Lineage-specific, fast-evolving GATA-like gene regulates zygotic gene activation to promote endoderm specification and pattern formation in the Theridiidae spider

Sawa Iwasaki-Yokozawa[1], Ryota Nanjo[1,2], Yasuko Akiyama-Oda[1,3,4], Hiroki Oda[1,2,*]

[1]Laboratory of Evolutionary Cell and Developmental Biology, JT Biohistory Research Hall, Takatsuki, Osaka 569-1125, Japan

[2]Department of Biological Sciences, Graduate School of Science, Osaka University

[3]PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan

[4]Department of Microbiology and Infection Control, Faculty of Medicine, Osaka Medical and Pharmaceutical University

*Correspondence to: Hiroki Oda (hoda@brh.co.jp)

Abstract

Background: The process of early development varies across the species-rich phylum Arthropoda. Owing to the limited research strategies for dissecting lineage-specific processes of development in arthropods, little is known about the variations in early arthropod development at molecular resolution. The Theridiidae spider, *Parasteatoda tepidariorum,* has its genome sequenced and could potentially to contribute to dissecting early embryonic processes.

Results: We present genome-wide identification of candidate genes that exhibit locally restricted expression in germ-disc forming stage embryos of *P. tepidariorum*, based on comparative transcriptomes of isolated cells from different regions of the embryo. A subsequent pilot screen by parental RNA interference identifies three genes required for body axis formation. One of them is a GATA-like gene that has been fast evolving after duplication and divergence from a canonical GATA family gene. This gene is designated *fuchi nashi* (*fuchi*) after its knockdown phenotypes, where the cell movement toward the formation of a germ disc was reversed. *fuchi* expression occurs in cells outside a forming germ disc and persists in the endoderm. Transcriptome and chromatin accessibility analyses of *fuchi* pRNAi embryos suggest that early *fuchi* activity regulates chromatin state and zygotic gene activation to promote endoderm

specification and pattern formation. We also show that there are many uncharacterized genes regulated by *fuchi*.

Conclusions: Our genome-based research using an arthropod phylogenetically distant from *Drosophila* identifies a lineage-specific, fast-evolving gene with key developmental roles in one of the earliest, genome-wide regulatory events, and allows for molecular exploration of the developmental variations in early arthropod embryos.

Background

The early processes of development and reproductive strategies of animals vary among species. In certain metazoan phyla, including Chordata and Arthropoda, a high degree of early developmental variations among species is observed, despite the identification of similar morphological traits and gene expression patterns during mid-embryogenesis in each phylum [1,2,3,4,5,6,7]. Mechanisms of animal evolution that account for varied developmental trajectories preceding the phylotypic period are poorly understood.

The species-rich phylum Arthropoda, comprising four major groups (Chelicerata, Myriapoda, Crustacea, and Hexapoda), conspicuously shows the diversification of early developmental processes without the disruption of stable embryonic traits, such as the elongating germ band and repetitive body units. Genetic mutation screens and a range of molecular and genetic techniques are feasible for the model organism *Drosophila melanogaster*, which has contributed to the identification of numerous essential genes and their interactions in body axis formation, germ layer formation, and segmentation. The *Drosophila* paradigm of genetic programs for early embryogenesis has allowed for candidate-gene approaches toward elucidating the evolution of developmental mechanisms in other insects or arthropod species. Such candidate-gene approaches, however, are biased with respect to the detection of key developmental genes; having relatively low consideration for lineage-specific, fast evolving, or orphan genes [8,9]. Conversely, comparative genomics in Arthropoda and other metazoan phyla has revealed lineage-specific genomic events, including gain/loss, expansion, or divergence of a gene or a gene family [10,11,12,13]. *bicoid, spätzle*, and *gurken* in *Drosophila*, which are key regulators in the organization of the major embryonic axes, are examples of lineage-specific or fast-evolving genes [14,15,16]. Considering these *Drosophila* cases and similar ones in different phyla [17,18,19], an

64 unbiased way of identifying key developmental genes is required. Therefore, several
65 studies that used non-*Drosophila* insect species have performed genetic mutation or
66 gene knockdown screens by assessing larval cuticle phenotypes [20,21,22,23].
67 However, the identification and characterization of novel genes with key functions in
68 specific processes of early embryogenesis are hardly feasible in any non-*Drosophila*
69 arthropods, particularly in non-insect arthropods.

70 In arthropods, many model species have continued to emerge, with the genome
71 sequenced [10,24]. There have been numerous technical merits with respect to the
72 study of early embryogenesis in one of them, the common house spider—*Parasteatoda*
73 *tepidariorum*, and distinct differences from *Drosophila* have been observed in the early
74 embryogenesis mechanism [25]. These differences include early completion of
75 cellularization [26,27] and involvement of Hedgehog signaling in establishing the
76 polarity of the first embryonic axis and regulating symmetry-breaking movement of the
77 Dpp signaling center during the orthogonalization of the first and second embryonic
78 axes [28,29,30]. In the *P. tepidariorum* experimental system, genetic regulations during
79 early embryonic processes can be investigated using a simple gene knockdown
80 technique, that is, parental RNA interference (pRNAi) [29]. Research using *P.*
81 *tepidariorum* has been empowered by genomic and bioinformatic resources
82 [25,31,32,33,34,35,36], thus providing the foundation for identifying key genes
83 involved in *P. tepidariorum* development.

84 The egg of *P. tepidariorum* is spherical in shape (Fig. 1). As the development
85 starts, the egg nuclei synchronously divide, approaching the surface of the egg (stage
86 1). Cellular organization of the embryo is established by the time the nuclei increase to
87 16 [26]. The initial blastoderm comprises approximately 64 cells (stage 2; 11 h AEL),
88 showing spherical symmetry at the morphological level. This symmetry is broken at
89 approximately 15 h AEL by the emergence of uneven cell densities toward the
90 formation of a germ disc within a hemisphere of the egg (stage 3) (Fig. 1; Additional
91 file 1: Movie S1), which side is designated the embryonic side. At the pole on this side,
92 a small number of cells is apically constricted, thus forming a blastopore [25,28].
93 Zygotic gene activities are required for the process of germ-disc formation [37].
94 Although most blastoderm cells participate in the formation of the germ disc, cells
95 derived from around the abembryonic pole do not, and remain in the non-germ-disc

region (Fig. 1; see also Additional file 1: Movie S1). The formed germ disc is sharply demarcated by the difference in cell density and comprises a single layer of more than 1,000 cuboidal epithelial cells [38]. Cells at the center of the germ disc, corresponding to the blastopore, internalize to become the central endoderm (cEND) cells and the cumulus mesenchymal (CM) cells [28,32]. The latter cell populations are clusterd and act as a dynamic source of Dpp signaling to promote the development of the dorsal side of the embryo [28,29] (Fig. 1). Cells at and near the rim of the germ disc internalize to become the peripheral endoderm (pEND) and mesoderm (pMES) cells [32]. A line of cells along the rim of the germ disc express specific genes, including a *P. tepidariorum hedgehog* homolog (*Pt-hh*), that contributes to the anterior-posterior patterning of the embryonic field [30,39,40]. The germ disc is transformed into a bilaterally symmetric germ band through orchestrated cell rearrangements, while being progressively patterned [29,38] (Fig. 1).

The formation of the sharply demarcated germ disc is an early morphogenetic event in the embryogenesis of Theridiidae and some other Araneoidea spiders [27,28, 37,41,42,43,44], but it is not a general feature of spider (Araneae) embryonic development [45,46,47]. There are many other variations in early developmental processes among spider species, including the presence/absence of a visible cumulus and various modes of gastrulation [44,48]. The spider lineage provides diverse developmental trajectories prior to reaching the embryonic traits typical of arthropods.

To explore the variation in the process of early embryonic development across the phylum Arthropoda at molecular resolution, we conducted comparative transcriptomes of cells isolated from different regions of the embryo at the germ-disc forming stage in *P. tepidariorum*. Through a pRNAi-based functional screen of differentially expressed gene candidates, we identified a lineage-specific GATA-like gene, *fuchi nashi* (*fuchi*), whose knockdown hindered the completion of germ-disc formation. We obtained genome-wide datasets that represented gene expressions and chromatin accessibilities affected by *fuch*i knockdown in early embryos. Our findings showed that *fuchi* regulates zygotic gene activation to promote endoderm specification and pattern formation in the early stage of the Theridiidae spider embryo. This study provides new molecular clues for exploring the mechanisms for diversifying early developmental trajectories in the phylum Arthropoda.

Results

Comparative transcriptome analyses of cells isolated from different regions of early *P. tepidariorum* embryos

To search for genes with locally-restricted expression in *P. tepidariorum* embryos at stage 3, we performed RNA sequencing (RNA-seq) of small cell populations (approximately 10-30 cells) isolated from central (c), intermediate (i), and peripheral (p) regions of the nascent germ disc using glass capillary needles (Fig. 2A, B). RNA-seq reads were mapped to the *P. tepidariorum* reference genome Ptep1.0 (GCA_000365465.1), and then the reads were counted against the AUGUSTUS gene models (aug3.1) [31]. The read counts were compared using edgeR [49] by setting the following three combinations: comparison I, c cells versus i/p cells; comparison II, p cells versus c/i cells; and comparison III, i cells versus c/p cells (Fig. 2C-E). Based on these comparisons, we genome-widely identified candidates of differentially expressed genes (DEGs) (Additional file 2: Tables S1-S3). Using the values of false discovery rate (FDR) and fold change (FC), the candidate DEGs were prioritized (lower FDR with $\log_2 FC < -10$ [for comparisons I and II] or $\log_2 FC > 10$ [for comparison III]), selecting the top 10 genes from comparison I as Group C (Fig. 2C), the top 5 genes from comparison II as Group P (Fig. 2D), and the top 5 genes from comparison III as Group CP (Fig. 2E). The top-ranked *g4238* in Group C was listed in Group CP, despite its much stronger expression in c cells than in other cells. To avoid redundancy, *g4238* was removed from Group CP. Therefore, a total of 19 genes were systematically selected as high-priority candidate genes that might exhibit locally-restricted expression in the stage-3 embryo.

To validate our gene-selection strategy, we examined the expression patterns of the 19 selected genes in stage-3 embryos using chromogenic whole-mount in situ hybridization (WISH) (Additional file 3: Fig. S1). Three of the ten Group C genes (*g4238*, *g14287*, and *g16467*) showed specific expression at the embryonic pole. Furthermore, three of the other Group C genes (*g15167*, *g5118*, *g20850*) and one of the four Group CP genes (*g26874*) showed specific expression at the embryonic pole and in a broad area on the abembryonic side. No other Group CP and Group P genes, however, showed specific detectable signals.

160 Additionally, similar comparative transcriptome analyses of isolated cells were
161 applied to stage-4 and early stage-5 embryos. The resulting DEG lists (Additional file
162 2: Tables S4, S5) included some of the DEGs identified by the analyses of the stage-3
163 samples (e.g., *g4238*, and *g132*), as well as genes that had been known to show region-
164 specific expression at the corresponding and/or later stages (eg., *Pt-lab1* [*g7954*], *Pt-*
165 *BarH1* [*g8250*], *Pt-prd2* [*g18397*], and *Pt-hh* [*g4322*]).
166 These data suggest the effectiveness of our gene-selection strategy based on
167 comparative transcriptome analyses of isolated cells from early *P. tepidariorum*
168 embryos.
169
170 **Pilot functional screen identifies three genes required for germ-disc formation**
171 **and/or axis formation**
172 To identify genes with key functions in the early embryonic development of *P.*
173 *tepidariorum*, we screened the 19 candidate DEGs (from the stage-3 samples) using
174 parental RNAi (pRNAi). Double-stranded RNA (dsRNA) synthesized for each gene,
175 as well as that for a non-spider control gene *green fluorescent protein* (*gfp*), was
176 injected into at least two adult females. Aliquots of eggs deposited by the females were
177 placed in oil to monitor the developmental process from early stages under the
178 stereomicroscope. Through this screening, we identified three genes whose knockdown
179 resulted in uncommon defects associated with germ disc formation (*g26874* and *g7720*)
180 and/or cumulus movement (*g26874*, *g7720*, and *g4238*) (Additional file 4: Table
181 S6)[50]. To verify the RNAi specificity, we injected two or three dsRNAs prepared
182 from non-overlapping regions of each positive gene and confirmed that the same
183 phenotypes were produced depending on the genes (Fig. 3A; Additional file 3: Fig. S2;
184 Additional file 5: Movie S2)[50]. In addition, we confirmed through WISH that the
185 transcript levels of the target genes were reduced in the corresponding pRNAi embryos
186 (Fig. 3B).
187 Time-lapse microscopy and cell tracking revealed that *g26874* pRNAi
188 embryos showed normal blastoderm formation (stage 2) and subsequent initiation of
189 cell density shift toward the formation of a germ disc, in a rather normal way, but they
190 failed to stabilize the boundary of the forming germ disc (25 h AEL), with the cell
191 density shift being reversed (Fig. 3A; Additional file 5: Movie S2; Additional file 6:

192 Movie S3). In all of such severely affected *g26874* pRNAi embryos, the initial cell
193 thickening at the embryonic pole took place. In 64% of the *g26874* pRNAi embryos (*n*
194 = 101 from 10 egg sacs), however, CM cell migration was faint or failed to occur. The
195 defects in germ-disc formation were not recovered at later stages, showing no clear
196 sign of cell movements toward the formation of a germ band. In the most severe cases
197 of *g26874* pRNAi embryos, cells that had initially shifted toward forming a germ disc
198 reversed their movement to spread across the embryo, followed by abrupt breakdown
199 of the surface cell layer and extrusion of the yolk materials (Additional file 5: Movie
200 S2)[50].
201 *g7720* pRNAi embryos showed normal blastoderm formation followed by cell-
202 density shift (stages 2 and 3), but its development was arrested around the end of stage
203 3, showing a gradually degenerating germ disc (Additional file 5: Movie S2)[50].
204 Although cumulus-like cell thickening was formed at the embryonic pole, it persisted
205 there, with the overall morphology failing to develop further.
206 *g4238* pRNAi embryos were morphologically indistinguishable from wild-
207 type and *gfp* pRNAi (control) embryos until early stage 5. The earliest visible
208 difference of *g4238* pRNAi embryos from normal embryos was their failure to initiate
209 migration of the CM cells at early stage 5, followed by their disassembly and dispersal
210 (Additional file 5: Movie S2)[50]. In many of the embryos, surface cells around the
211 closed blastopore formed a tail-like protrusion. Despite experiencing cumulus defects,
212 some *g4238* pRNAi embryos displayed asymmetric cell movements toward the
213 formation of a germ band.
214
215 **Molecular characterization of the three identified genes**
216 Developmental transcript profiling in whole wild-type embryos, based on public
217 RNA-seq datasets [36], revealed that there is little or no maternal supply of *g26874*
218 and *g4238* transcripts and increasing levels of the zygotic transcripts prior to peaking
219 at stage 4 (Fig. 3C). In contrast, substantial levels of *g7720* transcript appeared to be
220 maternally supplied, with the zygotic transcript expressed at rather uniform levels
221 through development.
222 Reciprocal blastp/tblastn searches using *Drosophila* and mouse reference
223 sequence (RefSeq) protein databases at National Center for Biotechnology

224 Informataion (NCBI) and the *P. tepidariorum* aug3.1 transcript sequences [31]
225 revealed that *g7720* encoded an orthologue of *Drosophila* proximal sequence element
226 A (PSEA)-binding protein 49kD (Pbp49) and mouse small nuclear RNA-activating
227 complex subunit 3 (Snapc3) [51], and *g4238* an orthologue of *Drosophila* Ets98B (D-
228 ets4) and mouse SAM pointed domain-containing Ets transcription factor (Fig. 3D)
229 [52]. *g4238* was identical to *Pt-Ets4*, as reported in a previous study [53], and
230 consistent with our observations.
231 In contrast to the *g4238 and g7720* cases, reciprocal blastp/tblastn searches
232 failed to confirm the presence of *Drosophila* and mouse orthologues for *g26874*. The
233 best-hit *Drosophila* and mouse proteins were Pannier (Pnr), a GATA transcription
234 factor, and GATA-4, respectively (Fig. 3D), but these proteins were significantly
235 closer to GATA family members encoded by *g8336* and several other genes in the *P.*
236 *tepidariorum* genome. Based on these results, along with those obtained from
237 comprehensive sequence analyses described below, we concluded that *g26874*
238 encodes a novel GATA-like gene, which we named "*fuchi nashi*" (abbreviated *fuchi*)
239 after its knockdown phenotypes. This Japanese word means rimless.
240
241 ***fuchi* is a lineage-specific, fast-evolving GATA-like gene**
242 The family of GATA transcription factors are characterized by the possession of a
243 specific DNA-binding domain, which comprises two GATA-type zinc finger (ZF)
244 motifs (ZF1 and ZF2), typically $CX_2CX_{17}CX_2C$, and two basic regions followed by
245 each ZF motif (BR1 and BR2) (Fig. 4A, B)[54,55]. Fuchi had this GATA-type DNA-
246 binding domain, although its sequence was highly divergent from those of *Drosophila*
247 Pnr and mouse GATA-4. To comprehensively understand the diversity of GATA family
248 proteins among metazoans and the phylogenetic origin of *fuchi*, we exhaustively
249 collected and manually aligned amino acid sequences of GATA family proteins from *P.*
250 *tepidariorum*, other spider and non-spider chelicerate species as well as from other
251 selected arthropod and non-arthropod metazoan species using publicly available
252 bioinformatic resources (Additional file 7: Tables S7, S8). The resulting sequence
253 alignment revealed that the metazoan GATA family members were classified under
254 canonical and noncanonical types. The canonical type was characterized by a BR1 of
255 29 amino acid residues with high sequence similarity, and the noncanonical type by a

256 BR1 of more, or less, than 29 amino acid residues with a varying degree of sequence
257 divergence. The canonical GATA family members included proteins encoded by five *P.*
258 *tepidariorum* genes (designated Pt-GATA1 to Pt-GATA5, encoded by *g20514*, *g1834*,
259 *g8336*, *g25261*, and *g8337*, respectively) and their counterparts from other Araneae
260 species, as well as *Drosophila* Serpent (Srp), the only *Nematostella vectensis* GATA
261 (Nv-GATA), two echinoderm GATAs, and mouse GATA-1 to GATA-6 (Fig. 4B-D;
262 Additional file 7: Table S8). Phylogenetically widespread distribution of the specific
263 sequence features suggested that the canonical GATA family members represent the
264 ancestral state for metazoan GATA proteins.

265 The sequence alignment of the noncanonical GATA family members, including
266 Fuchi, guided the identification of at least three groups specifically found within the
267 Araneae lineage, based on gap insertion patterns, cysteine residue spacing, and
268 sequence signatures (Fig. 4C, D; Additional file 7: Table S8). These three groups were
269 designated as GATA-like 1 to 3 (GATAL1, GATAL2, and GATAL3). GATAL1, which
270 included a *P. tepidariorum* protein encoded by *g26871* (Pt-GATAL1), was relatively
271 close to the canonical GATA family members at the sequence level, but they had a BR1
272 of 34 amino acid residues. GATAL2, which included Fuchi and another *P. tepidariorum*
273 protein encoded by *g26875* (designated Pt-GATAL2b), was specifically identified by a
274 signature sequence QCV(K/R)CG at the N-terminal end of ZF1, varied cysteine residue
275 spacing in ZF1 ($CX_2CX_{17\text{-}20}CX_2C$), and a highly varied number (23-44) of amino acid
276 residues in BR1. GATAL3, which was only found in spiders of RTA clade, was
277 specifically identified by conserved unique sequences and cysteine residue spacing
278 ($CX_2CX_{18}CX_2C$) in ZF1. The signature sequence unique to GATAL2, QCV(K/R)CG,
279 corresponded to ECVNCG, conserved in most canonical GATA family members
280 (including Pt-GATA1 to 4) and GATAL1, to QC(A/V)(M/V)CG, conserved in Pt-
281 GATA5 and its counterparts, and to ECANCG, conserved in GATAL3.

282 To investigate the phylogenetic relationship between canonical and
283 noncanonical GATA family proteins, we performed maximum likelihood (ML) analysis
284 using unambiguously aligned 75 amino acid sites from a total of 118 proteins with both
285 ZF1 and ZF2 representing the typical cysteine residue spacing ($CX_2CX_{17}CX_2C$) (Fig.
286 4E, F; Additional file 7: Table S8). Results showed Araneae-specific expansions of
287 GATA protein family members, through gene duplication and divergence. ML tree

288 topology supported the orthology of each of the GATA5, GATAL1, and GATAL2
289 groups. Furthermore, branch lengths indicated that the sequences of the GATAL2
290 proteins had been evolving much faster than those of the canonical GATA family and
291 GATAL1 proteins (Fig 4F). Owing to their atypical cysteine residue spacing in ZF1
292 ($CX_2CX_{18\text{-}20}CX_2C$), Fuchi and other Araneoidea GATAL2 proteins could not be
293 included in the ML analysis (Fig. 4C; Additional file 7: Table S8). Importantly,
294 GATAL2 proteins with the typical ZF1 ($CX_2CX_{17}CX_2C$), which were used in the ML
295 analysis, were present in a broader phylogenetic range (including Leptonetidae,
296 Eresidae, and RTA clade) (Fig. 4C, E, F; Supplementary Table 4). This observation
297 suggested that variously divergent sequences in ZF1 of the Araneoidea GATAL2
298 proteins, except the specific signature sequence, were likely due to rapid sequence
299 evolution from the GATAL2 ancestral state.
300 Unlike the other two of the three genes judged as positive in the functional
301 screen, *fuchi* was characterized as a lineage-specific gene. We therefore decided to
302 focus our later analyses on this gene.
303
304 ***fuchi* is expressed in endoderm cells originating from both sides of the early**
305 **embryo**
306 Among the eight GATA family genes in *P. tepidariorum*, *fuchi* was the only gene
307 which was transcribed at substantial levels during the germ-disc forming stage (stage
308 3) (Fig. 3C; Additional file 3: Fig. S3). We investigated the expression of *fuchi*
309 transcript in early embryos at cellular resolution, using conventional WISH and the
310 combination of FISH with antibody staining for β-catenin, which serves as a marker
311 for regions of cell-cell contact (Fig. 5). Owing to technical limitations in the fixation
312 of stage-1 embryos, the earliest stage of development examined was stage 2, the
313 blastoderm stage (13 h AEL), when there is no morphological sign of asymmetry
314 reflecting the future embryonic axes. At this stage, localized signals for *fuchi*
315 expression were detected on one side of the embryo (Fig. 5A, A', B, B') but could not
316 be related with the axis that would emerge a little later (15 h AEL).
317 At early stage 3 (15 h AEL), the embryo begins to display morphological
318 asymmetries, with an emerging embryonic-abembryonic (Em-Ab) axis (Fig. 1). At this
319 stage and later, the *fuchi* transcript was detected in a small number of cells condensed

320 at the embryonic pole (Fig. 5I, I', I") and in a larger number of cells showing a
321 progressively wider distribution on the abembryonic side of the embryo (Fig. 5J, J', J").
322 Using a flat-preparation of a representative early stage 3 (15 h AEL) embryo stained
323 for *fuchi* transcript, the number of *fuchi*-expressing cells was counted. There were 8
324 and 20 *fuchi*-expressing cells around the embryonic pole and on the abembryonic side,
325 respectively, among a total of approximately 300 cells (Additional file 3: Fig. S4).
326 Notably, signals for *fuchi* transcript in the nuclei were, in many cases, observed as
327 paired dots (Fig. 5I', J'), indicating that the zygotic transcription was already active at
328 early stage 3. In the equatorial area of stage 3 embryos (15 and 21 h AEL), there was a
329 transition from cells expressing *fuchi* transcript, on the abembryonic side, to those not
330 expressing it, on the embryonic side (Fig. 5J, J', J", K, K', K"). In the embryo at 21 h
331 AEL, the *fuchi*-expressing cells had larger apical surfaces than the *fuchi*-negative cells
332 (Fig. 5K').

333 In embryos at stage 4 (25 and 27 h AEL), most *fuchi*-expressing cells on the
334 abembryonic side had further enlarged surfaces with less tensed lines of cell-cell
335 contact, which contrasted with a tightly packed organization of *fuchi*-negative cells
336 constituting the nascent germ disc on the embryonic side (Fig. 5D, L, M, M', M"). A
337 few *fuchi*-expressing cells, however, were identified as part of the germ disc at its
338 border (Fig. 5M', M"). At the blastopore, cEND cells and CM cells are internalized
339 (Fig. 1). *fuchi* expression was observed in all these internalizing and internalized cells
340 (Fig. 5L, L', L").

341 In embryos at stage 5 (32 and 38 h AEL), all cells internalized through the
342 blastopore continued to express *fuchi* (Fig. 5E, E', F, F', N, N1', N1", O, O1").
343 Peripheral populations of *fuchi*-expressing cells appeared to be internalized a little later
344 than the central populations. At early stage 5 (32 h AEL) *fuchi*-expressing cells on the
345 embryo surface were found not only in the non-germ disc region but also at the rim of
346 the germ disc (Fig. 5N, N2', N2"). At late stage 5 (38 h AEL), there were *fuchi*-
347 expressing cells that were internalized/internalizing from or through the rim of the germ
348 disc (Fig. 5F, F', O, O2', O2"). Using a DNA stain probe, SPY555-DNA, we tracked
349 some cell nuclei in the germ-disc/non-germ-disc transition area from early stage 5 to
350 stage 8 in a live embryo (Additional file 4: Movie S4), which directly visualized two
351 phases of cell internalization. Phase 1 of cell internalization occurred from mid to late

352 stage 5, where cells that were not part of the germ disc were approaching the rim of the
353 germ disc and then internalizing to below the surface cell layer of the germ disc. These
354 were most likely to be *fuchi*-expressing cells. Phase 2 of cell internalization occurred
355 subsequently, where cells at the rim of the germ disc were internalizing. Double
356 staining for *fuchi* and *Pt-hh* transcripts revealed that certain cells at and near the rim of
357 the germ disc at early stage 5 expressed both *fuchi* and *Pt-hh* (Fig. 5P, P'). Upon
358 internalization, however, *fuchi*-expressing cells were exclusively *Pt-hh*-negative (Fig.
359 5Q, Q').
360 To determine whether *fuchi*-expressing cells internalized from the peripheral
361 side of the germ disc are endoderm or mesoderm, double staining of late stage 5 and
362 stage 6 embryos for *fuchi* and *Pt-fkh* (a marker for pMES plus pEND; [56]) or *003_J01*
363 (a marker for pMES)[57] was performed. The use of *003_J01* enabled us to identify
364 pMES cells prior to internalization, which was not the case with *Pt-twi*, another pMES
365 marker [32,56,58]. Observations indicated that all internalized, *fuchi*-expressing cells
366 were *Pt-fkh*-positive and most of them were *003_J01*-negative (Additional file 3: Fig.
367 S5). It was also indicated that *003_J01*-positive cells appeared to be internalized later
368 than *fuchi*-expressing cells, which presumably corresponded to Phase 2 of cell
369 internalization. *fuchi* expression in embryos at stages 6 and 7 never showed a spatially
370 periodic pattern along the emerging body axis (Fig. 5G, G', H, H'), contrasting with
371 mesoderm marker expression [58]. Taken together, these data suggested that *fuchi* is
372 expressed in endoderm cells originating from both sides of the early embryo. In
373 addition, *fuchi* expression occurred in association with extraembryonic differentiation
374 (Fig. 5G, G', H, H'). This *fuchi* expression was already initiated in the germ disc at late
375 stage 5 (Fig. 5O, O1").
376
377 ***fuchi* knockdown prevents the demarcation of a forming germ disc**
378 To investigate the defects following *fuchi* knockdown, we analyzed the expression of
379 *Pt-hh*, *Pt-fkh*, and *fuchi* transcripts in wild-type and *fuchi* pRNAi embryos at stage 4
380 and early stage 5 at cellular resolution. In normal germ-disc development, a line of cells
381 expressing *Pt-hh* and *Pt-fkh* was organized to demarcate the boundary of the forming
382 germ disc during stage 4 and early stage 5 (Fig. 6A, A', C, C', E, E', G, G'). In the
383 recognizable germ-disc/non-germ-disc transition area of *fuchi* pRNAi embryos at stage

384 4, the initial *Pt-hh* expression was strongly suppressed, while *Pt-fkh* expression was
385 detectably initiated in certain cells with *fuchi*-positive nuclei and their neighboring cells
386 (Fig. 6B, B', D, D'). At early stage 5, faint *Pt-hh* and *Pt-fkh* expression were detected
387 in small numbers of cells, but they did not contribute to shaping the germ disc (Fig. 6F,
388 F', H, H'). In *fuchi* pRNAi embryos at a later stage (late stage 5), there was no clear
389 separation between cell populations on the embryonic and abembryonic sides, with cell
390 internalizations being highly limited (Fig. 6I, I1', I1", I2', I2"). The only cells situated
391 below the surface layer were cells at or near the embryonic pole (Fig. 6I1', I1"), which
392 were associated with β-catenin concentrations marking the CM cell cluster. These
393 observations suggested that *fuchi* knockdown prevented the demarcation of a forming
394 germ disc and interrupted the formation of the germ layers.
395
396 ***fuchi* is required for zygotic activation of endodermal and patterning genes**
397 To explore the role of *fuchi* in regulating early *P. tepidariorum* development, we
398 conducted comparative transcriptome analyses of *fuchi* pRNAi versus untreated
399 embryos at stages 2, 3, and early stage 5 using RNA-seq. These analyses identified 6,
400 139, and 785 DEGs for stages 2, 3 and early stage 5, respectively (FDR<0.01; Fig. 7A,
401 B; Additional file 9: Tables S9-S11). The experiments were validated by detecting
402 lowered levels of *fuchi* transcript and unaffected levels of the α-catenin (*catA*)
403 transcript in the *fuchi* pRNAi samples (Fig. 7C). Except for the *fuchi* gene itself, there
404 was no overlap between the stage-2 and stage-3 DEGs but there was an overlap of 93
405 genes between the stage-3 and early stage-5 DEGs (Fig. 7B; Additional file 9: Tables
406 S9-S11). According to NCBI's gene annotation, 26 of the 93 DEGs encoded
407 uncharacterized proteins. Transcripts for the stage-2 DEGs appeared to be maternally
408 supplied (Additional file 3: Fig. S6), which contrasted with those for highly ranked
409 genes from the stage-3 and early stage-5 DEGs lists (sorted by FDR value). They tended
410 to show a sharp increase in the transcript level during stages 2 to 4, with no or little
411 transcript supply from the mother (Fig. 7C, D). All the top 20 genes from the stage-3
412 DEG list were included in the stage-5 DEG list (Fig. 7D; Additional file 9: Table S10).
413 Blastx searches against mouse and *Drosophila* RefSeq protein databases resulted in no
414 hits for seven of the top 10 DEGs (E-value < 1e-5) (Fig. 7C). Transcripts from 8 of the
415 top 10 DEGs (*g1125*, *g2114*, *g20850*, *g17950*, *g12453*, *g27539*, *g486*, and *g26914*)

416 were detected in all or part of the internalized endoderm at late stage 5, and those from
417 3 of the 8 genes (*g17950*, *g12453*, *g20850*) were detected in cells on the abembryonic
418 side at stage 3 (Additional file 3: Fig. S7). Notably, the *g20850* was identified as one
419 of the Group C genes in the initial comparative transcriptome analysis of isolated cells
420 (Fig. 2C; Additional file 3: Fig. S1).
421 The stage-3 *fuchi* pRNAi DEG list also included several genes previously
422 characterized as key regulators of embryonic patterning in *P. tepidariorum*, such as *Pt-*
423 *hh* (*g4322*)[30], *Pt-sog* (*g23966*)[29], and *Pt-Delta* (*g25248*)[32] (Additional file 3:
424 Fig. S8; Additional file 9: Table S10). Despite the effects of *fuchi* knockdown on the
425 expression of these patterning genes, certain patterning events involving *Pt-msx1* and
426 *Pt-Delta* expression were initiated around the blastopore in *fuchi* pRNAi embryos
427 (Additional file 3: Fig. S8). *fuchi* knockdown led to a lack of circular patterns of *Pt-hh*
428 and *Pt-otd* expression associated with anterior patterning, but its effects on the *Pt-hh*
429 and *Pt-otd* expression levels in whole embryos at early stage 5 were limited.
430 Taken together, these results suggested that *fuchi* is required for the zygotic
431 activation of a set of endodermal genes and of some patterning genes at stage 3 or
432 shortly before. In addition, *Pt-GATA4* (*g25261*), the likely counterpart of *Drosophila*
433 *srp* (Feitosa et al., 2017), and *Pt-GATA5* (*g8337*) were included in the stage-5 *fuchi*
434 pRNAi DEG list but not in the stage-3 one (Additional file 9: Tables S9, S10).
435
436 ***fuchi* is involved in regulating chromatin accessibility in specific genomic regions**
437 **in early embryos**
438 Certain GATA family members act as pioneer transcription factors that can influence
439 chromatin structure [59]. Therefore, we considered the possibility that *fuchi* controls
440 the transcription of downstream genes through chromatin structure regulation. To
441 investigate this possibility, we conducted assay for transposase-accessible chromatin
442 using sequencing (ATAC-seq) of wild-type (untreated), *fuchi* pRNAi, and *Pt-hh* pRNAi
443 embryos at stage 3 (18h AEL), obtaining genome-wide datasets on chromatin
444 accessibility. Read counts in ATAC-seq peaks were subjected to comparative analyses,
445 and 317 genomic regions that were differentially accessible between the *fuchi* pRNAi
446 and untreated samples but only 3 between the *Pt-hh* pRNAi and untreated samples were
447 identified (Fig. 8A; Additional file 10; Tables S12, S13; Additional file 11; Tables S14,

448 S15; FDR<0.05). This indicated that the *Pt-hh* pRNAi samples served as a negative
449 control for assessing the impact of *fuchi* knockdown. The two comparisons revealed
450 only one shared genomic region (Fig. 8A; Additional file 10; Table S13; Additional file
451 11; Table S15), which potentially reflected a nonspecific effect stemming from pRNAi
452 treatment. Therefore, 316 genomic regions where chromatin accessibility was
453 significantly affected following *fuchi* knockdown in stage-3 embryos were identified
454 (Fig. 8B).

455 Most of these regions (276/316, 93%) showed suppressed accessibility upon
456 *fuchi* knockdown (Fig. 8A, B), and 30 of them were found to be located closer to (or
457 within) genes included in the *fuchi* pRNAi DEG lists (for stage-3 and/or early stage-5)
458 than to any other annotated genes (Fig. 8B; Additional file 10: Table S13). Moreover,
459 motif discovery search revealed that motif sequences RNWGATAAGAVW, similar to
460 typical GATA-binding motif sequences, were significantly enriched among the
461 sequences of the *fuchi*-dependent chromatin accessibility peaks (Fig. 8C). *fuchi*-
462 dependent chromatin accessibility peaks containing these motif sequences were
463 actually present within or close to some of the genes identified by the comparative
464 transcriptome analyses of *fuchi* pRNAi versus untreated embryos (Fig. 8B).

465 Notable examples of such genes were the top-ranked gene from the stage-3
466 *fuchi* pRNAi DEG list (*g1125*), the key patterning gene *Pt-hh* (*g4322*), and the
467 canonical GATA family member *Pt-GATA5* (*g8337*) (Fig. 9A, B, C). *g1125* was
468 predicted to encode an uncharacterized product containing two potential
469 transmembrane domains (XP_015915605.1; Fig. 9D). Genes encoding homologous
470 proteins to this product were detected in the genomes of at least three other Araneoidea
471 species, *Araneus ventricosus* (GBO06802.1), *Argiope bruennichi* (KAF8764632.1),
472 and *Oedothorax gibbosus* (KAG8179838.1)(Fig. 9D), but any non-Araneoidea protein
473 sequences that showed similarity to these products (E-value < 1) were not found in the
474 NCBI non-redundant protein sequences collection (March 7, 2022). *Pt-GATA5*
475 transcript, like *fuchi*, was observed in cEND, CM, and pEND cells at stage 5 and widely
476 distributed endoderm cells at stage 7 (Fig. 9E, F). Altogether, these results suggested
477 that *fuchi* is required for establishing "open" states of chromatin structure in specific
478 genomic regions, some of which are associated with endodermal and patterning genes
479 expressed in the early stages.

480

481 **Detection of consistent endodermal gene expression based on single-nucleus and**
482 **single-cell RNA-seq datasets**

483 In a recent separate study, we performed single-nucleus and single-cell RNA-seq
484 analysis of late stage 5 embryos in *P. tepidariorum* [60]. In this study, clearly isolated
485 endodermal cell clusters, which were marked by a known early endoderm marker
486 012_A08 [56], were identified (Additional file3: Fig. S9). Strong *fuchi* expression was
487 detected in these clusters, with weaker levels of *fuchi* expression being additionally
488 detected in mesodermal (marked by *Pt-twi*)[58] and dorsal germ-disc (marked by *Pt-*
489 *GATA1*)[60] cell populations. At the same time, the endoderm-specific expression of
490 *Pt-GATA4* and *Pt-GATA5* was confirmed. In contrast to *fuchi*, and the two canonical
491 GATA genes, *Pt-hh* expression was missing or largely reduced in the endodermal cell
492 populations at late stage 5. These data were consistent with the staining data shown in
493 Fig. 5 and Fig. S5 (Additional file 3). The single-nucleus and single-cell RNA-seq
494 datasets further confirmed that most of the high-ranking genes from stage-3 *fuchi*
495 pRNAi DEG list showed endoderm-specific expression at late stage 5 (Additional file
496 3: Fig. S10).

497

498 **Discussion**

499 We conducted genome-based research using the spider *P. tepidariorum* to explore the
500 variations in the mechanisms of early embryonic development in the phylum
501 Arthropoda. A key initial attempt in this study was to analyze transcriptomes of cells
502 isolated from different regions of the early spider embryo. This allowed for the
503 genome-wide identification of candidate genes whose expression might be locally
504 restricted along the emerging first embryonic axis, without relying on knowledge from
505 other organisms. Although similar approaches to identifying localized transcripts in
506 early animal embryos have been reported in non-arthropod animals [17,61,62], early
507 arthropod embryos are rarely studied in such a way.

508 A benefit of using *P. tepidariorum* is the applicability of a simple gene-
509 knockdown technique, that is, pRNAi. This allowed us to carry out a functional screen
510 of some of the listed genes. Gene-specific embryonic phenotypes were obtained using
511 3 of the 19 genes tested (15%). Compared with similar screening efforts we previously

512 made following sequence-based gene selection, this small proportion is not surprising
513 [30]. Although similar RNAi-based gene knockdown is available in certain arthropods,
514 such as *Tribolium castaneum*, it is important to note that the *P. tepidariorum* system
515 allows for the easy monitoring of early embryonic developmental processes, even
516 following pRNAi treatment. Time-lapse recording of the phenotypes from early stages
517 gave sharable, rich, objective information on gene functions and the timing of their
518 actions. Early defects often lead to reduced characteristic morphologies at later stages.
519 Exclusion of common spontaneous defects is required for effective functional
520 screening. Direct detection of the earliest developmental defects following knockdown
521 treatments in the initial screening could simplify the structuring of our interests.
522 Despite the small scale of the functional screen we conducted, incomparable with the
523 large-scale, unbiased RNAi screen in *Tribolium* [23], our results may encourage
524 additional gene-function screens in *P. tepidariorum*.

525 Unexpectedly, our pRNAi screen found that the cell movement toward the
526 formation of a germ disc was reversed following the knockdown of a GATA-like
527 transcription factor gene, which we named *fuchi nashi* (*fuchi*). This phenotype suggests
528 the presence of a shape-stabilizing phase in the germ-disc development, which likely
529 corresponds to stage 4, during which a line of cells along the border is organized to
530 gain specific gene expressions. *fuchi* expression, however, is initiated much earlier than
531 this stage, in cells outside a forming germ disc, and later continues in all internalizing
532 and internalized endoderm cells. Our comparative transcriptome and chromatin
533 accessibility analyses of *fuchi* pRNAi versus untreated embryos suggested that early
534 *fuchi* activities have broad impacts on zygotic gene activation around stage 3. Lack of
535 the *fuchi* activities appeared to hinder the specification of endoderm fates. The
536 abnormal reversed cell movements in *fuchi* pRNAi embryos could be, at least in part,
537 explained if we assume that the different cell types being specified exhibit different
538 mechanical properties. The earliest morphological defects of *fuchi* pRNAi embryos
539 possibly result from a failure to establish the differences in adhesion tension between
540 cell populations.

541 We showed that early *fuchi* activities influence the expression of some of the
542 key signaling genes involved in embryonic development and pattern formation. These
543 include *Pt-hh*, *Pt-Delta*, and *Pt-sog*. *Pt-hh* and *Pt-Delta*, down-regulated in response to

544 *fuchi* knockdown at stage 3, are known to show expression patterns similar to *fuchi* in
545 early embryos (stages 3 and 4) [30,32], and *Pt-sog*, up-regulated oppositely, to be
546 expressed in broad domains of the stage-5 germ disc [29]. Notably, early *Pt-hh*
547 expression initiated from the abembryonic side contributes to the formation of the
548 global polarity in the germ disc and subsequent anterior patterning. In *fuchi* pRNAi
549 embryos, however, no patterning events that should occur from the abembryonic side
550 of the egg were identified. These findings suggest that early *fuchi* activities specify
551 endoderm fates and promote pattern formation in a wide area of the embryo, at the
552 same time. Similar GATA factor-mediated regulation of *hh* expression has been
553 documented in several mammalian and *Drosophila* developmental events [63,64,65].
554 The family of GATA transcription factors is known to regulate cell fate
555 specification and cell differentiation during various developmental processes in a wide
556 range of animals [67]. Our functional and phylogenetic characterization of *fuchi*
557 provide the opportunity to reconsider the relationship between the evolution of GATA
558 family members and their developmental roles. We defined "canonical" members of
559 the GATA family, as they have perfectly alignable, close sequences to each other
560 throughout the double zinc-finger DNA-binding domain regardless of the phylogenetic
561 distance of the species. The canonical members include all six GATA paralogs in
562 vertebrates, GATA-1 to GATA-6, the only *Nematostella* GATA, *Drosophila* Serpent
563 (Srp), and five *P. tepidariorum* GATA paralogs, Pt-GATA1 to Pt-GATA5. The
564 vertebrate GATAs are phylogenetically classified into two subgroups: GATA-1/2/3
565 and GATA-4/5/6, while Srp is closer to GATA-4/5/6 [55,68]. *srp* functions for
566 endoderm specification in early *Drosophila* embryos [69], and vertebrate GATA-4/5/6
567 genes have developmental roles in endodermal cell lineages [70,71,72,73]. Notably, in
568 mammals, GATA-6, in cooperation with GATA-4, is essential in the specification of
569 the primitive and definitive endoderm, through chromatin accessibility regulation
570 [74,75]. Based on these lines of evidence, the idea that the roles of GATA-4/5/6 genes
571 in endoderm specification and development are conserved across bilaterians has been
572 widely accepted. Consistent with this idea, single-cell and single-nucleus
573 transcriptomes indicated that the likely *srp* ortholog *Pt-GATA4* (*Pt-srp*) appears to be
574 expressed in an endoderm-specific manner at least at late stage 5 of *P. tepidariorum*
575 development (Additional file 3: Fig. S9), although a previous report mentioned that

576 "*Pt-srp* expression starts at stage 8 in the presumptive extraembryonic cells." [76]
577 Evidently, developmental transcriptomes gave a quantitative indication that there is
578 little *Pt-GATA4* expression at stage 4 and earlier but substantial expression at late stage
579 5 and later (Additional file 3: Fig. S3). We suggested that instead of *Pt-GATA4*, *fuchi*
580 and *Pt-GATA5* are key GATA family genes contributing to endoderm specification and
581 development in the earlier stages of *P. tepidariorum* development. Although *Pt-GATA5*
582 is a canonical member of the GATA family, it appears to have no orthologous
583 counterpart outside the Araneae. Fuchi is not a canonical GATA factor, having a highly
584 diverged sequence with retaining the dual zinc fingers. *fuchi* appears to act on *Pt-*
585 *GATA5*, possibly through chromatin accessibility regulation. The *fuchi*-dependent
586 ATAC-seq peak detected in the *Pt-GATA5* locus is located in an intron between exons
587 encoding the two zinc fingers. Sequential regulation of multiple GATA factors in early
588 endoderm specification and the subsequent differentiation process is a shared feature
589 among distantly related bilaterians, including mouse, *C. elegans* and *Drosophila*
590 [75,77,78]. However, the genetic components responsible for this regulation exhibit
591 no orthologous relationships between the species, even within the Arthropoda,
592 suggesting independent modifications of the GATA-mediated regulatory system for
593 endoderm specification and development in the respective animal lineages.
594 *fuchi* and its orthologs (GATAL2) were found in sub lineages of the Araneae,
595 including many Araneomorphae taxa but not non-Araneomorphae. These GATA-like
596 genes, regardless of the cysteine spacing patterns in ZF1, have been rapidly evolving
597 their sequences compared to most other GATA and GATA-like genes. Despite this
598 divergence trend, we could define the likely orthologous GATAL2 group using the
599 conservative signature sequence QCV(K/R)CG. Because the amino acid residues
600 homologous to this sequence are involved in binding to specific cofactors, such as the
601 vertebrate Friend of GATA (FOG) and *Drosophila* U-shaped (Ush), in canonical
602 GATA factors [79,80,81], Fuchi possibly binds to a unique cofactor via this sequence.
603 In relation with GATA cofactors, one of two *P. tepidariorum* homologs of *Drosophila*
604 Ush (*g14866*; the other homolog is *g16893*) was included in both the gene lists obtained
605 from the comparative RNA-seq and ATAC-seq analyses using stage-3 *fuchi* pRNAi
606 embryos (Fig. 8; Additional file 9: Table S10; Additional file 10: Table S13). Certain
607 GATA cofactors are known to exert an inhibitory effect to complicate the GATA

608 factor-mediated transcriptional regulation [63,64,82]. Such an inhibitory effect of a
609 GATA cofactor potentially accounts for the transient *Pt-hh* expression in the early
610 abembryonic cell population. High degrees of sequence divergence in the DNA-
611 binding domain of Fuchi and other GATAL2 proteins may be an indication of evolving
612 interactions with cofactors and DNA-binding sites. Previous studies have suggested
613 that N-terminal zinc fingers (ZF1) influence selectivity of DNA-binding sequences
614 [81,83]. Alternatively, the sequence divergences might have simply resulted from
615 reducing degrees of functional constrains. Future studies will focus on testing these
616 possibilities.

617 The phylogenetic origin of the *fuchi* activity in early endoderm specification
618 and those of the regulatory connection between *fuchi* and *Pt-GATA5*, is an intriguing
619 issue that provides molecular clues regarding the diversification mechanism of the
620 process of early embryonic development. The formation of a sharply demarcated germ
621 disc is not a common feature of early spider development. In Theridiidae spider
622 embryos, the formation of the germ disc involves an appearance of two distinct
623 endoderm or endoderm-like cell populations derived from both polar regions of the
624 early embryo, and the formed sharp boundary of the germ disc serves as sites of
625 endoderm and mesoderm cell internalization, as well as sites of sending patterning
626 signals [30,32,41,44]. These developmental strategies have potential merits of gaining
627 parallel processes, which may allow for a more rapid development. In the embryos of
628 many other spider species, including three distinct examples from outside the
629 Araneoidea, *Cupiennius salei* [45], *Pholcus phalangioides* [46], and *Hasarius adansoni*
630 [84], cell movements causing the germ layers only occur from one of the polar regions,
631 which corresponds to the embryonic pole or the center of the germ disc in the *P.*
632 *tepidariorum* embryo. The technical term "germ disc" results in potential confusions in
633 documenting the Theridiidae embryonic development. This term is, in certain cases,
634 used to indicate the multicell-layered region of the embryo that results from germ layer
635 formation in spiders and other arthropods [85]. The initial germ disc in Theridiidae
636 embryos, however, is a form of a single epithelial cell layer that precedes the formation
637 of the germ layers [26,27,37]. *fuchi* serves as a molecular marker for the cells outside
638 the forming germ disc. *fuchi* homologs may aid in identifying the existence of an
639 evolutionarily conserved abembryonic cell type in other spider species. Complicatedly,

640 the *P. tepidariorum* genome has a paralog of *fuchi* that was presumably generated by
641 recent tandem gene duplication. No corresponding paralog was detected in other
642 Theridiidae and non-Theridiidae spiders. The *fuchi* paralog, *Pt-GATAL2b*, shows little
643 expression at least during the early and mid embryogenesis of *P. tepidariorum*. Studies
644 on the expression and function of *fuchi*/*Pt-GATAL2b* and *Pt-GATA5* orthologs in the
645 early embryos of other spider species enable us to further address the problems on the
646 early developmental variations among the Araneae lineages.
647 This study has produced genome-wide datasets related to molecular and
648 genetic regulation of early embryonic development in *P. tepidariorum*, which can be
649 integrated with single-nucleus and single-cell RNA-seq datasets established in another
650 study [60]. These data resources, as well as the research strategies presented, will
651 facilitate our efforts to dissect the early developmental processes at molecular
652 resolution and establish an independent paradigm of genetic programs for early
653 arthropod development. Importantly, many of the genes revealed in this study encode
654 proteins not yet characterized, thus underscoring the importance of taking genome-
655 based approaches to studying the evolution of animal development.
656
657 **Conclusions**
658 Using a Theridiidae spider empowered by genome sequencing, we combined
659 comparative transcriptomes of isolated cells from different regions of early embryos
660 and a pRNAi-based functional screen of genes. This research strategy allowed us to
661 identify a lineage-specific, fast-evolving GATA-like gene, *fuchi*, that is essential for
662 early embryonic development in the spider. *fuchi* is expressed in future endodermal cell
663 poupulations in the early embryo. Further genome-wide analyses suggest that *fuchi*
664 regulates chromatin state and zygotic gene activation to promote endoderm
665 specification and pattern formation. The presented datasets have rich information about
666 the molecular regulation of early embryonic development in the Theridiidae spider.
667 The presence of many uncharacterized genes under the control of *fuchi* has been
668 revealed. Our genome-based research using a chelicerate arthropod phylogentically
669 distant from *Drosophila* provides the foundation for molecular and genetic exploration
670 of the variations in early development across Arthropoda.
671

672 **Materials and Methods**

673 **Spiders**

674 Laboratory stocks of *P. tepidariorum* were maintained at 25 °C in 16 h light/8 h dark
675 cycles. Developmental stages of the embryo were determined according to the previous
676 descriptions [28,58]. To determine the precise time of egg laying, female behaviors
677 were recorded every 5 min using a trail camera. The animal experimentation described
678 was approved by the institutional committee for animal care and use (No. 2020-1) and
679 conducted according to JT Biohistory Research Hall Regulation on Animal
680 Experimentation.

681

682 **Cell isolation**

683 Embryos at stage 3 were dechorionated with 50% commercial bleach for 1 min and
684 rinsed several times with distilled water. After removing water, the embryos were
685 placed on a glass slide with double-sticky tape and immediately covered with
686 halocarbon oil 700 (Sigma-Aldrich). Glass capillaries (2-000-075; Drummond) were
687 prepared in advance by pulling with a puller (PN-3; Narishige). The capillary tip was
688 broken using forceps, just before use, and filled with lysis buffer supplied in Dynabeads
689 mRNA DIRECT Kit (Ambion) by capillary action. Upon the manual manipulation of
690 the capillary under the stereomicroscope, approximately 10–30 cells were sucked from
691 a central, an intermediate, or a peripheral region of the stage-3 embryo and transferred
692 into 0.2 ml tubes containing 10 µl of lysis buffer. The samples were stored at -80 °C
693 until they were retrieved for RNA extraction.

694

695 **RNA-seq of isolated cells**

696 Poly(A) mRNA was extracted using the Dynabeads mRNA DIRECT Kit (Ambion).
697 After fragmentation of the extracted RNA for 5 min at 94 °C, first-strand cDNA
698 synthesis and subsequent cDNA amplification by 22 cycles of polymerase chain
699 reaction (PCR) were performed using the SMARTer stranded RNA-seq kit (Takara).
700 The synthesized first-strand cDNA and amplified cDNA were purified using Ampure
701 XP Beads (Beckman Coulter). Constructed libraries for sequencing were quantified
702 using the Agilent 2100 BioAnalyzer high sensitivity DNA kit (Agilent). Two or three
703 libraries for each region of the stage-3 embryo were sequenced in single-end runs in

704 the antisense direction using the MiSeq reagent kit V3 (150 cycles) on the Illumina
705 MiSeq platform. Over 10 million reads were obtained from each library. Following
706 essentially the same procedure, RNA-seq of cells from central and peripheral regions
707 of stage 4 and 5 embryos was performed.
708
709 **Processing of RNA-seq data**
710 The adapter and primer sequences, as well as the first three bases derived from the
711 SMARTer Stranded Oligo, were trimmed from the MiSeq raw reads using CLC
712 Genomics Workbench 7.0.3 (Qiagen). Quality trimming was also performed at the
713 following parameter settings: quality score, limit = 0.05; trim ambiguous nucleotides,
714 maximum number of ambiguities = 2; and filter on length, discard reads below length
715 = 30. Trimmed reads were aligned to the *Parasteatoda tepidariorum* genome (Ptep_1.0,
716 GCA_000365465.1) using the BLAT algorithm [86] in the DNA Data Bank of Japan
717 (DDBJ) Read Annotation Pipeline (https://p.ddbj.nig.ac.jp/pipeline/). Output
718 alignments were filtered using the PERL script filterPSL.pl, which is accessible from
719 the AUGUSTUS 3.0.1 scripts folder (http://github.com/Gaius-Augustus/Augustus),
720 with the following settings: 60% coverage; 90% identity; uniqueness threshold 0.96.
721 The filtered alignments (pslx-format) were converted to sam-format. Based on these
722 sam-format data, the number of mapped reads was counted against the AUGUSTUS
723 gene models (aug3.1; https://i5k.nal.usda.gov/content/data-downloads)[31]) using
724 htseq-count v.0.6.1p1 [87] with default settings. The RNA-seq datasets, including the
725 gene count matrix, were deposited in the NCBI Gene Expression Omnibus (GEO)
726 database (GSE193511).
727
728 **Comparative analyses of RNA-seq gene count datasets**
729 To assess the consistency of RNA-seq gene count datasets between biological
730 replicates of each sample group, the coefficients of determination were calculated using
731 the gene count matrix, and confirmed to be exceed 0.85. Out of the 27,990 annotated
732 genes, 10,862 genes showed one or more counts per million (cpm) for at least one of
733 the eight samples (p1-p3, i1-i3, c1, and c2). Count data matrices on these genes (10,862
734 genes × 8 samples) were used for comparative analyses. Differentially expressed genes
735 (DEGs) were identified using a Bioconductor package edgeR (version 3.8.6) [49]. The

736 following three types of comparisons were set up: comparison I, c1 and c2 vs. p1-p3
737 and i1-i3; comparison II, p1-p3 vs. c1, c2, and i1-i3; and comparison III, i1-i3 vs. c1,
738 c2, and p1-p3. Two criteria were applied to determine the high-priority candidate
739 genes: first, $\log_2 FC < -10$ in comparisons I and II, and $\log_2 FC > 10$ in comparison III;
740 and second, lower FDR values.
741
742 **cDNA cloning**
743 Full-length or partial cDNAs of newly identified genes were obtained from our
744 collections of expressed sequence tag clones previously described [32,39], or were
745 isolated by PCR amplification. cDNA fragments cloned in pBluescript II SK(+)
746 (Agilent), pTriplEx2 (Takara) or pZL1 (Invitrogen) were used to synthesize dsRNA for
747 pRNAi and/or RNA probes for in situ hybridization. The primers and cDNA clones
748 used are summarized in Table S16 (Additional file 12).
749
750 **Parental RNAi**
751 For dsRNA synthesis, DNA templates with the T7 promoter sequence at both ends were
752 prepared by PCR using appropriate primers depending on the plasmid type used.
753 dsRNAs were synthesized using a Megascript Kit (Ambion) as described previously
754 [29]. One to two µl of the dsRNA solution at a concentration of 2.0 µg/µl was injected
755 into the abdomen of adult females four times at 2–3 day-intervals using pulled glass
756 capillary tubes. In the pilot pRNAi screen, dsRNA against each gene was injected into
757 two or four females. When egg laying took place, more than 10 eggs were randomly
758 selected from each egg sac and transferred into halocarbon oil 700, which allowed us
759 to nondestructively monitor the embryonic development through cleared chorion.
760 For three genes (*g26874*, *g7720*, *g4238*) that were judged as positive in the
761 pilot screen, 2 or 3 additional dsRNAs were prepared from non-overlapping regions of
762 each of the cDNAs, and each dsRNA was injected into at least 2 females. Development
763 of eggs laid by the injected females were monitored using time-lapse recording, or with
764 occasional visual inspection, under the stereo microscope. *gfp* dsRNA was used as the
765 control.
766
767 **Time-lapse microscopy of live embryos**

768 Embryos dechorionated with 50 % commercial bleach were placed on a glass slide with
769 double-sticky tape and covered with halocarbon oil 700. When the observation was
770 started from stage 1, the dechorionation step was omitted to prevent possible effect on
771 the development. Images were taken every 5 or 10 min using a stereo microscope
772 (SZX12, Olympus) equipped with a color CCD camera (C7780-10, Hamamatsu
773 Photonics) controlled by AquaCosmos software (Hamamatsu Photonics) or other stereo
774 microscopes (M165C, Leica; or SZX7, Olympus) equipped with a color CMOS camera
775 (WRAYCAM-G200 or -NF300, WRAYMER).
776 For live imaging of nuclei around the germ-disc/non-germ-disc transition area,
777 dechorionated stage-4 embryos were placed on a glass slide with double-sticky tape,
778 and an embryo appropriately oriented was selected and skewered with a thin glass
779 needle to prevent the embryo from rotating during later development. A 1,000× stock
780 solution of a cell-permeable dye SPY555-DNA (Spirochrome), which was prepared in
781 50 µl of dimethylsulfoxide, was microinjected into the perivitelline space of the
782 skewered egg. The injecting volume should be low to avoid affecting the development
783 of the embryo from being affected. Approximately 1 hour after dye injection, the
784 embryo started to be examined under an Olympus BX50 fluorescence microscope
785 equipped with a cooled CCD camera (CoolSNAP HQ, Roper Scientific) controlled by
786 MetaMorph.version 6.1 (Universal Imaging). Using 10 × objective lens (UPlanAPO
787 10×/0.40, Olympus), bright-field and fluorescence 500 × 500 pixel images were
788 acquired every 5 min at exposure times of 10 msec and 100 msec, respectively. To
789 reduce the excitation light, two neutral density filters (U-ND6-2 and U-ND25-2,
790 Olympus) were used. A single focal plane, which was manually adjusted during
791 approximately 30 h of observation, was recorded at each time point. The observed
792 embryo was checked for its normal development. Time-lapse images were analyzed
793 using Imaris version 7.6.5, where cells were manually tracked, and ImageJ version
794 1.51d.
795
796 **Molecular phylogenetic analysis of GATA family members**
797 To collect amino acid sequences containing GATA-type DNA binding domain from
798 many arachnid species and other arthropod and nonarthropod metazoan representative
799 species, we used nucleotide sequence resources presented in Table S7 (Additional file

800 7), which included coding sequences from sequenced genomes, transcriptome
801 assemblies, and RNA-seq raw data, retrieved from GenBank (NCBI) or GigaScience
802 database. RNA-seq of *Hasarius adansoni* mid-stage embryos was performed in this
803 study, as described below. The RNA-seq data were subjected to de novo assembling
804 using Trinity (version 2.11.0). To identify amino acid sequences coding for GATA-
805 type DNA binding domains in the retrieved or assembled nucleotide sequences, blastp
806 searches were performed using the Fuchi amino acid sequence as a query with a cut-
807 off e-value of 1e-2, followed by the systematic selection of sequences that contained at
808 least one GATA-type zinc-finger motif sequence ($CX_2CX_{16\text{-}20}CX_2C$). The identified
809 amino acid sequences were manually aligned and classified based on the detection of
810 signature sequences, excluding GATA-like proteins that had only one GATA-type zinc
811 finger motif. For molecular phylogenetic analysis, 75 amino acid sites from 118
812 sequences, which were unambiguously aligned, were used (Additional file 7: Table S8).
813 The molecular phylogeny was inferred using the Maximum Likelihood (ML) method
814 and JTT matrix-based model [88] on the software MEGA11 [89].
815
816 **RNA-seq of whole embryos**
817 For RNA-seq analyses of *fuchi* pRNAi embryos, mated adult females were injected
818 with approximately 1–2 µl of *fuchi* dsRNA solution (2 µg/µl) four times at 2- to 3-day
819 intervals. For RNA extraction, 50–100 embryos at stage 2, 3, and early stage 5 that
820 were derived from egg sacs produced by the injected females 2 days before (untreated)
821 and 18–19 days after the first injection of the dsRNA were used. Two biological
822 replicates were prepared for each sample type. The mRNA extraction and library
823 construction for sequencing were performed as previously described [36]. The libraries
824 were quantified using the Library quantification kit (Takara) and Thermal cycler Dice
825 Realtime TP800 (Takara) and sequenced in single-end runs (150 cycles) in the
826 antisense direction on the Illumina MiSeq platform. The sequence reads were processed
827 as previously described [36], and the datasets were deposited in the GEO database
828 (GSE193650). Estimates of DEGs between untreated and *fuchi* pRNAi embryos at each
829 stage were performed using EdgeR (version 3.8.6).
830 For the *H. adansoni* RNA-seq, poly(A) mRNA was extracted using a
831 QuickPrep Micro mRNA purification kit (GE Healthcare) from 20 sibling mid-stage

832 embryos with extending limbs. Library construction for sequencing was performed as
833 previously described [36]. The library was sequenced in a paired-end run (150 × 2
834 cycles) on the Illumina Miseq platform, and the raw reads (SRR17326784) and the *H.*
835 *adansoni* transcriptome assembly (GJQJ00000000) were deposited in GenBank.
836
837 **ATAC-seq of whole embryos**
838 ATAC-seq was performed following the procedure described by Buenrostro et al.
839 (2015) [90] with modifications considering the procedure reported by Haines and
840 Eisen (2018) [91]. In brief, 40 whole embryos at stage 3 (18 h AEL) were
841 homogenized using disposable pellet pestles (12-141-368; Fisher Scientific) and 1.5
842 mL Eppendorf tubes (0030125150) in 50 µl lysis buffer (10mM Tris, 10mM NaCl,
843 3mM $MgCl_2$, adjusted to pH 7.4 with HCl) containing 0.15 mM spermidine, after
844 which 50 µl of lysis buffer and 1 µl of 10% Nonidet P-40 were added as the pestle
845 was rinsed. After a 10 min incubation, nuclei were spun down at 800 g and
846 resuspended in 17.5 µl water, followed by the addition of 25 µl of 2× TD buffer
847 (Illumina) and 7.5 µl of Nextra Tn5 transposase (Illumina). Transposed DNA was
848 purified using MinElute Reaction Cleanup Kit (Qiagen) and amplified with Illumina
849 Nextera Transposase Adaptors index PCR primers, and the amplified DNA was
850 purified using Ampure XP Beads (Beckman Coulter). Constructed libraries were
851 paired-end sequenced (151 bp reads × 2) on the Illumina HiSeq X Ten system.
852 Comparative analyses of ATAC-seq datasets from three different sample
853 types (wild-type, *fuchi* pRNAi, and *Pt-hh* pRNAi) were performed basically
854 following the procedure described by Reske et al. (2020)[92] with slight
855 modifications based on guidelines provided by Gaspar (2019)[93] and Delisle et al.
856 (2021)[94]. Two biological replicates were obtained from each sample type. ATAC-
857 seq raw reads were trimmed for adaptor and primer sequences and low-quality
858 sequences using Trim Galore! with default settings. Trimmed reads were then aligned
859 to the *P. tepidariorum* genome (Ptep_2.0) using bowtie2 v2.2.5 with default settings.
860 Reads mapped with low quality scores (less than 30) were removed using samtools
861 (version 1.9). PCR duplicates were removed using Picard MarkDuplicates (version
862 2.23.4). The alignments of filtered reads to the genome were saved as bam files. Peak
863 calling was performed using the ATAC-seq mode of Genrich (version 6.0).

Differential accessibility analysis was performed using a Bioconductor package, *csaw* [95]. Peak regions called in both replicates of each sample type were extracted, and the peak regions extracted from the three sample types were merged. Reads were counted in peaks, with low abundance peaks (log CPM < -3) removed. Read counts were subject to normalization with the trimmed mean of *M* values (TMM) method [96]. Differential accessibility was estimated using the glmQLFit and glmQLFTest functions in EdgeR (version 3.32.0). Peaks with FDR < 0.05 were considered as significant. To discover motif sequences in the identified differential peak regions, STREME [97] was used.

Wild-type whole embryos at stages 1 (9 h AEL) and 2 (12 h AEL) were similarly analyzed by ATAC-seq to detect genomic regions differentially accessible between the different stages (stages 1 to 3), but no clear results were obtained. All the ATAC-seq datasets were deposited in the GEO database (GSE193870).

Staining of embryos

Antisense RNA probes for in situ hybridization were prepared by *in vitro* transcription using Digoxigenin- or Dinitrophenyl-conjugated UTP (DIG, Roche 11277073910; DNP, PerkinElmer NEL555001EA) as previously described [56]. Whole-mount in situ hybridization with chromogenic substrate was performed in the previously described manner [29] with the exception of an additional tyramide-based signal amplification (TSA) process described previously [56]. In brief, after washing DIG-conjugated RNA probes, the samples were incubated with anti-DIG-POD antibody (11633716001, Roche; 1:500 dilution) and then reacted with DNP-tyramide (NEL747, PerkinElmer). In the detection step, anti-DNP-AP antibody (MB-3100, Vector; 1:50) and NBT/BCIP were used. Most samples were counterstained with 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich).

Double-color FISH was performed as previously described [56]. To visualize the DIG- and DNP-conjugated probes, anti-DIG-POD (Roche 11633716001; 1:500) and anti-DNP-HRP (PerkinElmer FP1129; 1:100) were used in combination with DyLight488-tyramide and DyLight680-tyramide, respectively.

Single- and double-color FISH were combined with immunostaining for β-catenin in certain cases (Figs. 5 and 6), where Proteinase K treatment prior to

896 hybridization was omitted. After completing the FISH procedure, the embryos were
897 incubated with rabbit anti-β-catenin antibody (C2206, Sigma-Aldrich 1:500) and then
898 with Cy3-conjugated anti-rabbit IgG antibody (AP187C, Chemicon; 1:200), followed
899 by counterstaining with DAPI.
900
901 **Image acquisition and processing**
902 Images for chromogenically stained WISH samples were obtained using a stereo
903 microscope (SZX12, Olympus), a fluorescence unit BT-ExSMOP (Biotools), and a
904 CCD camera (C7780-10, Hamamatsu Photonics). For confocal microscopy of
905 fluorescently stained embryos, the whole embryos were individually mounted on glass
906 slides with spacers of approximately 100–500 μm, which were adjusted depending on
907 the sample using sticky tape (CT-18S, Nichiban) and silicone vacuum grease (335148,
908 Beckman). To view different sides of a single embryo, it was rotated after being
909 mounted on the glass slide. The embryos were examined using a Leica TCS SPE
910 confocal system equipped with four laser sources (405, 488, 532, and 635 nm), which
911 were used to excite DAPI, DyLight488, Cy3, and DyLight680, respectively. Acquired
912 confocal stacks were processed and analyzed using Imaris version 7.6.5 (Bitplane) and
913 ImageJ version 1.51d. The Imaris snapshot feature and oblique slicer tool were used to
914 obtain images capturing the embryonic regions of interest. Linear signals of cell-cell
915 adherens junctions visualized by β-catenin staining were traced in Adobe Photoshop
916 CS5 ver. 12.0.4.
917
918 **Availability of data and materials**
919 Raw and processed data for the comparative RNA-seq and ATAC-seq analyses are
920 accessible from NCBI GEO (GSE193511, GSE193650, and GSE193870). The data
921 aligned to the reference genome can be viewed on our web database
922 (https://www.brh2.jp). The nucleotide sequences determined for *fuchi* and *Pt-GATA5*
923 are available under the accessions LC671429 and LC671430 in the
924 DDBJ/EMBL/GenBank International Nucleotide Sequence Databases. The sources of
925 sequence data used in the molecular phylogenetic analysis are listed in Additional file
926 7: Tables S7 and S8. Supplementary movies showing the pilot pRNAi screen and the
927 pRNAi specificity validation are available on figshare [50].

928

929 **Abbreviations**

930 ***AEL:*** After egg laying

931 ***ATAC-seq:*** Assay for transposase-accessible chromatin using sequencing

932 ***BR:*** Basic region

933 ***CPM:*** Counts per million

934 ***DAPI:*** 4',6-diamidino-2-phenylindole

935 ***DDBJ:*** DNA Data Bank of Japan

936 ***DEG:*** Differentially expressed gene

937 ***dsRNA:*** Double stranded RNA

938 ***FC:*** Fold change

939 ***FDR:*** False discovery rate

940 ***GEO:*** Gene Expression Omnibus

941 ***ML:*** Maximum likelihood

942 ***NCBI:*** National Center for Biotechnology Informataion

943 ***PCR:*** Polymerase chain reaction

944 ***pRNAi:*** Parental RNA interference

945 ***RNA-seq:*** RNA sequencing

946 ***RPKM:*** reads per kilobase of transcript, per million mapped reads

947 ***WISH:*** Whole-mount in situ hybridization

948 ***ZF:*** Zinc finger

949

950 **Acknowledgements**

951 We thank Akiko Noda for technical assistance, and other members of JT Biohistory

952 Research Hall for discussion.

953

954 **Funding**

955 This work was supported by JSPS Kakenhi (26440130, 17K07418, 20K06676 to YA;

956 15K07139 to HO), JST PRESTO (JPMJPR2041 to YA), and the core budget of JT

957 Biohistory Research Hall.

958

959 **Contributions**

960 S.I., Y.A., and H.O. designed the study. S.I. created all the RNA-seq datasets. S.I.
961 analyzed the RNA-seq datasets with help from Y.A. S.I. performed the pRNAi screen
962 and the gene expression and phenotype analyses. S.I. and H.O. performed the
963 phylogenetic analysis. R.N. created the ATAC-seq datasets. R.N. analyzed the ATAC-
964 seq datasets with help from Y.A. and H.O. S.I. wrote the first draft of the manuscript.
965 H.O. wrote the final version of the manuscript with input from all authors.
966
967 **Ethics approval and consent to participate**
968 Not applicable
969
970 **Consent for publication**
971 Not applicable
972
973 **Competing interests**
974 The authors declare that they have no competing interests.
975
976 **Supplementary Information**
977 **Additional file 1: Movie S1.** Time-lapse video of sibling *Parasteatoda tepidariorum*
978 embryos. The left embryo is the same as shown in Figure 1. Time after egg laying
979 (AEL) is indicated. Scale bar, 100 µm.
980 **Additional file 2: Table S1.** List of DEG candidates identified by comparative
981 transcriptome analysis of c versus i/p cells from stage-3 embryo (comparison I). **Table**
982 **S2.** List of DEG candidates identified by comparative transcriptome analysis of p
983 versus c/i cells from stage-3 embryo (comparison II)**. Table S3.** List of DEG candidates
984 identified by comparative transcriptome analysis of i versus c/p cells from stage-3
985 embryo (comparison III)**. Table S4.** List of DEG candidates indentified by comparative
986 transcriptome analysis of c versus p cells from stage-4 embryo**. Table S5.** List of DEG
987 candidates identified by comparative transcriptome analysis of c versus p cells from
988 early stage-5 embryo.
989 **Additional file 3: Figure S1.** Chromogenic WISH of stage-3 embryos using probes for
990 19 selected DEG candidates. **Figure S2.** Validation of the specificity of the RNAi
991 effects for *g26874*, *g7720*, and *g4238*. **Figure S3.** Developmental transcript profiling

992 of *P. tepidariorum* GATA family genes in wild-type embryos. **Figure S4.** Cell counting
993 in a stage-3 embryo stained for *fuchi*. **Figure S5.** *fuchi* expression in pEND cells.
994 **Figure S6.** Characterization of 5 DEG candidates from the comparative transcriptome
995 analysis of stage-2 *fuchi* pRNAi versus untreated embryos. **Figure S7.** Characterization
996 of the top-10 DEG candidates from the comparative transcriptome analysis of stage-3
997 *fuchi* pRNAi versus untreated embryos. **Figure S8.** Effects of *fuchi* RNAi on
998 expression of selected genes. **Figure S9.** Single-nucleus transcriptome data showing
999 endodermal and mesodermal cell populations in late stage-5 embryos. **Figure S10.**
1000 Single-nucleus and single-cell transcriptome data showing cell populations expressing
1001 the genes regulated by *fuchi*.
1002 **Additional file 4: Table S6.** Summary of a pilot pRNAi screen of 19 selected genes.
1003 **Additional file 5: Movie S2.** Time-lapse video showing phenotypes *gfp*, *g26874*,
1004 *g7720*, and *g4238* pRNAi embryos. Time after egg laying (AEL) and time after start of
1005 recording are indicated. Movie starts are adjusted by the timing of blastoderm
1006 formation. Some of the embryos are the same as those shown in Figure 3A. Scale bar,
1007 100 μm.
1008 **Additional file 6: Movie S3.** Tracking of cells in wild-type and *g26874* pRNAi
1009 embryos. In the *g26874* pRNAi embryo, cells on the abembryonic side initially shift
1010 toward forming a germ disc but, later, they return back. Time after start of recording is
1011 indicated.
1012 **Additional file 7: Table S7.** List of bioinformatic resources used for phylogenetic
1013 characterization of Fuchi and other GATA family members in spiders. **Table S8.** Amino
1014 acid sequence alignment and classification of GATA family members from spiders and
1015 other metazoans.
1016 **Additional file 8: Movie S4.** Time-lapse video showing cell internalizations taking
1017 place around the germ-disc rim. Nuclei labeled by SPY555-DNA are presented in white.
1018 The germ disc on the embryonic side is to the lower right, while the non-germ-disc
1019 region on the abembryonic side is to the upper left. Internalizing cells are tracked. Cells
1020 from the non-germ-disc region are moving to below the germ-disc epithelium. Later,
1021 cells at the germ-disc rim internalize. Time [d : h : m] after start of recording is indicated.
1022 **Additional file 9: Table S9.** List of DEGs identified by comparative transcriptome
1023 analysis of *fuchi* pRNAi versus untreated embryos at stage 2. **Table S10.** List of DEGs

1024 identified by comparative transcriptome analysis of *fuchi* pRNAi versus untreated
1025 embryos at stage 3. **Table S11.** List of DEGs identified by comparative transcriptome
1026 analysis of *fuchi* pRNAi versus untreated embryos at early stage 5.
1027 **Additional file 10: Table S12.** Data from comparative analysis of read counts in
1028 extracted ATAC-seq peaks between *fuchi* pRNAi and wild-type embryos at stage 3
1029 using edgeR. **Table 13.** List of differential ATAC-seq peaks between *fuchi* pRNAi and
1030 wild-type embryos at stage 3 (FDR < 0.05).
1031 **Additional file 11: Table S14.** Data from comparative analysis of read counts in
1032 extracted ATAC-seq peaks between *Pt-hh* pRNAi and wild-type embryos at stage 3
1033 using edgeR. **Table S15.** List of differential ATAC-seq peaks between Pt-hh pRNAi
1034 and wild-type embryos at stage 3 (FDR < 0.05).
1035 **Additional file 12: Table S16.** Primers and cDNA clones used for RNAi and in situ
1036 hybridization.
1037

1038 **References**

1039 1. Raff RA. The Shape of Life: Genes, Development, and the Evolution of Animal Form. Chicago: Univ.
1040 Chicago Press; 1996.

1041 2. Duboule D. Temporal colinearity and the phylotypic progression: a basis for the stability of a
1042 vertebrate Bauplan and the evolution of morphologies through heterochrony. Dev Suppl. 1994;135–42.

1043 3. Galis F, van Dooren TJM, Metz JAJ. Conservation of the segmented germband stage: robustness or
1044 pleiotropy? Trends Genet. 2002;18:504–9.

1045 4. Sander K. The evolution of patterning mechanisms: gleanings from insect embryogenesis and
1046 spermatogenesis. In: Goodwin BC Hold N Wylie CC. (eds) Development and evolution. Cambridge:
1047 Cambridge University Press; 1983. p. 137–60.

1048 5. Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, et al. Gene expression
1049 divergence recapitulates the developmental hourglass model. Nature. 2010;468:811–4.

1050 6. Richardson MK. Heterochrony and the Phylotypic Period. Dev Biol. 1995;172:412–21.

1051 7. Irie N, Kuratani S. Comparative transcriptome analysis reveals vertebrate phylotypic period during
1052 organogenesis. Nat Commun. 2011;2:248.

8. Domazet-Loso T, Tautz D. An Evolutionary Analysis of Orphan Genes in Drosophila. Genome Res. 2003;13:2213–9.

9. Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. More than just orphans: are taxonomically-restricted genes important in evolution? Trends in Genetics. 2009;25:404–13.

10. Thomas GWC, Dohmen E, Hughes DST, Murali SC, Poelchau M, Glastad K, et al. Gene content evolution in the arthropods. Genome Biol. 2020;21:15.

11. Morino Y, Hashimoto N, Wada H. Expansion of TALE homeobox genes and the evolution of spiralian development. Nat Ecol Evol. 2017;1:1942–9.

12. Holland LZ, Albalat R, Azumi K, Benito-Gutiérrez È, Blow MJ, Bronner-Fraser M, et al. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res. 2008;18:1100–11.

13. Khalturin K, Anton-Erxleben F, Sassmann S, Wittlieb J, Hemmrich G, Bosch TCG. A Novel Gene Family Controls Species-Specific Morphological Traits in Hydra. PLOS Biol. Public Library of Science; 2008;6:e278.

14. Stauber M, J\textbackslashHackle H, Schmidt-Ott U. The anterior determinant bicoid of Drosophila is a derived Hox class 3 gene. Proc Natl Acad Sci U A. 1999;96:3786–9.

15. Lynch JA, Peel AD, Drechsler A, Averof M, Roth S. EGF signaling and the origin of axial polarity among the insects. Curr Biol. 2010;20:1042–7.

16. Lima LF, Torres AQ, Jardim R, Mesquita RD, Schama R. Evolution of Toll, Spatzle and MyD88 in insects: the problem of the Diptera bias. BMC Genomics. 2021;22:562.

17. Nishida H, Sawada K. macho-1 encodes a localized mRNA in ascidian eggs that specifies muscle fate during embryogenesis. Nature. 2001;409:724–9.

18. Fekany K, Yamanaka Y, Leung T, Sirotkin HI, Topczewski J, Gates MA, et al. The zebrafish bozozok locus encodes Dharma, a homeodomain protein essential for induction of gastrula organizer and dorsoanterior embryonic structures. Development. 1999;126:1427–38.

1078 19. Bae S, Reid CD, Kessler DS. Siamois and Twin are redundant and essential in formation of the
1079 Spemann organizer. Dev Biol. 2011;352:367–81.

1080 20. Pultz MA, Zimmerman KK, Alto NM, Kaeberlein M, Lange SK, Pitt JN, et al. A Genetic Screen for
1081 Zygotic Embryonic Lethal Mutations Affecting Cuticular Morphology in the Wasp Nasonia vitripennis.
1082 Genetics. 2000;154:1213–29.

1083 21. Sulston IA, Anderson KV. Embryonic patterning mutants of Tribolium castaneum. Development.
1084 1996;122:805–14.

1085 22. Maderspacher F, Bucher G, Klingler M. Pair-rule and gap gene mutants in the flour beetle Tribolium
1086 castaneum. Dev Genes Evol. 1998;208:558–68.

1087 23. Schmitt-Engel C, Schultheis D, Schwirz J, Ströhlein N, Troelenberg N, Majumdar U, et al. The
1088 iBeetle large-scale RNAi screen reveals gene functions for insect development and physiology. Nat
1089 Commun. 2015;6:7822.

1090 24. Consortium i5K. The i5K Initiative: advancing arthropod genomics for knowledge, human health,
1091 agriculture, and the environment. J Hered. 2013;104:595–600.

1092 25. Oda H, Akiyama-Oda Y. The common house spider Parasteatoda tepidariorum. EvoDevo. 2020;11:6.

1093 26. Kanayama M, Akiyama-Oda Y, Oda H. Early embryonic development in the spider Achaearanea
1094 tepidariorum: microinjection verifies that cellularization is complete before the blastoderm stage.
1095 Arthropod Struct Dev. 2010;39:436–45.

1096 27. Suzuki H, Kondo A. Early embryonic development, including germ-disk stage, in the Theridiid
1097 spider Achaearanea japonica (Bös. et Str.). J Morphol. 1995;224:147–57.

1098 28. Akiyama-Oda Y, Oda H. Early patterning of the spider embryo: a cluster of mesenchymal cells at
1099 the cumulus produces Dpp signals received by germ disc epithelial cells. Development. 2003;130:1735–
1100 47.

1101 29. Akiyama-Oda Y, Oda H. Axis specification in the spider embryo: dpp is required for radial-to-axial
1102 symmetry transformation and sog for ventral patterning. Development. 2006;133:2347–57.

1103 30. Akiyama-Oda Y, Oda H. Cell migration that orients the dorsoventral axis is coordinated with
1104 anteroposterior patterning mediated by Hedgehog signaling in the early spider embryo. Development.
1105 2010;137:1263–73.

1106 31. Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, et al. The house spider
1107 genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biol. 2017;15:62.

1108 32. Oda H, Nishimura O, Hirao Y, Tarui H, Agata K, Akiyama-Oda Y. Progressive activation of Delta-
1109 Notch signaling from around the blastopore is required to set up a functional caudal lobe in the spider
1110 Achaearanea tepidariorum. Development. 2007;134:2195–205.

1111 33. Posnien N, Zeng V, Schwager EE, Pechmann M, Hilbrant M, Keefe JD, et al. A comprehensive
1112 reference transcriptome resource for the common house spider Parasteatoda tepidariorum. PLoS One.
1113 2014;9:e104885.

1114 34. Kono N, Nakamura H, Ito Y, Tomita M, Arakawa K. Evaluation of the impact of RNA preservation
1115 methods of spiders for de novo transcriptome assembly. Mol Ecol Resour. 2016;16:662–72.

1116 35. Sasaki M, Akiyama-Oda Y, Oda H. Evolutionary origin of type IV classical cadherins in arthropods.
1117 BMC Evol Biol. 2017;17:142.

1118 36. Iwasaki-Yokozawa S, Akiyama-Oda Y, Oda H. Genome-scale embryonic developmental profile of
1119 gene expression in the common house spider Parasteatoda tepidariorum. Data Brief. 2018;19:865–7.

1120 37. Pechmann M. Formation of the germ-disc in spider embryos by a condensation-like mechanism.
1121 Front Zool. 2016;13:35.

1122 38. Hemmi N, Akiyama-Oda Y, Fujimoto K, Oda H. A quantitative study of the diversity of stripe-
1123 forming processes in an arthropod cell-based field undergoing axis formation and growth. Dev Biol.
1124 2018;437:84–104.

1125 39. Kanayama M, Akiyama-Oda Y, Nishimura O, Tarui H, Agata K, Oda H. Travelling and splitting of
1126 a wave of hedgehog expression involved in spider-head segmentation. Nat Commun. 2011;2:500.

1127 40. Akiyama-Oda Y, Oda H. Hedgehog signaling controls segmentation dynamics and diversity via
1128 msx1in a spider embryo. Sci Adv. 2020;6.

41. Montgomery J TH. The development of Theridium, an Aranead, up to the stage of reversion. J Morphol. 1909;20:297–352.

42. Chaw RC, Vance E, Black SD. Gastrulation in the spider Zygiella x-notata involves three distinct phases of cell internalization. Dev Dyn. 2007;236:3484–95.

43. Mittmann B, Wolff C. Embryonic development and staging of the cobweb spider Parasteatoda tepidariorum C. L. Koch, 1841 (syn.: Achaearanea tepidariorum; Araneomorphae; Theridiidae). Dev Genes Evol. 2012;222:189–216.

44. Edgar A, Bates C, Larkin K, Black S. Gastrulation occurs in multiple phases at two distinct sites in Latrodectus and Cheiracanthium spiders. Evodevo. 2015;6:33.

45. Wolff C, Hilbrant M. The embryonic development of the central American wandering spider Cupiennius salei. Front Zool. 2011;8:15.

46. Turetzek N, Prpic N-MM. Observations on germ band development in the cellar spider Pholcus phalangioides. Dev Genes Evol. 2016;226:413–22.

47. Pechmann M. Embryonic development and secondary axis induction in the Brazilian white knee tarantula Acanthoscurria geniculata, C. L. Koch, 1841 (Araneae; Mygalomorphae; Theraphosidae). Dev Genes Evol. 2020;230:75–94.

48. Sekiguchi K. Reduplication in spider eggs produced by centrifugation. Sci Rep Tokyo Bunrika Daigaku Sec B. 1957;8:227–80.

49. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

50. Iwasaki-Yokozawa S, Akiyama-Oda Y, Oda H. Supplementary movies for “Lineage-specific, fast-evolving GATA-like gene regulates zygotic gene activation to promote endoderm specification and pattern formation in the Theridiidae spider” [Internet]. figshare. 2022 [cited 2022 Jun 8]. Available from: https://doi.org/10.6084/m9.figshare.c.5963661.v1

51. Su Y, Song Y, Wang Y, Jessop L, Zhan L, Stumph WE. Characterization of a Drosophila Proximal-Sequence-Element-Binding Protein Involved in Transcription of Small Nuclear RNA Genes. Eur J Biochem. 1997;248:231–7.

1156 52. Chen T, Bunting M, Karim FD, Thummel CS. Isolation and characterization of five Drosophila genes
1157 that encode an ets-related DNA binding domain. Dev Biol. 1992;151:176–91.

1158 53. Pechmann M, Benton MA, Kenny NJ, Posnien N, Roth S. A novel role for Ets4in axis specification
1159 and cell migration in the spider Parasteatoda tepidariorum. Elife. 2017;6.

1160 54. Tang Y, Wei Y, He W, Wang Y, Zhong J, Qin C. GATA transcription factors in vertebrates:
1161 evolutionary, structural and functional interplay. Mol Genet Genomics. 2014;289:203–14.

1162 55. Patient RK, McGhee JD. The GATA family (vertebrates and invertebrates). Curr Opin Genet Dev.
1163 2002;12:416–22.

1164 56. Akiyama-Oda Y, Oda H. Multi-color FISH facilitates analysis of cell-type diversification and
1165 developmental gene regulation in the Parasteatoda spider embryo. Dev Growth Differ. 2016;58:215–24.

1166 57. Oda H, Akiyama-Oda Y. Microarray data on the comparison of transcript expression between normal
1167 and Pt-DeltaRNAi embryos in the common house spider Parasteatoda tepidariorum. Data Brief.
1168 2019;25:104350.

1169 58. Yamazaki K, Akiyama-Oda Y, Oda H. Expression patterns of a twist-related gene in embryos of the
1170 spider Achaearanea tepidariorum reveal divergent aspects of mesoderm development in the fly and
1171 spider. Zool Sci. 2005;22:177–85.

1172 59. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. Genes Dev.
1173 2014;28:2679–92.

1174 60. Akiyama-Oda Y, Akaiwa T, Oda H. Reconstruction of the Global Polarity of an Early Spider Embryo
1175 by Single-Cell and Single-Nucleus Transcriptome Analysis. submitted.

1176 61. Rebagliati MR, Weeks DL, Harvey RP, Melton DA. Identification and cloning of localized maternal
1177 RNAs from xenopus eggs. Cell. 1985;42:769–77.

1178 62. Sasai Y, Lu B, Steinbeisser H, Geissert D, Gont LK, De Robertis EM. Xenopus chordin: a novel
1179 dorsalizing factor activated by organizer-specific homeobox genes. Cell. 1994;79:779–90.

1180 63. Kozhemyakina E, Ionescu A, Lassar AB. GATA6 is a crucial regulator of Shh in the limb bud. PLoS
1181 Genet. 2014;10:e1004072.

1182 64. Tokusumi Y, Tokusumi T, Stoller-Conrad J, Schulz RA. Serpent, Suppressor of Hairless and U-
1183 shaped are crucial regulators of hedgehog niche expression and prohemocyte maintenance during
1184 Drosophila larval hematopoiesis. Development. 2010;137:3561–8.

1185 65. Xuan S, Sussel L. GATA4 and GATA6 regulate pancreatic endoderm identity through inhibition of
1186 hedgehog signaling. Development. 2016;143:780–6.

1187 66. Fujikura J, Yamato E, Yonemura S, Hosoda K, Masui S, Nakao K, et al. Differentiation of embryonic
1188 stem cells is induced by GATA factors. Genes Dev. 2002;16:784–9.

1189 67. Tremblay M, Sanchez-Ferras O, Bouchard M. GATA transcription factors in development and
1190 disease. Development. 2018;145:dev164384.

1191 68. Gillis WQ, Bowerman BA, Schneider SQ. The evolution of protostome GATA factors: molecular
1192 phylogenetics, synteny, and intron/exon structure reveal orthologous relationships. BMC Evol Biol.
1193 2008;8:112.

1194 69. Reuter R. The gene serpent has homeotic properties and specifies endoderm versus ectoderm within
1195 the Drosophila gut. Development. 1994;120:1123–35.

1196 70. Morrisey EE, Tang Z, Sigrist K, Lu MM, Jiang F, Ip HS, et al. GATA6 regulates HNF4 and is
1197 required for differentiation of visceral endoderm in the mouse embryo. Genes Dev. 1998;12:3579–90.

1198 71. Shivdasani RA. Molecular Regulation of Vertebrate Early Endoderm Development. Dev Biol.
1199 2002;249:191–203.

1200 72. Tam PP, Kanai-Azuma M, Kanai Y. Early endoderm development in vertebrates: lineage
1201 differentiation and morphogenetic function. Curr Opin Genet Dev. 2003;13:393–400.

1202 73. Fukuda K, Kikuchi Y. Endoderm development in vertebrates: fate mapping, induction and regional
1203 specification. Dev Growth Differ. 2005;47:343–55.

1204 74. Frankenberg S, Gerbe F, Bessonnard S, Belville C, Pouchin P, Bardot O, et al. Primitive Endoderm
1205 Differentiates via a Three-Step Mechanism Involving Nanog and RTK Signaling. Dev Cell.
1206 2011;21:1005–13.

1207 75. Heslop JA, Pournasr B, Liu J-T, Duncan SA. GATA6 defines endoderm fate by controlling chromatin
1208 accessibility during differentiation of human-induced pluripotent stem cells. Cell Rep. 2021;35:109145.

1209 76. Feitosa NM, Pechmann M, Schwager EE, Tobias-Santos V, McGregor AP, Damen WGM, et al.
1210 Molecular control of gut formation in the spider Parasteatoda tepidariorum. Genesis. 2017;55.

1211 77. Maduro MF, Rothman JH. Making Worm Guts: The Gene Regulatory Network of the Caenorhabditis
1212 elegans Endoderm. Dev Biol. 2002;246:68–85.

1213 78. Murakami R, Okumura T, Uchiyama H. GATA factors as key regulatory molecules in the
1214 development of Drosophila endoderm. Dev Growth Differ. 2005;47:581–9.

1215 79. Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, et al. FOG, a Multitype Zinc Finger
1216 Protein, Acts as a Cofactor for Transcription Factor GATA-1 in Erythroid and Megakaryocytic
1217 Differentiation. Cell. 1997;90:109–19.

1218 80. Crispino JD, Lodish MB, MacKay JP, Orkin SH. Use of Altered Specificity Mutants to Probe a
1219 Specific Protein–Protein Interaction in Differentiation: the GATA-1:FOG Complex. Mol Cell.
1220 1999;3:219–28.

1221 81. Waltzer, Lucas, Bataille, Laetitia, Peyrefitte, Sandrine, Haenlin, Marc. Two isoforms of Serpent
1222 containing either one or two GATA zinc fingers have different roles in Drosophila haematopoiesis.
1223 EMBO J. 2002;21:5477–86.

1224 82. Chlon TM, Doré LC, Crispino JD. Cofactor-mediated restriction of GATA-1 chromatin occupancy
1225 coordinates lineage-specific gene expression. Mol Cell. 2012;47:608–21.

1226 83. Pedone, Paolo, V., Omichinski, James, G., Nony, Pascale, Trainor, Cecelia, Gronenborn, Angela, M.,
1227 Clore, G.Marius, et al. The N-terminal fingers of chicken GATA-2 and GATA–3 are independent
1228 sequence-specific DNA binding domains. EMBO J. 1997;16:2874–82.

1229 84. Oda H, Iwasaki-Yokozawa S, Usui T, Akiyama-Oda Y. Experimental duplication of bilaterian body
1230 axes in spider embryos: Holm's organizer and self-regulation of embryonic fields. Dev Genes Evol.
1231 2020;230:49–63.

1232 85. Holm Å. Experimentelle Untersuchungen über die Entwicklung und Entwicklungsphysiologie des
1233 Spinnenembryos. Zool BiDr Upps. 1952;29:293–424.

1234 86. Kent WJ. BLAT—The BLAST-Like Alignment Tool. Genome Res. 2002;12:656–64.

1235 87. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing
1236 data. Bioinformatics. 2015;31:166–9.

1237 88. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein
1238 sequences. Bioinformatics. 1992;8:275–82.

1239 89. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11.
1240 Mol Biol Evol. 2021;38:3022–7.

1241 90. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin
1242 Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109:21.29.1-9.

1243 91. Haines JE, Eisen MB. Patterns of chromatin accessibility along the anterior-posterior axis in the
1244 early Drosophila embryo. PLoS Genet. 2018;14:e1007367.

1245 92. Reske JJ, Wilson MR, Chandler RL. ATAC-seq normalization method can significantly affect
1246 differential accessibility analysis and interpretation. Epigenetics Chromatin. 2020;13:22.

1247 93. Gaspar JM. ATAC-seq Guidelines [Internet]. Harv. FAS Inform. 2019 [cited 2022 Apr 26]. Available
1248 from: https://informatics.fas.harvard.edu/./atac-seq-guidelines.html

1249 94. Delisle, L., Doyle, M., Heyl, F. Galaxy Training: ATAC-Seq data analysis [Internet]. 2021 [cited
1250 2022 Apr 26]. Available from: https://training.galaxyproject.org/training-
1251 material/topics/epigenetics/tutorials/atac-seq/tutorial.html

1252 95. Lun ATL, Smyth GK. csaw: a Bioconductor package for differential binding analysis of ChIP-seq
1253 data using sliding windows. Nucleic Acids Res. 2016;44:e45.

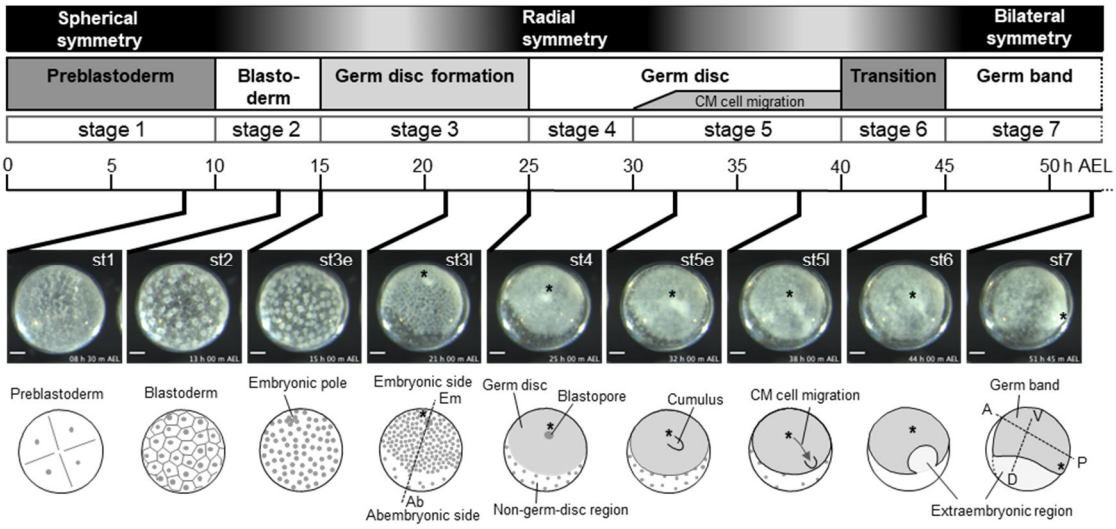
1254 96. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of
1255 RNA-seq data. Genome Biol. 2010;11:R25.

1256 97. Bailey TL. STREME: accurate and versatile sequence motif discovery. Bioinformatics.
1257 2021;37:2834–40.

98. Fernández R, Kallal RJ, Dimitrov D, Ballesteros JA, Arnedo MA, Giribet G, et al. Phylogenomics, Diversification Dynamics, and Comparative Transcriptomics across the Spider Tree of Life. Curr Biol. 2018;28:1489-1497.e5.

99. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.

100. Akiyama-Oda Y, Akaiwa T, Oda H. Supplementary Data for "Reconstruction of the Global Polarity of an Early Spider Embryo by Single-Cell and Single-Nucleus Transcriptome Analysis." figshare; 2022 [cited 2022 Jun 8]; Available from: https://doi.org/10.6084/m9.figshare.c.5963571.v1



1268

1269 **Fig. 1. Early embryonic processes of *Parasteatoda tepidariorum*.** Selected time-lapse
1270 images of a live wild-type embryo (Additional file 1: Movie S1) are shown along with
1271 their timelines (time [h] after egg laying [AEL]), stages of development, and
1272 illustrations of morphological characteristics. The morphology of the developing *P.*
1273 *tepidariorum* embryo undergoes symmetry transitions as shown (top). An axis of radial
1274 symmetry (embryonic-abembryonic axis) and two axes of bilateral symmetry (anterior-
1275 posterior [A-P] and dorsal-ventral [D-V] axes) are indicated by broken lines. Asterisks
1276 mark cells around the embryonic pole, cells at the closed blastopore, and cells at the
1277 posterior pole, which are in the same lineage. Scale bars,100 µm.
1278

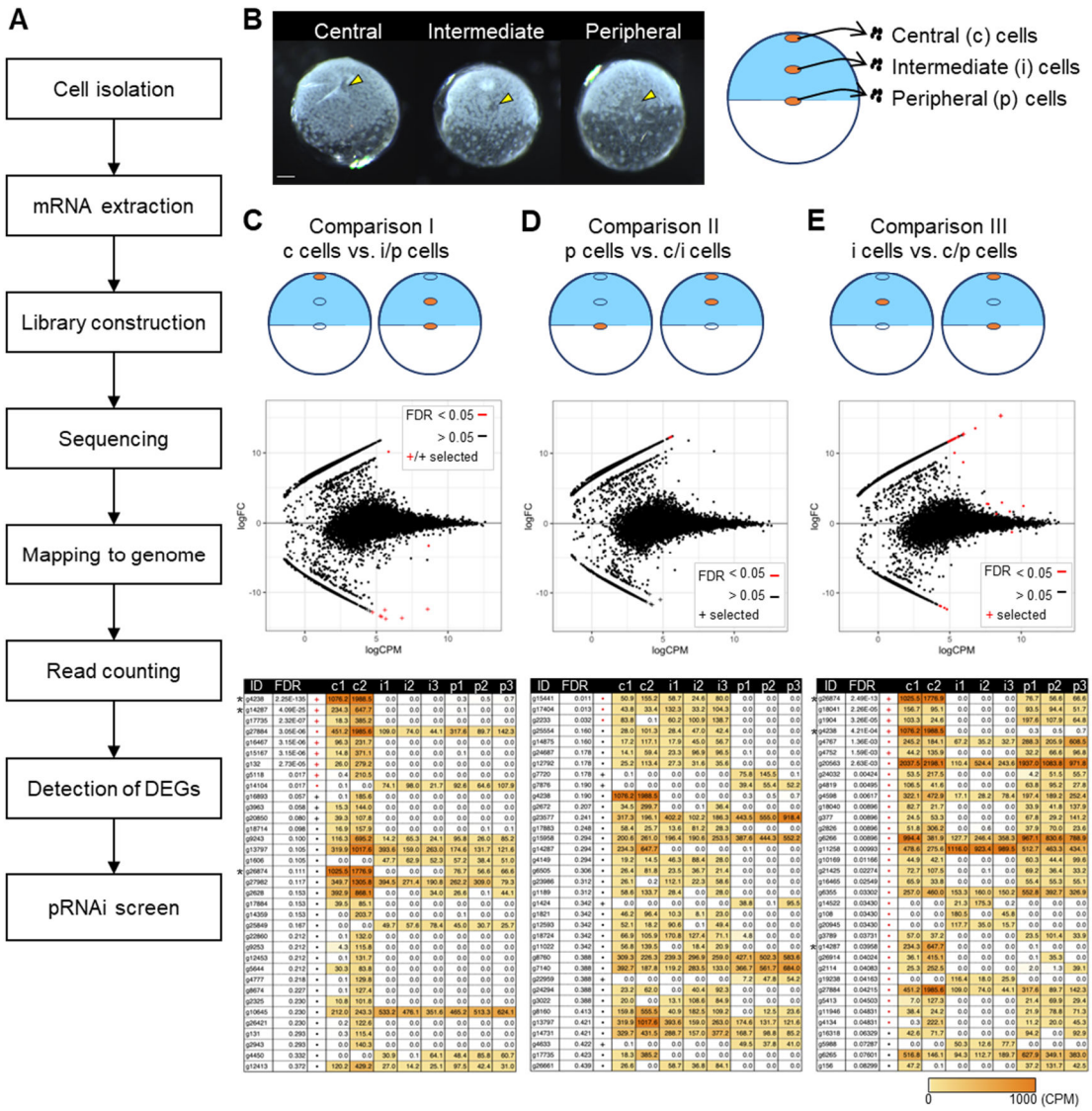
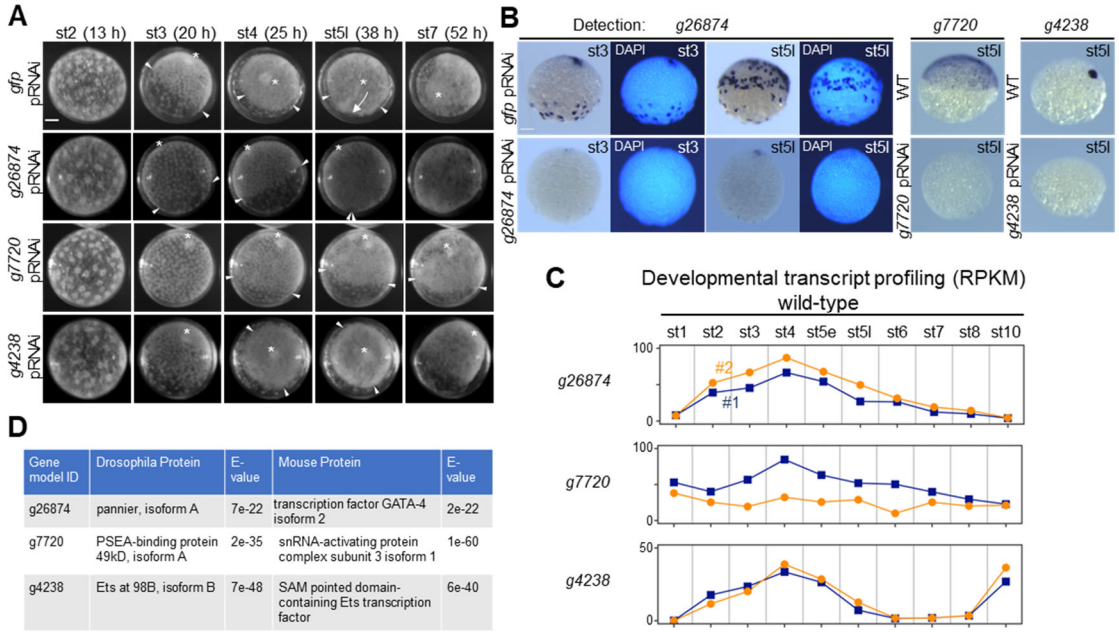


Fig. 2. Genome-wide identification of candidate genes with locally restricted expression in stage-3 embryo. (A) Outline of the experimental procedure. **(B)** Images showing live stage-3 embryos following isolation of cells from 3 different regions (yellow arrowheads; central [c], intermediate [i], and peripheral [p]) of the embryo, along with the schematic showing the isolated cells used for RNA-seq. Scale bar, 100 µm. **(C–E)** Three types of comparison using RNA-seq datasets from the c, i, and p cells (comparison I: c cells versus i/p cells [C]; comparison II: p cells versus c/i cells [D]; comparison III: i cells versus c/p cells [E]) detected candidates of differentially expressed genes (DEGs). Top: schematic showing the grouping of the datasets for comparison. Middle: MA-plot of $\log_2$ fold-change (FC) versus $\log_2$ average expression level (CPM; count per million) from 10,862 genes. Bottom: lists of DEG candidates

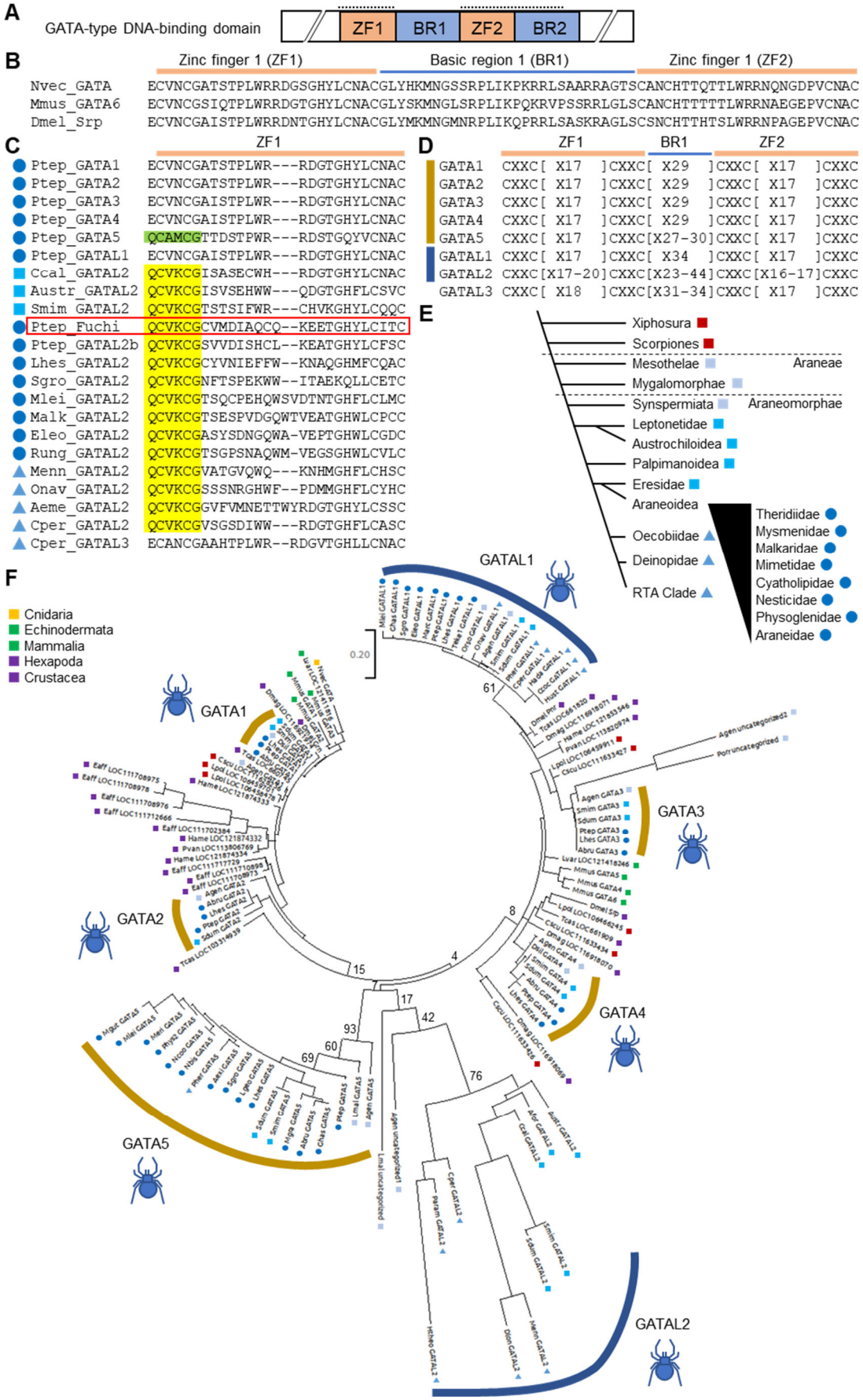
1291 with normalized expression levels (CPM) from biological replicates of the three sample
1292 types (c1, c2, i1, i2, i3, p1, p2, and p3), which are sorted by FDR values. Genes with
1293 FDR < 0.05 are highlighted in red, and candidates of DEGs that were selected for a
1294 pilot pRNAi screen are indicated by plus signs in the MA-plots and tables. Each table
1295 displays the top 35 genes; the full lists are presented in Additional file 2: Tables S1–
1296 S3. Note that genes marked by asterisks appear in multiple tables.
1297



D

Gene model ID	Drosophila Protein	E-value	Mouse Protein	E-value
g26874	pannier, isoform A	7e-22	transcription factor GATA-4 isoform 2	2e-22
g7720	PSEA-binding protein 49kD, isoform A	2e-35	snRNA-activating protein complex subunit 3 isoform 1	1e-60
g4238	Ets at 98B, isoform B	7e-48	SAM pointed domain-containing Ets transcription factor	6e-40

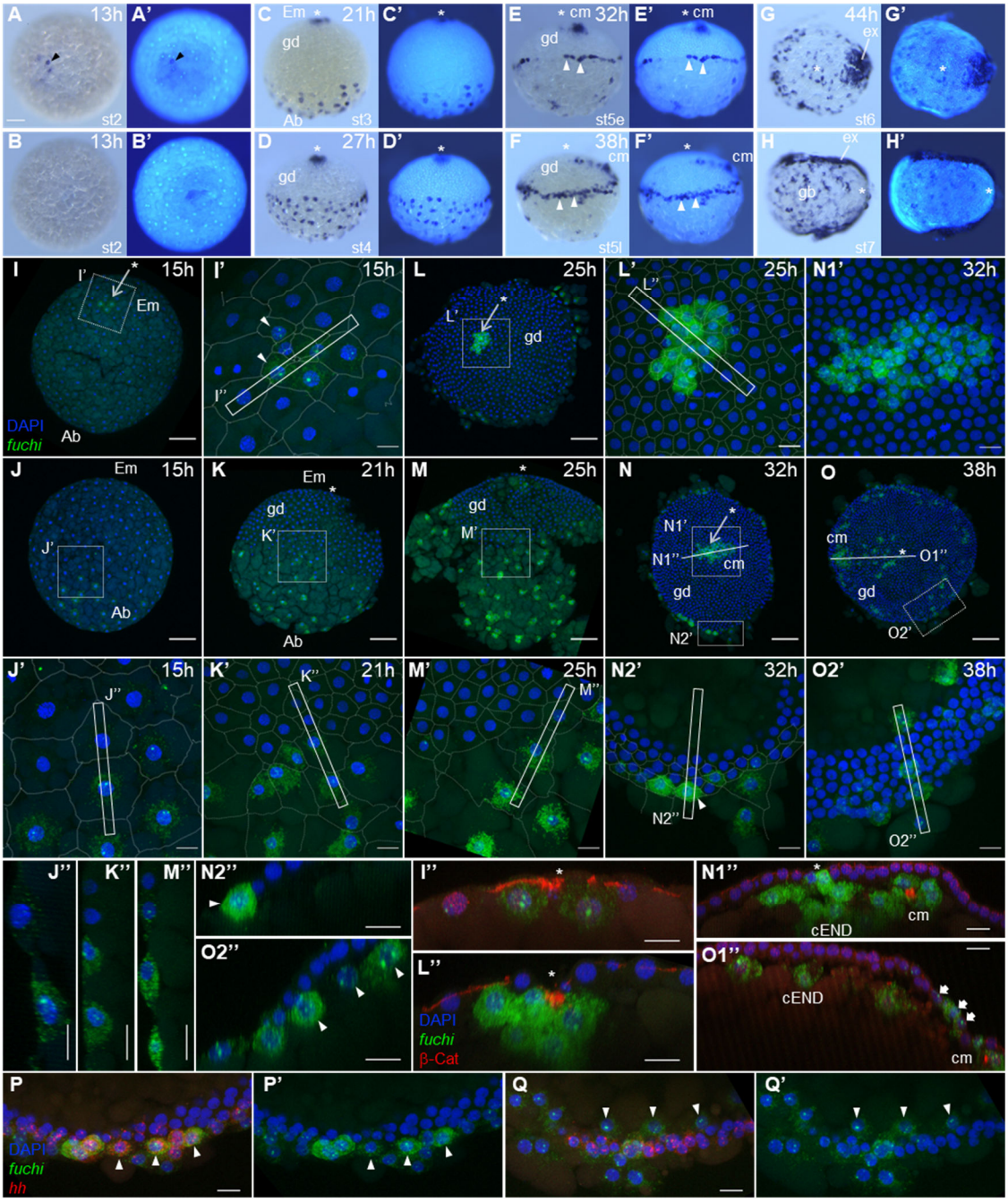
Fig. 3. Identification of three genes in a pilot pRNAi screen of the DEG candidates. **(A)** Images showing the development of live *gfp* pRNAi, *g26874* pRNAi, *g7720* pRNAi, and *g4238* pRNAi embryos. These images are related to Movie S2 (Additional file 4). Stages of development and time (h) after egg laying (AEL) are shown. Asterisks indicate the embryonic pole, closed blastopore, and posterior pole. Arrowheads indicate the rim of the forming/formed germ disc, or corresponding sites in morphologically affected embryos. The arrow indicates CM cell migration, which did not occur in the *g26874* pRNAi, *g7720* pRNAi, and *g4238* RNAi embryos. **(B)** Effect of *g26874*, *g7720*, or *g4238* pRNAi on the expression of the corresponding target gene at stage 3 and/or late stage 5, as revealed by WISH. *gfp* pRNAi or wild-type (WT) embryos were stained as controls. Counterstains with DAPI were displayed for the comparison between *g26874* and *gfp* pRNAi embryos. **(C)** Graphs showing developmental profiling of the transcript levels (RPKM; reads per kilobase of transcript, per million mapped reads) for *g26874*, *g7720*, and *g4238* in wild-type embryos, based on public datasets [36]. Two biological replicates are individually shown. **(D)** Results of blast searches against *D. melanogaster* and mouse RefSeq protein databases using predicted amino acid sequences of *g26874*, *g7720*, and *g4238* as queries. The top-hit protein from each search is shown. Scale bars, 100 µm in A and B.


A
GATA-type DNA-binding domain
ZF1
BR1
ZF2
BR2
B
Zinc finger 1 (ZF1)
Basic region 1 (BR1)
Zinc finger 1 (ZF2)
Nvec_GATA ECVNCGATSTPLWRRDGSGHYLCNACGLYHKMNGSSRPLIKPKRRLSAARRAGTSCANCHTTQTTLWRRNQNGDPVCNAC
Mmus_GATA6 ECVNCGSIQTPLWRRDGTGHYLCNACGLYSKMNGLSRPLIKPQKRVPSSRRLGLSCANCHTTTTTLWRRNAEGEPVCNAC
Dmel_Srp ECVNCGAISTPLWRRDNTGHYLCNACGLYMKMNGMNRPLIKQPRRLSASKRAGLSCSNCHTTHTSLWRRNPAGEPVCNAC
C
ZF1
Ptep_GATA1 ECVNCGATSTPLWR---RDGTGHYLCNAC
Ptep_GATA2 ECVNCGATSTPLWR---RDGTGHYLCNAC
Ptep_GATA3 ECVNCGAISTPLWR---RDGTGHYLCNAC
Ptep_GATA4 ECVNCGAISTPLWR---RDGTGHYLCNAC
Ptep_GATA5 QCAMCGTTDSTPWR---RDSTGQYVCNAC
Ptep_GATAL1 ECVNCGAISTPLWR---RDGTGHYLCNAC
Ccal_GATAL2 QCVKCGISASECWH---RDGTGHYLCNAC
Austr_GATAL2 QCVKCGISVSEHWW---QDGTGHFLCSVC
Smim_GATAL2 QCVKCGTSTSIFWR---CHVKGHYLCQQC
Ptep_Fuchi QCVKCGCVMDIAQCQ--KEETGHYLCITC
Ptep_GATAL2b QCVKCGSVVDISHCL--KEATGHYLCFSC
Lhes_GATAL2 QCVKCGCYVNIEFFW--KNAQGHMFCQAC
Sgro_GATAL2 QCVKCGNFTSPEKWW--ITAEKQLLCETC
Mlei_GATAL2 QCVKCGTSQCPEHQWSVDTNTGHFLCLMC
Malk_GATAL2 QCVKCGTSESPVDGQWTVEATGHWLCPCC
Eleo_GATAL2 QCVKCGASYSDNGQWA-VEPTGHWLCGDC
Rung_GATAL2 QCVKCGTSGPSNAQWM-VEGSGHWLCVLC
Menn_GATAL2 QCVKCGVATGVQWQ---KNHMGHFLCHSC
Onav_GATAL2 QCVKCGSSSNRGHWF--PDMMGHFLCYHC
Aeme_GATAL2 QCVKCGGVFVMNETTWYRDGTGHYLCSSC
Cper_GATAL2 QCVKCGVSGSDIWW---RDGTGHFLCASC
Cper_GATAL3 ECANCGAAHTPLWR--RDGVTGHLLCNAC
D
ZF1
BR1
ZF2
GATA1 CXXC[X17]CXXC[X29]CXXC[X17]CXXC
GATA2 CXXC[X17]CXXC[X29]CXXC[X17]CXXC
GATA3 CXXC[X17]CXXC[X29]CXXC[X17]CXXC
GATA4 CXXC[X17]CXXC[X29]CXXC[X17]CXXC
GATA5 CXXC[X17]CXXC[X27-30]CXXC[X17]CXXC
GATAL1 CXXC[X17]CXXC[X34]CXXC[X17]CXXC
GATAL2 CXXC[X17-20]CXXC[X23-44]CXXC[X16-17]CXXC
GATAL3 CXXC[X18]CXXC[X31-34]CXXC[X17]CXXC
E
Xiphosura
Scorpiones
Mesothelae
Araneae
Mygalomorphae
Synspermiata
Araneomorphae
Leptonetidae
Austrochiloidea
Palpimanoidea
Eresidae
Araneoidea
Oecobiidae
Deinopidae
RTA Clade
Theridiidae
Mysmenidae
Malkaridae
Mimetidae
Cyatholipidae
Nesticidae
Physoglenidae
Araneidae
F
Cnidaria
Echinodermata
Mammalia
Hexapoda
Crustacea
GATAL1
GATA1
GATA3
GATA2
GATA4
GATA5
GATAL2
0.20
61
8
4
15
17
42
93
60
69
76

1319

1320 **Fig. 4. Phylogenetic characterization of Fuchi and other spider GATA family**
1321 **proteins. (A)** Schematic of GATA-type DNA-binding domain. This domain comprises
1322 two zinc finger motifs (ZF1 and ZF2) and two basic regions (BR1 and BR2). **(B)**
1323 Amino acid sequence alignment of the ZF1-BR1-ZF2 region (plus one adjacent residue
1324 on the N-terminal side) of three canonical GATA family members from cnidarian,
1325 mouse and *Drosophila* (Nvec_GATA, *Nematostella vectensis* GATA; Mmus_GATA6,
1326 *Mus musculus* GATA-6; Dmel_Srp, *Drosophila melanogaster* Serpent). **(C)** Amino acid
1327 sequence alignment of the ZF1 (plus one flanking residue on the N-terminal side) of
1328 various GATA family members from *P. tepidariorum* and other spider species. There
1329 are five canonical (Ptep_GATA1 to GATA5) and three noncanonical (Ptep_GATAL1,
1330 Fuchi, GATAL2b) GATA family members in *P. tepidariorum*. The GATAL2 group,
1331 including Fuchi, is characterized by the signature sequence highlighted in yellow, and
1332 the GATA5 group by the signature sequence highlighted in green. There are two types
1333 of GATAL2 proteins; the ZF1 of one type is aligned with that of the canonical GATA
1334 without gap but the ZF1 of the other type is not. **(D)** Classification of spider GATAs
1335 based on cysteine residue spacing patterns in the ZF1-BR1-ZF2 region. GATA5s in
1336 some spider species had a BR1 of slightly varied length. **(E)** Diagram showing
1337 phylogeny of spider taxa, based on a recent study [98]. **(F)** Maximum likelihood tree
1338 of canonical and non-canonical GATA family members from Araneae, non-Araneae
1339 chelicerates, non-chelicerate arthropods, non-arthropod bilaterians, and cnidarian,
1340 using 75 amino acid sites unambiguously alignable. GATAL2 proteins with imperfectly
1341 aligned ZF1, including Fuchi, were excluded from this analysis. Bootstrap values (100
1342 replicates) are presented at selected nodes. Color geometric codes indicate taxonomic
1343 groups of the sequence source. Details of species names and sequences are available in
1344 Tables S7 and S8 (Additional file 7).
1345
1346

1347

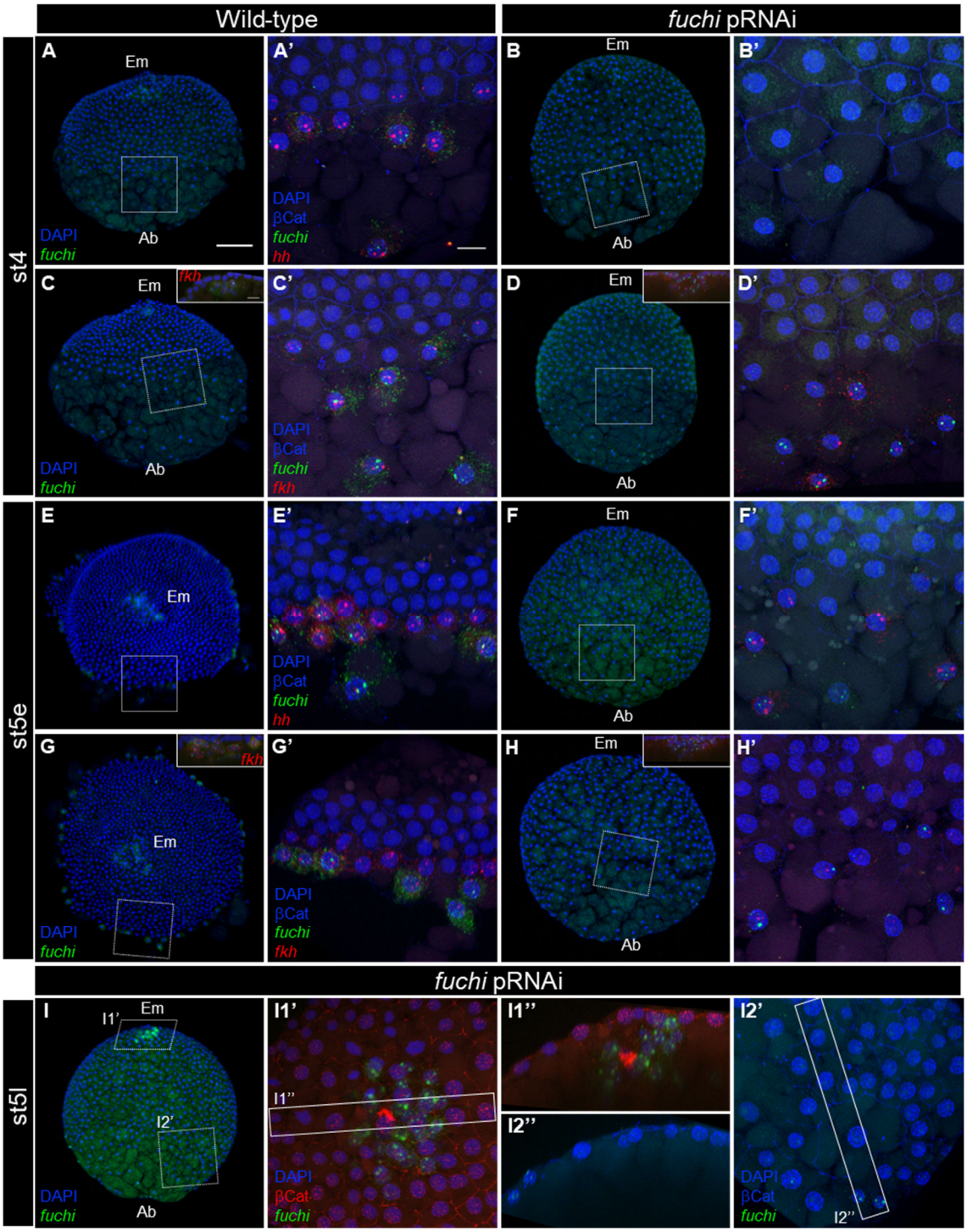


1348

1349 **Fig. 5. Expression of *fuchi* transcript in specific cells on the embryonic and**
1350 **abembryonic sides of early embryos. (A–H)** Chromogenic WISH of wild-type
1351 embryos at indicated stages using probes for *fuchi* transcript (left), counter-stained for
1352 DNA (right). **(A, B)** Opposite sides of the same embryo. Arrowhead indicates the signal.
1353 **(C–F)** Lateral view. The forming/formed germ disc (gd) is up. Arrowheads denote cells
1354 expressing *fuchi* at or near the germ-disc rim. **(G, H)** Forming germ band (gb) with
1355 expanding extraembryonic region (ex). Asterisks mark the embryonic pole, closed

1356 blastopore, or posterior pole. **(I–O)** Reconstruction of confocal stacks showing wild-
1357 type embryos fluorescently stained for *fuchi* transcript (green) and DNA (blue) at 15 (I,
1358 J), 21 (K), 25 (L, M), 32 (N), and 38 (O) h AEL. Asterisks mark the embryonic pole or
1359 closed blastopore. I and J show the same embryo from different directions. **(I'–M', N1',**
1360 **N2', O2')** High magnification of the boxed areas in I–O. Cell-cell junctions visualized
1361 by staining for β-catenin are traced in I'–M', and N2'. **(I"–M", N2", O2")** Optical
1362 sections corresponding to the rectangular regions in I'–M', N2', and O2'. Arrowheads
1363 indicate *fuchi*-expressing cells that appear to be internalizing or internalized. **(N1",**
1364 **O1")** Optical sections corresponding to the lines in N and O. Signals for β-catenin are
1365 overlaid in red in I", L", N1", and O1". β-Catenin concentrations mark the cumulus
1366 (cm). Fat arrows indicate *fuchi* expression in the future extraembryonic region. **(P, P',**
1367 **Q, Q')** Simultaneous detection of *fuchi* (green) and *Pt-hh* (red) transcripts and DNA
1368 (blue) in cells at or near the germ-disc rim at early (P, P') and late (Q, Q') stage 5.
1369 Arrowheads indicate cells expressing both *fuchi* and *Pt-hh* before internalization and
1370 cells expressing only *fuchi* after internalization. Em, embryonic side; Ab, abembryonic
1371 side. Scale bars, 100 µm in A, I–O; 20 µm in other panels.
1372
1373

1374

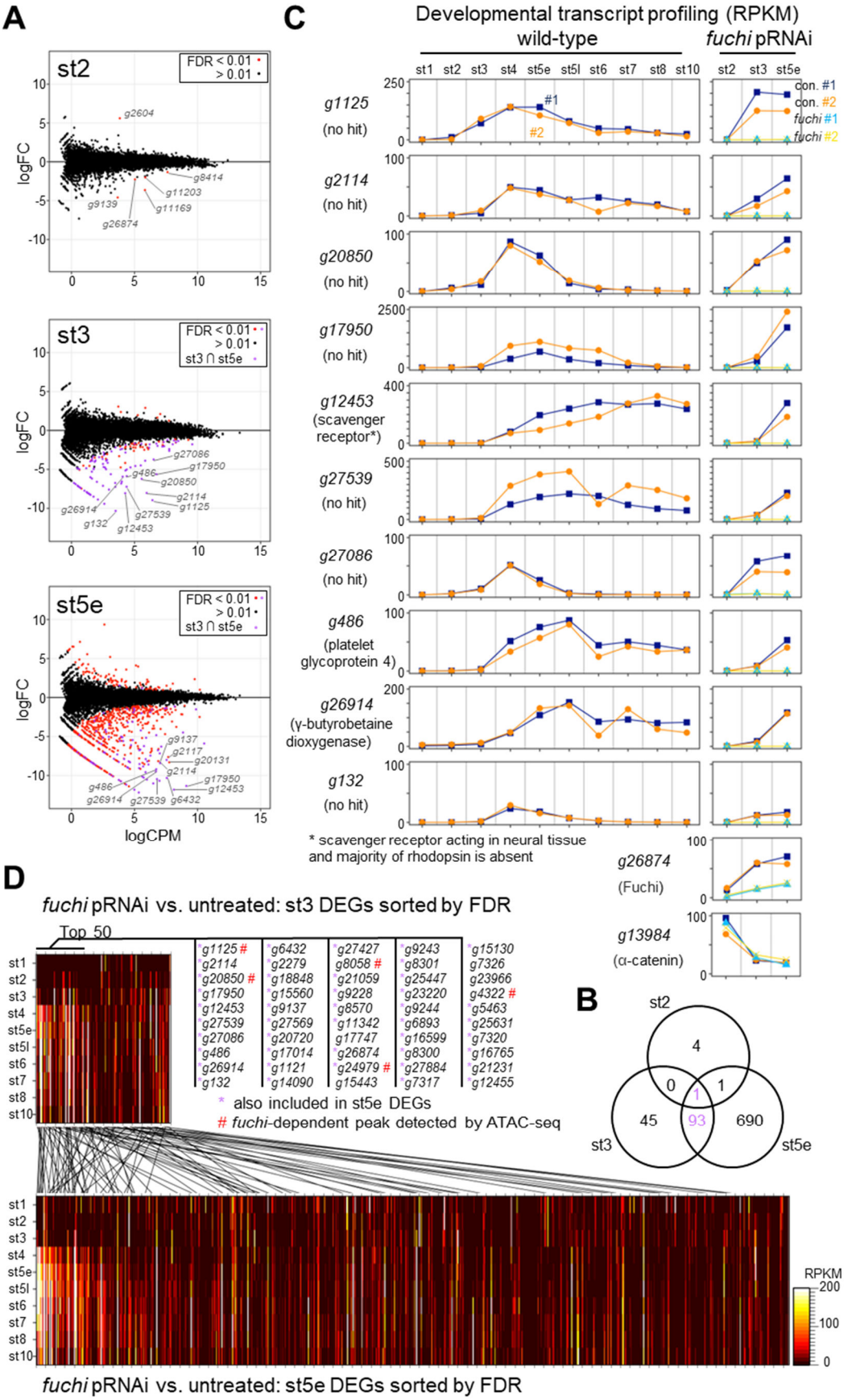


1375

1376 **Fig. 6. Defects of *fuchi* pRNAi embryos in demarcating a forming germ disc. (A–**
1377 **H)** Reconstructions of confocal stacks showing wild-type (A, C, E, G) and *fuchi* pRNAi
1378 (B, D, F, H) embryos stained for *fuchi* (green, A–H) and *Pt-hh* (red, A, B, E, F) or *Pt-*
1379 *fkh* (red, C, D, G, H) transcripts, DNA (blue, A–H), and β-catenin (blue, A'–H'; not
1380 shown in A-H) at stage 4 (A–D) and early stage 5 (E–H). In A–D, F, H, and I, the

1381 embryonic side (Em) is to top, with the abembryonic (Ab) to bottom. In E and G, the
1382 embryonic side is viewed from the top. Optical sections show the presence of specific
1383 cells expressing *fuchi* at the embryonic pole (insets, C, D, G, H). The boxed areas in
1384 A-H are magnified in A'-H', respectively. *fuchi* transcript was detectable even in cells
1385 on the abembryonic side of *fuchi* pRNAi embryos, although the signals were restricted
1386 to within the nuclei. Cells expressing *Pt-hh* and *Pt-fkh* transcripts at reduced levels
1387 were observed at regions corresponding to the the germ-disc rim and nearby
1388 abembryonic region in *fuchi* pRNAi embryos at early stage 5, but these cells did not
1389 serve as rim cells demarcating a forming germ disc. (I) Reconstructions of confocal
1390 stacks showing a *fuchi* pRNAi embryo stained for *fuchi* transcript (green, I, I1', I2'),
1391 DNA (blue, I, I1', I2'), and β-catenin (red, I1'; blue, I2') at late stage 5. (I1', I2') High
1392 magnification of the boxed areas in I as indicated. (I1", I2") Cross-sections of the
1393 rectangular regions in I1' and I2'. The cluster of CM cells were eventually internalized,
1394 but with few cells internalized in other regions. Note that the cells spread over the
1395 abembryonic side with epithelial integrity retained to a certain extent. Scale bars, 100
1396 μm in A; 20 μm in A', inset of C.
1397

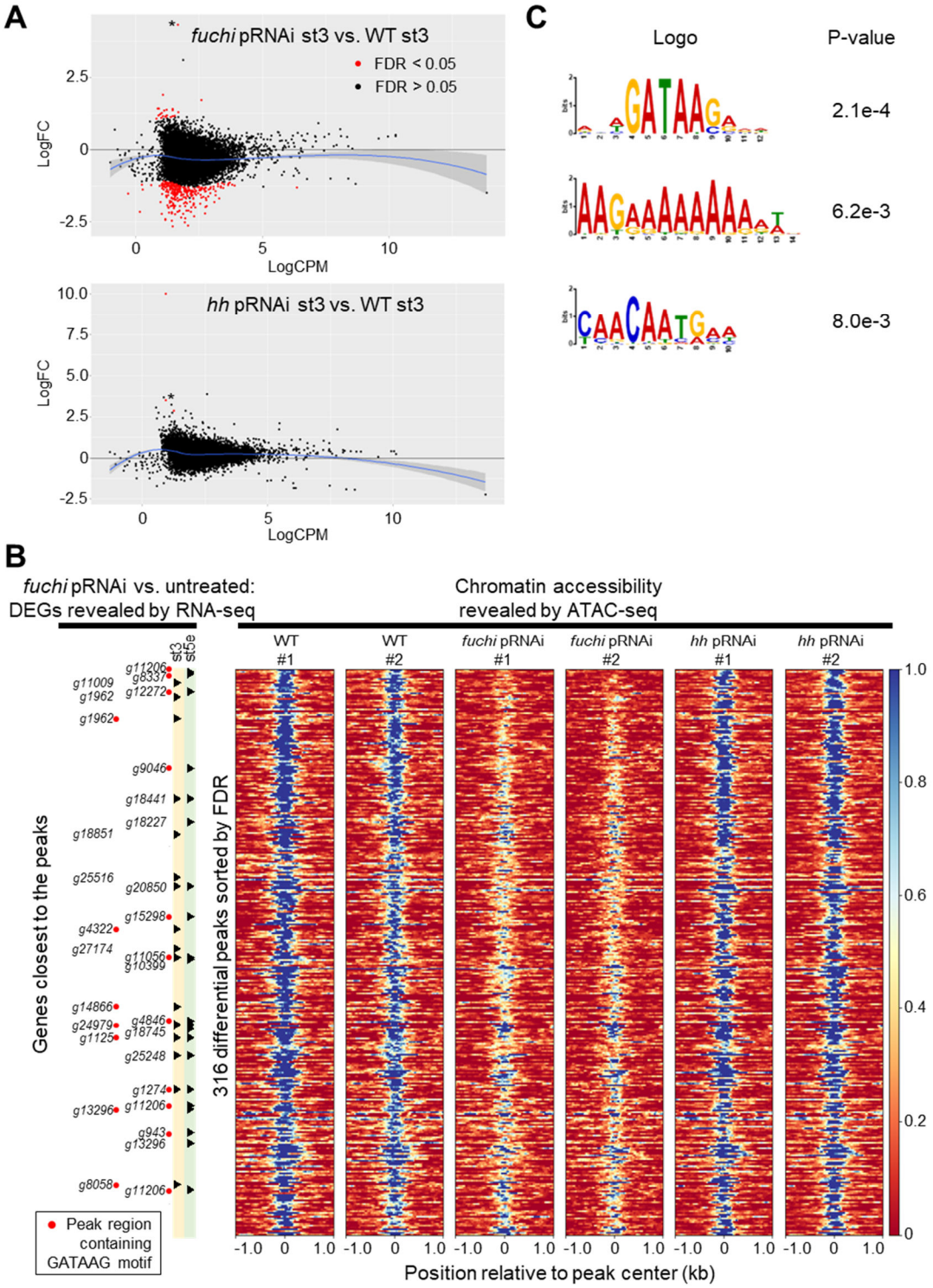
1398


A
st2
FDR < 0.01
> 0.01
logFC
g2604
g8414
g11203
g9139
g11169
g26874
st3
st3 ∩ st5e
g27086
g17950
g486
g20850
g2114
g1125
g26914
g27539
g132
g12453
st5e
g9137
g2117
g20131
g17950
g12453
g6432
logCPM
C
Developmental transcript profiling (RPKM)
wild-type
fuchi pRNAi
st1 st2 st3 st4 st5e st5l st6 st7 st8 st10
st2 st3 st5e
con. #1
con. #2
fuchi #1
fuchi #2
g1125
(no hit)
g2114
(no hit)
g20850
(no hit)
g17950
(no hit)
g12453
(scavenger receptor*)
g27539
(no hit)
g27086
(no hit)
g486
(platelet glycoprotein 4)
g26914
(γ-butyrobetaine dioxygenase)
g132
(no hit)
* scavenger receptor acting in neural tissue and majority of rhodopsin is absent
g26874
(Fuchi)
g13984
(α-catenin)
D
fuchi pRNAi vs. untreated: st3 DEGs sorted by FDR
Top 50
g1125 #
g2114
g20850 #
g17950
g12453
g27539
g27086
g486
g26914
g132
g6432
g2279
g18848
g15560
g9137
g27569
g20720
g17014
g1121
g14090
g27427
g8058 #
g21059
g9228
g8570
g11342
g17747
g26874
g24979 #
g15443
g9243
g8301
g25447
g23220
g9244
g6893
g16599
g8300
g27884
g7317
g15130
g7326
g23966
g4322 #
g5463
g25631
g7320
g16765
g21231
g12455
* also included in st5e DEGs
# fuchi-dependent peak detected by ATAC-seq
B
st2
st3
st5e
4
0
1
1
45
93
690
RPKM
200
100
0
fuchi pRNAi vs. untreated: st5e DEGs sorted by FDR

1399

1400 **Fig. 7. Genome-wide identification of genes whose expression levels are affected**
1401 **following *fuchi* knockdown. (A)** MA-plots of $\log_2$ fold-change (FC) versus $\log_2$
1402 average expression level (CPM) showing DEGs between *fuchi* pRNAi and control
1403 (untreated) embryos at stage 2 (st2), stage 3 (st3), and early stage 5 (st5e). Genes with
1404 FDR < 0.01 are highlighted in red or purple, with the purple genes included in both
1405 stage-3 and early stage-5 DEG lists. **(B)** Venn diagram showing the numbers of the
1406 DEGs from the three sample types and their overlaps. Note that *fuchi* (*g26874*) itself is
1407 included in all the DEG lists. **(C)** Graphs showing developmental profiling of the
1408 transcript levels for the top 10 DEGs from the stage-3 samples in wild-type embryos
1409 and the effects of *fuchi* pRNAi on the transcript levels. Top hit proteins that resulted
1410 from blastx searches against mouse and *Drosophila* RefSeq protein databases are
1411 shown, although with no hits obtained with seven of the ten genes (E-value < 1e-5).
1412 Data on *fuchi* (*g26874*) and α-catenin (*g13984*) validated the experiments. **(D)** Heat
1413 maps showing developmental profiling of the transcript levels for all the DEGs from
1414 the stage-3 (upper) and early stage-5 (lower) samples in wild-type embryos (sorted by
1415 FDR values). The top 50 DEGs from stage-3 are listed, where the genes included in
1416 stage-5 DEGs are indicated by purple asterisks. Genomic regions differentially
1417 accessible between wild-type and *fuchi* pRNAi embryos at stage 3 were found to be
1418 located within or close to some of the DEGs marked by "#"(related to Fig. 8 and
1419 Additional file 10: Table S13).
1420

1421

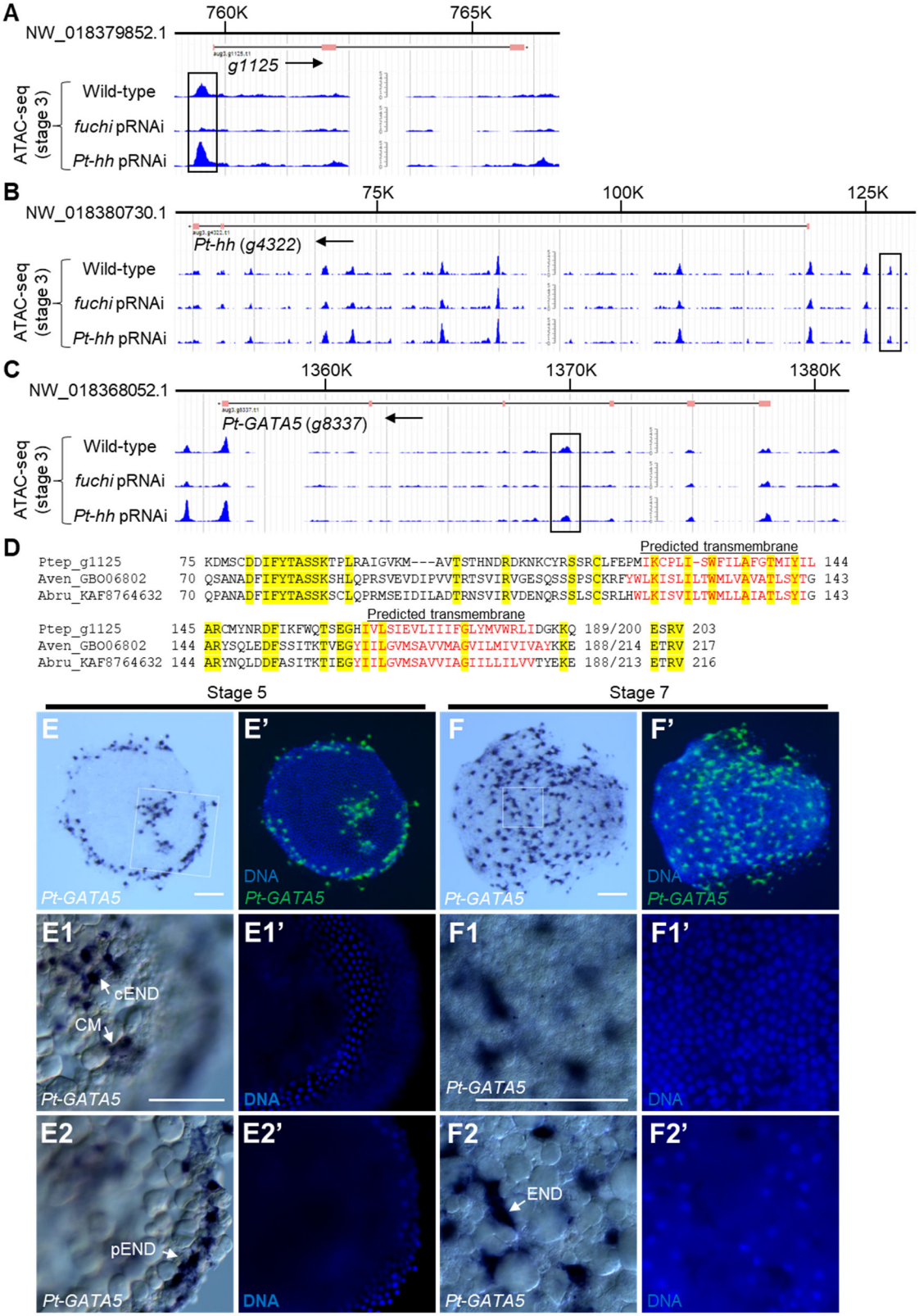


1422

1423 **Fig. 8. Effect of *fuchi* knockdown on chromatin accessibility in specific genomic**
1424 **regions in stage-3 embryos. (A)** Distributions of differentially accessible genomic
1425 regions identified by comparisons between ATAC-seq datasets from *fuchi* (upper) or
1426 *Pt-hh* (lower) pRNAi and wild-type embryos. MA-plots represent $\log_2$ fold-change

1427 (FC) versus $\log_2$ average ATAC signal abundance (CPM). Two biological replicates
1428 were obtained from each sample type and analyzed. Significant differentially accessible
1429 regions (FDR < 0.05) are highlighted in red. Blue lines are loess fits to each distribution
1430 with 95% confidence intervals shaded in gray. Asterisks indicate positive peaks at the
1431 same genomic region. Detection of this region might not be specific to *fuchi* activity.
1432 **(B)** Heat maps showing chromatin accessibilities in the 316 genomic regions identified
1433 as *fuchi*-dependent (sorted by FDR values). Data from the biological replicates are
1434 individually shown. Some of the identified regions are located close to or within the
1435 loci of DEGs identified by the RNA-seq analyses, as shown on the left side. **(C)** Motif
1436 sequences enriched in the *fuchi*-dependent peaks, as revealed by a motif discovery tool
1437 STREME [97]. Peak regions containing detected motif sequences,
1438 RNWGATAAGAVW, are indicated in B (red dots).
1439

1440



1441

1442

1443

1444 **Fig. 9. Three notable examples of genes revealed by combined analyses of the**
1445 **RNA-seq and ATAC-seq datasets. (A–C)** Normalized ATAC-seq signal of stage-3
1446 wild-type, *fuchi* pRNAi, and *Pt-hh* pRNAi embryo at the *g1125* (A), *Pt-hh* (*g4322*) (B),
1447 and *Pt-GATA5* (*g8337*) (C) loci. Line with ticks at the top represents the scaffold
1448 sequence, below which annotated exons (pink boxes) and introns (black lines) are
1449 shown. *fuchi*-dependent peak regions are indicated by open rectangles. Arrows indicate
1450 the orientation of the gene. **(D)** Amino acid sequence alignment of the *g1125* protein
1451 and its homologs detected in the genomes of two other spider species *Araneus*
1452 *ventricosus* (GBO06802.1) and *Argiope bruennichi* (KAF8764632.1). Residues
1453 identical across all proteins are presented with a yellow background. Residues shown
1454 in red indicate regions predicted as transmembrane domains by InterProScan version
1455 5.53-87.0 [99]. **(E, F)** Detection of *Pt-GATA5* transcript and DNA (E', F') in stage-5 (E)
1456 and stage-7 (F) embryos using chromogenic WISH. **(E1, E2, F1, F2)** High
1457 magnification of the boxed areas in E and F, along with counter staining for DNA (E1',
1458 E2', F1', F2'). The focal plane is at two different depths; the plane shown in E2 and F2
1459 is deeper than that in E1 and F1, respectively. Signals are observed in cEND, CM, and
1460 pEND cells, which are located below the germ-disc epithelial layer, and in evenly
1461 distributed endoderm (END) cells, which are located below the germ-band epithelial
1462 layer. Scale bars, 100 μm.
1463
1464