurons in the retrovesicular ganglion (RVG).

729 The expression patterns are determined by colocalization with neuron subtype-specific

730 reporters. Filled boxes represent positive expression, while empty boxes indicate no detectable

731 expression.

732 **B:** CFI-1 binding signal at the *glr-4* locus. (Bottom) Design of an endogenously tagged, nuclear

733 localized fluorescent reporter allele of *glr-4* (*2xNLS::mScarlet::SL2::glr-4*). The reporter cassette

734 is inserted immediately downstream of the ATG of *glr-4* with CRISPR/Cas9.

735 **C:** Representative fluorescent micrographs showing that the expression of

736 2xNLS::mScarlet::SL2::GLR-4 is unaffected in SAB neurons (arrowheads) in *cfi-1(-)* mutants.

737 Bright white signal to the left of the SAB neurons indicates *glr-4* expression in head neurons.

738 **D:** *2xNLS::mScarlet::SL2::glr-4* expression in the ventral cord motor neurons (indicated by

739 arrowheads in the representative images) in WT and *cfi-1(-)* animals at L4 stage. Ectopic

740 expression of *glr-4* was detected in *cfi-1(-)* mutants. (Bottom) Each dot in the quantification

741 graph of total number of motor neurons (left) represents an individual animal. Each dot in the

742 fluorescence intensity quantification graph (right) represents an individual motor neuron that

743 shows *glr-4* expression. For simplicity, only motor neurons in the anterior VNC are included in

744 the fluorescence intensity quantification. p-values are indicated in the graphs. N ≥ 15.

745 **E:** RNA-seq data from the CeNGEN project showing expression of *glr-4* transcripts in SAB, DB,

746 VB, AS, VA, and DA motor neurons.

747 **F:** Single molecule fluorescent *in situ* hybridization (smFISH) shows mild ectopic expression of

748 *glr-4* transcripts in the motor neurons of L1 *cfi-1(-)* animals. Left: represented images showing

749 ectopic *glr-4* mRNA molecules (Cy5, red) in the nucleus (DAPI, blue) of a motor neuron in *cfi-1(-*

750 *)* mutants. Right: quantification of the number of *glr-4* transcripts detected in the anterior ventral

751 nerve cord in WT and *cfi-1(-)* animals. p-value is indicated in the graph. N ≥ 18.

752 **G:** RT-PCR from whole worm lysates showing upregulation of *glr-4* transcripts in *cfi-1(-)*

753 mutants.

H: Models summarizing the repressive regulation of glr-4 expression by CFI-1 in the SAB neurons versus ventral nerve cord motor neurons.

Figure 4. CFI-1 is sufficient to repress *glr-4* and continuously required to maintain this repression in ventral cord motor neurons.

A: Quantification of the number of motor neurons expressing *glr-4* in WT and *cfi-1(-)* animals at four developmental stages – L2, L3, L4, and day 2 adults (D2). p-values are indicated in the graph. N ≥ 12.

B: Diagram showing the timeline for the administration of auxin to induce degradation of CFI-1 protein in *kas16[mNG::AID::cfi-1]; otTi28[unc-11prom8+ehs-1prom7+rgef-1prom2::TIR1::mTurquoise2::unc-54 3'UTR]* animals. *otTi28* drives expression of TIR1 specifically in neurons.

C: Quantification graph comparing expression of the endogenous *glr-4* reporter between the ethanol group (control) and auxin group on *kas16[mNG::AID::cfi-1]; otTi28[unc-11prom8+ehs-1prom7+rgef-1prom2::TIR1::mTurquoise2::unc-54 3'UTR]* animals. Ectopic expression of *2xNLS::mScarlet::SL2::glr-4* in motor neurons was detected upon CFI-1 protein knock-down. p-values are indicated in the graph. N = 15.

D: CFI-1 is sufficient to repress *glr-4* expression. Left: representative images showing loss of *glr-4::tagRFP* expression in the SAB neurons (arrowheads) upon overexpression (OE) of *cfi-1*. Right: quantification of the number of SAB neurons expressing *glr-4::tagRFP* in wildtype and *cfi-1* overexpression animals. p-value is indicated in the graph. N ≥ 14.

E: Model summarizing the sufficiency of CFI-1 to repress *glr-4* in SAB neurons.

Figure 5. CFI-1 directly represses *glr-4* by binding to its promoter via a conserved binding motif

A: CFI-1 ChIP-seq tracks at the *glr-4* locus. (Bottom) Schematics of a series of *glr-4* transgenic reporter lines which contain different parts of the *cis*-regulatory regions upstream of the gene. CFI-1 recognizes a highly conserved binding motif shared by the ARID family (right). Using bioinformatic analysis, 13 of these binding motifs were identified in the *cis*-regulatory region upstream of *glr-4*, which overlap with the CFI-1 binding peaks.

B: Quantification of the number of motor neurons expressing the reporters of *glr-4* shown in panel A in WT and *cfi-1(-)* animals in day 2 adults. Only the reporters that contain the proximal *glr-4* promoter regions show ectopic expression in *cfi-1(-)* mutants. p-values are indicated in the

27

787   graph. All reporters were analyzed in the adult (D2), except the 4.9kb reporter which was

788   analyzed at the L4 stage. N ≥ 13.

789   **C:** Schematic depicting the design of the *2xNLS::mScarlet::SL2::glr-4*$^{11CFI\text{-}1\ sites\ MUT}$ allele. Point

790   mutations were introduced to the 11 CFI-1 binding motifs that fall in the proximal CFI-1 binding

791   peaks at the *glr-4* promoter.

792   **D:** Quantification of the number of motor neurons expressing WT *2xNLS::mScarlet::SL2::glr-4*

793   reporter and the *2xNLS::mScarlet::SL2::glr-4*$^{11CFI\text{-}1\ sites\ MUT}$ allele at day 2 adult stage. Ectopic

794   expression of the reporter was observed with the point mutation allele, which shows a total

795   number of *glr-4-* expressing motor neurons comparable to what was observed in *cfi-1(-)* null

796   mutants.

797

798   **Figure 6. CFI-1 represses *glr-4* by counteracting activation from the cholinergic terminal**

799   **selector UNC-3 and Hox proteins LIN-39 and MAB-5.**

800   **A:** Summary of the endogenous expression patterns of *cfi-1*/Arid3, *glr-4*/GRIK1, *unc-3*/COE, *lin-*

801   *39*/HOX, and *mab-5*/HOX in the five cholinergic motor neuron subtypes and in the SAB neurons

802   of the retrovesicular ganglion (RVG). Filled boxes represent positive expression, while empty

803   boxes indicate no detectable expression.

804   **B:** Graph summarizing quantification results of the expression of the 2xNLS::mScarlet::SL2::glr-

805   4 reporter in WT animals, *cfi-1(-)* mutants, *lin-39(-); mab-5(-)* double mutants, and *unc-3(-); lin-*

806   *39(-); mab-5(-)* triple mutants at the L4 stage. p-values are indicated in the graph. N = 15.

807   **C:** (Doitsidou et al.) ChIP-seq binding peaks for UNC-3 (in WT animals and *cfi-1(-)* mutants),

808   CFI-1, LIN-39, and MAB-5 at the *glr-4* locus. (Bottom) Schematic showing the details in the

809   design of *2xNLS::mScarlet::SL2::glr-4*$^{COE1\ MUT}$. Point mutations were introduced to the proximal

810   COE motif (UNC-3 binding site) to the *glr-4* gene.

811   **D:** Graph summarizing expression of the 2xNLS::mScarlet::SL2::glr-4 reporter in WT animals

812   and *unc-3(-)* mutants, and the expression of the 2xNLS::mScarlet::SL2::glr-4$^{COE1\ MUT}$ allele.

813   Quantification was conducted at the day 2 adult stage. P-values are indicated in the graph. N ≥

814   13.

815   **E:** Graph summarizing expression of the *2xNLS::mScarlet::SL2::glr-4* reporter in WT animals,

816   *cfi-1(-)* mutants, *unc-3(-)* mutants, and *cfi-1(-); unc-3(-)* double mutants and the expression of

817   the *2xNLS::mScarlet::SL2::glr-4*$^{COE1\ MUT}$ allele in WT animals and *cfi-1(-)* mutants in day 2 adults.

818   p-values are indicated in the graph. N ≥ 13.

819

820   **Figure 7. Protein motif analysis of CFI-1. A:** Schematics of the four CFI-1 cDNA constructs

821   tested for rescue effects including the WT cDNA, cDNA with deletion of the ARID domain

822   (ΔARID), cDNA with deletion of the eARID domain (ΔeARID), and cDNA with deletion of the

823   helix-turn-helix structure (ΔHTH).

824   **B:** Quantification results of the number of motor neurons showing expression of *glr-4::tagrfp* in

825   *cfi-1(ot786)* mutant animals expressing the rescue constructs.

826   **C:** Schematic model summarizing our findings in SAB and nerve cord motor neurons.

827

828

LEGENDS OF SUPPLEMENTARY FIGURES AND FILES

Supplementary Figure 1. Harsh touch assay on *C. elegans* animals carrying the *3xflag::cfi-1* allele used for ChIP-Seq.
Summary of the percentage of animals that show normal response to harsh touch stimulation for each genotype. Normal response is determined by forward locomotion upon harsh touch of the tail. See Materials and Methods for details. The null *ot786* allele of *cfi-1* is used as a positive control. N = 60.

Supplementary Figure 2. Expression of *glr-4* in nerve cord motor neurons.
A: Representative images showing colocalization analysis of *glr-4* expression and motor neuron subtype-specific reporters *unc-30::mNG* (marker for GABAergic motor neurons), *glr-5::gfp* (marker for VC neurons), and *cho-1::yfp* (marker for cholinergic motor neurons).
B: Summary of the expression of *glr-4* in the VNC motor neurons. Filled boxes indicate expression, while empty boxes indicate no detectable expression. Each row represents an individual worm scored at larval stage 4 (L4) for colocalization of *glr-4* and motor neuron subtype-specific reporters.
C: Representative images showing the colocalization analysis for *cfi-1* expression in RVG and ventral cord motor neurons. The *cho-1:: mCherry* reporter is expressed in cholinergic motor neurons.

Supplementary Figure 3. The expression of motor neuron terminal differentiation genes is not affected in *cfi-1* mutant animals.
A-F: Quantification of the number of motor neurons expressing fluorescent reporters for terminal differentiation genes in WT and *cfi-1 (ot786)* mutant animals. Stage: L4. P-value is indicated in the graph. N ≥ 10. **G**: All genes are bound by CFI-1 based on ChIP-seq data.

Supplementary File 1. Genes expressed in *C. elegans* neurons and bound by CFI-1.

Supplementary File 2. Protein ontology analysis on genes bound by CFI-1.

Supplementary File 3. Protein ontology analysis on genes expressed in IL2 neurons and bound by CFI-1.

Supplementary File 4. List of *C. elegans* strains used or generated for this study**.**

REFERENCES

AHN, S., YANG, H., SON, S., LEE, H. S., PARK, D., YIM, H., CHOI, H. J., SWOBODA, P. & LEE, J. 2022. The C. elegans regulatory factor X (RFX) DAF-19M module: A shift from general ciliogenesis to cell-specific ciliary and behavioral specialization. *Cell Rep,* 39**,** 110661.

AN, G., MINER, C. A., NIXON, J. C., KINCADE, P. W., BRYANT, J., TUCKER, P. W. & WEBB, C. F. 2010. Loss of Bright/ARID3a function promotes developmental plasticity. *Stem Cells,* 28**,** 1560-7.

BAUMANN, M., PONTILLER, J. & ERNST, W. 2010. Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview. *Mol Biotechnol,* 45**,** 241-7.

BERGHOFF, E. G., GLENWINKEL, L., BHATTACHARYA, A., SUN, H., VAROL, E., MOHAMMADI, N., ANTONE, A., FENG, Y., NGUYEN, K., COOK, S. J., WOOD, J. F., MASOUDI, N., CROS, C. C., RAMADAN, Y. H., FERKEY, D. M., HALL, D. H. & HOBERT, O. 2021. The Prop1-like homeobox gene unc-42 specifies the identity of synaptically connected neurons. *Elife,* 10.

BRAMSWIG, N. C., CALUSERIU, O., LUDECKE, H. J., BOLDUC, F. V., NOEL, N. C., WIELAND, T., SUROWY, H. M., CHRISTEN, H. J., ENGELS, H., STROM, T. M. & WIECZOREK, D. 2017. Heterozygosity for ARID2 loss-of-function mutations in individuals with a Coffin-Siris syndrome-like phenotype. *Hum Genet,* 136**,** 297-305.

BRENNER, S. 1974. The genetics of Caenorhabditis elegans. *Genetics,* 77**,** 71-94.

BROCKIE, P. J., MADSEN, D. M., ZHENG, Y., MELLEM, J. & MARICQ, A. V. 2001. Differential expression of glutamate receptor subunits in the nervous system of Caenorhabditis elegans and their regulation by the homeodomain protein UNC-42. *J Neurosci,* 21**,** 1510-22.

CHASNOV, J. R., SO, W. K., CHAN, C. M. & CHOW, K. L. 2007. The species, sex, and stage specificity of a Caenorhabditis sex pheromone. *Proc Natl Acad Sci U S A,* 104**,** 6730-5.

DI SALVIO, M., DI GIOVANNANTONIO, L. G., ACAMPORA, D., PROSPERI, R., OMODEI, D., PRAKASH, N., WURST, W. & SIMEONE, A. 2010. Otx2 controls neuron subtype identity in ventral tegmental area and antagonizes vulnerability to MPTP. *Nat Neurosci,* 13**,** 1481-8.

DITCH, L. M., SHIRANGI, T., PITMAN, J. L., LATHAM, K. L., FINLEY, K. D., EDEEN, P. T., TAYLOR, B. J. & MCKEOWN, M. 2005. Drosophila retained/dead ringer is necessary for neuronal pathfinding, female receptivity and repression of fruitless independent male courtship behaviors. *Development,* 132**,** 155-64.

DOITSIDOU, M., FLAMES, N., TOPALIDOU, I., ABE, N., FELTON, T., REMESAL, L., POPOVITCHENKO, T., MANN, R., CHALFIE, M. & HOBERT, O. 2013. A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in C. elegans. *Genes Dev,* 27**,** 1391-405.

DUPUY, D., LI, Q. R., DEPLANCKE, B., BOXEM, M., HAO, T., LAMESCH, P., SEQUERRA, R., BOSAK, S., DOUCETTE-STAMM, L., HOPE, I. A., HILL, D. E., WALHOUT, A. J. & VIDAL, M. 2004. A first version of the Caenorhabditis elegans Promoterome. *Genome Res,* 14**,** 2169-75.

ETCHBERGER, J. F., FLOWERS, E. B., POOLE, R. J., BASHLLARI, E. & HOBERT, O. 2009. Cis-regulatory mechanisms of left/right asymmetric neuron-subtype specification in C. elegans. *Development,* 136**,** 147-60.

FENG, W., LI, Y., DAO, P., ABURAS, J., ISLAM, P., ELBAZ, B., KOLARZYK, A., BROWN, A. E. & KRATSIOS, P. 2020. A terminal selector prevents a Hox transcriptional switch to safeguard motor neuron identity throughout life. *Elife,* 9.

GLENWINKEL, L., TAYLOR, S. R., LANGEBECK-JENSEN, K., PEREIRA, L., REILLY, M. B., BASAVARAJU, M., RAFI, I., YEMINI, E., POCOCK, R., SESTAN, N., HAMMARLUND, M., MILLER, D. M., 3RD & HOBERT,

31

O. 2021. In silico analysis of the transcriptional regulatory logic of neuronal identity specification throughout the C. elegans nervous system. *Elife,* 10.

GREGORY, S. L., KORTSCHAK, R. D., KALIONIS, B. & SAINT, R. 1996. Characterization of the dead ringer gene identifies a novel, highly conserved family of sequence-specific DNA-binding proteins. *Mol Cell Biol,* 16**,** 792-9.

HADER, T., WAINWRIGHT, D., SHANDALA, T., SAINT, R., TAUBERT, H., BRONNER, G. & JACKLE, H. 2000. Receptor tyrosine kinase signaling regulates different modes of Groucho-dependent control of Dorsal. *Curr Biol,* 10**,** 51-4.

HERRSCHER, R. F., KAPLAN, M. H., LELSZ, D. L., DAS, C., SCHEUERMANN, R. & TUCKER, P. W. 1995. The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. *Genes Dev,* 9, 3067-82.

HOBERT, O. 2002. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. *Biotechniques,* 32, 728-30.

HOBERT, O. 2008. Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc Natl Acad Sci U S A,* 105**,** 20067-71.

HOBERT, O. 2016a. A map of terminal regulators of neuronal identity in Caenorhabditis elegans. *Wiley Interdiscip Rev Dev Biol,* 5**,** 474-98.

HOBERT, O. 2016b. Terminal Selectors of Neuronal Identity. *Curr Top Dev Biol,* 116**,** 455-75.

HOBERT, O. & KRATSIOS, P. 2019. Neuronal identity control by terminal selectors in worms, flies, and chordates. *Curr Opin Neurobiol,* 56**,** 97-105.

IWAHARA, J. & CLUBB, R. T. 1999. Solution structure of the DNA binding domain from Dead ringer, a sequence-specific AT-rich interaction domain (ARID). *EMBO J,* 18**,** 6084-94.

IWAHARA, J., IWAHARA, M., DAUGHDRILL, G. W., FORD, J. & CLUBB, R. T. 2002. The structure of the Dead ringer-DNA complex reveals how AT-rich interaction domains (ARIDs) recognize DNA. *EMBO J,* 21**,** 1197-209.

JOHNSTON, R. J., JR. 2013. Lessons about terminal differentiation from the specification of color-detecting photoreceptors in the Drosophila retina. *Ann N Y Acad Sci,* 1293**,** 33-44.

KADKHODAEI, B., ITO, T., JOODMARDI, E., MATTSSON, B., ROUILLARD, C., CARTA, M., MURAMATSU, S., SUMI-ICHINOSE, C., NOMURA, T., METZGER, D., CHAMBON, P., LINDQVIST, E., LARSSON, N. G., OLSON, L., BJORKLUND, A., ICHINOSE, H. & PERLMANN, T. 2009. Nurr1 is required for maintenance of maturing and adult midbrain dopamine neurons. *J Neurosci,* 29**,** 15923-32.

KERK, S. Y., KRATSIOS, P., HART, M., MOURAO, R. & HOBERT, O. 2017. Diversification of C. elegans Motor Neuron Identity via Selective Effector Gene Repression. *Neuron,* 93**,** 80-98.

KORTSCHAK, R. D., TUCKER, P. W. & SAINT, R. 2000. ARID proteins come in from the desert. *Trends Biochem Sci,* 25**,** 294-9.

KOSHO, T., MIYAKE, N. & CAREY, J. C. 2014. Coffin-Siris syndrome and related disorders involving components of the BAF (mSWI/SNF) complex: historical review and recent advances using next generation sequencing. *Am J Med Genet C Semin Med Genet,* 166C**,** 241-51.

KRATSIOS, P., KERK, S. Y., CATELA, C., LIANG, J., VIDAL, B., BAYER, E. A., FENG, W., DE LA CRUZ, E. D., CROCI, L., CONSALEZ, G. G., MIZUMOTO, K. & HOBERT, O. 2017. An intersectional gene regulatory strategy defines subclass diversity of C. elegans motor neurons. *Elife,* 6.

KRATSIOS, P., PINAN-LUCARRE, B., KERK, S. Y., WEINREB, A., BESSEREAU, J. L. & HOBERT, O. 2015. Transcriptional coordination of synaptogenesis and neurotransmitter signaling. *Curr Biol,* 25**,** 1282-95.

KRATSIOS, P., STOLFI, A., LEVINE, M. & HOBERT, O. 2012. Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nat Neurosci,* 15**,** 205-14.

957 LESTARI, W., ICHWAN, S. J., OTSU, M., YAMADA, S., ISEKI, S., SHIMIZU, S. & IKEDA, M. A. 2012.
958   Cooperation between ARID3A and p53 in the transcriptional activation of p21WAF1 in response
959   to DNA damage. *Biochem Biophys Res Commun,* 417**,** 710-6.
960 LI, Y. & KRATSIOS, P. 2021. Transgenic reporter analysis of ChIP-Seq-defined enhancers identifies novel
961   target genes for the terminal selector UNC-3/Collier/Ebf. *MicroPubl Biol,* 2021.
962 LI, Y., OSUMA, A., CORREA, E., OKEBALAMA, M. A., DAO, P., GAYLORD, O., ABURAS, J., ISLAM, P.,
963   BROWN, A. E. & KRATSIOS, P. 2020. Establishment and maintenance of motor neuron identity via
964   temporal modularity in terminal selector function. *Elife,* 9.
965 LIN, C., SONG, W., BI, X., ZHAO, J., HUANG, Z., LI, Z., ZHOU, J., CAI, J. & ZHAO, H. 2014. Recent advances
966   in the ARID family: focusing on roles in human cancer. *Onco Targets Ther,* 7**,** 315-24.
967 LLORET-FERNANDEZ, C., MAICAS, M., MORA-MARTINEZ, C., ARTACHO, A., JIMENO-MARTIN, A.,
968   CHIRIVELLA, L., WEINBERG, P. & FLAMES, N. 2018. A transcription factor collective defines the
969   HSN serotonergic neuron regulatory landscape. *Elife,* 7.
970 MARQUES, F., SARO, G., LIA, A. S., POOLE, R. J., FALQUET, L. & GLAUSER, D. A. 2019. Identification of
971   avoidance genes through neural pathway-specific forward optogenetics. *PLoS Genet,* 15**,**
972   e1008509.
973 MATHELIER, A., ZHAO, X., ZHANG, A. W., PARCY, F., WORSLEY-HUNT, R., ARENILLAS, D. J., BUCHMAN, S.,
974   CHEN, C. Y., CHOU, A., IENASESCU, H., LIM, J., SHYR, C., TAN, G., ZHOU, M., LENHARD, B.,
975   SANDELIN, A. & WASSERMAN, W. W. 2014. JASPAR 2014: an extensively expanded and updated
976   open-access database of transcription factor binding profiles. *Nucleic Acids Res,* 42**,** D142-7.
977 MI, H., MURUGANUJAN, A., CASAGRANDE, J. T. & THOMAS, P. D. 2013. Large-scale gene function
978   analysis with the PANTHER classification system. *Nat Protoc,* 8**,** 1551-66.
979 MIYAKE, N., TSURUSAKI, Y. & MATSUMOTO, N. 2014. Numerous BAF complex genes are mutated in
980   Coffin-Siris syndrome. *Am J Med Genet C Semin Med Genet,* 166C**,** 257-61.
981 NELSON, C. E., HERSH, B. M. & CARROLL, S. B. 2004. The regulatory content of intergenic DNA shapes
982   genome architecture. *Genome Biol,* 5**,** R25.
983 NITTA, K. R., JOLMA, A., YIN, Y., MORGUNOVA, E., KIVIOJA, T., AKHTAR, J., HENS, K., TOIVONEN, J.,
984   DEPLANCKE, B., FURLONG, E. E. & TAIPALE, J. 2015. Conservation of transcription factor binding
985   specificities across 600 million years of bilateria evolution. *Elife,* 4.
986 PATSIALOU, A., WILSKER, D. & MORAN, E. 2005. DNA-binding properties of ARID family proteins. *Nucleic
987   Acids Res,* 33**,** 66-80.
988 PENG, G. H., AHMAD, O., AHMAD, F., LIU, J. & CHEN, S. 2005. The photoreceptor-specific nuclear
989   receptor Nr2e3 interacts with Crx and exerts opposing effects on the transcription of rod versus
990   cone genes. *Hum Mol Genet,* 14**,** 747-64.
991 POPOWSKI, M., TEMPLETON, T. D., LEE, B. K., RHEE, C., LI, H., MINER, C., DEKKER, J. D., ORLANSKI, S.,
992   BERGMAN, Y., IYER, V. R., WEBB, C. F. & TUCKER, H. 2014. Bright/Arid3A acts as a barrier to
993   somatic cell reprogramming through direct regulation of Oct4, Sox2, and Nanog. *Stem Cell
994   Reports,* 2**,** 26-35.
995 RATLIFF, M. L., TEMPLETON, T. D., WARD, J. M. & WEBB, C. F. 2014. The Bright Side of Hematopoiesis:
996   Regulatory Roles of ARID3a/Bright in Human and Mouse Hematopoiesis. *Front Immunol,* 5**,** 113.
997 RHEE, C., LEE, B. K., BECK, S., ANJUM, A., COOK, K. R., POPOWSKI, M., TUCKER, H. O. & KIM, J. 2014.
998   Arid3a is essential to execution of the first cell fate decision via direct embryonic and
999   extraembryonic transcriptional regulation. *Genes Dev,* 28**,** 2219-32.
1000 SAADAT, K., LESTARI, W., PRATAMA, E., MA, T., ISEKI, S., TATSUMI, M. & IKEDA, M. A. 2021. Distinct and
1001   overlapping roles of ARID3A and ARID3B in regulating E2Fdependent transcription via direct
1002   binding to E2F target genes. *Int J Oncol,* 58.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J. Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods,* 9**,** 676-82.

SHAHAM, S. & BARGMANN, C. I. 2002. Control of neuronal subtype identity by the C. elegans ARID protein CFI-1. *Genes Dev,* 16**,** 972-83.

SHANDALA, T., KORTSCHAK, R. D., GREGORY, S. & SAINT, R. 1999. The Drosophila dead ringer gene is required for early embryonic patterning through regulation of argos and buttonhead expression. *Development,* 126**,** 4341-9.

SHANDALA, T., KORTSCHAK, R. D. & SAINT, R. 2002. The Drosophila retained/dead ringer gene and ARID gene family function during development. *Int J Dev Biol,* 46**,** 423-30.

SHANDALA, T., TAKIZAWA, K. & SAINT, R. 2003. The dead ringer/retained transcriptional regulatory gene is required for positioning of the longitudinal glia in the Drosophila embryonic CNS. *Development,* 130**,** 1505-13.

SHANG, L., CHO, M. T., RETTERER, K., FOLK, L., HUMBERSON, J., ROHENA, L., SIDHU, A., SALIGANAN, S., IGLESIAS, A., VITAZKA, P., JUUSOLA, J., O'DONNELL-LURIA, A. H., SHEN, Y. & CHUNG, W. K. 2015. Mutations in ARID2 are associated with intellectual disabilities. *Neurogenetics,* 16**,** 307-14.

SIBBONS, J. 2004. *Identifying the roles of dead ringer in the Drosophila eye* Ph.D., The University of Adelaide.

SIPONEN, M. I., WISNIEWSKA, M., LEHTIO, L., JOHANSSON, I., SVENSSON, L., RASZEWSKI, G., NILSSON, L., SIGVARDSSON, M. & BERGLUND, H. 2010. Structural determination of functional domains in early B-cell factor (EBF) family of transcription factors reveals similarities to Rel DNA-binding proteins and a novel dimerization motif. *J Biol Chem,* 285**,** 25875-9.

SMITH, J. A., HOLDEN, K. R., FRIEZ, M. J., JONES, J. R. & LYONS, M. J. 2016. A novel familial autosomal dominant mutation in ARID1B causing neurodevelopmental delays, short stature, and dysmorphic features. *Am J Med Genet A,* 170**,** 3313-3318.

VALENTINE, S. A., CHEN, G., SHANDALA, T., FERNANDEZ, J., MISCHE, S., SAINT, R. & COUREY, A. J. 1998. Dorsal-mediated repression requires the formation of a multiprotein repression complex at the ventral silencer. *Mol Cell Biol,* 18**,** 6584-94.

WEBB, C. F., BRYANT, J., POPOWSKI, M., ALLRED, L., KIM, D., HARRISS, J., SCHMIDT, C., MINER, C. A., ROSE, K., CHENG, H. L., GRIFFIN, C. & TUCKER, P. W. 2011. The ARID family transcription factor bright is required for both hematopoietic stem cell and B lineage development. *Mol Cell Biol,* 31**,** 1041-53.

WEBB, C. F., SMITH, E. A., MEDINA, K. L., BUCHANAN, K. L., SMITHSON, G. & DOU, S. 1998. Expression of bright at two distinct stages of B lymphocyte development. *J Immunol,* 160**,** 4747-54.

WEIRAUCH, M. T., YANG, A., ALBU, M., COTE, A. G., MONTENEGRO-MONTERO, A., DREWE, P., NAJAFABADI, H. S., LAMBERT, S. A., MANN, I., COOK, K., ZHENG, H., GOITY, A., VAN BAKEL, H., LOZANO, J. C., GALLI, M., LEWSEY, M. G., HUANG, E., MUKHERJEE, T., CHEN, X., REECE-HOYES, J. S., GOVINDARAJAN, S., SHAULSKY, G., WALHOUT, A. J. M., BOUGET, F. Y., RATSCH, G., LARRONDO, L. F., ECKER, J. R. & HUGHES, T. R. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell,* 158**,** 1431-1443.

WILSKER, D., PATSIALOU, A., DALLAS, P. B. & MORAN, E. 2002. ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. *Cell Growth Differ,* 13**,** 95-106.

WILSKER, D., PROBST, L., WAIN, H. M., MALTAIS, L., TUCKER, P. W. & MORAN, E. 2005. Nomenclature of the ARID family of DNA-binding proteins. *Genomics,* 86**,** 242-51.

1049   ZHANG, F., BHATTACHARYA, A., NELSON, J. C., ABE, N., GORDON, P., LLORET-FERNANDEZ, C., MAICAS,
1050   M., FLAMES, N., MANN, R. S., COLON-RAMOS, D. A. & HOBERT, O. 2014. The LIM and POU
1051   homeobox genes ttx-3 and unc-86 act as terminal selectors in distinct cholinergic and
1052   serotonergic neuron types. *Development,* 141**,** 422-35.
1053   ZHANG, L., WARD, J. D., CHENG, Z. & DERNBURG, A. F. 2015. The auxin-inducible degradation (AID)
1054   system enables versatile conditional protein depletion in C. elegans. *Development,* 142**,** 4374-84.
1055   ZHENG, C., JIN, F. Q., TRIPPE, B. L., WU, J. & CHALFIE, M. 2018a. Inhibition of cell fate repressors secures
1056   the differentiation of the touch receptor neurons of Caenorhabditis elegans. *Development,* 145.
1057   ZHENG, W., YANG, X., HU, R., CAI, R., HOFMANN, L., WANG, Z., HU, Q., LIU, X., BULKLEY, D., YU, Y., TANG,
1058   J., FLOCKERZI, V., CAO, Y., CAO, E. & CHEN, X. Z. 2018b. Hydrophobic pore gates regulate ion
1059   permeation in polycystic kidney disease 2 and 2L1 channels. *Nat Commun,* 9**,** 2302.
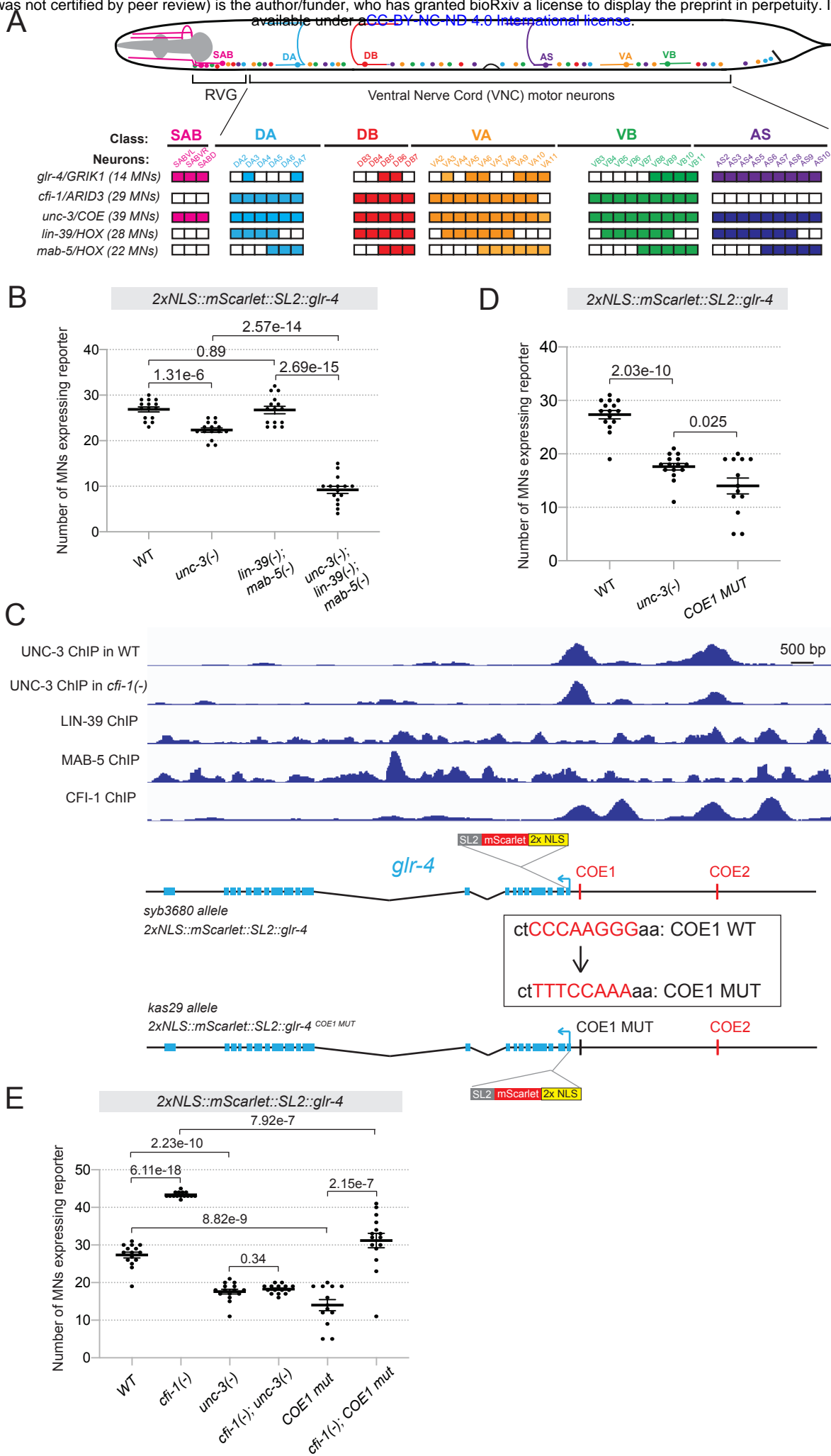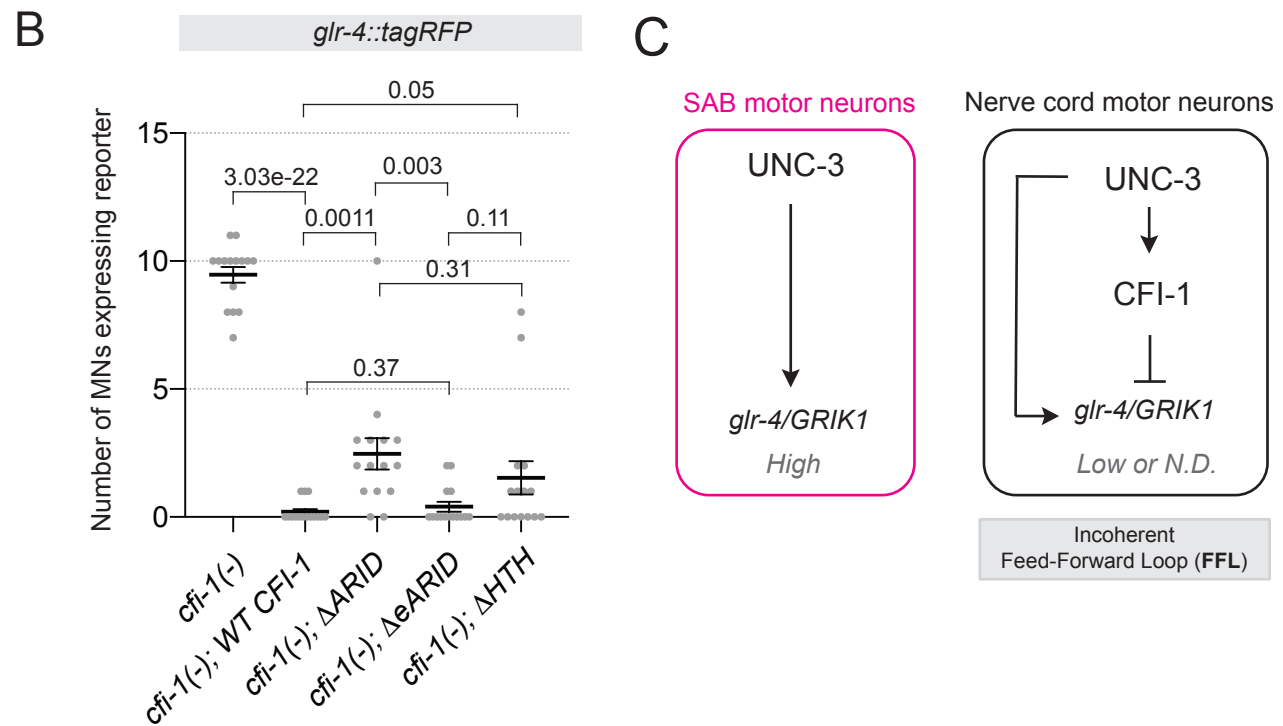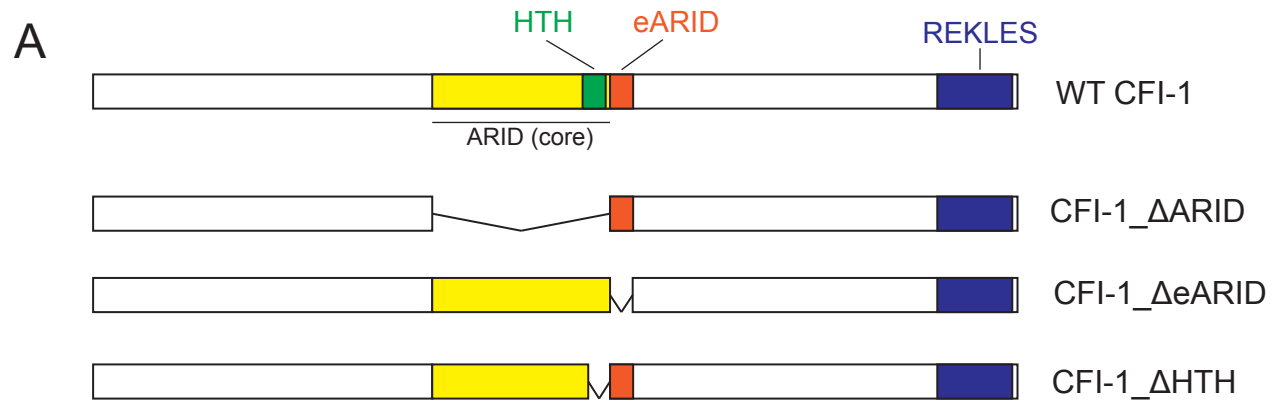1060
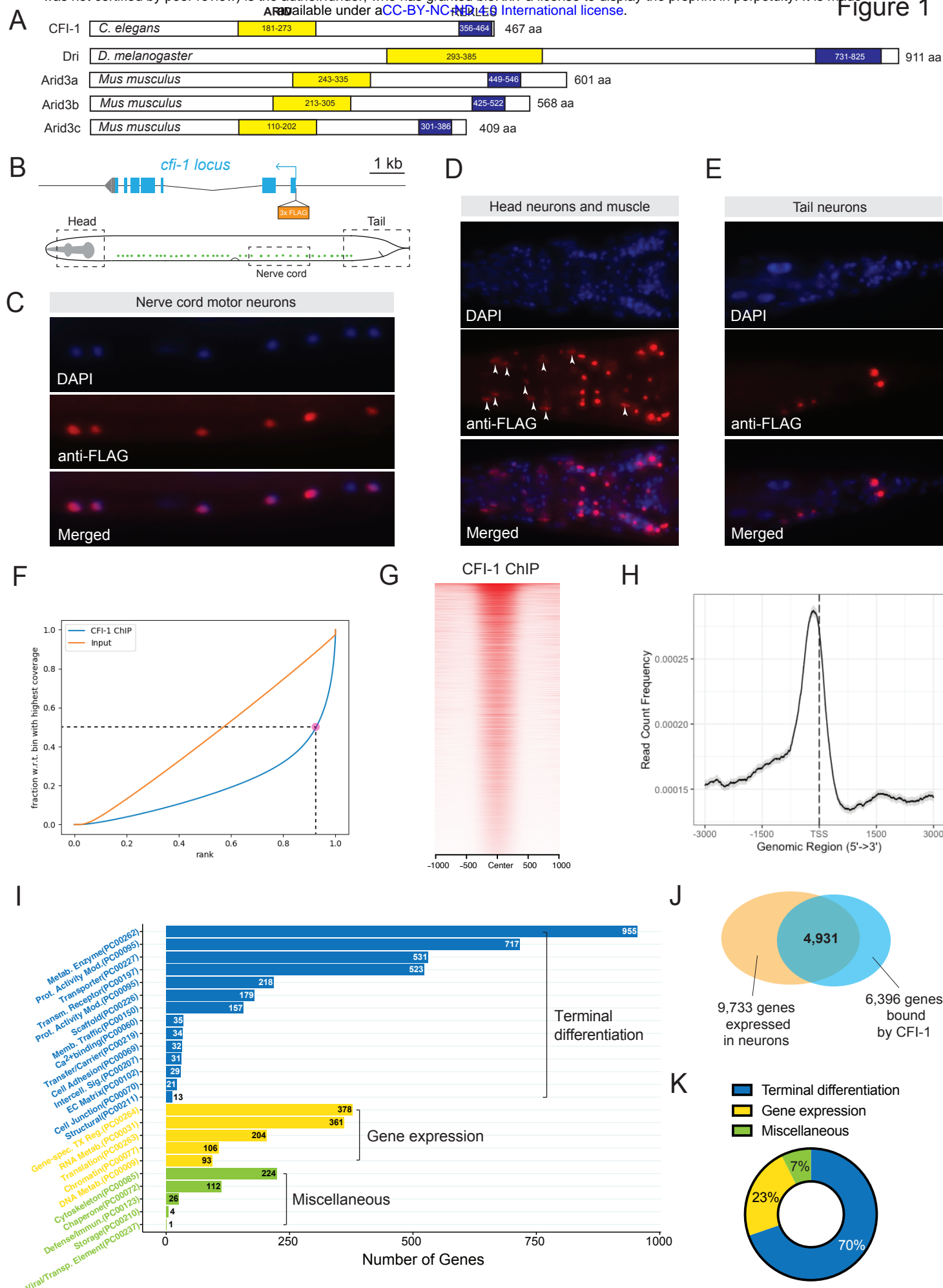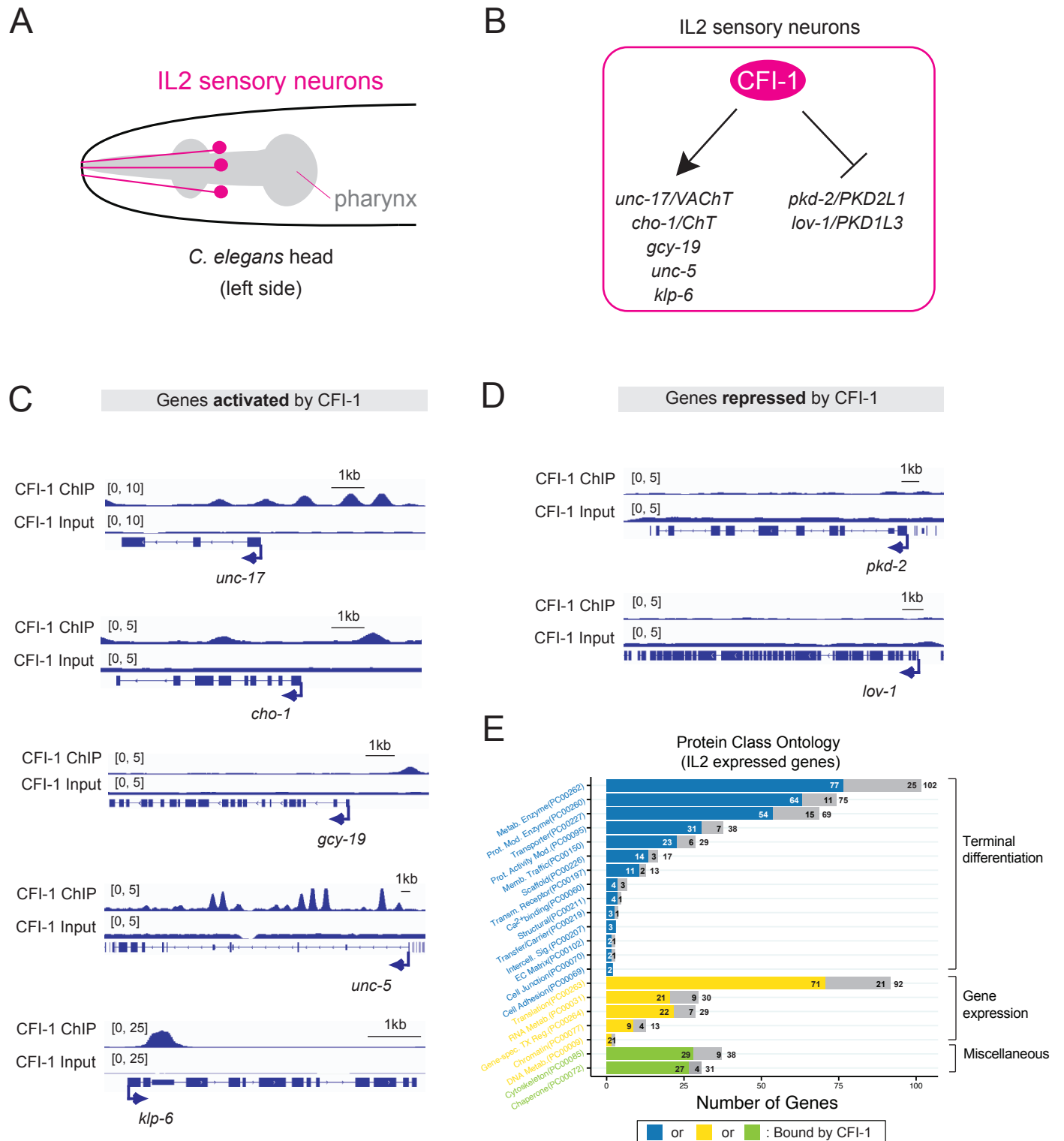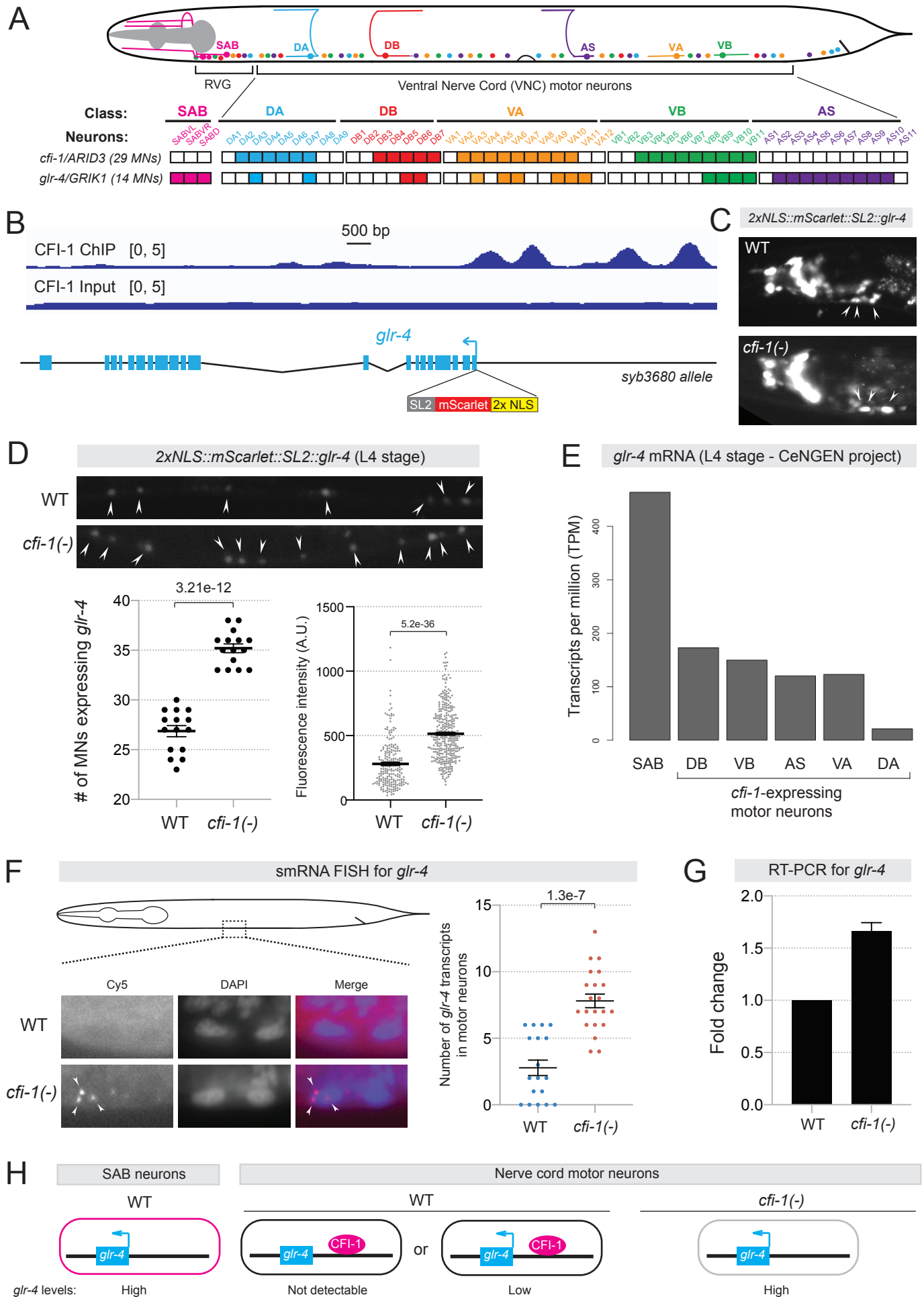
Figure 1
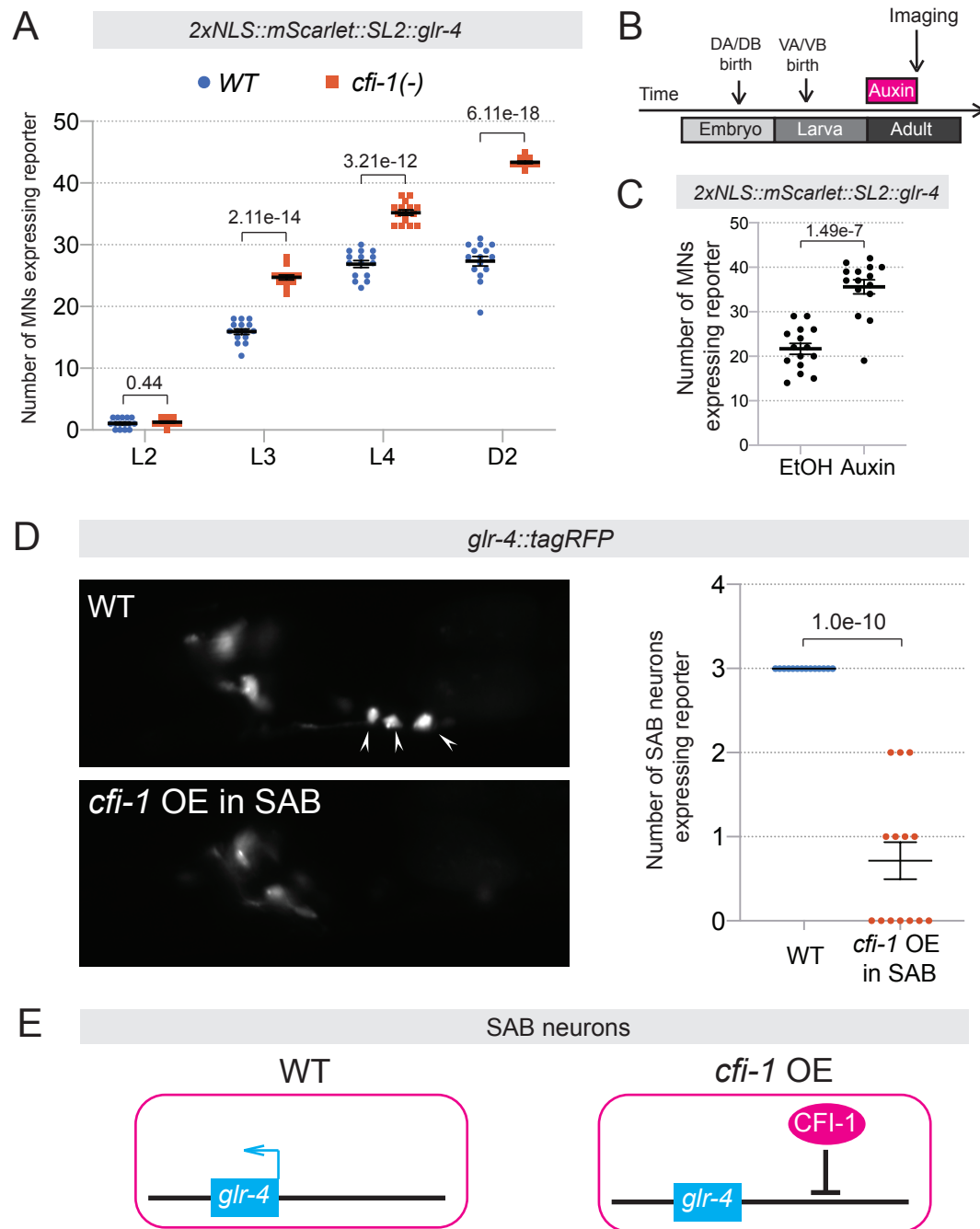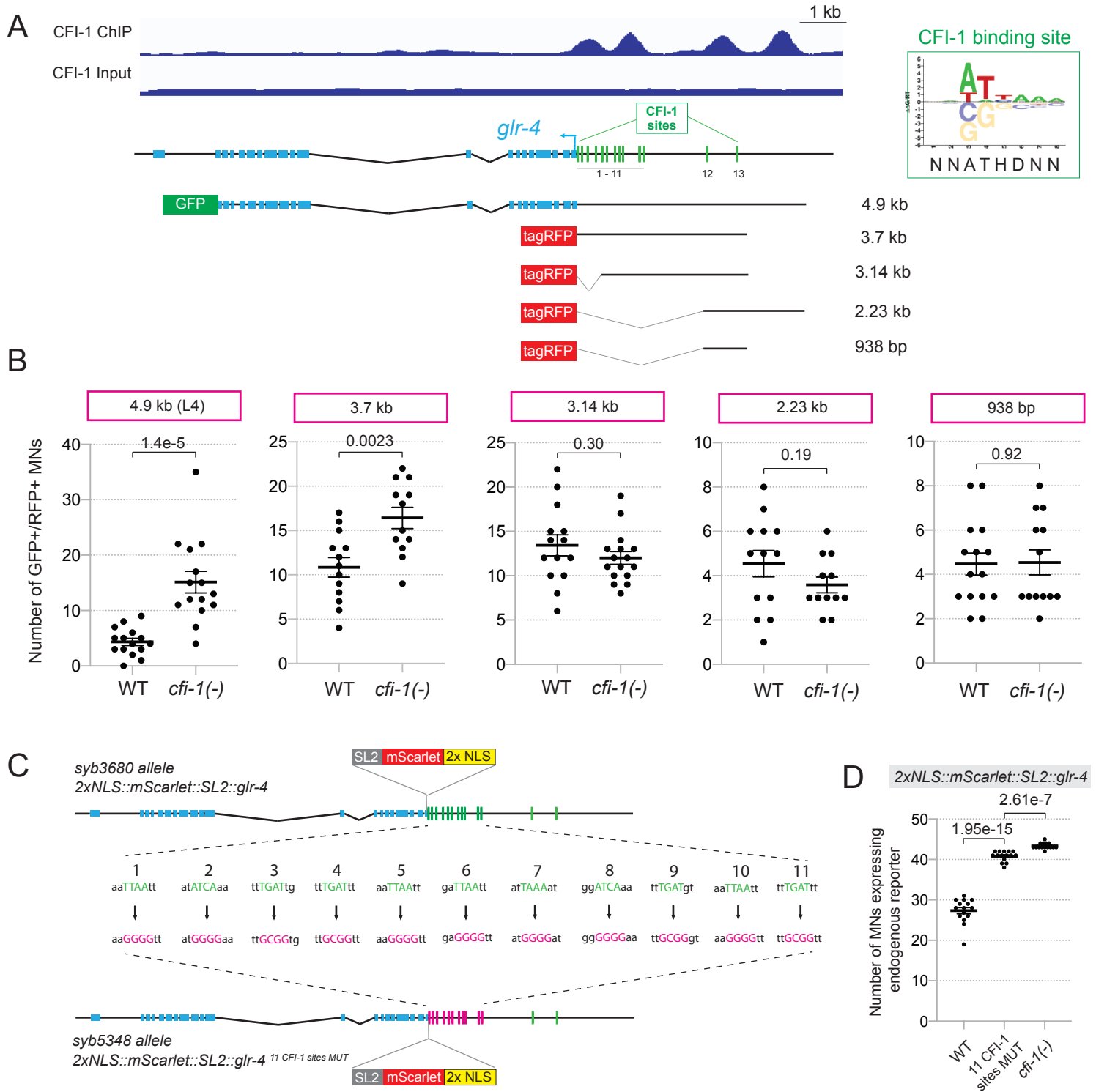
Figure 2

Figure 3

Figure 4

Figure 5

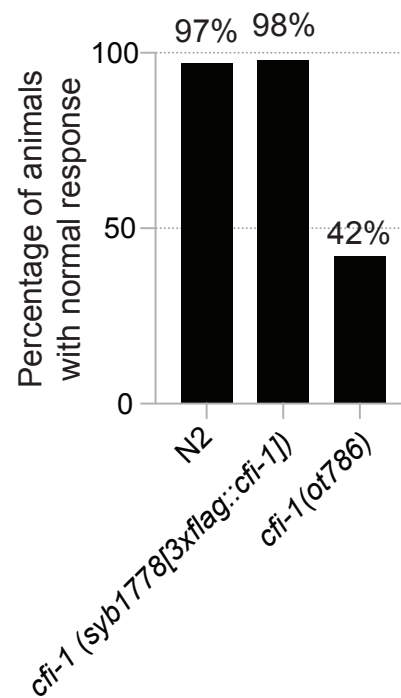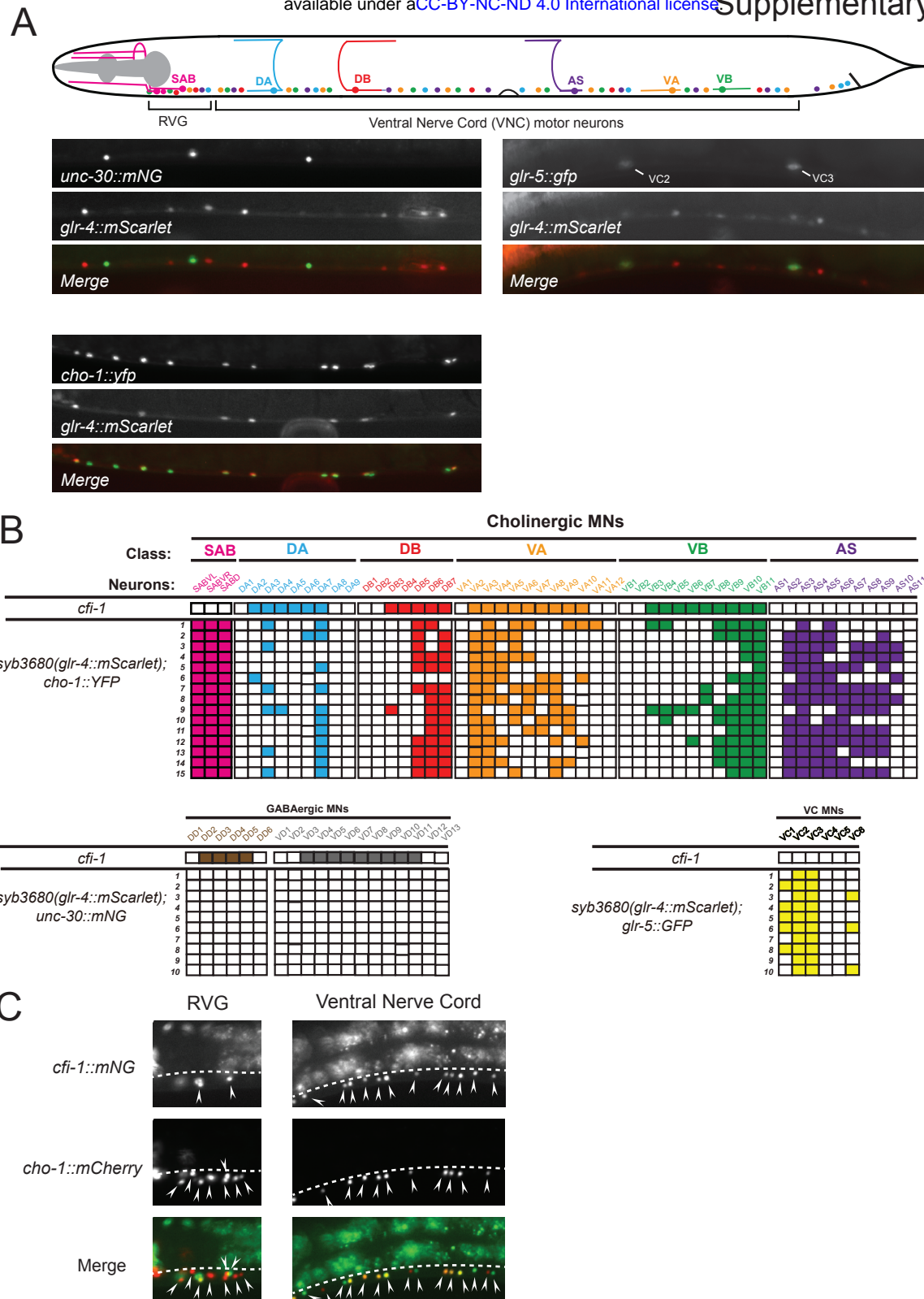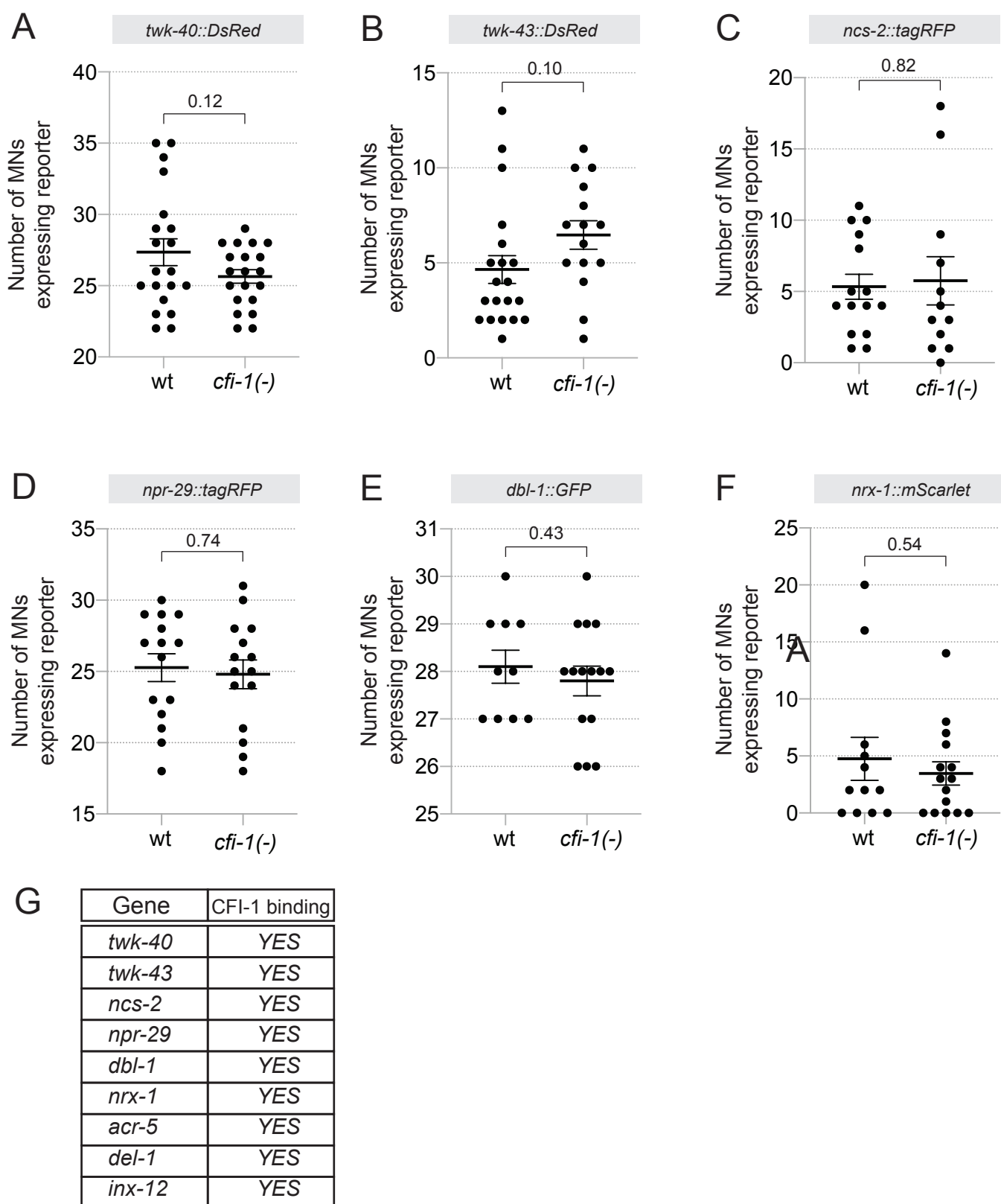Figure 6

Figure 7

Supplementary Figure 1. Harsh touch assay on *C. elegans* animals carrying the *3xflag::cfi-1* allele used for ChIP-Seq.
Summary of the percentage of animals that show normal response to harsh touch stimulation for each genotype. Normal response is determined by forward locomotion upon harsh touch of the tail. See Materials and Methods for details. The null *ot786* allele of *cfi-1* is used as a positive control. N = 60.

Supplementary Figure 2. Expression of *glr-4* in nerve cord motor neurons.
A: Worm schematic with motor neuron subtypes. Representative images showing colocalization analysis of *glr-4* expression and motor neuron subtype-specific reporters *unc-30::mNG* (marker for GABAergic motor neurons), *glr-5::gfp* (marker for VC neurons), and *cho-1::yfp* (marker for cholinergic motor neurons). **B**: Summary *glr-4* expression in VNC MNs. Filled boxes indicate expression, while empty boxes indicate no detectable expression. Each row represents an individual worm scored at larval stage 4 (L4) for colocalization of *glr-4* and motor neuron subtype-specific reporters. **C**: Representative images showing the colocalization analysis for *cfi-1* expression in RVG and ventral cord motor neurons. The *cho-1:: mCherry* reporter is expressed in cholinergic motor neurons.

Supplementary Figure 3. The expression of motor neuron terminal differentiation genes is not affected in cfi-1 mutant animals.

A-F: Quantification of the number of MNs expressing fluorescent reporters for terminal identity genes in WT and *cfi-1 (ot786)* mutant animals. Stage: L4. P-value is indicated in the graph. N = 10. **G**: All genes are bound by CFI-1 based on ChIP-seq data.