Functional divergence of Capicua isoforms explains differential tissue vulnerability in neurological disease

Hamin Lee[1,2,3], Esmeralda Villavicencio Gonzalez[1,2,3], Elias M. Rivera[2,3], Mark A. Durham[2,3,4,5], Ronald Richman[2,3], Elizabeth H.-Y. Chu[2,3], Kailey Xia[2,3], Hu Chen[3,6], Zhandong Liu[3,6], Surabi Veeraragavan[2,3], Binoy Shivanna[7], Huda Y. Zoghbi[1,2,3,4,6,7,8,9,10*]

[1]Program in Genetics and Genomics, Baylor College of Medicine, Houston, TX 77030, USA
[2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA.
[3]Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, Texas 77030, USA
[4]Program in Developmental Biology, Baylor College of Medicine, Houston, TX 77030, USA
[5]Medical Scientist Training Program, Baylor College of Medicine, Houston, TX 77030, USA
[6]Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA
[7]Department of Pediatrics, Texas Children's Hospital, Baylor College of Medicine, Houston, TX 77030, USA
[8]Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA
[9]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX 77030, USA
[10]Lead contact
[*]Correspondence: hzoghbi@bcm.edu

SUMMARY

Many neurological diseases attack specific brain regions despite widespread expression of the disease-related protein. Spinocerebellar ataxia type 1 (SCA1) primarily affects the cerebellum, though Ataxin-1 (ATXN1) is widely expressed. We previously showed that intensified interaction between mutant ATXN1 and Capicua (CIC) drives SCA1 pathogenesis in the cerebellum, whereas ATXN1 loss augments amyloid beta production in the hippocampus and cortex. CIC, however, forms a complex with ATXN1 and its paralog, Ataxin-1-like (ATXN1L)—yet knockout of either yields completely different phenotypes. To determine whether this could be due to CIC having two isoforms, we generated mice bearing either the long (CIC-L) or short (CIC-S) isoform. Loss of CIC-L led to cognitive deficits, whereas loss of CIC-S caused perinatal lethality, phenocopying, ATXN1 and ATXN1L knockout mice, respectively. Furthermore, CIC-L preferentially interacts with ATXN1, and CIC-S with ATXN1L. Our data underscore the importance of isoform–paralog interplay in studying regional vulnerability in neurodegenerative diseases.

Keywords

Capicua, CIC, ATXN1, ATXN1L, isoform, paralog, alloform, protein complex, neurodegeneration, neurodevelopment

INTRODUCTION

One of the puzzling features of many neurological diseases is that they seem to affect particular brain regions or cell types even though the disease-related protein is expressed widely throughout the brain and often in other tissues as well [1, 2]. Typically the disease ramifies outward from that locus of initial damage: Alzheimer's, for example, first targets the entorhinal cortex and hippocampus, whereas Parkinson's disease centers on the substantia nigra [1, 2]. The same regional specificity is seen in Mendelian disorders as well, which allow us to more readily trace their developmental origins. For example, spinocerebellar ataxia type 1 (SCA1) is caused by expansion of a polyglutamine repeat tract in the gene *ATAXIN-1* (*ATXN1*) [3, 4]. The ATXN1 protein is a transcriptional co-repressor that is expressed throughout the central nervous system (CNS) in both neurons and glia and is also expressed in organs such as the heart and liver [5]. In the CNS, it is most abundant in hippocampal pyramidal cells and cerebellar Purkinje cells, yet SCA1 pathology is largely restricted to the Purkinje cells and brainstem [6]. Interestingly, a lack of ATXN1 does not cause ataxia but rather learning and memory deficits [7-9]: copy number variation of *ATXN1* is associated with Alzheimer's disease risk [10, 11], and *Atxn1* knock-out in mice potentiates amyloid beta production [12]. If, however, we delete *Atxn1* and its paralog, *Ataxin-1*-like (*Atxn1l*), the double mutant mice develop a quite different constellation of abnormalities that includes hydrocephalus, omphalocele lung defects, and perinatal lethality [13]. Loss of *Atxn1l* by itself causes the same phenotypes but with reduced penetrance [13].

It is reasonable to assume that this combination of regional specificity independent of expression level and the very different tissues affected by loss of ATXN1L indicates that ATXN1 and ATXN1L have different native interactors. It was rather surprising, then, to find that they both form a complex with the transcriptional repressor Capicua (CIC) [13, 14], and both proteins are critical for maintaining CIC stability [13-16]. ATXN1 and ATXN1L share a conserved CIC-binding domain [16-18] and can substitute for each other in binding to CIC, exhibiting partial functional redundancy [8, 13, 19]. CIC is also broadly expressed [13, 14, 20] and interacts with ATXN1 throughout the brain [14]; the effect of the polyglutamine expansion in mutant ATXN1 is to intensify the effect of this interaction, leading to hyper-repression of CIC targets [21, 22]. The formation of the mutant ATXN1–CIC complex is necessary for Purkinje cell pathology in SCA1 [22, 23]. On the other hand, loss of ATXN1 reduces CIC levels, leading to derepression of CIC targets *Etv4* and *Etv5* [12]. These transcription factors have binding sites upstream of *Bace1* and increase its expression, which in turn promotes amyloidogenic processing of amyloid precursor protein (APP) and may be a risk factor for Alzheimer disease [12]. Notably, this upregulation occurs only in the cortex and hippocampus, not in the cerebellum [12].

CIC has broad developmental roles in both the peripheral tissues and the CNS. In mice, complete loss of CIC causes perinatal lethality, with hydrocephalus, omphalocele and impaired lung development [13, 24]—reminiscent of *Atxn1l* knockout (KO) mice. In humans, heterozygous loss-of-function variants cause CIC haploinsufficiency syndrome (CHS), a neurodevelopmental disorder characterized by intellectual disability,

autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), and seizures [14]. Dual loss of both ATXN1 and ATXN1L phenocopies the complete loss of CIC [13], and conditional KO in the brain causing a CHS-like phenotype [14]. How do these two paralogs, both acting through CIC, produce such distinct loss-of-function outcomes?

We hypothesized that the answer might arise from the fact that CIC exists in two isoforms generated by alternative promoter usage, namely, CIC-L (long) and CIC-S (short) [16]. Both isoforms share the HMG box and C1 DNA-binding domains as well as the ATXN1/ATXN1L-binding domain (ATXN1/1L-BD) [16, 25, 26]. The functions of the isoform-specific N-termini remain largely undefined, leaving open the possibility that the two isoforms have distinct biological roles. To determine whether isoform-specific functions contribute to the regional specificity of ATXN1 and ATXN1L gain- or loss-of-function phenotypes, we generated CIC isoform-specific KO mice, characterized their phenotypes, and evaluated transcriptomic changes. We then performed biochemical studies to assess the formation of different complexes. We found that, surprisingly, subtle differences in the relative abundance of CIC isoforms and ATXN1 paralogs at the protein level, together with the differential complexes they assemble, can dictate regional vulnerability.

RESULTS

CIC isoforms and ATXN1/ATXN1L show different spatiotemporal expression patterns

To investigate how the expression patterns of CIC isoforms relate to that of the CIC–ATXN1/1L complex, we profiled protein levels of each CIC isoform together with ATXN1 and ATXN1L in the cortex, cerebellum, and lung, of wild-type (WT) mice at four developmental stages: birth (P0), postnatal day 6 (P6), postnatal day 21 (P21), and 12 weeks (**Figure 1A, B**). These regions and developmental timepoints were selected based on prior evidence of tissue-specific changes due to disruption of the CIC–ATXN1/1L complex: in the forebrain, *Cic* knockout (KO) or *Atxn1/Atxn1l* double KO in mice leads to selective reduction of cortical layers 2–4 beginning around postnatal day 5 (P5), with behavioral abnormalities emerging by 3 months of age [14]. In the lung, *Atxn1l* KO in mice disrupts the transcriptional program of early alveolarization at P6 and results in reduced survival by P21 [13]. The cerebellum, which is highly sensitive to ATXN1 gain-of-function toxicity, was included as a comparator to these loss-of-function–vulnerable tissues [23].

ATXN1L and both CIC isoforms showed a similar pattern of expression in the lungs, with the highest levels at P6 and decreasing at later timepoints, whereas ATXN1 expression decreases over time after P0 (**Figure 1A, B**). Notably, ATXN1L levels are considerably higher in the lungs compared to the cortex and cerebellum, with peak expression at P6. In the cortex, CIC-L and CIC-S expression patterns diverged after P6: CIC-L showed a slight increase at P21 that was maintained at 12 weeks, while CIC-S markedly decreased between P6 and P21 and remained low at 12 weeks. In the cerebellum, all four proteins showed a pronounced increase from P6 to P21, followed by a decrease in CIC-S and ATXN1L at 12 weeks. In 12-week-old brains, we also looked at the

103 hippocampus, striatum, and brainstem and found that both CIC isoforms were most highly expressed in the
104 cerebellum and least expressed in the brainstem (**Figure S1**).
105 Thus, CIC, ATXN1, and ATXN1L show distinct tissue- and isoform-specific expression patterns in mice
106 from birth through young adulthood.
107
108 **CIC-S is critical for survival and for lung alveolarization**
109 To dissect the individual roles of each CIC isoform, we used CRISPR/Cas9 to generate *Cic-L*- and *Cic-S*-
110 specific KO mice by targeting the unique first coding exon of each isoform. For *Cic-S*-KO, we targeted *Cic-S*
111 exon 1 with a single guide RNA (sgRNA) and selected founder lines with deletions that removed the in-frame
112 downstream ATG (**Figure 2A, S2A** and **B**). For *Cic-L*-KO, we used two sgRNAs to introduce a deletion within
113 exon 2, resulting in a frameshift mutation and selective disruption of the long isoform. We confirmed the deletions
114 by Sanger sequencing (**Figure S2C**) and used western blot to validate complete loss of either CIC-L or CIC-S in
115 homozygous KO animals (**Figure 2B**).
116 Since germline KO of *Cic* results in perinatal lethality [14, 24], we sought to determine how loss of each
117 isoform affects survival. We used chi-square analysis to compare observed genotype frequencies to the expected
118 Mendelian ratios of WT (25%), heterozygous KO (50%), and KO (25%) at postnatal days 0, 6, and 21 (P0, P6,
119 and P21; **Tables 1** and **2**). Both *Cic-S* and *Cic-L*-KO animals exhibited the expected Mendelian ratios at P0,
120 indicating no embryonic lethality. At P6, neither group showed significant deviation from expected ratios, but by
121 P21, approximately 70% of *Cic-S*-KO had died (chi-square analysis, $p = 0.007$; **Table 1**) while *Cic-L*-KO animals
122 maintained normal survival (chi-square analysis, $p = 0.74$; **Table 2**).
123 We measured the weight and lifespan of surviving animals beyond P21. *Cic-S*-KO and *Cic-L*-KO mice
124 weighed less than WT throughout the 9-week period, although *Cis-S*-KO stopped gaining weight around week 7
125 while *Cic-L*-KO mice continued to grow, reaching near WT weight by week 9 (**Figure 3A**). *Cic-S*-KO mice
126 exhibited significant early lethality after 3 weeks, with 60% (15 of 25 animals) dying before 20 weeks, whereas
127 *Cic-L*-KO animals did not display reduced survival (**Figure 3B**). Among the *Cic-S*-KOs that died, 80% developed
128 hydrocephalus accompanied by kyphosis, emaciation, and lethargy prior to death. One animal experienced
129 spontaneous and ultimately fatal home-cage seizures (**Figure 3C**). The remaining two animals that died without
130 clear hydrocephalus or seizures showed further weight loss beyond their already reduced baseline prior to death.
131 Some *Cic-S*-KO mice survived up to 40 weeks without any of these features, however, suggesting incomplete
132 penetrance. Neither *Cic-L*-KO nor WT animals developed any noticeable phenotype during this period.
133 These abnormalities of the *Cic-S*-KO mice are strongly reminiscent of those previously reported in *Atxn1l*-
134 KO mice (hydrocephalus with incomplete penetrance, lower body weight compared to WT controls, and reduced
135 survival by P21) [13]. Because *Atxn1l*-KO mice also have impaired alveolarization mediated by loss of CIC
136 function that dysregulates extracellular matrix (ECM) remodeling and alveolarization during late alveolarization

stage (P17–P23), which was also present in another line of mice bearing a *Cic* loss-of-function allele [13], we performed hematoxylin and eosin (H&E) staining of lung tissue at P21 and quantified the mean linear intercepts (MLI). Despite potential survivor bias in *Cic-S*-KO animals at this stage, the MLI values were significantly greater in *Cic-S*-KO lungs than in WT littermates (*Cic-S*-KO, 29.73 ± 1.05 vs. WT littermate, 25.86 ± 0.71; p=0.038, unpaired t-test; **Figure 3D**) whereas MLI values of *Cic-L*-KO lungs were comparable to WT littermate controls (*Cic-L*-KO, 27.21 ± 1.36 vs. WT littermate, 28.14 ± 1.17; p=0.658 , unpaired t-test; **Figure 3E**). Therefore, loss of CIC-S, but not CIC-L, impairs alveolarization, supporting functional overlap between CIC-S and ATXN1L.

CIC-L is crucial for neurobehavioral function

Given the critical role of CIC in neurodevelopment [14, 27, 28], we performed a battery of behavioral assays in *Cic-S*-KO, *Cic-L*-KO, and WT littermates to assess anxiety-like behavior, general locomotor activity, motor coordination, and social interaction—phenotypes previously linked to CIC dysfunction in both people with CHS and in mouse models [14, 29]. *Cic-L*-KO mice displayed reduced anxiety-like behavior in the elevated plus maze assay, spending significantly more time in the open arms (**Figure 4A**), and hyperactivity in the open field test, with increased total distance traveled and higher speed compared to controls (**Figure 4B**). They had a shorter latency to fall from the rotarod (**Figure 4C**) and took more time to descend in the pole test (**Figure S3A**). In the fear-conditioning assay, *Cic-L*-KO mice showed deficits in learning and memory, freezing less in both contextual (**Figure 4D**) and cued fear tests (**Figure 4E**). In the three-chamber assay, both KO lines preferred to interact with the conspecific than the object (**Figure S3B**).

Cic-S-KO mice exhibited minimal behavioral abnormalities (**Figure 4A, B** and **C**). The only significant phenotype that *Cic-S*-KO mice displayed was increased freezing in the contextual test (**Figure 4D**) and pre-tone in the cued test (**Figure 4E**), which do not signify any learning or memory deficit. These data suggest that, for the animals that survive because of decreased penetrance, CIC-S is not critical for most neurobehavioral functions, with only mild effects on baseline freezing behavior.

Cic-L-KO mice phenocopied *Atxn1*-KO mice, which show learning deficits [7-9]. Collectively, the survival and neurobehavioral data point to a functional and spatio-temporal convergence between CIC-L and ATXN1 on the one hand, and CIC-S and ATXN1L on the other.

Region-specific transcriptomic overlap between CIC isoform and ATXN1 paralog KOs

To determine whether phenotypic similarities between these two pairs of mouse lines reflect shared transcriptional changes, we performed bulk RNA sequencing (RNA-seq) on tissues linked to the phenotype of each CIC isoform loss (P6 lung and 12-week cortex) and compared the differentially expressed genes (DEGs) previously identified in *Atxn1*-KO and *Atxn1l*-KO mice. We used the cerebellum as an internal control, since it

171 does not contribute to the CHS phenotype despite its great sensitivity to polyglutamine-expanded ATXN1 [23]
172 In the P6 lung, *Cic-S*-KO samples exhibited a greater number and magnitude of DEGs (p-adjusted < 0.05
173 and |log2 fold change (Log2FC)| > 0.25) than *Cic-L*-KO, with enrichment of upregulated genes consistent with
174 CIC's role as a transcriptional repressor (**Figure S4A** and **B**). These differences aligned with more pronounced
175 survival and developmental phenotypes in *Cic-S*-KO animals (**Table 1**, **Figure 3**). Gene set enrichment analysis
176 (GSEA) [30] revealed distinct GO term enriched for each isoform KO, with *Cic-S*-KO enriched for immune-
177 related terms (**Figure S4C**) and *Cic-L*-KO for chromosome segregation-related terms (**Figure S4D**).
178 Consistent with the phenotypic similarities, the *Cic-S*-KO DEGs showed stronger overlap with the *Atxn1l*-
179 KO DEGs [13] (Fisher's exact test, OR = 19.85, $p = 6.32 \times 10^{-132}$; **Figure S5A**) than *Cic-L*-KO (OR = 3.58, $p =$
180 1.7×10^{-12}; **Figure S5B**), with most shared genes showing concordant expression changes. Interestingly, although
181 *Atxn1*-KO lungs had only 34 DEGs, just over a third overlapped with the *Cic-S*-KO DEGs (OR = 4.74, p = $5.83 \times$
182 10^{-5}; **Figure S5C**) or the *Cic-L*-KO DEGs (Fisher's exact test, OR = 9.86, $p = 8.54 \times 10^{-8}$; **Figure S5D**). Given
183 the role of the *Etv–Mmp* axis in *Atxn1l*-KO alveolarization defects [13] and the fact that *Etv1/4/5* are direct CIC
184 targets [13, 22, 31, 32], we examined their expression specifically. *Etv4, Mmp8, Mmp9,* and *Mmp12* were strongly
185 upregulated in *Cic-S*-KO lungs, while only *Etv1* and *Mmp13* showed mild upregulation in *Cic-L*-KO (**Figure**
186 **S5E**). GSEA revealed enrichment of extracellular matrix (ECM) assembly-related pathways in *Cic-S*-KO,
187 whereas *Cic-L*-KO showed weak or non-significant enrichment (**Figure S5F**). These results suggest that CIC-S
188 and ATXN1L regulate similar genes in the lung during lung development.
189 In the 12-week-old cortex, there were more DEGs in the *Cic-L*-KO with larger fold-changes than in *Cic-S*-
190 KO (**Figure S6A, B**). Whereas *Cic-S*-KO top GSEA terms were mostly related to glia or to bioenergetics (**Figure**
191 **S6C, D**), *Cic-L*-KO terms involved synapse and neurotransmission-related functions (**Figure S6E, F**).
192 Comparison with *Atxn1*-KO cortex DEGs (*Xhako et al., unpublished data*) revealed greater overlap and stronger
193 enrichment with *Cic-L*-KO than *Cic-S*-KO, with more genes changing in the same direction (**Figure S7A, B**).
194 Enrichment analysis of these concordant DEGs showed that only *Cic-L*-KO shared synapse- and memory-related
195 GO terms with *Atxn1*-KO (**Figure S7C, D**). Consistent with previous findings that the *Etv4/Etv5–Bace1* axis is
196 upregulated in *Atxn1*-KO cortex [12], we observed similar upregulation of this axis in *Cic-L*-KO cortex (**Figure**
197 **S7E**). These findings support the idea that CIC-L and ATXN1 regulate similar gene networks that underlie
198 learning and memory.
199 In the cerebellum, a region generally more resistant to the loss of the CIC–ATXN1/1L complex but clearly
200 vulnerable to the gain of CIC-ATXN1 function [22, 23], we observed fewer DEGs in *Cic-L*-KO compared to the
201 cortex, but more DEGs in *Cic-S*-KO relative to its cortical counterpart (**Figure S8A and B**), potentially reflecting
202 higher CIC-S expression in this region at 12 weeks (**Figure S1**). The two KOs also showed distinct GO term
203 enrichments (**Figure S8C–F**), and DEG overlap with *Atxn1*-KO was stronger for *Cic-S*-KO (Fisher's exact test,
204 OR = 2.41, $p = 1.69 \times 10^{-17}$; **Figure S9A**) than for *Cic-L*-KO (Fisher's exact test, OR = 1.61, $p = 2.85 \times 10^{-8}$;

205 **Figure S9B**). *Etv4/5*, which were strongly upregulated in the cortex of *Cic-L*-KO, showed less upregulation in the
206 cerebellum, and *Bace1* remained unchanged (**Figure S9C**).
207 CIC-L and CIC-S, therefore, largely regulate different biological processes. Their respective transcriptional
208 convergences with ATXN1/1L are both region- and timepoint-dependent.
209
210 **CIC-S preferentially associates with ATXN1L, and CIC-L associates with ATXN1, in a region-specific**
211 **manner**
212 Given the region-specific transcriptomic overlap—between the cortex in *Cic-L*–KO and *Atxn1*–KO mice,
213 and the cerebellum in *Cic-S*–KO and *Atxn1*–KO mice—we sought to determine whether CIC isoform–paralog
214 pairs form distinct complexes across brain regions. We performed size exclusion chromatography (SEC) on
215 protein extracts from cortical and cerebellar tissue harvested from 2–3-month-old *Cic-S*-KO and *Cic-L*-KO mice
216 to examine the elution profiles of CIC, ATXN1, and ATXN1L.
217 In the WT cortex both ATXN1 and ATXN1L co-eluted with CIC in a high molecular weight complex (~1.7
218 MDa, fraction 12), as well as in distinct smaller complexes (ATXN1 peaking at fraction 15, ~300 kDa; ATXN1L
219 at fractions 16–17, ~170–88 kDa; **Figure 5A, B**), consistent with prior reports [14]. In contrast, both *Cic-S*-KO
220 and *Cic-L*-KO cortices showed less ATXN1 and ATXN1L in the high molecular weight fractions and a
221 concomitant increase in smaller fractions, suggesting that CIC is essential for maintaining their incorporation into
222 the larger complex (**Figure 5A, B**). Interestingly, in *Cic-L*-KO cortex, ATXN1 showed a stronger shift from the
223 large to small complex (large-to-small complex ratio: 0.41± 0.05, vs. WT 1.00± 0.08; $p = 0.003$, one-way
224 ANOVA) than in *Cic-S*-KO (0.67± 0.08; $p = 0.039$) (**Figure 5C**, left), indicating preferential association of
225 ATXN1 with CIC-L. In contrast, in *Cic-S*-KO cortex ATXN1L shifted markedly to smaller fractions (0.11± 0.03,
226 vs. WT 1.00± 0.03; $p = 0.0001$) and was nearly absent from the large complex. This shift was less pronounced in
227 *Cic-L*-KO (0.51± 0.13; $p = 0.006$; **Figure 5C**, right), suggesting a stronger association of ATXN1L with CIC-S.
228 These findings confirm isoform-specific associations within the CIC–ATXN1/ATXN1L complex in the cortex.
229 In the *Cic-S*-KO cerebellum, ATXN1L shifted markedly toward smaller fractions (0.07± 0.03, vs. WT
230 1.00± 0.08; $p = 0.0002$). The shift was more modest in *Cic-L*-KO (0.63± 0.13; $p = 0.042$; **Figure 5D–F**),
231 consistent with its preferential association with CIC-S observed in the cortex. In contrast to the cortex, ATXN1
232 also shifted in *Cic-S*-KO cerebellum (0.42± 0.06, vs. WT 1.00± 0.17; $p = 0.015$), while its distribution remained
233 largely unchanged in *Cic-L*-KO (0.90± 0.11; $p = 0.82$; **Figure 5D-F**).
234 The cerebellum has the highest level of CIC-S among all brain regions, as well as the highest total CIC
235 expression (**Figure S1A–C**). ATXN1 is expressed at lower levels, however, resulting in the highest CIC-to-
236 ATXN1 ratio in the cerebellum (**Figure S1D**), which may lead to a greater proportion of ATXN1 bound to CIC.
237 Supporting this hypothesis, comparison of WT cerebellar and cortical SEC patterns (**Figure S10**) revealed a
238 striking difference: the cerebellum exhibits a higher proportion of ATXN1 in the large complex and lower levels

239 in the small complex (**Figure S10C**), a pattern not observed for ATXN1L (**Figure S10D**). The relatively high
240 CIC-S levels may explain why CIC-S is especially important for ATXN1 large complex formation in the
241 cerebellum. Moreover, the combination of a high CIC/ATXN1 ratio likely predisposes the cerebellum to gain-of-
242 function effects of the complex. It is worth noting that cerebellar single-nucleus RNAseq [33] show that both *Cic*
243 and *Atxn1* are highly expressed in Purkinje cells, which are most vulnerable to SCA1 (**Figure S11**).
244 Overall, these results indicate that CIC isoforms and ATXN1 paralogs favor particular interactions
245 depending on the context. While the CIC-S–ATXN1L interaction appears strong and consistent across regions,
246 ATXN1 associates with different CIC isoforms in different tissues.
247
248 **The unique N-terminal of each CIC isoform mediates the ATXN1 paralog-specific interaction**
249 To understand the mechanism of isoform-specific complex formation of CIC with either ATXN1 or
250 ATXN1L, we first examined the interaction of CIC-L and CIC-S with ATXN1 and ATXN1L by performing
251 coimmunoprecipitation (co-IP) assays in WT cortical lysates. Despite comparable expression levels of CIC
252 isoforms in the cortex, ATXN1 immunoprecipitation showed strong enrichment of CIC-L relative to CIC-S
253 (**Figure 6A**). Conversely, ATXN1L immunoprecipitation demonstrated an even more pronounced enrichment of
254 CIC-S over CIC-L (**Figure 6B**), which may explain the more prominent ATXN1L shift observed in SEC (**Figure**
255 **5A-C**). These results support the hypothesis that ATXN1 preferentially interacts with CIC-L, whereas ATXN1L
256 favors CIC-S.
257 Since CIC-S and CIC-L share a canonical ATXN1/ATXN1L binding domain (ATXN1/1L-BD) that can
258 bind both ATXN1 and ATXN1L, we hypothesized that there should be isoform-specific binding sites outside the
259 ATXN1/1L-BD that mediate preferential interaction between CIC-S and ATXN1L, and CIC-L and ATXN1. To
260 test this possibility, we performed the analysis using human CIC, ATXN1, and ATXN1L sequences. The W37A
261 mutation in CIC-S has been previously shown to disrupt binding with the AXH domain of ATXN1[18]. Using the
262 corresponding mutation in CIC-L (W946A), we found that CIC-S W37A lost interaction with ATXN1 but CIC-L
263 W946A retained ATXN1 binding (**Figure 6C**). This suggests that CIC-L contains an additional ATXN1-binding
264 region. We saw the opposite with ATXN1L, despite CIC-S containing only 22 isoform-specific amino acids at its
265 N-terminus: ATXN1L still bound CIC-S W37A but failed to interact with CIC-L W946A (**Figure 6D**). These data
266 indicate that distinct, noncanonical regions in each isoform mediate selective interactions with ATXN1 and
267 ATXN1L.
268 To test whether the N-terminal region of CIC-L can mediate the selective interaction with ATXN1, we
269 expressed this region alone and performed co-IP. The CIC-L N-terminus, which spans 931 amino acids and does
270 not contain the ATXN1/1L-BD, interacted with ATXN1 but not with ATXN1L (**Figure 6E, F**). Thus, the N-
271 terminal region of CIC-L is sufficient to mediate interaction with ATXN1, but not ATXN1L, and may thereby
272 contribute to isoform-paralog specificity.

Together, these findings reveal a previously unrecognized mechanism by which CIC isoforms interact with different ATXN1 paralogs through isoform-specific domains, establishing a novel molecular basis for paralog-specific complex formation.

CIC-L expression is more dependent on ATXN1 than CIC-S

Given the preferential interaction between ATXN1 and CIC-L, and the known role of ATXN1 in maintaining CIC protein stability [16], we hypothesized that CIC-L would be more dependent on ATXN1. To test this, we measured CIC-L and CIC-S protein levels in WT and *Atxn1*-KO mice across brain regions sensitive to ATXN1 loss (cortex, hippocampus) or gain (cerebellum, brainstem). CIC-L levels were significantly lower in all four regions in *Atxn1*-KO mice, but CIC-S levels were more modestly reduced, and only in the cerebellum (**Figure S12A-D**). The cerebellar reduction in CIC-S was consistent with our SEC analysis showing a shift of CIC-S in ATXN1-containing fractions (**Figure 5D-F**). IP of ATXN1 confirmed a preferential association with CIC-L in all brain regions examined (**Figure S12E**). These findings indicate that the cortical and hippocampal abnormalities in *Atxn1*-KO mice may result primarily from the loss of CIC-L function.

DISCUSSION

The present study uncovers functionally distinct roles for the two CIC isoforms, CIC-L and CIC-S. The overlap of transcriptomic changes in *Cic-S*-KO and *Atxn1l*-KO mice explains why these two lines show the same early lethality, hydrocephalus, and lung alveolarization defects, while *Cic-L*-KO and *Atxn1*-KO mice share learning and memory deficits. These phenotypic parallels reveal an unexpected degree of specificity in CIC–ATXN1 complex formation, given that both isoforms share the canonical ATXN1/ATXN1L binding domain (ATXN1/1L-BD); the specificity is attributable to the unique N-terminal domains of each isoform that favor binding between CIC-S and ATXN1L, and between CIC-L and ATXN1. The sharp distinctions between the two sets of phenotypes also provides compelling *in vivo* support of the idea that different isoforms greatly expand protein interaction capacities and may indeed serve as "functional alloforms" [34].

Prior studies in *Drosophila* have shown that CIC isoforms play distinct roles during development: Cic-S is critical for early embryogenesis through its interaction with Groucho [35], while Cic-L is essential for oogenesis [36]. These findings support the idea that the divergent N-terminal regions of CIC isoforms confer isoform- and context-specific functions. However, *Drosophila* Cic-S has evolved on its own [35], with the species possessing only Atx-1 (containing the AXH domain) and lacking the paralog pairs found in mammals. Furthermore, there is no evidence for an endogenous functional complex between *Drosophila* Cic and Atx-1. By using a mammalian system in which both CIC isoforms and ATXN1 paralogs are highly conserved, we identified a previously unrecognized layer of regulation: isoform–paralog–specific interactions.

These isoform-specific roles are not only of evolutionary interest but also carry direct clinical relevance.

307 Recently, a CIC-L–specific mutation was reported in a patient with CHS [37], which is consistent with our finding
308 that CIC-L is critical for neurological function. There may be additional patients with CIC-L variants who remain
309 unrecognized. Furthermore, given the role of CIC-L in learning and memory and its regulation of the *Etv4/5–*
310 *Bace1* axis, CIC-L variants could, like *ATXN1* [10-12, 38, 39]. contribute to susceptibility to Alzheimer's disease.
311 More broadly, the expression of different isoforms can be dysregulated in disease, but gene-level expression
312 measures commonly used in biology average the translational potential of each isoform and therefore may not
313 accurately reflect protein concentration. It has also been proposed that isoform-specific therapeutics would
314 improve efficacy and reduce adverse effects [40].

315 Our findings help resolve two longstanding questions: (1) What underlies the regional vulnerability to
316 loss or gain of the CIC–ATXN1 complex despite its broad expression? (2) How does the loss of ATXN1 versus
317 ATXN1L lead to distinct phenotypes, despite both acting through CIC? We propose that the combination of
318 isoform–paralog–specific interactions and regional expression patterns determines the differential sensitivity to
319 complex disruption. Consistent with this model, the similarity in learning and memory deficits, along with
320 upregulation of the *Etv4/5–Bace1* axis, in both *Atxn1*-KO and *Cic-L*-KO mice suggests that these phenotypes
321 arise from preferential CIC-L–ATXN1 complex formation in the cortex and possibly the hippocampus.

322 Although we did not assess isoform-specific contributions to the cerebellar and brainstem pathology seen
323 in SCA1, the current study does shed light on this question. We find that total CIC levels and CIC/ATXN1 ratios
324 are highest in the cerebellum, with a greater proportion of ATXN1 being incorporated into the large complex. Our
325 prior work in SCA1 mice demonstrated that ~50% reduction of total CIC (affecting both isoforms) [21] or a
326 polyQ-ATXN1 mutation that would abolish binding to both isoforms fully rescues the Purkinje cell phenotype
327 [22, 23]. These findings indicate that total CIC levels bound to ATXN1, rather than isoform identity, are critical
328 for Purkinje cell degeneration. By contrast, CIC expression is lowest in the brainstem, which may be why loss of
329 the ATXN1-CIC interaction does not fully rescue brainstem SCA1 phenotypes [21, 22] and why CIC motifs are
330 less enriched in the brainstem's disease-associated gene network [41]. In SCA1, the brainstem's ATXN1 protein
331 complex composition is distinct from that in the cerebellum [42], indicating that brainstem pathology arises
332 through different mechanisms.

333 Our findings also allow us to draw some inferences that can be tested in the future. First, ATXN1 loss
334 does not fully recapitulate CIC-L deficiency: transcriptomic comparisons reveal only partial overlap between
335 *Atxn1*-KO and *Cic-L*-KO datasets, indicating that ATXN1 and CIC-L have functions outside of their shared
336 complex. Indeed, ATXN1 has interactors beyond CIC [22, 43-46], and such partners may act in compensatory
337 transcriptional or chromatin regulatory pathways, buffering CIC-dependent gene expression programs.

338 What are the ATXN1/ATXN1L-independent mechanisms that distinguish CIC-L from CIC-S? If isoform
339 differences were driven solely by interactions with ATXN1 and ATXN1L, *Cic-L*-KO and *Cic-S*-KO
340 transcriptomes would be expected to fully overlap, differing only in the magnitude of differential expression.

Instead, the partial overlap of DEGs may reflect isoform-specific interactors that influence target gene selection. Alternatively, the isoforms may differ in cell type expression, affecting distinct cellular populations. Divergent GO term enrichments support this notion. Profiling isoform-specific expression across cell types and applying single-cell transcriptomics to each isoform KO will be critical to dissect these contributions.

Finally, our findings highlight a broader principle that paralogous proteins and their isoform-specific partners can form distinct regulatory complexes with completely different biological functions. The CIC–ATXN1/1L complex may serve as a paradigm for understanding how paralog-isoform specific interactions shape regional vulnerability and disease susceptibility in other diseases such as those involving amyloid precursor protein [47-53], tau [54, 55], or α-synuclein [56], all of which exist within paralog families and have multiple isoforms.

Limitations of the study

Because the majority of *Cic-S*-KO animals die before P21, only the healthiest underwent neurobehavioral characterization. We therefore cannot rule out the possibility that young animals might also have neurological deficits. To better understand the role of CIC isoforms in the adult brain, future studies employing conditional knockout models of each isoform, as well as of *Atxn1l,* will be critical. Such models could uncover additional functions of CIC isoforms and context-dependent effects not captured in our germline KO approach.

ACKNOWLEDGEMENTS

We thank Ji-Yoen Kim, Sameer Bajikar, Dah-eun Chloe Chung, Harini Pallikarana, Sih-Rong Wu, and Xue Deng for their critical comments on the manuscript, and Harry Orr and Vicky Brandt for editorial input. We further thank the Baylor College of Medicine (BCM) Genetically Engineered Rodent Model (GERM) Core, Jan and Dan Duncan Neurological Research Institute (NRI) Rodent Neurobehavior Core, and Neuropathology Core. Funding: This work was funded by National Institute of Aging/NIH grant F31 AG077918 (H.L.), Cockrell Family Funding (H.L.), National Institute of Neurological Disorders and Stroke/NIH grant R01 NS027699 (H.Y.Z.), Freedom Together Foundation grant MR-2023-4260 (H.Y.Z.), Howard Hughes Medical Institute (H.Y.Z.), Chao Endowment (Z.L.), Huffington Foundation (Z.L.), National Institutes of Health Cancer Center Grant /NIH grant P30 CA125123 (BCM GERM Core), Cancer Biology Research Grant /NIH grant 1R50CA283804 (BCM GERM Core), Eunice Kennedy Shriver National Institute of Child Health and Human Development Intellectual and Developmental Disabilities Research Center/NIH grant P50HD103555 (NRI Rodent Neurobehavior Core and Neovascularization Core).

AUTHOR CONTRIBUTIONS

H.L. and H.Y.Z. conceived the study, designed the experiments, and analyzed and interpreted the data. H.L. generated mouse models. H.L., E.V.G., and E.M.R. performed protein biochemistry from mouse tissues, and H.L. performed cellular experiments. H.L. and E.H.-Y.C. performed RNA extractions, and H.C. performed alignment of raw sequencing data; H.L. performed RNA-seq analysis. H.C. and Z.L. supervised RNA-seq processing and analysis. H.L., E.H.-Y.C., and K.X. performed mouse genotyping, colony management, and survival and weight tracking. M.A.D. designed and generated the hCIC-L plasmid. H.L. and M.A.D. generated mutant constructs and performed additional cloning, with technical assistance from E.H.-Y.C. R.R. performed size-exclusion chromatography. S.V. performed mouse behavior experiments. B.S. processed lung tissue for histopathology and performed morphometric analysis. H.L. wrote the manuscript and prepared the figures with input from all authors. H.Y.Z. interpreted and reviewed all data and edited the manuscript.

DECLARATION OF INTERESTS

H.Y.Z. is on the scientific advisory board of Neurogene, Lyterian, the Column Group, and Cajal Therapeutics, and is a co-founder of Cajal Therapeutics. H.Y.Z. is also a director at Regeneron. None of these activities have any relevance to the work presented in this paper.

TABLES

Table 1. Number of surviving animals from *Cic*$^{S\text{-KO}/+}$ intercrosses at a given time point

Genotype	Observed (%) at P0	Observed (%) at P6	Observed (%) at P21
Cic$^{+/+}$	22 (24.4 %)	32 (28.1 %)	31 (29.8%)
Cic$^{S\text{-KO}/+}$	50 (55.6 %)	65 (57.0 %)	64 (61.5%)
Cic$^{S\text{-KO}/S\text{-KO}}$	18 (20.0 %)	17 (14.9 %)	9 (8.6 %)
Total	90	114	104
Chi-square*	ns (0.26)	ns (0.18)	** (0.007)

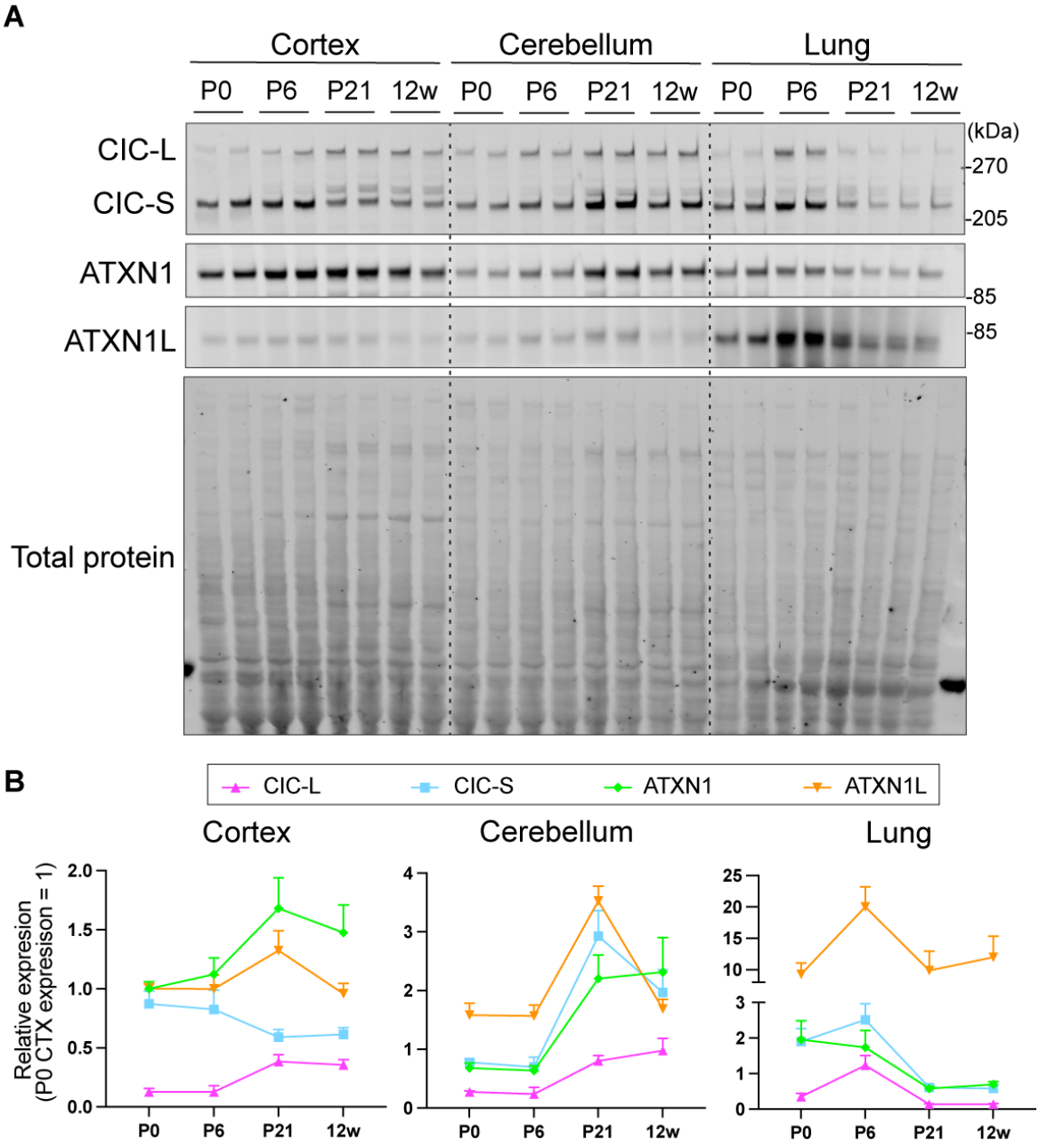
*Chi-square analysis of the observed ratio compared to the expected Mendelian ratio of *Cic*$^{+/+}$ (25%), *Cic*$^{S\text{-}KO/+}$ (50%), and *Cic*$^{S\text{-}KO/S\text{-}KO}$ (25%). Each time point is an independent cohort. ** $p<0.01$

Table 2. Number of surviving animals from *Cic*$^{L\text{-KO}/+}$ intercrosses at a given time point

Genotype	Observed (%) at P0	Observed (%) at P6	Observed (%) at P21
Cic$^{+/+}$	22 (20.9 %)	18 (25.7 %)	28 (26.7 %)
Cic$^{L\text{-KO}/+}$	53 (50.5 %)	38 (54.3 %)	54 (51.4 %)
Cic$^{L\text{-KO}/L\text{-KO}}$	30 (28.6 %)	14 (20.0 %)	23 (21.9 %)
Total	105	70	105
Chi-square*	ns (0.53)	ns (0.86)	ns (0.74)

*Chi-square analysis of the observed ratio compared to the expected Mendelian ratio of *Cic*$^{+/+}$ (25%), *Cic*$^{L\text{-}KO/+}$ (50%), and *Cic*$^{L\text{-}KO/L\text{-}KO}$ (25%). Each time point is an independent cohort.

400 **FIGURES**



401

402 **Figure 1. CIC-L and CIC-S show different spatiotemporal expression patterns in the mouse cortex,**
403 **cerebellum, and lung.**

404 (A) Representative immunoblot of CIC-L, CIC-S, and related proteins in wild-type (WT) mouse cortex,
405 cerebellum and lung lysates at indicated time points. P, postnatal day; w, week-old. Antibody against the
406 C terminus of CIC recognizes both isoforms of CIC.
407 (B) Quantification of CIC-L and CIC-S protein expression in WT cortex, cerebellum and the lung. Each
408 protein was normalized to total protein, and each timepoint was normalized to P0 cortex (CTX) value of

409 each protein. CIC isoforms were normalized to total CIC in P0 CTX. Data are displayed as mean ± SEM
410 from n = 4 for each time point.

411

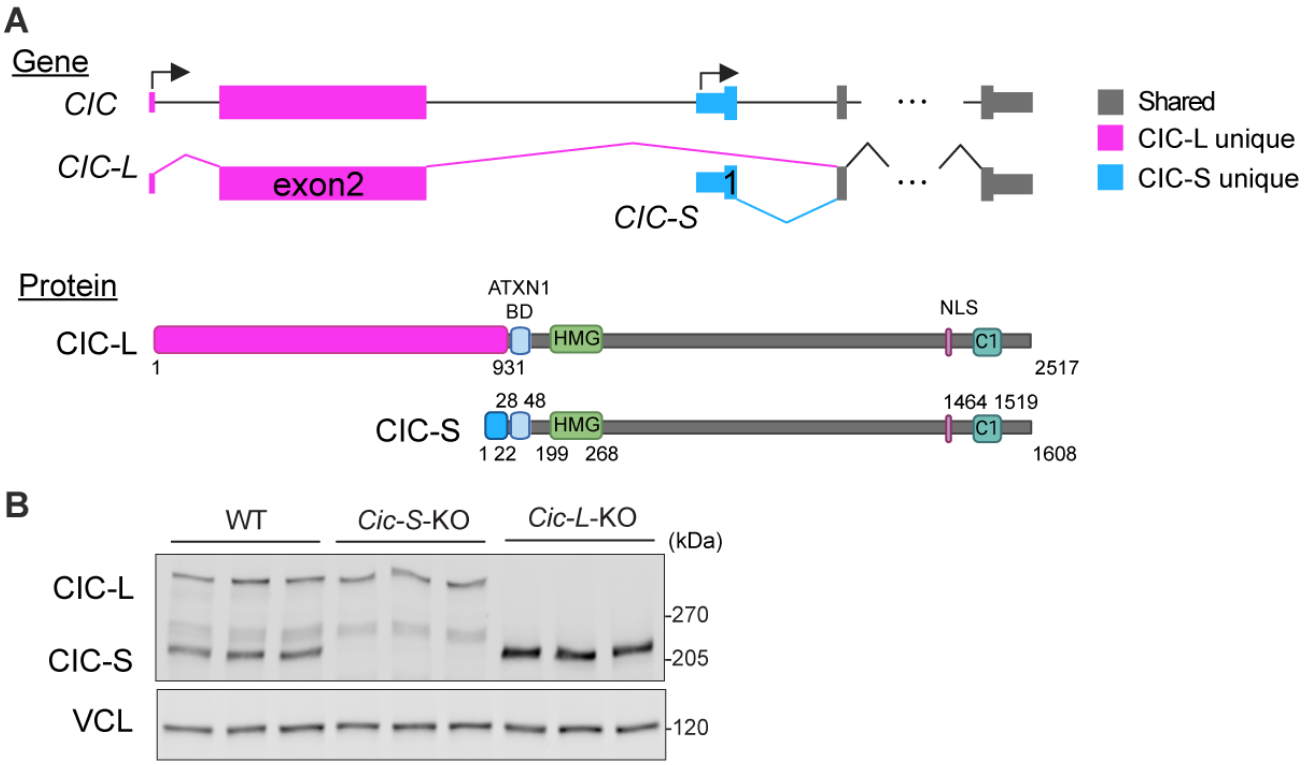


Figure 2. *Cic-L*-KO and *Cic-S*-KO mice were generated using CRISPR.

(A) Gene and protein structure of human CIC-L and CIC-S. ATXN1/1L BD: ATXN1/1L binding site, HMG: HMG box, C1: C1 domain, NLS: Nuclear localization signal

(B) Immunoblot of CIC-L and CIC-S of cortex lysate of WT, *Cic-S*-KO and *Cic-L*-KO mice at 1 month of age. Vinculin (VCL) as loading control. n =3 per genotype.

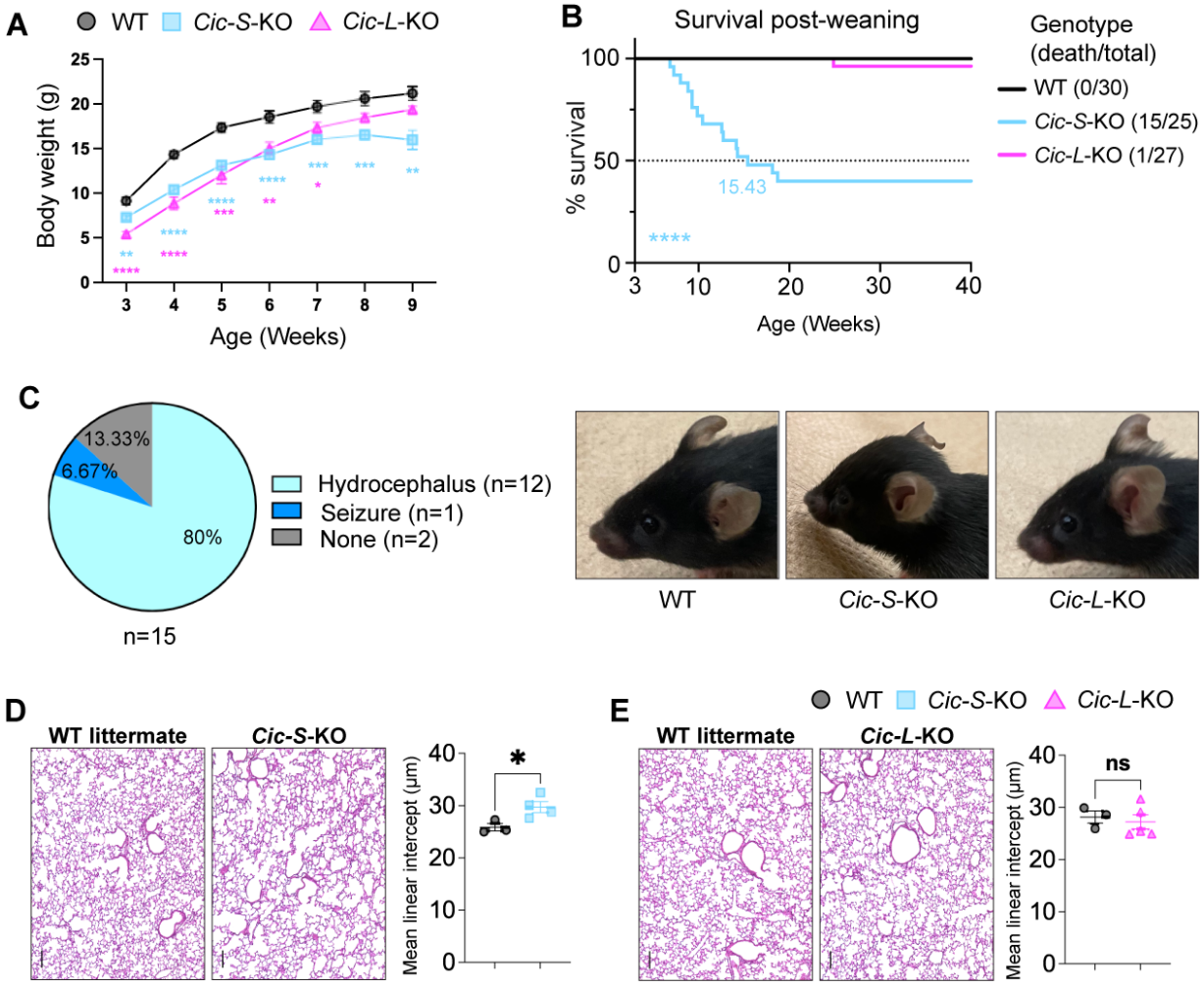


Figure 3. *Cic-S*-KO mice display early lethality and lung alveolarization defects.

(A) Weekly body weights of *Cic-S*-KO, *Cic-L*-KO, and their WT littermate controls starting from week 3. n =16 for each genotype, 8 males and 8 females. All data are represented as mean ± SEM. Two-way ANOVA with repeated measures, with Tukey's multiple comparisons. In each case, *, **, ***, and **** denote $p < 0.05$, $p < 0.01$, $p < 0.001$, and $p < 0.0001$ respectively.

(B) Kaplan-Meier curves for survival post-weaning (3 weeks) of WT, *Cic-S*-KO and *Cic-L*-KO mice. n = 30, 25, and 27 respectively). Mantel-Cox test. **** denote $p < 0.0001$.

(C) Left: Phenotypes observed before lethality of *Cic-S*-KO animals (n = 15). Right: Representative photographs of the heads of WT, *Cic-S*-KO and *Cic-L*-KO mice at 9 weeks of age; *Cic-S*-KO mice show dome-shaped heads due to hydrocephalus.

(D and E) Lung tissues from *Cic-S*-KO (D) and *Cic-L*-KO (E) mice and their WT littermates were collected at P21 for morphometric analysis. Left: Representative hematoxylin and eosin (H&E)-stained lung sections. Scale bar, 100 µm. Right: Quantification of alveolar airspace by mean linear intercept (MLI). Data are presented as mean ± SEM (n = 3–5 per group). Unpaired t-test. In each case, *, and ns denote $p < 0.05$, and $p > 0.05$, respectively.

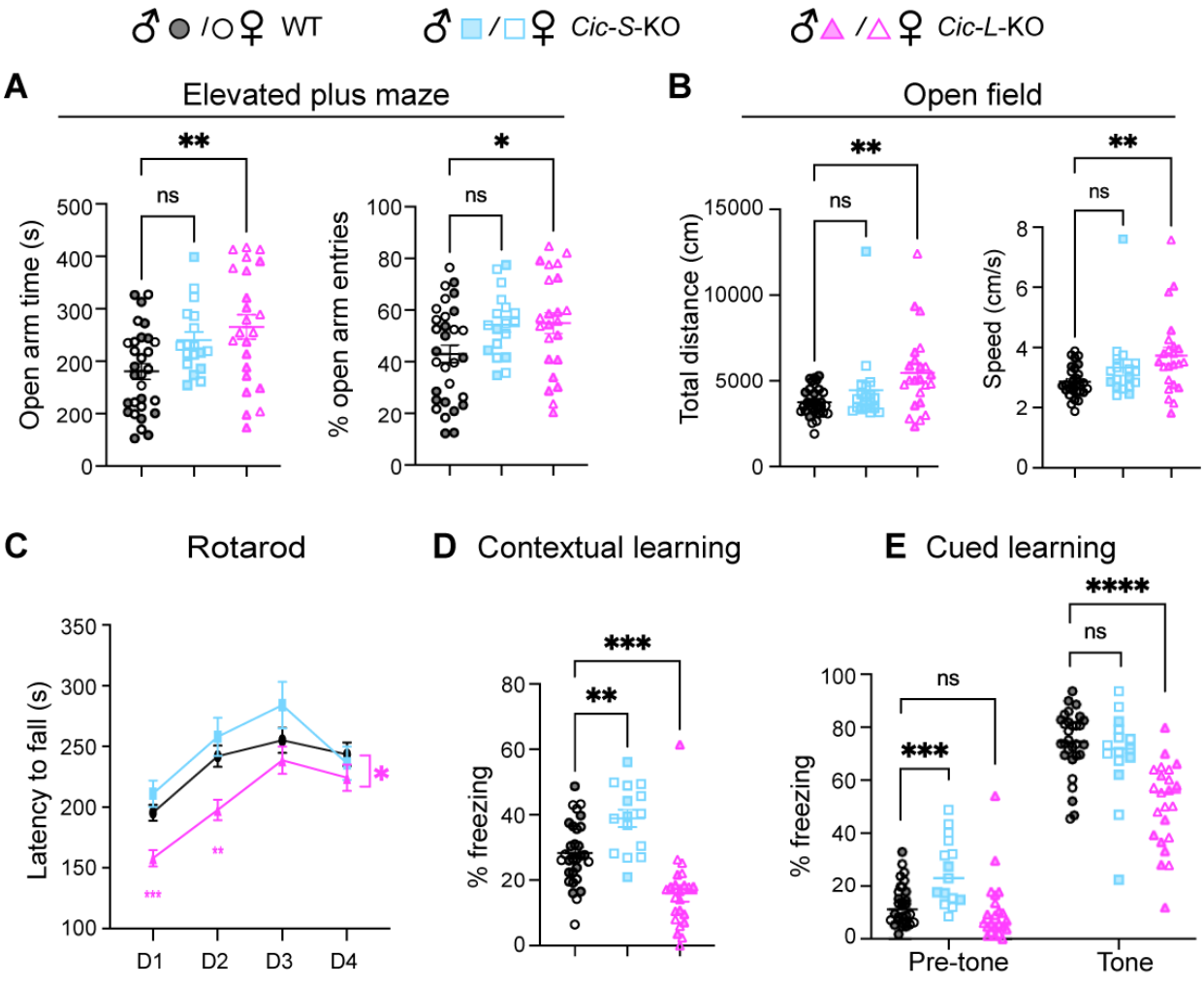


Figure 4. *Cic-L*-KO mice display neurobehavioral deficits.

(A-E) Behavior characterization of *Cic-S*-KO, *Cic-L*-KO, and their WT littermate controls was performed starting at 8 weeks of age.

(A) Anxiety-like behavior measured with an elevated plus maze (EPM) at 8-10 weeks of age.
(B) Activity measured with open field assay (OFA) at 9-11 weeks of age.
(C) Motor coordination measured with rotarod assay at 9-11 weeks of age. Data points indicate male and female combined mean ± SEM. (D = Day)

(D and E) Learning and memory assessed with fear conditioning assay (FC) at 19-21 weeks of age.

For (A) and (B), n = 30 (14 males, 16 females) WT, n = 18 *Cic-S*-KO (5 males, 13 females), and n = 23 *Cic-L*-KO (12 males, 11 females) were analyzed. Due to the early lethality of *Cic-S*-KO mice, n = 17 (5 males, 12 females) and n = 15 (4 males, 11 females *Cic-S*-KO mice were analyzed for (C), (D) and (E) respectively. One-way ANOVA with Tukey's multiple comparisons test were used for (A), (B), (D) and (E), two-way ANOVA with repeated measures, with Tukey's multiple comparisons for (C). In each case, *, **, ***, **** and ns denote $p < 0.05$, $p < 0.01$, $p < 0.001$, $p < 0.0001$, and $p > 0.05$, respectively. All data are represented as mean ± SEM.

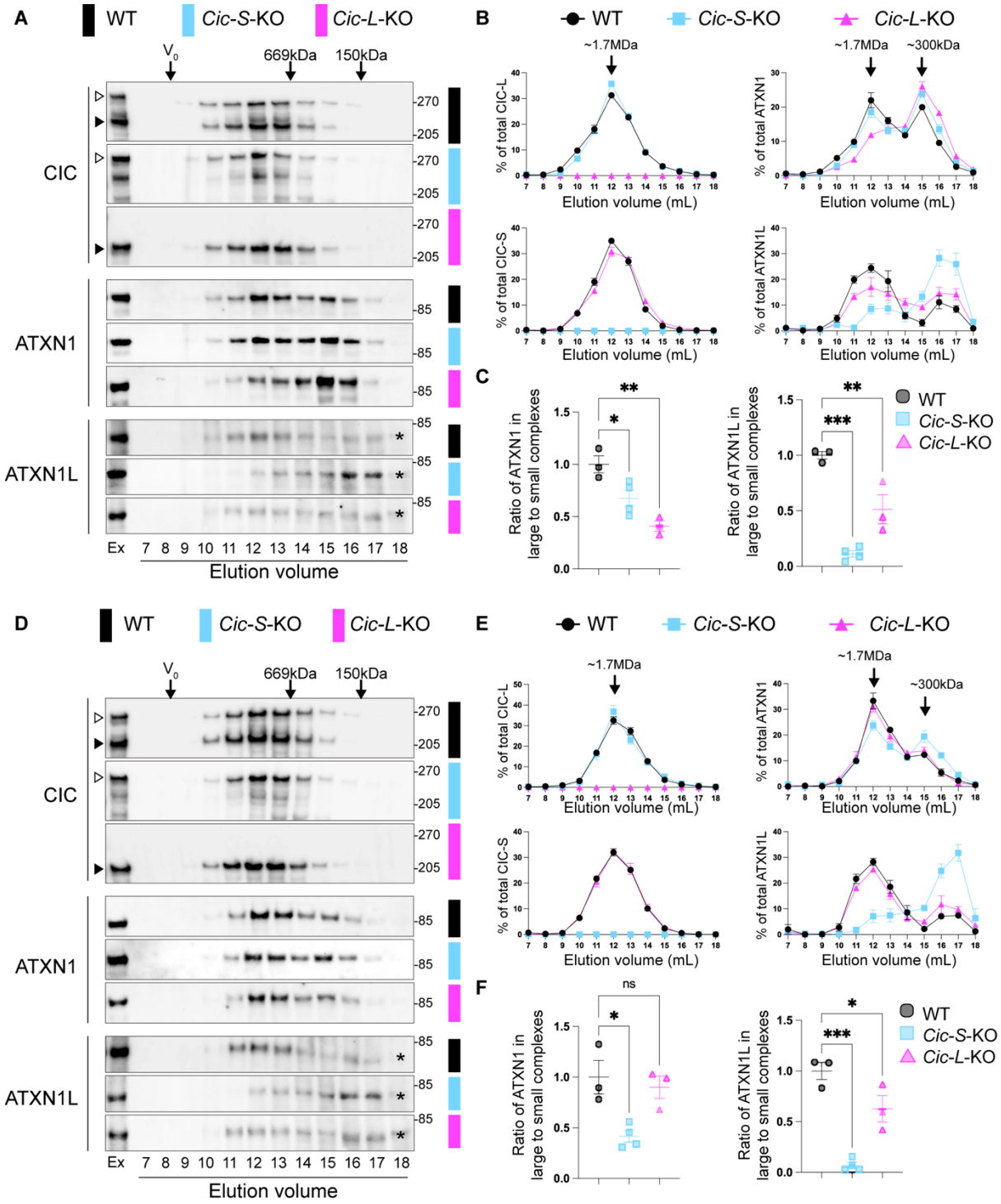


Figure 5. Differential complex formation of ATXN1 and ATXN1L in *Cic-S*-KO and *Cic-L*-KO in the cortex and the cerebellum

(A-C and D-F) Representative immunoblot of size exclusion chromatography (SEC) of WT, *Cic-L*-KO, and *Cic-S*-KO cortex (A) and cerebellum (D) lysates probed for CIC-L, CIC-S, ATXN1, and ATXN1L. The column void volume (V_0), gel-filtration standards thyroglobulin (669 kDa) and ADH (150 kDa), and

elution volume (ml) of each collected fraction are indicated. Ex, sample extract; open arrow head, CIC-L; closed arrow head, CIC-S; *, non-specific band.

(B and E) SEC elution profile of each protein of WT, *Cic-L*-KO, and *Cic-S*-KO cortex (B) and cerebellum (E). The percentage of each protein (mean ± SEM) in each fraction was determined from n = 3–4 independent cortical extracts. Protein abundance in each fraction was quantified by densitometry, with the signal in each fraction normalized to the total signal across all fractions for each genotype.

(C and F) Quantification of the ratio of ATXN1 and ATXN1L in large to small molecular weight fractions in the cortex (C) and the cerebellum (F). Left: Ratio of ATXN1 in the large (peak elution fractions 11–13) to small (peak elution fractions 15–17) complexes, normalized to WT. Right: Ratio of ATXN1L in the large (peak elution fractions 11–13) to small (peak elution fractions 16–17) complexes, normalized to WT. Statistical significance was assessed by one-way ANOVA. In each case, *, **, and *** denote $p < 0.05$, $p < 0.01$, $p < 0.001$, respectively. All data are represented as mean ± SEM.

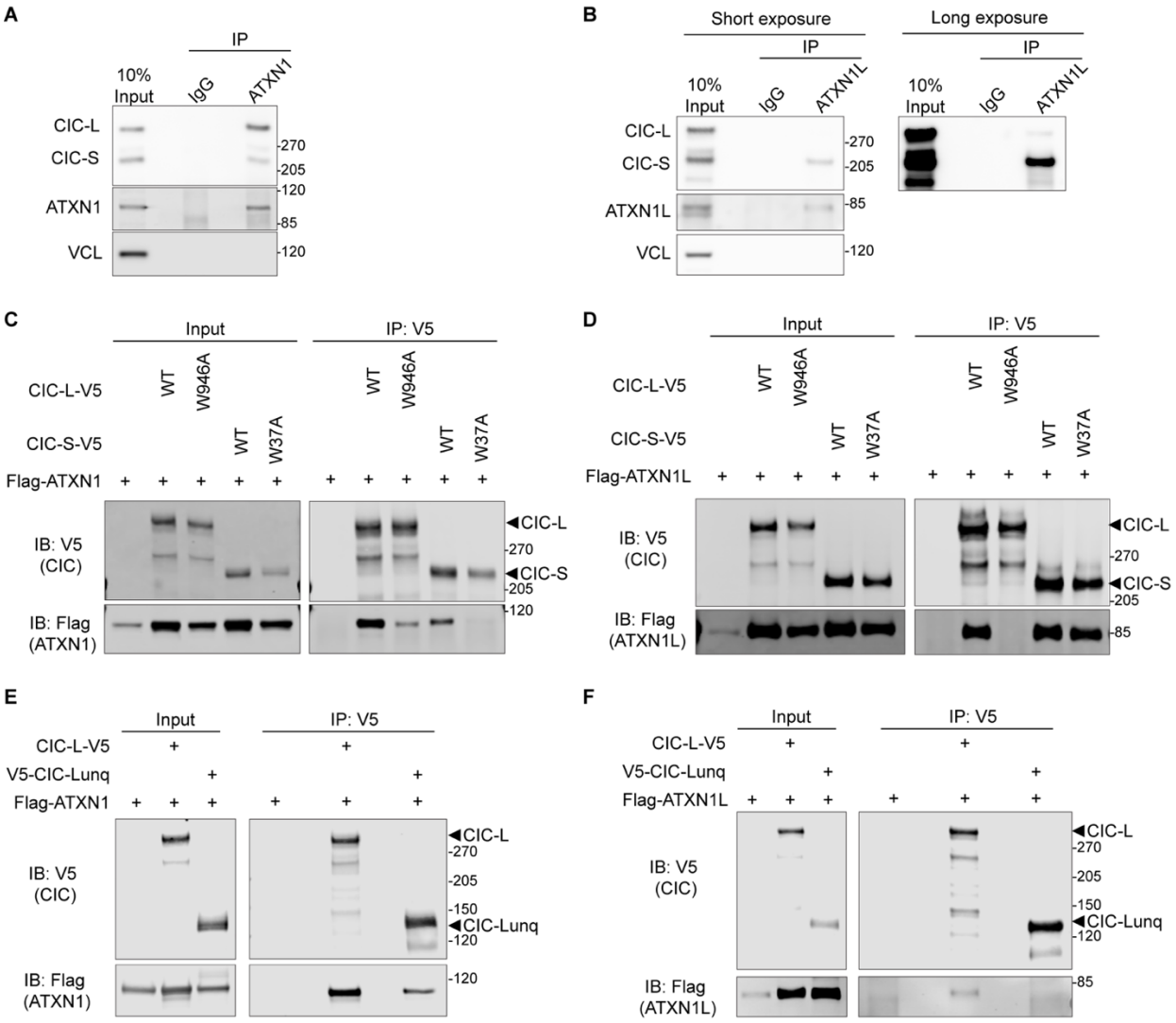


Figure 6. The unique domain of CIC-L interacts exclusively with ATXN1, while the unique domain of CIC-S interacts exclusively with ATXN1L.

(A and B) Representative immunoblots of immunoprecipitation (IP) experiments from WT cortex. (A) IP of ATXN1 and co-IPed CIC-L and CIC-S. (B) IP of ATXN1L and co-IPed CIC-L and CIC-S.

(C and D) Representative immunoblots of IP experiments from HEK293T co-transfected with V5-tagged CIC-L or CIC-S, either WT or with an ATXN1/1L binding mutation (W946A or W37A), and Flag-tagged ATXN1 (C) or ATXN1L (D). IP of V5 and co-IPed Flag.

(E and F) Representative immunoblots of IP experiments from HEK293T co-transfected with V5-tagged CIC-L or CIC-L unique N-terminal (CIC-Lunq), and Flag-tagged ATXN1(C) or ATXN1L (D). IP of V5 and co-IPed Flag.

482 **EXPERIMENTAL MODEL DETAILS**

483 **Mouse husbandry**

484 All mice were backcrossed to C57BL/6J background. All mice were housed and maintained in the animal
485 facilities at Baylor College of Medicine (BCM) on a 12-hour light-dark cycle. BCM Institutional Animal Care &
486 Use Committee (IACUC) approved all animal care and experimental procedures.

487

488 **HEK293T cell line**

489 Human embryonic kidney cell T (HEK293T) (ATCC, CRL-3216) were cultured in Dulbecco's modified Eagle's
490 medium (DMEM, Corning, MT10013CV) supplemented with 10% fetal bovine serum (Gibco A4736201) and 1X
491 antibiotic-antimycotic (Gibco, 15240096). Cell cultures were maintained at 37 °C in a humidified 5% CO_2
492 incubator.

493

494 **METHOD DETAILS**

495 **Generation of *Cic-L*-KO and *Cic-S*-KO mice**

496 *Cic-L*-KO and *Cic-S*-KO alleles were generated using CRISPR/Cas9 genome editing. For *Cic-L*-KO, two crRNAs
497 targeting exon 2 of *Cic-L*, the first coding exon unique to the CIC-L isoform, were designed (5′-
498 CGCACCTGGCGCACGACTGC-3′ and 5′-TCCAGCCGCACCGACTGCTG-3′; IDT) to generate a large
499 deletion. For Cic-S-KO, a single crRNA targeting exon 1 of Cic-S was designed (5′-
500 ACGCGGGTATCAGGGGCCTG-3′; IDT).

501 crRNA, tracrRNA, and Cas9 protein were prepared as described previously [22] with modifications to total
502 reaction volume, concentration and delivery method. Final concentrations were 6.6 μM crRNA, 6.6 μM
503 tracrRNA, and 6 μM Cas9 in $T_{10}E_{0.1}$ buffer. For *Cic-L*-KO, each crRNA was diluted to 3.3 μM. The mixture was
504 electroporated into zygotes by the Genetically Engineered Rodent Model (GERM) Core at Baylor College of
505 Medicine (BCM).

506 Founders were screened by PCR using primers flanking the expected deletion sites (*Cic-L*-KO: 5′-
507 TCCTCCTCTACTGACACGGC-3′, 5′-TAGGACAAGCAGCTCATCACC-3′; *Cic-S*-KO: 5′-
508 GGCCAAGGAGCGGGTTAC-3′, 5′-TCCTAGTGCAGTACCGTCCA-3′). PCR products were TOPO cloned
509 (Invitrogen) and analyzed by Sanger sequencing for frameshift deletions. *Cic-S*-KO founders were selected for

510 deletions eliminating the downstream ATG in exon 1. Founder lines were backcrossed to C57BL/6J WT mice
511 (Jackson Laboratory), and germline transmission was confirmed by sequencing of F1 progeny.

512 Immunoblotting was performed on homozygous KO animals to confirm the loss of the corresponding CIC
513 isoform. One independent founder line was established for each isoform KO and maintained by successive
514 backcrossing (>5 generations) to C57BL/6J prior to experimental use.

515

516 **Genotyping of *Cic-L*-KO and *Cic-S*-KO mice**

517 Genotyping was performed by PCR using allele-specific primers. For *Cic-L*, WT and KO alleles were
518 distinguished using two primer sets: *Cic-L*-WT (5′-ACTGGAGAAGGGAGCTGCAC-3′ and 5′-
519 CTGTCCCTCGAAGTCTCATCG-3′) and *Cic-L*-KO (5′-TCCTCCTCTACTGACACGGC-3′ and 5′-
520 TAGGACAAGCAGCTCATCACC-3′). For *Cic-S*, a single primer pair (5′-GGCCAAGGAGCGGGTTAC-3′ and
521 5′-TCCTAGTGCAGTACCGTCCA-3′) was used, and WT versus KO alleles were distinguished based on product
522 size.

523

524 **Lung Histopathology**

525 The lungs were harvested for morphometry studies on postnatal day (P) 21, as described before [57]. Both lungs
526 of the experimental mouse (n=3-5/group) were inflated at 25 cm pressure for at least 10 minutes via tracheal
527 cannulation with 4% paraformaldehyde. The inflated lungs were then fixed overnight with 4% paraformaldehyde
528 at room temperature with gentle agitation. The fixed lungs were stored in 70% ethanol in a refrigerator until
529 further processing for lung histopathology. The lungs were then paraffin-embedded, sliced into 5 µm thin sections,
530 and stained with H&E. We examined at least five random non-overlapping fields of 10X H&E stained images
531 from both lungs of each animal to quantify alveolarization by mean linear intercepts (MLIs). To calculate the
532 MLI, grids of horizontal and vertical lines were placed on the H&E stained lung images, and the number of times
533 the lines intersected with the lung tissue was counted. The MLI was determined by dividing the total length of the
534 grid lines by the number of intersects.

535

536 **Behavioral Assays**

537 All behavioral analyses were conducted during the light phase of a 12-hour light-dark cycle by an experimenter
538 blinded to mouse genotype. Mice had *ad libitum* access to food and water except during testing. Age-matched
539 mice were used in all experiments, and wild-type (WT) littermates from both C*ic-S*-KO and *Cic-L*-KO breeding

were pooled as controls. Mice were habituated in the test room for at least 30 min prior to each assay. Tests were performed under room lighting of 150 lux with background white noise at 60 dB.

Elevated plus maze: Mice (8-10 weeks of age) were placed in the center of an elevated plus maze consisting of two open arms (25 × 8 cm, with 0.5 cm raised edges) and two closed arms (25 × 8 cm, enclosed by 15 cm high walls), elevated 50 cm above the floor. Behavior was recorded for 10 min using a camera and ANY-maze video tracking system (Stoelting, Wood Dale, IL, USA). Measures included time spent, distance travelled and number of entries in open and closed arms.

Open field assay: Mice (9-11 weeks of age) were placed in the center of a 40 × 40 × 30 cm chamber equipped with photobeams (AccuScan, Columbus, OH, USA) and allowed to explore for 30 min testing period. Locomotor activity was automatically analyzed using AccuScan Fusion software (Omnitech). Parameters included total distance traveled and mean motor speed.

Rotarod: This assay was performed as described previously [22]. Mice (9-11 weeks of age) were placed on a rotating rod (3 cm diameter, 30 cm length; Ugo Basile) and tested for four consecutive days, with four trials per day and a minimum 30-min inter-trial interval. Each trial lasted a maximum of 10 min. The rod accelerated linearly from 5 to 40 rpm over the first 5 min, and then remained at 40 rpm for the final 5 min. Latency to fall was recorded, with two consecutive passive rotations scored as a fall.

Pole test: Mice (15-17 weeks of age) were placed head-upward at the top of a vertical threaded metal pole. Time to descend to the base was recorded with a 60 s cutoff. If a mouse turned but fell before reaching the floor, the time to fall was recorded. Mice unable to turn downward were scored as 60 s.

Conditioned fear: On day 1, mice (19-21 weeks of age) were placed in a conditioning chamber (Actimetrics, Coulbourn instruments, Wilmette, IL) and presented with two conditioned stimulus (CS)- unconditioned stimulus (US) pairings, programmed using the FreezeFrame software (Actimetrics). The trial comprised of 2 min of no sound, followed by 30 s of sound at 80 dB (CS), which was followed by a mild 0.8 mA foot shock for 2 s (US). The tone-shock pairing was repeated after 2 minutes. Twenty-four hours later, contextual memory was assessed by returning the mice to the same chamber for 5 min and quantifying freezing behavior. One hour later, cued fear memory was assessed in a novel chamber by monitoring freezing for 3 min without stimulus, followed by 3 min with the tone. Freezing was recorded and quantified using FreezeFrame software (Actimetrics).

Three-chamber sociability assay: Partner mice (matched for age, sex, and genetic background) were habituated to the three-chamber apparatus (length: 54 cm; width: 40 cm; height: 23 cm) divided into three chambers by removable partitions, for 1 h per day on two consecutive days prior to testing, while placed inside wire cups. On the test day, following habituation, experimental mice (12-15 weeks of age) were placed in the central chamber

571 and allowed to freely explore for 10 min. A novel partner mouse was then placed inside a wire cup in either the
572 left or right chamber, while a novel object was placed inside a wire cup in the opposite chamber. Placement of the
573 partner mouse was randomized across trials. Experimental mice were allowed to explore for 10 min, and
574 interaction time with each wire cup was manually scored.

575

576 **RNA extraction**

577 Cortical and cerebellar tissues from 12-week-old mice, and lung tissues from P6 mice, were dissected, flash-
578 frozen in liquid nitrogen, and stored at −80 °C until further processing. Total RNA was isolated using the
579 miRNeasy Mini Kit (Qiagen, 217004) according to the manufacturer's instructions. RNA concentration and purity
580 were measured using a NanoDrop 1000 (Thermo Fisher).

581

582 **RNA sequencing and analysis**

583 RNA samples were submitted to Azenta Life Sciences for RNA integrity assessment, library preparation, and
584 sequencing. Libraries were generated following polyA selection for mRNA and sequenced on Illumina NovaSeq
585 or HiSeq platforms, yielding ~ 30 million 150-bp paired-end reads per sample.

586 For each sample, the raw reads from multiple lanes were merged, and then were aligned to the Mus musculus
587 reference transcriptome (GRCm38.p6, GENCODE M18) using STAR v2.7 with "-quantMode GeneCounts" to
588 generate raw read counts [58].

589 Differential gene expression analysis was performed using DESeq2 v1.44.0 [59]. Principal component analysis
590 (PCA) was performed on the top 500 most variable genes. Genes were considered differentially expressed if
591 adjusted $p < 0.05$ and |log2 fold change| > 0.25.

592 Previously published lung P6 microarray data of *Atxn1*-KO and *Atxn1l*-KO were obtained from *Lee et al., 2011*.
593 Differentially expressed genes (DEGs) reported in that dataset were used for comparison; the original publication
594 defined DEGs using thresholds of $p < 0.05$ and |fold change| > 1.2. For *Atxn1*-KO RNA-seq data from cortex and
595 cerebellum (*Xhako et al., unpublished data*), read count data were processed with the same DESeq2 pipeline, and
596 limma v3.60.6 [60] was used for batch correction prior to PCA.

597 Gene set enrichment analysis (GSEA) for Gene Ontology (GO) terms was conducted on all ranked genes using
598 fgsea v1.30.0, with neuro/glia-related GO terms selected based on keywords including "neuron," "neuro,"
599 "nervous," "synapse," "axon," "glia," "glial," "astrocyte," "oligodendrocyte," "microglia," "myelin," and "brain."
600 [61] Overlapping differentially expressed genes were analyzed for GO term enrichment using clusterProfiler

601 v4.12.6 [62], with the background universe defined as all genes detected in the RNA-seq dataset for the
602 corresponding tissue.

603

604 **Protein extraction for size exclusion chromatography**

605 Cortical or cerebellar tissues were collected from two 2–3-month-old mice per genotype and pooled per biological
606 replicate. Tissues were homogenized in TX buffer (50 mM Tris, pH 8.0; 75 mM NaCl; 0.5% Triton X-100;
607 Sigma) supplemented with freshly added protease inhibitor (GenDEPOT, P3100-020) and phosphatase inhibitors
608 (GenDEPOT, P3200-020) using a Dounce homogenizer. Homogenates were extracted at 4 °C for 20 min with
609 rotation and clarified by centrifugation at 20,000 × g for 10 min. Supernatants were re-centrifuged, filtered
610 through 0.45 μm centrifugal filters (Millipore), and protein concentrations were determined by BCA assay
611 (Pierce).

612

613 **Size exclusion chromatography**

614 Gel filtration chromatography was performed on a Superose 6 10/300 GL column (Cytiva) using an AKTA Pure
615 system at 4 °C. The column was equilibrated in 50 mM Tris (pH 8.0), 50 mM NaCl, 0.1% Triton X-100. Column
616 calibration was performed with thyroglobulin (667 kDa), apoferritin (440 kDa), alcohol dehydrogenase (150
617 kDa), BSA (66 kDa), and cytochrome C (14.3 kDa). Fractions (1 mL) were collected, supplemented with protease
618 inhibitor (GenDEPOT, P3100-020) and phosphatase inhibitors (GenDEPOT, P3200-020), and analyzed by SDS-
619 PAGE on NuPAGE 3–8% Tris-Acetate gels (Invitrogen).

620

621 **Cloning**

622 The hCIC-S cDNA was obtained from a previously described plasmid [16]. The unique N-terminal region of
623 hCIC-L was synthesized (GenScript) and cloned into a PU57 backbone. Because direct amplification of the ~4.8
624 kb CIC-Shared region was hindered by high GC content at the 3′ end, a two-step cloning strategy was employed:
625 the 5′ and 3′ CIC fragments were first assembled into pENTR223 containing a synthetic multiple cloning site with
626 an AvrII site, followed by insertion of the ~4 kb central CIC fragment by PCR to generate full-length CIC-Shared.

627 A W37A point mutation (W37A in CIC-S, corresponding to W946A in CIC-L) was introduced into hCIC-Shared
628 in pENTR223 using gene blocks and Gibson assembly (New England Biolabs (NEB), E2621L) after XhoI (5′
629 end) and StuI (N-terminal) digestion. CIC-L and CIC-S entry clones were generated by Gibson assembly of the
630 respective unique 5′ regions into the CIC-Shared backbone, followed by addition of C-terminal V5 tags. The

631 hCIC-L–unique sequence was cloned into pENTR223 with an N-terminal V5 tag and a stop codon after amino
632 acid 931. All entry clones were recombined into pcDNA3.1-DEST (derived from pcDNA3.1-3×FLAG-DEST,
633 with stop codons preventing FLAG expression) using Gateway LR recombination (Thermo Fisher).

634 The hATXN1L cDNA (Horizon Discovery) was cloned with an added N-terminal FLAG tag into pLV-EF1α-
635 IRES-Puro [63] using PCR amplification and Gibson assembly. The FLAG-hATXN1 coding sequence was
636 obtained from the previously described pcDNA1-ATXN1 plasmid [18] and cloned into pcDNA3.1 by Gibson
637 assembly.

638

639 **Transfection and co-immunoprecipitation in HEK293T cells**

640 HEK293T cells were transfected with the indicated plasmids using TransIT-293 transfection reagent (Mirus Bio)
641 according to the manufacturer's instructions and incubated for 48–72 h. Cells were washed with PBS and either
642 processed immediately or stored at –80 °C. Frozen cells were thawed on ice and lysed in cold lysis buffer (75 mM
643 NaCl, 5 mM $MgCl_2$, 50 mM Tris-HCl (pH 8.0), 0.5% Triton X-100, with protease and phosphatase inhibitors) for
644 30–60 min at 4 °C with rotation. Lysates were cleared by centrifugation at 20,000 × g for 20 min at 4 °C.

645 For co-immunoprecipitation, lysates from one well of a 6-well plate were incubated with 1 µg V5 antibody
646 (Invitrogen) pre-bound to 15 µL Protein G Dynabeads (Fisher Scientific, Cat. #10-009-D) for 2–4 h at 4 °C with
647 rotation. Beads were washed four times with lysis buffer, and bound proteins were eluted in 1× NuPAGE Sample
648 Reducing Agent and NuPAGE LDS Sample Buffer (Invitrogen), boiled for 10 min, and analyzed by SDS-PAGE
649 on NuPAGE 3–8% Tris-Acetate gels (Invitrogen).

650

651 **Immunoprecipitation in mouse brain tissue**

652 Cortices from 12-week-old WT mice were homogenized in either cold NEMT buffer (50 mM Tris-HCl pH 7.5,
653 150 mM NaCl, 1 mM EDTA, 0.5% NP-40) or lysis buffer (75 mM NaCl, 5 mM $MgCl_2$, 50 mM Tris-HCl pH 8.0,
654 0.5% Triton X-100), each supplemented with freshly added protease inhibitor (GenDEPOT, P3100-020) and
655 phosphatase inhibitors (GenDEPOT, P3200-020). Tissue homogenization was performed using an electric pestle
656 (handheld polytron, WPR, 47747-370), followed by mechanical shearing through a 27G syringe. Lysates were
657 incubated on ice for 20 min and clarified by centrifugation at 20,000 × g for 20 min at 4 °C. Ten percent of the
658 supernatant was reserved as input, and the remaining lysate was precleared with Dynabeads Protein G (Fisher
659 Scientific, 10-009-D) for 30 min at 4 °C. Precleared lysates were incubated for 2–3 h at 4 °C with 1–2 µg of
660 antibody—anti-ATXN1 (534, in-house [64]), anti-ATXN1L (881, in-house), or mouse IgG control—conjugated to
661 30 µL Dynabeads Protein G. Beads were washed four times with the corresponding lysis buffer and final wash

662 with PBS. Bound proteins were eluted in 1× NuPAGE Sample Reducing Agent and NuPAGE LDS Sample Buffer
663 (Invitrogen), boiled for 10 min, and analyzed by SDS-PAGE on NuPAGE 3–8% Tris-Acetate gels (Invitrogen).

664 For *Atxn1*-KO experiments, cortex, hippocampus, cerebellum, and brainstem were dissected from 4-week-old WT
665 and *Atxn1*-KO mice and processed using the ATXN1 immunoprecipitation protocol described above.

666

667 **Protein extraction for immunoblotting**

668 Brain and lung tissues were collected at postnatal day zero (P0), P6, P21, and 12 weeks and snap-frozen in liquid
669 nitrogen and kept in -80°C until processing. For protein extraction, frozen tissues were homogenized in cold lysis
670 buffer (75 mM NaCl, 5 mM $MgCl_2$, 50 mM Tris pH 8.0, 0.5% Triton X-100, supplemented with freshly added
671 protease inhibitor (GenDEPOT, P3100-020) and phosphatase inhibitors (GenDEPOT, P3200-020) using an
672 electric pestle (handheld polytron, WPR, 47747-370), followed by mechanical shearing through a 27G syringe.
673 Lysates were incubated on ice for 20 min and centrifuged at 20,000 × g for 20 min at 4 °C. Protein concentrations
674 in the supernatant were measured using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific), and
675 samples were normalized to equal concentrations.

676

677 **Western blot analysis**

678 For immunoblotting, samples were mixed with NuPAGE Sample Reducing Agent and NuPAGE LDS Sample
679 Buffer (Invitrogen), boiled for 10 min, and run on NuPAGE 3–8% Tris-Acetate gels (Invitrogen). Proteins were
680 transferred to nitrocellulose membranes (Amersham, #10600004) using transfer buffer containing 10% methanol.
681 Membranes were blocked with either 5% skim milk in Tris-buffered saline with 0.1% Tween® 20 (TBST) or
682 Intercept (TBS) Blocking Buffer (LI-COR Biosciences) for 1 h at room temperature and incubated overnight at
683 4 °C with primary antibodies diluted in 3% BSA in TBST. The following primary antibodies were used: anti-CIC
684 (rabbit, in-house, 1:1,000), anti-ATXN1[5] (rabbit, in-house, 11750VII, 1:2,000), anti-ATXN1L (mouse, 674, in-
685 house, 1:1,000), anti-Vinculin (mouse, Sigma-Aldrich, 1:5,000), anti-Flag (Sigma-Aldrich, 1:4000), and anti-V5
686 (Invitrogen, 1:2,000).

687 After washing in TBST, membranes were incubated for 1 h at room temperature with either HRP-conjugated
688 secondary antibodies (Jackson ImmunoResearch, 1:10,000 in 5% milk/TBST) or fluorophore-conjugated
689 secondary antibodies (LI-COR Biosciences, 1:10,000 in Intercept Blocking Buffer). HRP blots were developed
690 using ECL (Cytiva, RPN2236) or SuperSignal Femto (Thermo Fisher Scientific, 34095) and imaged with a GE
691 Amersham Imager 680 or ChemiDoc MP Imaging System or LI-COR Odyssey M system. Fluorescent blots were
692 imaged directly using the LI-COR Odyssey M system.

693

694 **Statistical analysis**

695 All statistical analyses were performed using GraphPad Prism 9.0, except for RNA-seq analyses. Survival of
696 animals was compared to the expected Mendelian ratio using a chi-square test. Comparisons between two groups
697 were performed using unpaired t-tests. Comparisons involving more than two groups were analyzed using one-
698 way or two-way ANOVA, followed by Tukey's multiple comparisons test. Repeated-measures data were analyzed
699 using two-way repeated-measures ANOVA with Tukey's multiple comparisons test. Survival curves were
700 analyzed using the Mantel–Cox test. All reported p values are from post hoc comparisons. Data are presented as
701 mean ± SEM. Statistical significance was defined as $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), and $p < 0.0001$
702 (****).

703 RNA-seq analyses were performed in R using DESeq2, as described above. Fisher's exact test was used to
704 calculate odds ratios and p-values for gene set overlaps, and pairwise comparisons between genotypes for box
705 plots were performed using two-tailed t-tests with Benjamini–Hochberg correction for multiple comparisons.

706 Sample sizes were based on prior studies and standard practice in the field. Biological replicates refer to
707 independent animals. Experimental analyses were performed in a blinded manner whenever possible.

708

SUPPLEMENTARY FIGURES

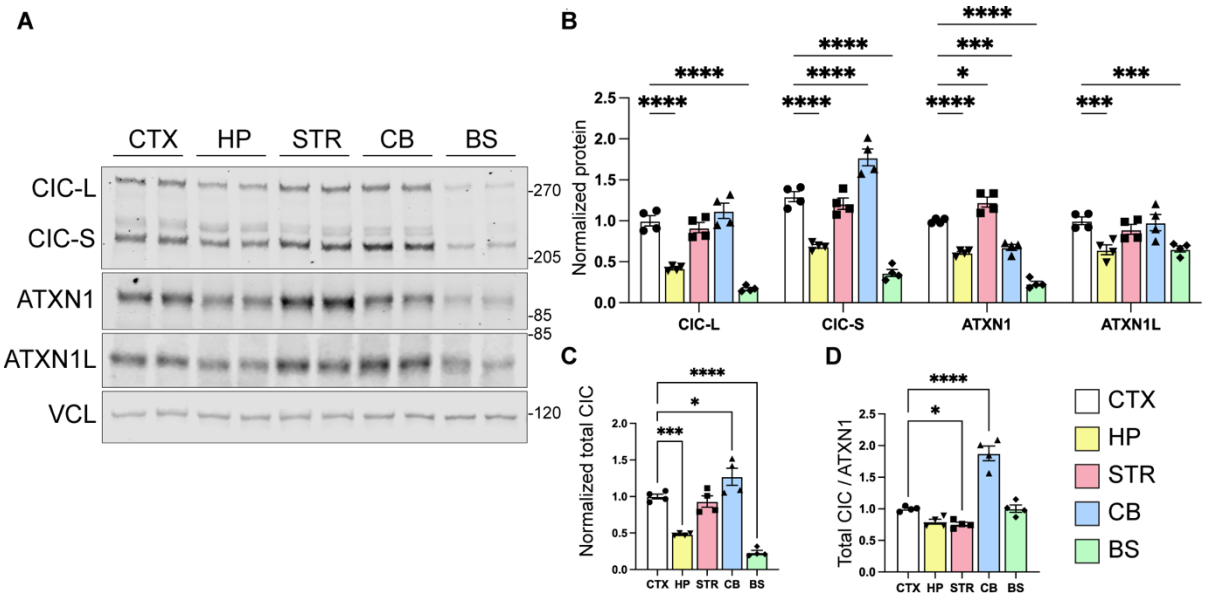


Figure S1. CIC-L, CIC-S, ATXN1, and ATXN1L expression in different brain regions.

(A) Representative immunoblot of CIC-L, CIC-S, ATXN1, and ATXN1L in 12-week-old WT mouse brain region lysates at the indicated brain regions (CTX: cortex, HP: hippocampus, STR: striatum, CB: cerebellum, BS: brainstem).

(B and C) Quantification of each protein's expression. (C) Total CIC combining CIC-L and CIC-S. Protein levels were normalized to vinculin (VCL), and values at each region were normalized to the cortex value for the respective protein.

(D) Total CIC was normalized to ATXN1 for each region. Mean of each region compared to the normalized cortical expression of each protein. One-way ANOVA. In each case, *, **, ***, and **** denote $p < 0.05$, $p < 0.01$, $p < 0.001$, and $p < 0.0001$ respectively. All data are represented as mean ± SEM from n = 4 per region.

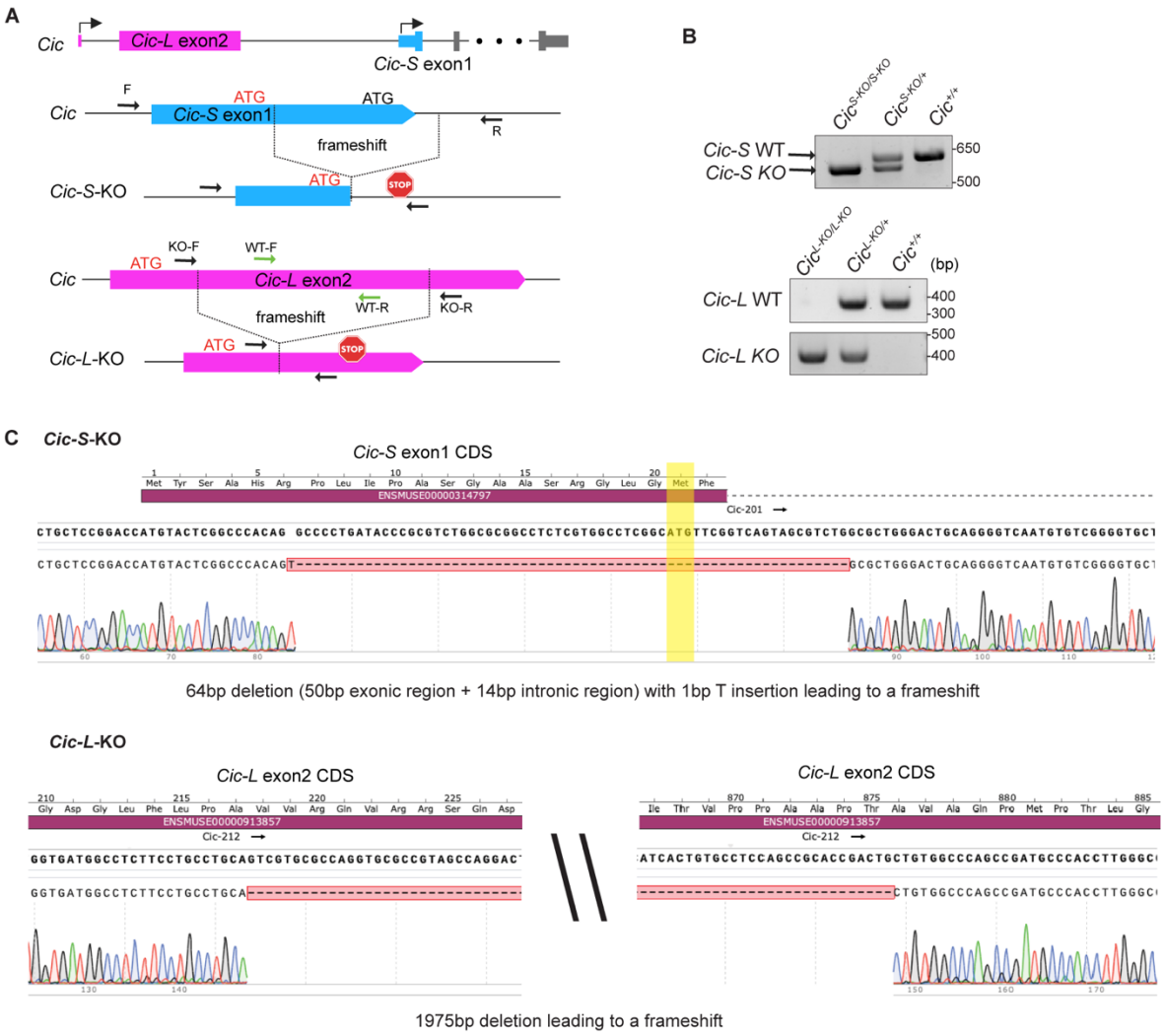


Figure S2. *Cic-S*-KO and *Cic-L*-KO were generated by frameshift indels.

(A) Schematic of *Cic-L*-KO and *Cic-S*-KO allele generation. *Cic-L*-KO was generated by a 1975bp deletion of the first coding exon of *Cic-L* (*Cic-L* exon2). The deletion is downstream of the start codon and leads to a frameshift of *Cic-L*. *Cic-S*-KO was generated by an indel that spans the *Cic-S* unique exon1 and the downstream intron. The deletion is downstream of the start codon and spans 50bp of the exon including the downstream start codon and 14bp of the intron with a one bp insertion. Overall, this leads to a frameshift of *Cic-S*. The start codon is marked with red and the downstream ATG is marked with black. The PCR primers for genotyping are marked as arrows.

(B) Representative gel image of genotyping PCR reaction. *Cic-L* KO allele (383bp), *Cic-L* KO allele (403bp), *Cic-S* WT allele (633bp), and *Cic-S* KO allele (570bp).

(C) Sanger sequencing confirming a frameshift indel of *Cic-S*-KO (top), and *Cic-L*-KO (bottom) genomic DNA.

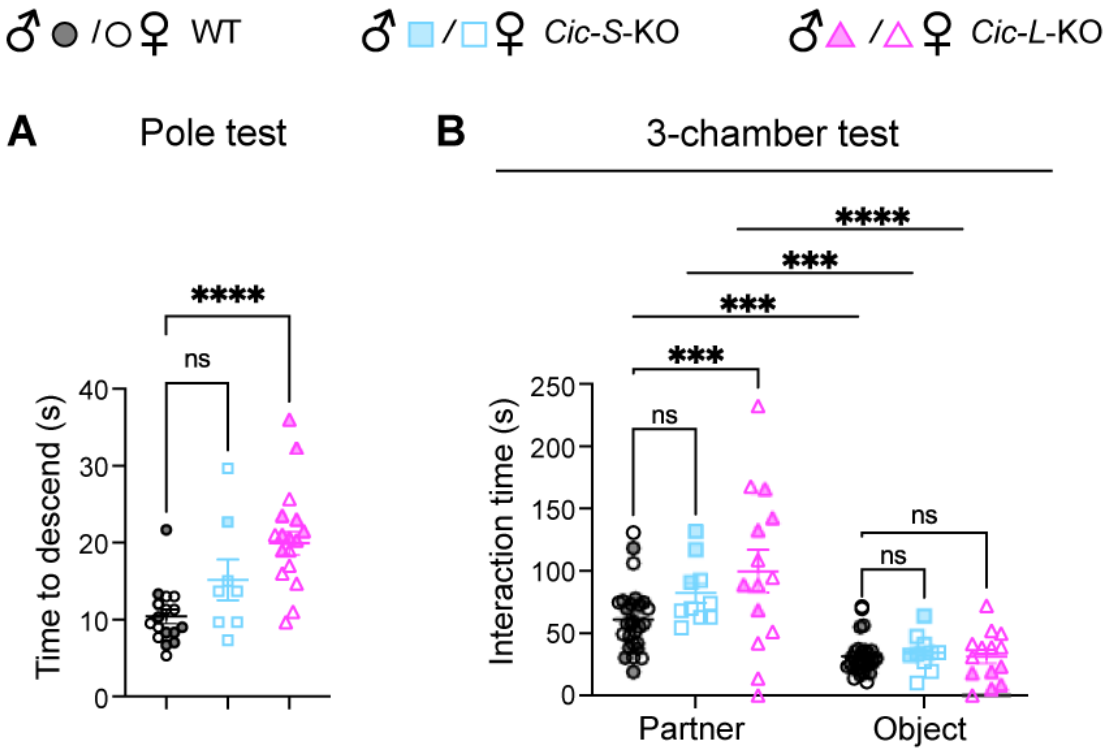


Figure S3. *Cic-L*-KO shows motor deficits in the pole test and more time spent with partner during a social interaction assay.

(A) Fine motor behavior measured with pole test at 15-17 weeks of age.
(B) Social behavior measured with the three-chamber assay at 12-15 weeks of age.
For (A) n = 17 (8 males, 9 females) WT, n = 8 *Cic-S*-KO (1 males, 7 females), and n =16 *Cic-L*-KO (9 males, 7 females) were analyzed with one-way ANOVA with Tukey's multiple comparisons. For (B) n = 27 (12 males, 15 females) WT, n = 10 *Cic-S*-KO (3 males, 7 females), and n =14 *Cic-L*-KO (5 males, 9 females) were analyzed with two-way ANOVA. In each case, ***, **** and ns denote $p < 0.001$, $p < 0.0001$, and $p > 0.05$, respectively. All data are represented as mean ± SEM.

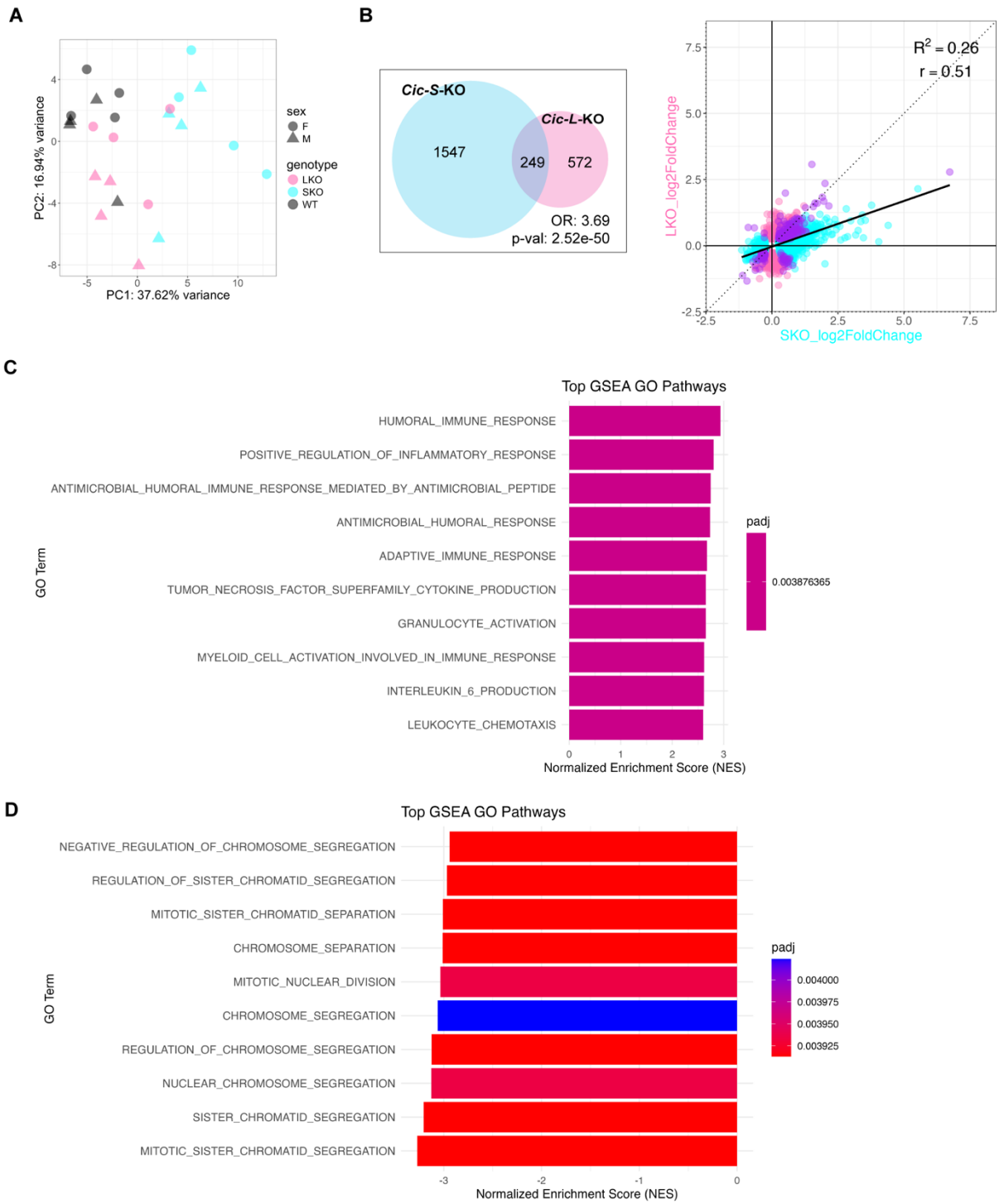


Figure S4. *Cic-S*-KO shows greater number and magnitude of DEGs in the lungs compared to *Cic-L*-KO DEGs.

(A) Principal component analysis (PCA) of RNA-seq data from P6 lungs (WT, *Cic-S*-KO [SKO], *Cic-L*-KO [LKO]) using the top 500 most variable genes.

(B) Left: Venn diagram showing overlap of differentially expressed genes (DEGs; padj < 0.05 & |log2 fold change (log2FC)| >0.25) between *Cic-S*-KO and *Cic-L*-KO. Odds ratio (OR) and Fisher's exact *p*-value

are indicated. A total of 15473 genes were analyzed. Right: Scatterplot comparing log2FC of DEGs *Cic-S*-KO versus *Cic-L*-KO. Pearson correlation coefficient (r) and R^2 are shown. Overlapping DEGs marked in purple.

(C and D) Top gene ontology (GO) enrichment terms for biological processes of *Cic-S*-KO (C) and *Cic-L*-KO (D).

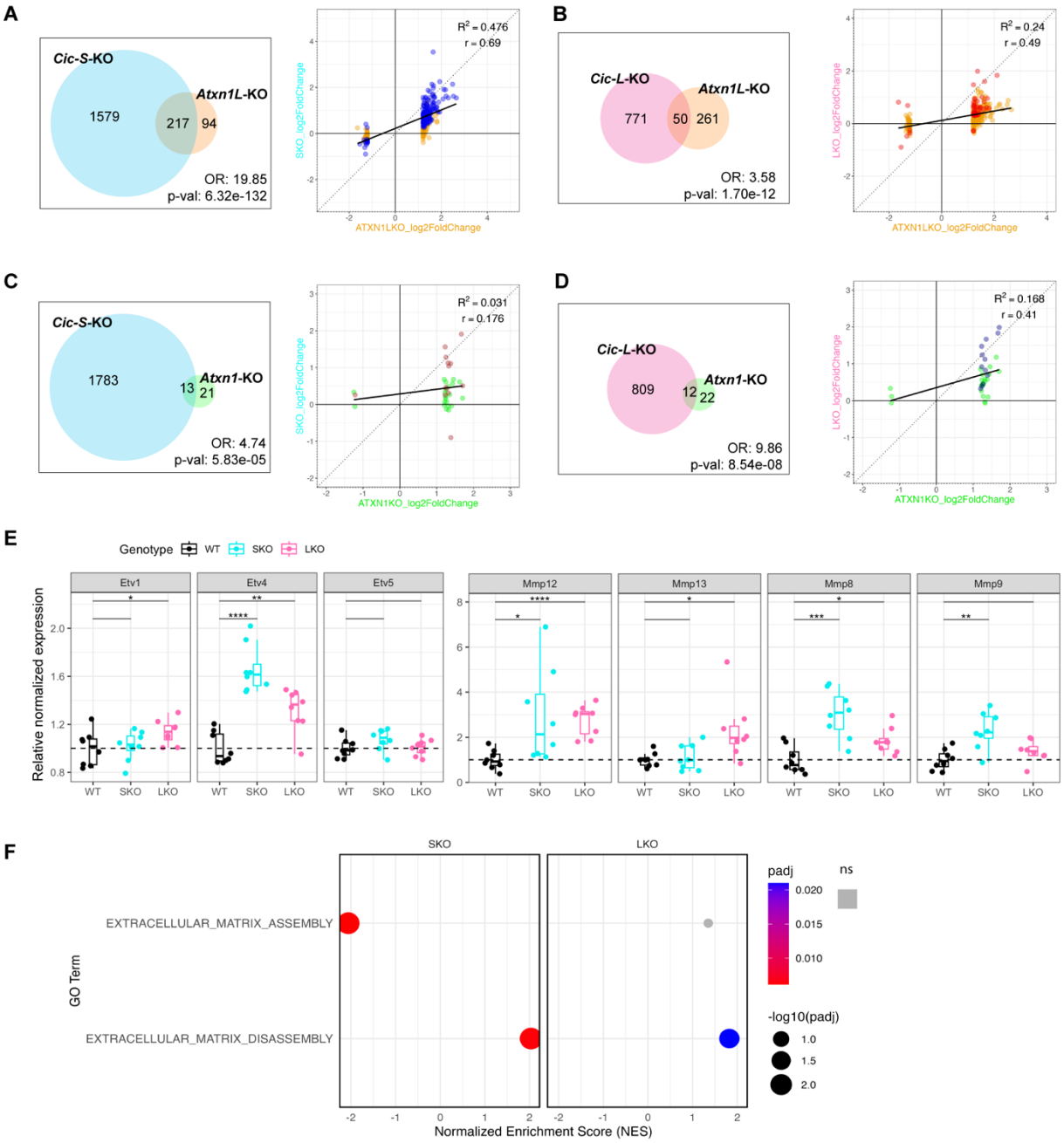


Figure S5. *Cic-S*-KO DEGs show greater overlap with *Atxn1l*-KO DEGs than *Cic-L*-KO DEGs in the lungs.

(A and B) Left: Venn diagrams showing overlap between DEGs in *Cic-S*-KO [SKO] (A) or *Cic-L*-KO [LKO] (B) and DEGs from *Atxn1l*-KO lungs reported in *Lee et al., 2011*. Odds ratio (OR) and Fisher's exact test p-value are indicated. A total of 15473 genes were analyzed. Right: Scatterplot comparing log2FC of DEGs of *Cic-S*-KO (A) or *Cic-L*-KO (B) versus *Atxn1l*-KO. Pearson correlation coefficient (r) and R^2 are shown. Overlapping DEGs marked in blue (A) or red (B).

(C and D) Left: Venn diagrams showing overlap between DEGs in *Cic-S*-KO (A) or *Cic-L*-KO (B) and DEGs from *Atxn1*-KO lungs reported in *Lee et al., 2011*. OR and Fisher's exact *p*-value are indicated. A total of

771 15473 genes were analyzed. Right: Scatterplot comparing log2FC of DEGs of *Cic-S*-KO (C) or *Cic-L*-KO
772 (D) versus *Atxn1*-KO. Pearson correlation coefficient (r) and R^2 are shown. Overlapping DEGs marked in
773 brown (C) or navy (D).

774 (E) Boxplot of *Etv-Mmp* genes. Statistical analysis was performed using pairwise two-tailed t-tests with
775 Benjamini–Hochberg correction for multiple comparisons. In each case, *, **, ***, and **** denote $p <$
776 0.05, $p < 0.01$, $p < 0.001$, and $p < 0.0001$, respectively. Non-significant comparisons ($p > 0.05$) are not
777 shown.

778 (F) Enrichment of GO of extracellular matrix assembly and disassembly terms of *Cic-S*-KO (Left) or *Cic-L*-
779 KO (Right) gene set enrichment analysis (GSEA).

780

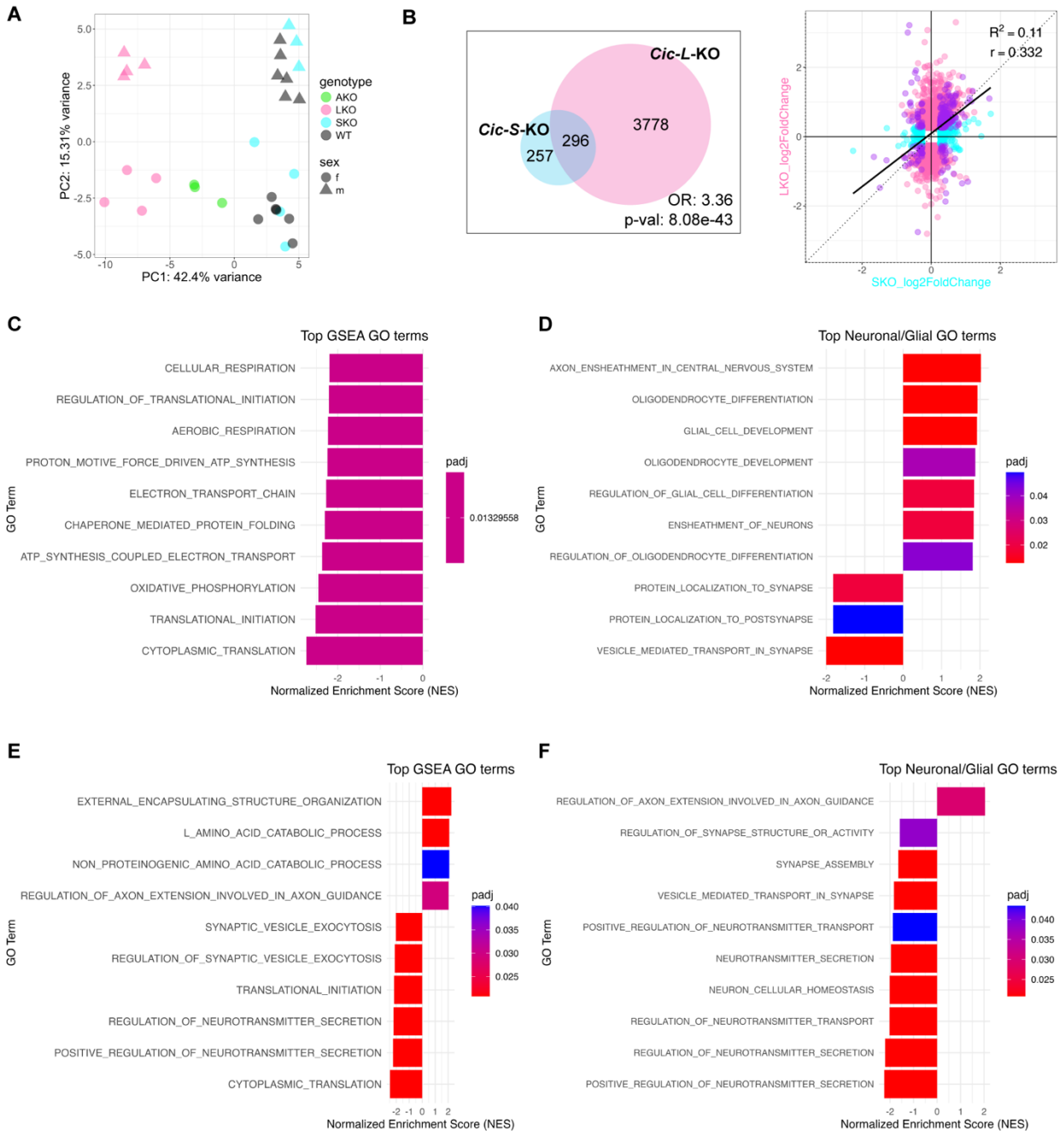


Figure S6. *Cic-L*-KO shows greater number and magnitude of DEGs than *Cic-S*-KO DEGs in the cortex.

(A) Principal component analysis (PCA) of RNA-seq data from 12-week-old cortex of WT, *Cic-S*-KO [SKO], *Cic-L*-KO [LKO] of this study and 10-week-old cortex from *Xhako et al., in preparation*. *Atxn1*-KO [AKO] and WT using the top 500 most variable genes.

(B) Left: Venn diagram showing overlap of DEGs between *Cic-S*-KO and *Cic-L*-KO. OR and Fisher's exact *p*-value are indicated. A total of 15354 genes were analyzed. Right: Scatterplot comparing log2FC of DEGs *Cic-S*-KO versus *Cic-L*-KO. Pearson correlation coefficient (r) and R^2 are shown. Overlapping DEGs marked in purple.

790 (C and D) Top gene ontology (GO) terms of all biological processes (C) and neuro- or glia- related terms (D)
791 of *Cic-S*-KO gene set enrichment analysis (GSEA).
792 (E and F) Top enriched GO terms of all biological process (E) and neuro- or glia- related -terms (F) of *Cic-L*-
793 KO GSEA.

794

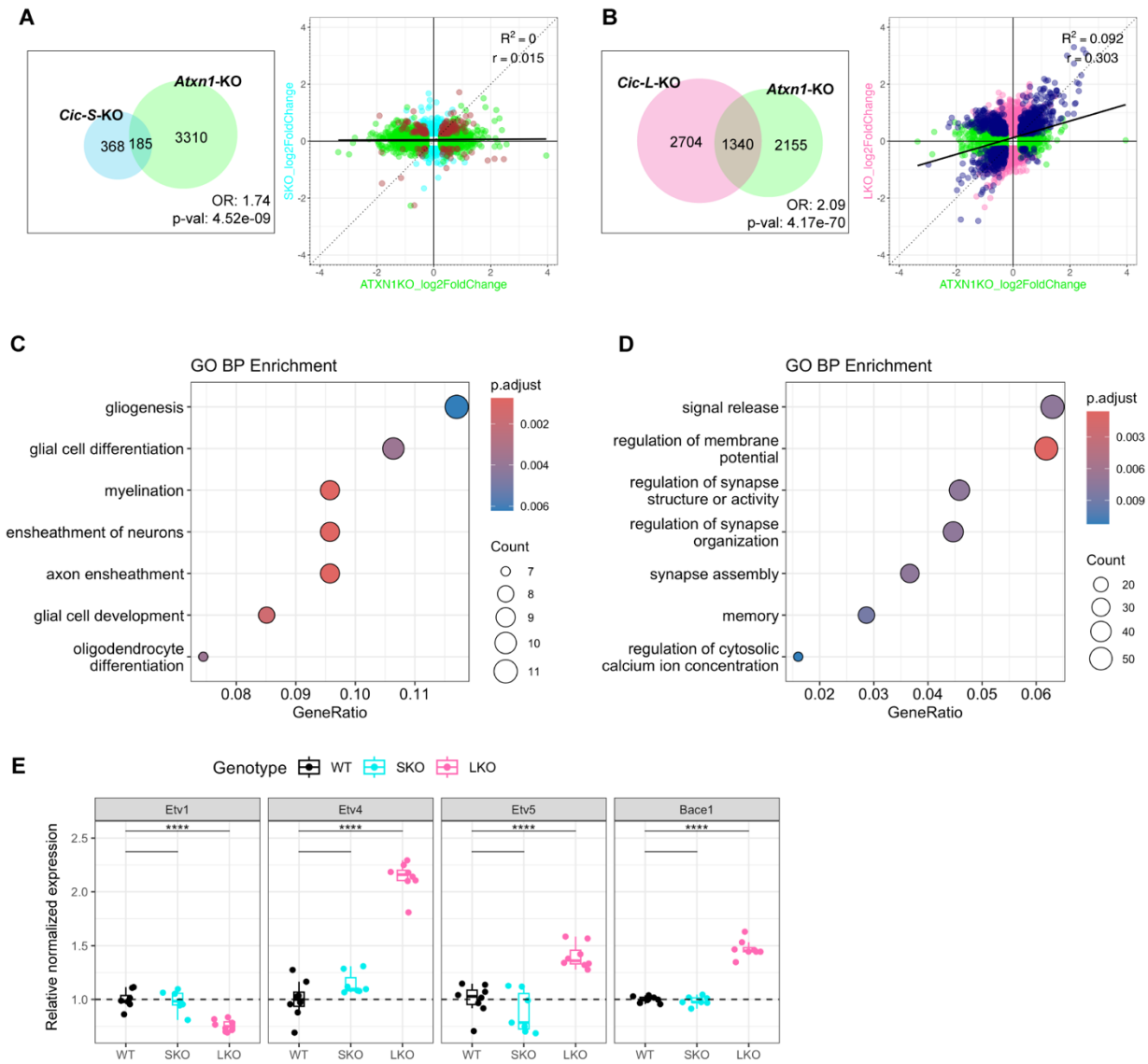
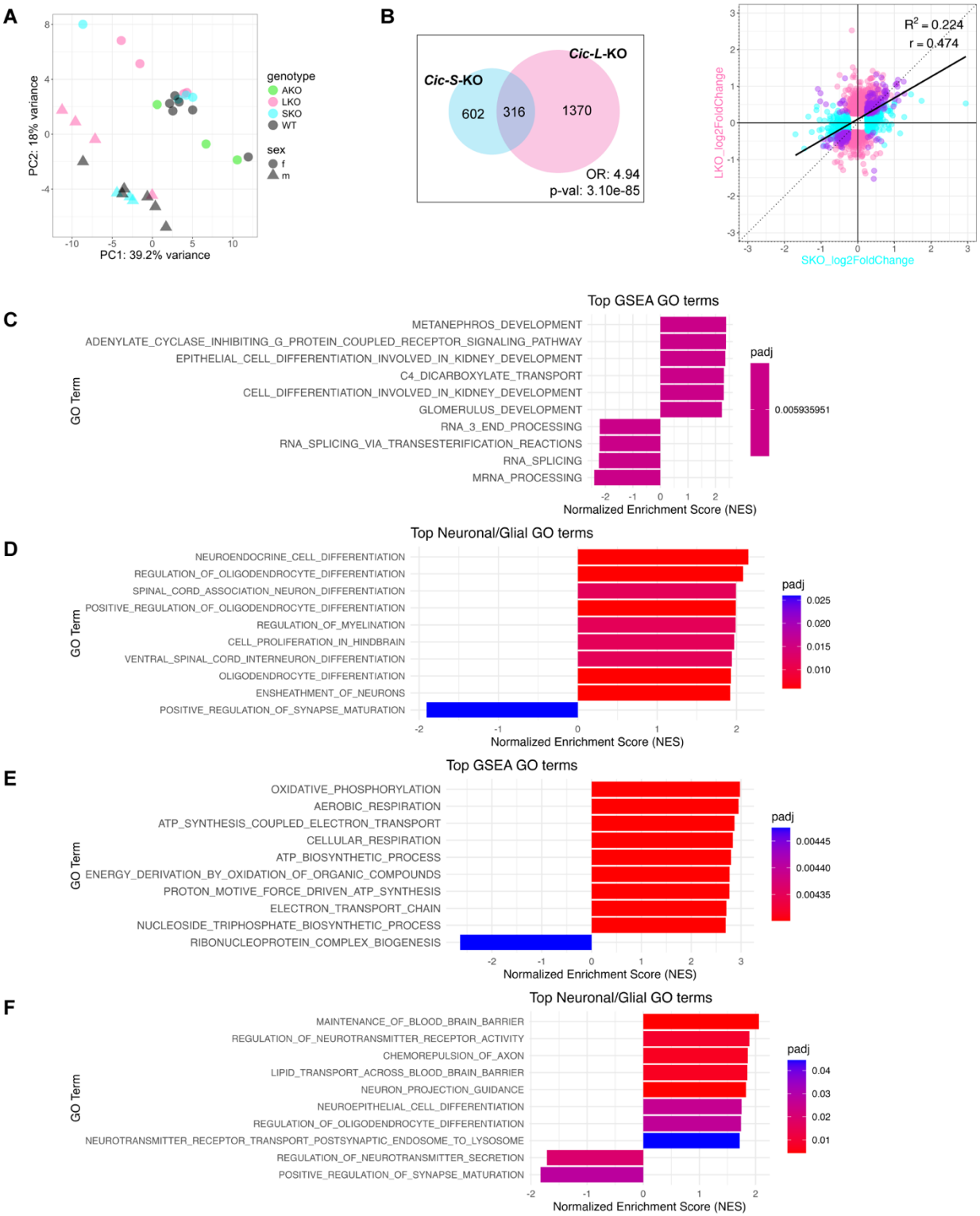


Figure S7. *Cic-L*-KO DEGs show greater overlap with *Atxn1*-KO DEGs than *Cic-S*-KO in the cortex.

(A and B) Left: Venn diagrams showing overlap between DEGs in *Cic-S*-KO [SKO] (A) or *Cic-L*-KO [LKO] (B) and DEGs (padj < 0.05 & |log2FC| >0.25) from *Atxn1*-KO cortex from *Xhako et al., in preparation*. OR and Fisher's exact *p*-value are indicated. A total of 15290 genes were analyzed. Right: Scatterplot comparing log2FC of DEGs of *Cic-S*-KO (A) or *Cic-L*-KO (B) versus *Atxn1*-KO. Pearson correlation coefficient (r) and R^2 are shown. Overlapping DEGs marked in brown (A) or navy (B).

(C and D) Top enriched gene ontology terms of biological process of the DEGs that overlap and change in the same direction between *Atxn1*-KO and *Cic-S*-KO (C), or *Atxn1*-KO and *Cic-L*-KO (D).

(E) Boxplot of *Etv1/4/5-Bace1*. Two-tailed t-tests with Benjamini–Hochberg correction for multiple comparisons. In each case, **** denotes $p < 0.0001$. Non-significant comparisons ($p > 0.05$) are not shown.



807

808 **Figure S8. *Cic-L*-KO shows greater number and magnitude of DEGs in the cerebellum than *Cic-S*-KO.**

809 (A) Principal component analysis (PCA) of RNA-seq data from 12-week-old cerebellum of WT, *Cic-S*-KO
810 [SKO], *Cic-L*-KO [LKO] of this study and 10-week-old cerebellum from *Xhako et al., in preparation*.
811 *Atxn1*-KO [AKO] and WT using the top 500 most variable genes.
812 (B) Left: Venn diagram showing overlap of DEGs between *Cic-S*-KO and *Cic-L*-KO. OR and Fisher's exact
813 *p*-value are indicated. A total of 15354 genes were analyzed. Right: Scatterplot comparing log2FC of
814 DEGs *Cic-S*-KO versus *Cic-L*-KO. Pearson correlation coefficient (r) and R^2 are shown. Overlapping
815 DEGs marked in purple.
816 (C and D) Top gene ontology (GO) terms of all biological process (C) and neuro- or glia- related terms (D) of
817 *Cic-S*-KO gene set enrichment analysis (GSEA).
818 (E and F) Top enriched GO terms of all biological process (E) and neuro- or glia- related -terms (F) of *Cic-L*-
819 KO GSEA.

820

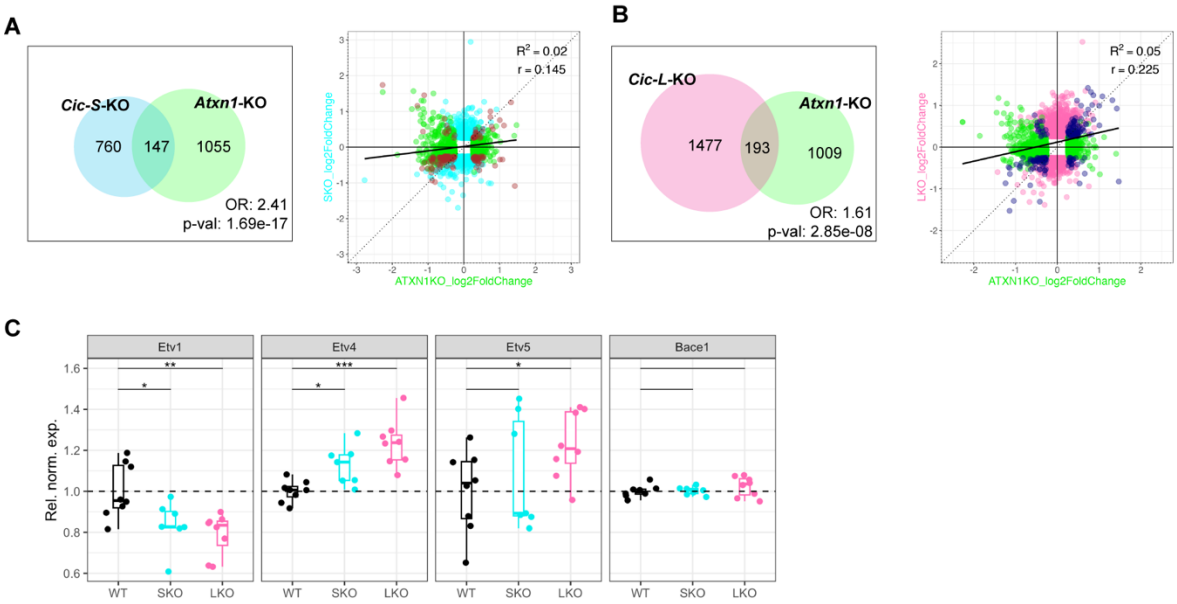
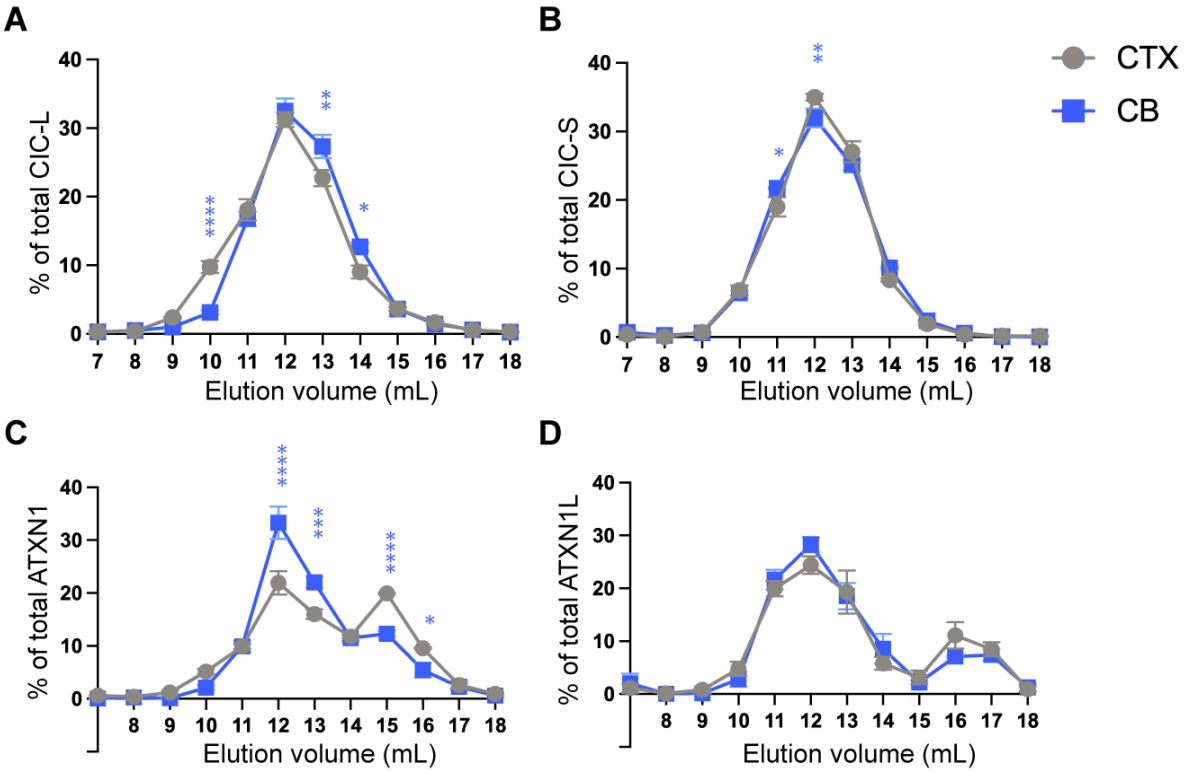


Figure S9. *Cic-S*-KO DEGs show greater overlap in the cerebellum with *Atxn1l*-KO than *Cic-L*-KO DEGs.

(A and B) Left: Venn diagrams showing overlap between DEGs in *Cic-S*-KO [SKO] (A) or *Cic-L*-KO [LKO] (B) and DEGs (padj < 0.05 & |log2FC| >0.25) from *Atxn1*-KO cerebellum from *Xhako et al., in preparation*. OR and Fisher's exact *p*-value are indicated. A total of 15290 genes were analyzed. Right: Scatterplot comparing log2FC of DEGs of *Cic-S*-KO (A) or *Cic-L*-KO (B) versus *Atxn1*-KO. Pearson correlation coefficient (r) and R^2 are shown. Overlapping DEGs marked in brown (A) or navy (B).

(C) Boxplot of *Etv1/4/5-Bace1*. Two-tailed t-tests with Benjamini–Hochberg correction for multiple comparisons. In each case, *, **, and ***, denote $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. Non-significant comparisons ($p > 0.05$) are not shown.



832

833 **Figure S10. CIC isoforms form different complexes with ATXN1 and ATXN1L in the WT cortex and**
834 **cerebellum (related to Figure 5).**

835 (A-D) SEC elution profile of each protein of WT cortex and cerebellum. The percentage of each protein
836 (mean ± SEM) in each fraction was determined from n = 3–4 independent cortical extracts. Protein
837 abundance in each fraction was quantified by densitometry, with the signal in each fraction normalized to
838 the total signal across all fractions for both regions. Statistical significance was assessed by two-way
839 ANOVA. In each case, *, **, ***, and **** denote $p < 0.05$, $p < 0.01$, $p < 0.001$, $p < 0.0001$ respectively.
840 All data are represented as mean ± SEM.

841

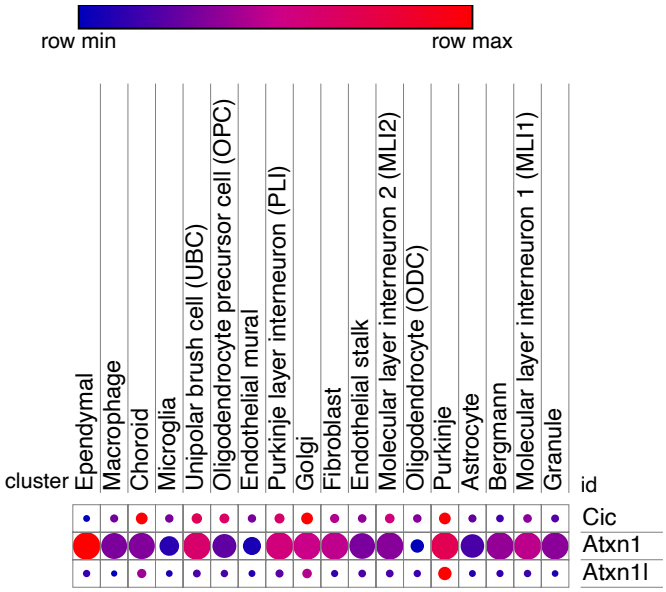
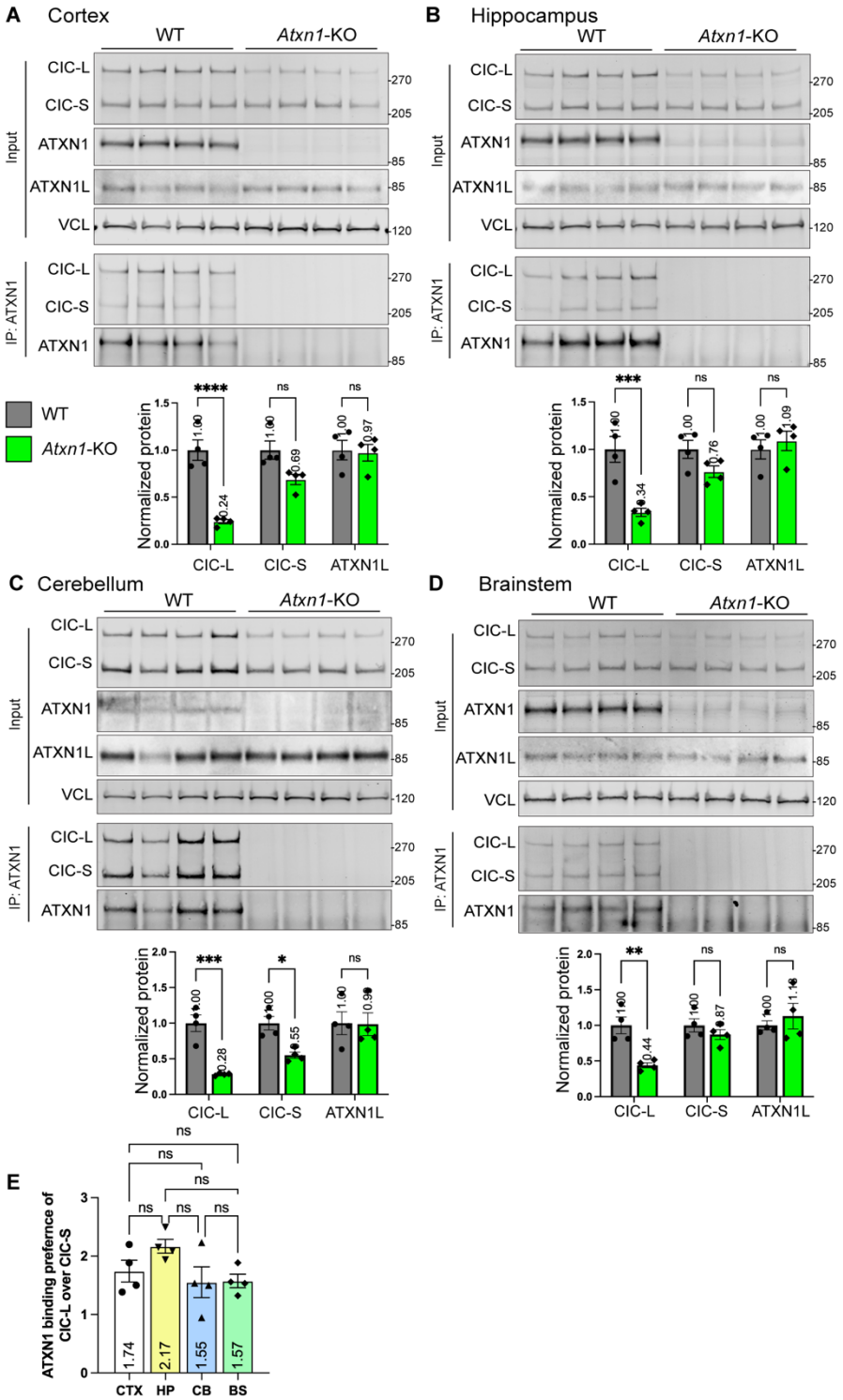


Figure S11. ***Cic, Atxn1,*** **and** ***Atxn1l*** **expression in mouse cerebellum at P60.**



845

846 **Figure S12. CIC-L is more dependent on ATXN1 than CIC-S in all brain regions**

847 (A-D) Representative western blot images (top) and quantification of input (bottom) of ATXN1, ATXN1L and
848 CIC expression in WT mice and Atxn1-KO mice of four different brain regions. Protein levels were
849 normalized to VCL, then relative levels were normalized to WT expression.

850 (E) Binding preference of CIC-L/CIC-S in ATXN1 IP normalized to the ratio in input.

851 CTX, cortex; HP, hippocampus; CB, cerebellum; BS, brainstem; Data=mean ± SEM; n = 4 per region. One-
852 way ANOVA; *p<0.05; **p<0.01; ***p<0.001; ***p<0.0001.

853

854 **REFERENCES**

855 1. Fu, H., et al., *Selective vulnerability in neurodegenerative diseases.* Nature Neuroscience 2018 21:10,
856 2018. **21**(10).
857 2. Kampmann, M., *Molecular and cellular mechanisms of selective vulnerability in neurodegenerative*
858 *diseases.* Nat Rev Neurosci, 2024. **25**(5): p. 351-371.
859 3. Orr, H.T., et al., *Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1.* Nat
860 Genet, 1993. **4**(3): p. 221-6.
861 4. Banfi, S., et al., *Identification and characterization of the gene causing type 1 spinocerebellar ataxia.* Nat
862 Genet, 1994. **7**(4): p. 513-20.
863 5. Servadio, A., et al., *Expression analysis of the ataxin-1 protein in tissues from normal and spinocerebellar*
864 *ataxia type 1 individuals.* Nat Genet, 1995. **10**(1): p. 94-8.
865 6. Zoghbi, H.Y. and H.T. Orr, *Spinocerebellar ataxia type 1.* Semin Cell Biol, 1995. **6**(1): p. 29-35.
866 7. Matilla, A., et al., *Mice lacking ataxin-1 display learning deficits and decreased hippocampal paired-*
867 *pulse facilitation.* J Neurosci, 1998. **18**(14): p. 5508-16.
868 8. Crespo-Barreto, J., et al., *Partial loss of ataxin-1 function contributes to transcriptional dysregulation in*
869 *spinocerebellar ataxia type 1 pathogenesis.* PLoS Genet, 2010. **6**(7): p. e1001021.
870 9. Asher, M., et al., *Cerebellar contribution to the cognitive alterations in SCA1: evidence from mouse*
871 *models.* Hum Mol Genet, 2020. **29**(1): p. 117-131.
872 10. Swaminathan, S., et al., *Genomic Copy Number Analysis in Alzheimer's Disease and Mild Cognitive*
873 *Impairment: An ADNI Study.* Int J Alzheimers Dis, 2011. **2011**: p. 729478.
874 11. Swaminathan, S., et al., *Analysis of copy number variation in Alzheimer's disease: the NIALOAD/*
875 *NCRAD Family Study.* Curr Alzheimer Res, 2012. **9**(7): p. 801-14.
876 12. Suh, J., et al., *Loss of Ataxin-1 Potentiates Alzheimer's Pathogenesis by Elevating Cerebral BACE1*
877 *Transcription.* Cell, 2019. **178**(5): p. 1159-1175 e17.
878 13. Lee, Y., et al., *ATXN1 protein family and CIC regulate extracellular matrix remodeling and lung*
879 *alveolarization.* Dev Cell, 2011. **21**(4): p. 746-57.
880 14. Lu, H.C., et al., *Disruption of the ATXN1-CIC complex causes a spectrum of neurobehavioral phenotypes*
881 *in mice and humans.* Nat Genet, 2017. **49**(4): p. 527-536.
882 15. Wong, D., et al., *TRIM25 promotes Capicua degradation independently of ERK in the absence of*
883 *ATXN1L.* BMC Biol, 2020. **18**(1): p. 154.
884 16. Lam, Y.C., et al., *ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1*
885 *neuropathology.* Cell, 2006. **127**(7): p. 1335-47.
886 17. Mizutani, A., et al., *Boat, an AXH domain protein, suppresses the cytotoxicity of mutant ataxin-1.* EMBO
887 J, 2005. **24**(18): p. 3339-51.
888 18. Kim, E., et al., *Structural basis of protein complex formation and reconfiguration by polyglutamine*
889 *disease protein Ataxin-1 and Capicua.* Genes Dev, 2013. **27**(6): p. 590-5.
890 19. Bowman, A.B., et al., *Duplication of Atxn1l suppresses SCA1 neuropathology by decreasing*
891 *incorporation of polyglutamine-expanded ataxin-1 into native complexes.* Nat Genet, 2007. **39**(3): p. 373-
892 9.
893 20. Kim, E., et al., *Deficiency of Capicua disrupts bile acid homeostasis.* Sci Rep, 2015. **5**: p. 8272.
894 21. Fryer, J.D., et al., *Exercise and genetic rescue of SCA1 via the transcriptional repressor Capicua.*
895 Science, 2011. **334**(6056): p. 690-3.
896 22. Coffin, S.L., et al., *Disruption of the ATXN1-CIC complex reveals the role of additional nuclear ATXN1*
897 *interactors in spinocerebellar ataxia type 1.* Neuron, 2023. **111**(4): p. 481-492 e8.
898 23. Rousseaux, M.W.C., et al., *ATXN1-CIC Complex Is the Primary Driver of Cerebellar Pathology in*
899 *Spinocerebellar Ataxia Type 1 through a Gain-of-Function Mechanism.* Neuron, 2018. **97**(6): p. 1235-
900 1243 e5.
901 24. Simon-Carrasco, L., et al., *Inactivation of Capicua in adult mice causes T-cell lymphoblastic lymphoma.*
902 Genes Dev, 2017. **31**(14): p. 1456-1468.
903 25. Lee, Y., *Regulation and function of capicua in mammals.* Exp Mol Med, 2020. **52**(4): p. 531-537.

26. Jimenez, G., S.Y. Shvartsman, and Z. Paroush, *The Capicua repressor--a general sensor of RTK signaling in development and disease.* J Cell Sci, 2012. **125**(Pt 6): p. 1383-91.
27. Hwang, I., et al., *CIC is a critical regulator of neuronal differentiation.* JCI Insight, 2020. **5**(9).
28. Ahmad, S.T., et al., *Capicua regulates neural stem cell proliferation and lineage specification through control of Ets factors.* Nat Commun, 2019. **10**(1): p. 2000.
29. Tan, Q. and H.Y. Zoghbi, *Mouse models as a tool for discovering new neurological diseases.* Neurobiol Learn Mem, 2019. **165**: p. 106902.
30. Subramanian, A., et al., *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences, 2005. **102**(43).
31. Dissanayake, K., et al., *ERK/p90(RSK)/14-3-3 signalling has an impact on expression of PEA3 Ets transcription factors via the transcriptional repressor capicua.* Biochem J, 2011. **433**(3): p. 515-25.
32. Weissmann, S., et al., *The Tumor Suppressor CIC Directly Regulates MAPK Pathway Genes via Histone Deacetylation.* Cancer Res, 2018. **78**(15): p. 4114-4125.
33. Kozareva, V., et al., *A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types.* Nature, 2021. **598**(7879): p. 214-219.
34. Yang, X., et al., *Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing.* Cell, 2016. **164**(4): p. 805-17.
35. Fores, M., et al., *Origins of context-dependent gene repression by capicua.* PLoS Genet, 2015. **11**(1): p. e1004902.
36. Rodriguez-Munoz, L., et al., *Noncanonical function of Capicua as a growth termination signal in Drosophila oogenesis.* Proc Natl Acad Sci U S A, 2022. **119**(31): p. e2123467119.
37. Sharma, S., et al., *Novel CIC variants identified in individuals with neurodevelopmental phenotypes.* Hum Mutat, 2022. **43**(7): p. 889-899.
38. Bertram, L., et al., *Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE.* Am J Hum Genet, 2008. **83**(5): p. 623-32.
39. Bettens, K., et al., *Follow-up study of susceptibility loci for Alzheimer's disease and onset age identified by genome-wide association.* J Alzheimers Dis, 2010. **19**(4): p. 1169-75.
40. Kjer-Hansen, P., T.G. Phan, and R.J. Weatheritt, *Protein isoform-centric therapeutics: expanding targets and increasing specificity.* Nat Rev Drug Discov, 2024. **23**(10): p. 759-779.
41. Friedrich, J., et al., *Antisense oligonucleotide-mediated ataxin-1 reduction prolongs survival in SCA1 mice and reveals disease-associated transcriptome profiles.* JCI Insight, 2018. **3**(21).
42. Jafar-Nejad, P., et al., *Regional rescue of spinocerebellar ataxia type 1 phenotypes by 14-3-3epsilon haploinsufficiency in mice underscores complex pathogenicity in neurodegeneration.* Proc Natl Acad Sci U S A, 2011. **108**(5): p. 2142-7.
43. Tsuda, H., et al., *The AXH domain of Ataxin-1 mediates neurodegeneration through its interaction with Gfi-1/Senseless proteins.* Cell, 2005. **122**(4): p. 633-44.
44. Lim, J., et al., *Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1.* Nature, 2008. **452**(7188): p. 713-8.
45. Serra, H.G., et al., *RORalpha-mediated Purkinje cell development determines disease severity in adult SCA1 mice.* Cell, 2006. **127**(4): p. 697-708.
46. Chen, H.K., et al., *Interaction of Akt-phosphorylated ataxin-1 with 14-3-3 mediates neurodegeneration in spinocerebellar ataxia type 1.* Cell, 2003. **113**(4): p. 457-68.
47. Kang, J., et al., *The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor.* Nature, 1987. **325**(6106): p. 733-6.
48. Ponte, P., et al., *A new A4 amyloid mRNA contains a domain homologous to serine proteinase inhibitors.* Nature, 1988. **331**(6156): p. 525-7.
49. Tanzi, R.E., et al., *Protease inhibitor domain encoded by an amyloid protein precursor mRNA associated with Alzheimer's disease.* Nature, 1988. **331**(6156): p. 528-30.
50. Kitaguchi, N., et al., *Novel precursor of Alzheimer's disease amyloid protein shows protease inhibitory activity.* Nature, 1988. **331**(6156): p. 530-2.

954 51. Wasco, W., et al., *Identification of a mouse brain cDNA that encodes a protein related to the Alzheimer*
955 *disease-associated amyloid beta protein precursor.* Proc Natl Acad Sci U S A, 1992. **89**(22): p. 10758-62.
956 52. Wasco, W., et al., *Isolation and characterization of APLP2 encoding a homologue of the Alzheimer's*
957 *associated amyloid beta protein precursor.* Nat Genet, 1993. **5**(1): p. 95-100.
958 53. Sprecher, C.A., et al., *Molecular cloning of the cDNA for a human amyloid precursor protein homolog:*
959 *evidence for a multigene family.* Biochemistry, 1993. **32**(17): p. 4481-6.
960 54. Chung, D.C., et al., *Cellular and pathological heterogeneity of primary tauopathies.* Mol Neurodegener,
961 2021. **16**(1): p. 57.
962 55. Wang, Y. and E. Mandelkow, *Tau in physiology and pathology.* Nature Reviews Neuroscience, 2016.
963 **17**(1): p. 22-35.
964 56. Rhinn, H., et al., *Alternative α-synuclein transcript usage as a convergent mechanism in Parkinson's*
965 *disease pathology.* Nat Commun, 2012. **3**: p. 1084.
966 57. Shrestha, A.K., et al., *Consequences of early postnatal lipopolysaccharide exposure on developing lungs*
967 *in mice.* Am J Physiol Lung Cell Mol Physiol, 2019. **316**(1): p. L229-l244.
968 58. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
969 59. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq*
970 *data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
971 60. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray*
972 *studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
973 61. Korotkevich, G., V. Sukhov, and A. Sergushichev, *Fast gene set enrichment analysis.* bioRxiv, 2019: p.
974 060012.
975 62. Xu, S., et al., *Using clusterProfiler to characterize multiomics data.* Nat Protoc, 2024. **19**(11): p. 3292-
976 3320.
977 63. Hayer, A., et al., *Engulfed cadherin fingers are polarized junctional structures between collectively*
978 *migrating endothelial cells.* Nat Cell Biol, 2016. **18**(12): p. 1311-1323.
979 64. Gennarino, V.A., et al., *A Mild PUM1 Mutation Is Associated with Adult-Onset Ataxia, whereas*
980 *Haploinsufficiency Causes Developmental Delay and Seizures.* Cell, 2018. **172**(5): p. 924-936 e11.

981