An atlas of spider development at single-cell resolution provides new insights into arthropod embryogenesis

Daniel J. Leite[1,2¶*], Anna Schönauer[1¶], Grace Blakeley[1&], Amber Harper[1&], Helena Garcia-Castro[1&], Luis Baudouin-Gonzalez[3], Ruixun Wang[4], Naïra Sarkis[4], Alexander Günther Nikola[5], Ventaka Sai Poojitha Koka[5], Nathan J. Kenny[1,6], Natascha Turetzek[5], Matthias Pechmann[4], Jordi Solana[1*], Alistair P. McGregor[1,2*]

[1] Department of Biological and Medical Sciences, Oxford Brookes University, Oxford OX3 0BP, UK.

[2] Department of Biosciences, Durham University, Durham DH1 3LE, UK.

[3] Oxford University Museum of Natural History, University of Oxford, Oxford OX1 3PW, UK

[4] Institute for Zoology, Biocenter, University of Cologne, Zuelpicher Str. 47b, 50674 Cologne, Germany

[5] Evolutionary Ecology, Faculty of Biology, Biocenter, Ludwig-Maximilians-University Munich, Planegg-Martinsried, Germany

[6] Department of Biochemistry Te Tari Matū Koiora, University of Otago, Dunedin, Aotearoa New Zealand.

* Corresponding authors
Email: daniel.j.leite@durham.ac.uk (DJL)
Email: jsolana@brookes.ac.uk (JS)
Email: alistair.mcgregor@durham.ac.uk (APM)

[¶] These authors contributed equally to this work

[&]These authors contributed equally to this work

Abstract

Spiders are a diverse order of chelicerates that diverged from other arthropods over 500 million years ago. Research on spider embryogenesis has made important contributions to understanding the evolution of animal development. In particular, studies of the common house spider *Parasteatoda tepidariorum* using developmental candidate gene approaches have provided key insights into the regulation and evolution of many processes including axis formation, segmentation and patterning. However, there remains a paucity of knowledge about the cells that build spider embryos, their gene expression profiles and fate. Single-cell transcriptomic analyses have been revolutionary in describing these complex landscapes of cellular genetics in a range of animals. Therefore, we carried out single-cell RNA sequencing of *P. tepidariorum* embryos at stages 7, 8 and 9, which encompass the establishment and patterning of the body plan, and initial differentiation of many tissues and organs. We identified 23 cell clusters marked by many developmental toolkit genes, as well as a plethora of non-candidate genes not previously investigated. We found many Hox genes were markers of cell clusters, and Hox gene paralogs often were present in different clusters. This provided further evidence of sub- and/or neo-functionalisation of these important developmental genes after the whole genome duplication in the arachnopulmonate ancestor. We also examined the spatial expression of marker genes for each cluster to generate a comprehensive cell atlas of these embryonic stages. This revealed new insights into the cellular basis and genetic regulation of head patterning, hematopoiesis, limb development, gut development and posterior segmentation. This atlas will serve as a platform for future analysis of spider cell specification and fate, and the evolution of these processes among animals at cellular resolution.

Introduction

Studying and comparing embryogenesis among animals has helped to elucidate how animal development is regulated by toolkit genes and provided insights into ancestral mechanisms, and how these processes evolve [1]. To gain deeper knowledge of these processes requires data on differential gene expression at cellular resolution during embryogenesis, to allow us to understand the distinct cell types that build animals. This can now be achieved by using single-cell analysis [2-5]. However, this approach has only been applied to a few models and there is therefore a great need to perform this work in additional key taxa [6-13].

Arthropods are a vastly diverse phylum that displays a wonderous array of morphological and behavioural differences [14]. Studies of mandibulates, particularly insects, not least the fruit fly *Drosophila melanogaster,* have provided exceptionally detailed insights into embryogenesis, however, it remains imperative to carry out broad taxon sampling in order to capture differences between lineages and to infer evolutionary trajectories and outcomes. Since chelicerates, including spiders, represent an outgroup to mandibulate arthropods, the study of their development offers a unique perspective to better understand the evolution of embryogenesis among arthropods and other animals more generally [15].

The common house spider *Parasteatoda tepidariorum* has already proven to be a powerful model for understanding the genetic regulation of key processes during spider embryogenesis, as well as specific spider innovations [15-19]. Previous studies of *P. tepidariorum* embryogenesis have provided significant insights into the regulation and evolution of axis formation [20, 21], patterning [22-25], germ layer specification [26-28], development of the nervous system [29], eye specification [30-32], appendage patterning [33-36], germ line formation [37] and segmentation [27, 38-45].

During development, *P. tepidariorum* initially forms a radially symmetrical germ disc in one hemisphere of the embryo with the extra-embryonic tissue in the other [19, 21, 25, 46, 47]. In the centre of the germ disc, beneath the epithelial layer, the cumulus forms from a group of mesenchymal cells. The cumulus has been likened to an organiser as migration of

these cells to the rim of the germ disc breaks radial symmetry, giving rise to the germ band with both antero-posterior (A-P) and dorso-ventral (D-V) axes [19, 45, 48, 49]. Therefore, generation of the bilaterally symmetrical germ band by stage 7 is a key point in embryogenesis from when the spider body is built and patterned by several important processes.

The spider body is composed of the anterior prosoma (cephalothorax) and posterior opisthosoma (abdomen). Formation of these two body tagmata requires different modes of segmentation. The formation of prosomal segments are somewhat simultaneously controlled by genes including *hedgehog*, *orthodenticle, msx1* and *Sox21b-1* [40, 42, 45, 50]. Opisthosomal segments on the other hand are added sequentially from a posterior segment addition zone (SAZ), regulated by *caudal* and dynamic interplay between the Wnt and Delta-Notch signalling pathways [27, 38, 41]. During the formation of segments through stages 7 to 9, the expression of the Hox genes becomes refined into specific domains to determine segmental identity and appendage location [22, 24].

Prosomal limb buds become apparent from stage 8 and subsequently elongate and differentiate into the chelicerae, pedipalps and walking legs with their associated sensory organs [15, 19, 29]. The opisthosomal limb buds also appear during these stages and differentiate into organs like the book lungs and spinnerets [15, 19]. Furthermore, *twist* and *forkhead* regulate the differentiation of the mesoderm during stages 7 to 9 in cells that were internalised earlier in embryogenesis, as well as new cells from posteriorly added segments [26, 27]. Neurogenesis also begins at stage 8, patterning regions across the head, brain, central (CNS), and peripheral nervous systems (PNS) [21, 31, 32].

Genomic and transcriptomic analysis of spiders and other chelicerates has also captured evolutionary events, which are likely impactful on the regulation and evolution of embryogenesis. Analysis of the genomes of the spider *P. tepidariorum* and the scorpion *Centruroides sculpturatus* revealed that there was a whole genome duplication (WGD) in the ancestor of arachnopulmonates (spiders, scorpions and their relatives) [22, 23, 51, 52]. WGDs have occurred in several animal lineages, including vertebrates, but are still enigmatic. They give rise to retained duplicated genes (called ohnologs), which may diverge in function and/or

share ancestral functions [53-55]. Among the retained ohnologs in *P. tepidariorum* there are many developmental genes and most strikingly two copies of every Hox gene, except *fushi taratzu* (*ftz*) [22, 23, 51, 52]. Expression of all 19 Hox genes was previously characterised during stages 7 to 9 and showed differences suggestive of sub- or neo-functionalisation in germband cells after WGD [22].

Our understanding of spider development and how it compares to other animals has mainly relied on analysis of well-known developmental candidate genes. However, the use of transcriptomic techniques has identified non-candidate genes instrumental in *P. tepidariorum* development, such as the role of *Ets-4* in cumulus integrity [20, 23, 50, 51, 56, 57]. This highlights their utility and the need for further non-candidate gene approaches that will provide new insights into spider development [20, 23, 50, 51, 56, 57]. Furthermore, we still lack a general understanding of the many cells that build *P. tepidariorum* embryos: their spatial distribution, differentiation, and the combinations of genes they express.

Single-cell RNA sequencing (scRNA-seq) transcriptomics provide an unbiased view of gene expression during embryogenesis at cellular resolution and additionally allows the characterisation of cell types and their differentiation based on marker genes [5, 58]. scRNA-seq has been successfully applied to provide a deeper understanding of embryogenesis and organogenesis in important invertebrate model systems like the fruit fly *Drosophila*, the nematode *Caenorhabditis elegans* and the ascidian *Ciona intestinalis* as well as vertebrates like the frog *Xenopus tropicalis*, mice and zebrafish [6-13]. Therefore, this approach offers the interrogation of developmental processes in a multi-dimensional landscape and can aid the discovery of new genes in developmental processes beyond classical candidate gene approaches.

Here we have also taken advantage of our recent advances in cell dissociation, combining ACetic-MEthanol (ACME) dissociation [3] and the SPLiT-seq scRNA-seq [59] technology to describe at cellular resolution spider embryogenesis at stages 7, 8 and 9. This has allowed us to define cell types and capture new candidate genes involved in several important developmental processes during these key stages of embryogenesis. These data

reveal novel insights into the Hox patterning, head and CNS development, germ layer specification and the cellular basis and regulation of posterior segment addition, as well as new perspectives on the role of the so called extra-embryonic cells. Our study and resources provide a platform to more broadly understand spider development at cellular resolution and are an important point for comparison to cell types and developmental processes in other arthropods and animals.

Results

Single-cell sequencing of three stages of spider embryogenesis

To interrogate spider development at single-cell resolution we sequenced single-cells from three embryonic stages (stages 7, 8.1 and 9.1) of *P. tepidariorum* (Fig 1A) [19]. We focussed our analysis on these stages because they mark the onset and/or continued progress of key developmental processes including segmentation, initiation of nervous system development, development and differentiation of germ layers, appendage development and patterning of the embryo along the AP and DV axes [18, 19, 21-23, 26-28, 31-34, 38-45, 48, 60-62].

We carried out separate ACME dissociations for each developmental stage (Fig 1B) [3]. ACME allowed early fixation of cells and therefore reduced the stress responses that occur with enzymatic dissociations. We then subjected these samples to SPLiT-seq barcoding [59], processing all stages in parallel in the same plate (Fig 1B). For each stage we obtained two libraries, corresponding to the 4^{th} round of barcoding of the SPLiT-seq process, containing transcriptomes from a total of approximately 30,000 cells. The cells had low (<1%) mitochondrial expression, indicating minimal cell stress and transcriptional noise in the dataset (Fig 1C). After filtering (see Materials and Methods) based on UMI counts, number of genes expressed per cell, mitochondrial expression and doublet removal, the total processed dataset contained 18,516 cells. This represented 4824, 4833, 8859 cells at stages 7, 8 and 9 respectively, with median UMI count per cell of 1468, 1656, and 1344, and a median of 674, 713, and 563 genes quantified per cell. Stage 7 and 8 were comparable in these metrics, whereas stage 9 had fewer UMI and gene counts per cell, but more cells overall (Fig 1C).

Processing of each stage separately (Fig 1D) revealed only minor differences in the contribution of informative principal components (S1 Fig). Only library two showed increasing number of informative components from stages 7 to 9, indicating a slight trend of increased data structure, perhaps due to differentiation during development (S1 Fig). To integrate these stages reciprocal principal component analysis (PCA) integration was used, where anchors between pairs of datasets were determined by projection into each other's reduced PCA space

with mutual nearest neighbours constrained on the anchors [63]. This helped to overcome issues regarding separation of stage 9 from stages 7 and 8, potentially due to the differences seen in UMI and features counts per cell (Fig 1C). With this integrated data, clusters were determined to acquire a conservative estimate of cell type diversity (Fig 1E). Given that there are no previous assessments of spider cell type diversity across different stages during embryogenesis, we estimated the clustering resolution for individual stages and for the integrated data, which revealed 23 clusters (see Materials and Methods) (Fig 1E, 1F). There was considerable variability in the abundance of cells from a given stage within each cluster (Fig 1G), ranging from 2415 (13.0 %) cells in cluster 0 to 280 (1.5 %) cells in cluster 22 (Fig 1H).

Marker genes for each of the 23 clusters were predicted with an in-cluster-versus-all-others approach, including genes that were expressed in at least 40% of the cells in their respective cluster and a return threshold of *p*-value of 1^{-5}. Numbers of marker genes ranged from 20 to 222 (Fig 1H) and related to some diversity in the clustering stability. A particularly unstable cluster represented by the fewest markers was cluster 0. This cluster contained ~13% of the cells in the dataset and these were dispersed across the Uniform Manifold Approximation and Projection (UMAP) (Fig 1E, 1H, S2A Fig). It is possible that these are stem or pluripotent cells, expressing few variable features, or potentially noisy artefactual cells. However, interestingly, the top marker for cluster 0 was a PRDM gene, *hamlet* (*ham*) [aug3.g11431], which has known functions in stem cell maintenance, neural stem cell and endoderm fate, and therefore tentatively suggests that these are possibly stem cells [64, 65] (S2B, S2C Fig).

Overall, the clustering of these data showed sufficient structure and information that allowed us to interrogate cell types and characterise marker genes in *P. tepidariorum* with respect to key processes during spider embryogenesis.

Hox markers in the single-cell data are consistent with A-P cell identity and potential sub- and/or neo-functionalisation

We found all 19 Hox genes in *P. tepidariorum* expressed in the scRNA-seq data and 13 were markers of 14 clusters (Fig 2A). When we grouped cell clusters using the similarity of their Hox gene expression patterns, three major groups emerged. The most anterior Hox genes, *labial* (*lab*) and *proboscipedia* (*pb*) Hox genes, which are expressed in pedipalps were most strongly expressed in clusters 13 and 21 [22, 35]. *Deformed* (*Dfd*) and *Sex combs reduced* (*Scr*), which are expressed in legs were most strongly expressed in cluster 14 [22]. The posterior Hox genes *fushi tarazu* (*ftz*) *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal-A* (*abdA*) and *Abdominal-B* (*AbdB*), which are expressed in the opisthosoma, were most strongly expressed in clusters 3 and 5 [22]. There was also a group of other clusters that expressed Hox genes comparatively more weakly. These distinctions match the boundaries in the domains of Hox expression between the prosoma and opisthosoma [22] (S3A Fig). This shows that our scRNA-seq data was able to detect clear patterns of Hox regulated cell type identity during embryogenesis (Fig 2). We next assessed whether Hox genes display cellular evidence of sub- and/or neofunctionalization between Hox ohnologs.

While the average expression of some pairs of Hox ohnologs across clusters in our scRNA-seq data was generally similar (*lab*, *Scr*, *Antp*, *abdA*), we also quantified ohnolog pairs without correlated expression across cell types (*Hox3* and *AbdB*) (Fig 2C, 2D). We also found variation in their temporal and spatial expression during embryogenesis consistent with previous gene expression studies using in situ hybridisation (ISH) showing (Fig 2B) [22]. For example, *pb*, *Dfd* and *Antp* ohnologs all show differences in their temporal expression (Fig 2B). We also observed that for the *pb* and *Scr* duplicates the paralog with more spatially restricted embryonic expression was a marker for fewer clusters in the scRNA-seq data (Fig 2A, 2D). For example, *pb-A* expression is mostly limited to pedipalpal expression and was predominantly in clusters 13 and 21, whereas *pb-B* is expressed in L1-L4 as well as the pedipalps and a marker for clusters 0, 7, 9 and 13 (Fig 2A, 2D). *Scr-A* is expressed in a distal

domain of L3-L4 legs and mostly relates to cluster 14, whereas *Scr-B* is more broadly expressed in L2-L4 and was expressed in multiple clusters in addition to cluster 14 (Fig 2A, 2D). *Antp* duplicates also have different spatial domains across the A-P axis (Fig 2A, 2D). *Antp-A* is expressed throughout the opisthosoma and predominantly in the SAZ, whereas *Antp-B* is restricted to the first two opisthosomal segments and L4 [22, 24]. In the scRNA-seq data *Antp-A* was a marker for clusters 1, 3 and 5, but *Antp-B* and was a marker for clusters 5, 6 and 7, highlighting that cluster 5 was likely the overlap in opisthosomal segments but other cell types relate to their mutually exclusive expression.

We have therefore captured single-cell transcriptomic evidence that most Hox genes are markers of distinct cell clusters. Hox genes appeared to relate to clusters along the A-P axis, with divergence in cluster expression between ohnologs reflecting likely sub- and/or neo-functionalization.

Identification of four cell clusters with different expression dynamics underlying spider precheliceral region patterning

The specification and regulation of the cells that later make up the different regions of the adult spider precheliceral region are poorly understood. The spider precheliceral region is specified and differentiate at the anterior of the germ band where no Hox genes are expressed during stages 7 to 9 [22, 35]. We observed that clusters 8, 10, 16 and 19 have some of the lowest Hox expression relative to other clusters in the scRNA-seq data (Fig 2A) suggesting that they correspond to cells of the precheliceral region.

orthodenticle-1 (*otd1*) and *hedgehog* (*hh*) are necessary for spider head development and are both initially expressed at the anterior rim of the germ disc [42, 61]. Their expression then migrates posteriorly, defining the most posterior margin of the precheliceral region and abutting the anterior of the cheliceral segment [42, 61]. *otd1* [g5047] was the top marker for clusters 10 and a marker for 19, while *hh* [g23071] was a marker of cluster 10 and was also expressed in 19 (Fig 3A). This suggests that clusters 10 and 19 relate to cells of the forming

precheliceral region. We therefore analysed two other markers from cluster 10 and four from cluster 19 (Fig 3). The cluster 10 markers *lim1a* [g12191] and *Pax6.2* [g12868] were both expressed at stage 7 at the anterior rim of the germ band (Fig 3A, 3B, S4A Fig). Expression of both genes subsequently migrates posteriorly, like *otd1*, and by stage 9 they are both expressed at the anterior boundary of the cheliceral segment but restricted to a ventral domain (Fig 3C). Of the cluster 19 markers tested (*Pax6.1* [g12873], *SoxE2* [g18856], *Tbx3* [aug3.g3745] and an *optomotor-blind* like gene [aug3.g3790]), *Pax6.1* and *Tbx3* were expressed earlier than the other markers. At stage 7, *Pax6.1* was expressed in a stripe along the anterior rim, and *Tbx3* had faint expression at the lateral edges of the germ band close to the anterior rim (Fig 3C). Like cluster 10 markers, their expression subsequently migrated towards the posterior of the precheliceral region (Fig 3C). However, in contrast to cluster 10 genes, expression of all four cluster 19 markers became limited to the posterior dorsal region of the precheliceral region by stage 9 (Fig 3C). This suggests that the broad region marked by *otd1* expression encompasses cluster 10 and 19 cells with ventral and dorsal identity respectively and potentially correspond to distinct regions of the hindbrain.

Cluster 16 cells were not marked by *otd1* expression, however we analysed three markers, *six3.1* [g1245], *six3.2* [g25543], and *visual system homeobox (vsx)* [aug3.g27186], which were all predominantly expressed in the precheliceral region (Fig 3B, 3C). *six3.1* was initially expressed in a stripe along the anterior rim of the germ band at stage 7, while *six3.2* appeared at the anterior rim at stage 8, and finally *vsx* by stage 9 (Fig 3C). Unlike *otd1*, all three genes maintained their expression at the anterior of the precheliceral region, indicating why *otd1* was not a marker for cluster 16 (Fig 3C). Cluster 16 cells therefore likely represent a distinct anterior region of the neurogenic ectoderm that could contribute to structures including the forebrain [66, 67].

We carried out double fluorescent ISH of precheliceral markers to better distinguish expression dynamics and the mutually exclusive regions of expression. *Pax6.1* (cluster 19) is initially maintained at the anterior limit of the *Pax6.2* (cluster 10) expression domain (S4C Fig). By stage 9.1 *Pax6.2* is more ventrally restricted and *Pax6.1* is more dorsally restricted,

overlapping by only a few cells, highlighting their distinct D-V domains of expression in the developing pre-cheliceral region (S4C Fig). *Pax6.2* remains anterior to the travelling and splitting *hh* stripe and at stage 9.1 their expression domains overlap at the anterior border of the cheliceral segment (S4A Fig). This was further supported by *Pax6.1* with *otd1* double FISH, showing that the posterior *Pax6.1* expression domain overlaps with *otd1* (S4B Fig).

Many cluster 8 markers, such as *otd1*, *Pax6.1*/*2*, *six3.1*, were also markers of cluster 10, 16, and 19, indicating a possible clustering issue (Fig 3B). Additionally, one of the top markers for cluster 8, *tailless* (*tll*) [g18090] was also a marker for clusters 10, 16 and 19, albeit with less significance (Fig 3B). Expression of *tll* starts at the anterior rim of the germband, however it maintains the anterior rim boundary of expression as the domain broadens towards the posterior margin of the precheliceral region (Fig 3C). *ham* was another cluster 8 marker that was not in clusters 10, 16, or 19, and has been shown to function in neural stem cells and their differentiation [64, 65] (S2 Fig), Cluster 8 therefore may represent a stem cell type that is widely distributed throughout the developing precheliceral region and may contribute to neurogenic tissues [68].

Overall, our scRNA-seq data provides new insights into the genetic regulation of precheliceral patterning. It suggests that distinct combinatorial gene expression fields mark different regions of the precheliceral region at stage 7, such as the forebrain and hindbrain, that later become regionalised during stages 8 and 9.

Two clusters were related to earlier defined D-V patterning

As in other arthropods, *decapentaplegic* (*dpp*) [g29377] regulates dorsal identity in spider embryos [21]. We found that *dpp* was a marker of cluster 11, along with *BMPR* [aug3.g2323], *noggin* [g27229] and *Tbx3* (S5A, S5B Fig). Like *dpp*, *BMPR* was expressed in the cumulus and both *BMPR* and *noggin* are subsequently expressed in the dorsal field (S5C Fig). Later, at stages 8 and 9, *noggin* was expressed broadly around the dorsal periphery of the germ band, while *BMPR* expression was restricted to dorsal domains in each appendage (S5C Fig).

Tbx3 was expressed in the precheliceral region and at the ventral midline, as well as in dorsal regions of prosomal appendages, like *BMPR* (S5C Fig).

Formation and patterning of the ventral midline are critical steps in the development of the nervous system. Ventral patterning begins prior to cumulus migration in *P. tepidariorum,* identified by the expression of *sog* [g13327] and *forkhead* (*fkh*) [g3001] [21]. The strongest expression of *sog* and *fkh* was present in cluster 4 (S5A, S5B Fig). We therefore assayed the expression of five other cluster 4 marker genes (*Nkx6.2* [g12201], *RGMA* [g28941], *LRR2* [g7463], *vitK-C* [g11868] and *ham*). While the onset of expression of these markers varied, they were all expressed along the ventral midline by stage 9 (S2C, S5C Fig). At the stage the expression of these cluster 4 markers was similar to *sog*, with differences in the presence/absence of expression in the precheliceral region (S5B, S5C Fig). As with *sog*, expression of all cluster 4 markers was excluded from the posterior SAZ and the peripheral cells of the germ band. Collectively, expression of known and new markers clearly identify cluster 4 as cells originating from the ventral midline.

The extra-embryonic contains endodermal gut cells and newly discovered cell types

P. tepidariorum endodermal cells that contribute to the gut have been shown to arise from the so called extra-embryonic region and express genes *serpent* [g7067] and *hepatocyte-nuclear factor-4* (*hnf-4*) [g4057] [28]. Despite these genes being exclusively expressed in cluster 22, neither were markers for this cluster (S6A, S6B Fig). We therefore analysed three cluster 22 markers *GPCPD* [g1958], *HSP* [g14898] and *myo* [g6459] (S6B Fig). Like *serpent* and *hnf-4*, all three of these genes were expressed in the extra-embryonic region (S6C Fig). Interestingly, cluster 22 contained the highest number of significant marker genes (Fig 1H). This suggests that these early endodermal cells are very distinct from the other cells and offers many further candidates to better understand endoderm development.

We also identified two cell clusters that were marked by mostly unstudied genes expressed around the dorsal of the germ band and in extra-embryonic cells. Cluster 17 (Fig 4A) was marked by *dpp*, and two other genes, *CAD-like* [g6385] and *GATA2-like* [g4744], which were all expressed in the cumulus, indicating dorsal identity (Fig 4B, 4C) [21, 25, 50]. *CAD-like* was also expressed in large cells surrounding the germ disc and in the extra-embryonic region (Fig 4C). Subsequently, *CAD-like* expressing cells are observed beneath the germ band at stage 8, but by stage 9 expression of this gene was again only observed in peripheral cells of the germ band (Fig 4C). *GATA2-like* expression was also seen in the dorsal field, and subsequently in all dorsal regions around the germ band, in the SAZ, and later in appendages. Other cluster 17 markers (*ush* [aug3.g16893], *cytochrome p450* [g29100], *Tbx20* [g5619] and *platelet glycoprotein* [g4985]) were all expressed in the dorsal region around the germ band and the extra-embryonic region (Fig 4C). These results suggest that cluster 17 may relate to cells from the dorsal region of the germ band and the extra-embryonic region.

Two well-known mesodermal genes *Mef2.1* [g3542] and *Mef2.2* [g24898], were markers of cluster 20 (Fig 4B) [69, 70]. Expression of these genes around the head at stage 9 appears to be within the dorsal extremes of the germ band (Fig 4C). We also explored the expression of three other cluster 20 markers, *hemocyanin A* [g11873], *C-ets1* [g472], and *DNA directed RNA pol* [g17128]. All three markers showed expression extending from the dorsal regions around the head into the extra-embryonic region (Fig 4C). We also assessed the expression of two other hemocyanin genes that were not markers (but expressed in cluster 20), *hemocyanin B* [g13621] and *hemocyanin C* [g22680] and found they had very similar expression to *hemocyanin A* (Fig 4B, 4C). These cluster 20 markers therefore revealed a previously unobserved cell type, which given the function of the orthologs of the marker genes in *Drosophila*, likely correspond to hemocytes [71-74]. It also suggests that either dorsal cells are migrating across the extra-embryonic region earlier than dorsal closure at stage 13 [19], or that extra-embryonic cells are being recruited to dorsal tissues of the embryo proper.

Diversity in clusters that relate to prosomal and opisthosomal appendages

As described above, Hox gene expression suggested clusters 13 and 14 correspond to the developing pedipalps and L3 - L4 legs respectively (Fig 2). However, we identified additional clusters likely associated with the developing appendages (S7A Fig).

Cluster 9 markers (S7B Fig) showed expression patterns that related to all appendages, including the previously characterised genes *Distal-less* (*Dll*) [g10793] [39] and *sp6-9* [g22966] [30, 75], which are expressed in all prosomal appendages. The marker of cluster 9 *Tbx3* also showed expression in all of the developing prosomal and opisthosomal appendages at stage 8 and 9 (Fig 3C). In addition, *Tbx3* and *Dll* were also expressed in opisthosomal segments, which suggest they are also involved in the development of appendages of this tagma like the book lungs and spinnerets (Fig 3C) [39]. This is consistent with the serial homology of prosomal and opisthosomal appendages [reviewed in 76].

We also identified several clusters that related to appendage mesoderm cells. The cluster 12 marker *hunchback* (*hb*) [g27583], has been previously shown to be required for the development of the L1 and L2 segments (S7B, S7C Fig) [60]. Cluster 12 was also marked by mesodermal genes *FGFR1* [g11749] and *FGFR2* [g3961] [49].This suggests that cluster 12 was specific to L1 and L2 and possibly mesodermal. Two clusters, 18 and 21, which shared markers *SoxD2*, *FGFR1* and *FGFR2* that are expressed in mesodermal cells of prosomal segments (S7B, S8C Fig) [43, 49]. Cluster 21 was also marked by the Hox genes *lab-A*/*B*, *pb-A* and *Hox3-B*, suggesting that these cells relate to pedipalps, while cluster 18 was marked by *Scr-A* suggesting these cells correspond to L2-L4, although another marker of cluster 18 and 21 marker gene we assayed, *Integrin-α8* [g23098], was expressed in all four walking legs (S7C Fig). We also identified cluster 15 as a peculiar appendage mesoderm related cell type. The markers g13852 and g12621 both showed early expression in many germ band cells and some extra embryonic cells (S7C Fig). However, by stage 8 expression was restricted to appendage segments and by stage 9 all prosomal limb buds (S7C Fig). We analysed another cluster 12 marker *Notch2* [g30344] and observed that this gene was first expressed in a single broad domain in the anterior region of the germ disc that splits and, like *hb*, was subsequently

expressed in L1 and L2 from stage 8 (S7C Fig) [60]. Hence, clusters 12, 15, 18 and 21 are likely related to a diversity of mesodermal cells in developing prosomal appendages.

Taken together, clusters 9, 13, 14, 15, 18 and 21 therefore represent a diverse range of appendage related cells including generic cells presumably found in all appendages, as well as cells specific to particular segments also marked by Hox gene expression.

Two clusters correspond to the peripheral nervous system cells distributed across the embryo

Appendages and organs require the innervation of the PNS to the CNS. We identified two clusters that relate to PNS cell types (S8A Fig). For one of these, cluster 7, we analysed five markers, *Awh* [g21739], *netrin* [g18008], *latrophilin cirl* [g14495], g15398 and *ham*, in addition to the previously published expression of *Emx3/4* [g27623/g27624] (S2, S8B, S8C Fig) [23]. All these genes were expressed in the precheliceral region and in both the prosoma and opisthosoma, extending from the midline into appendages (S8C Fig). Given the function of *ham* in the PNS and of *Awh* in (motor) neuron cell types, this cluster possibly relates to PNS cells that innervate appendages [64, 65, 77, 78]. Cluster 6 also likely corresponds to another PNS cell type. Markers of this cluster included both *Scr* and *Dfd* duplicates, suggesting cluster 6 cells are located in walking leg segments. The other cluster 6 markers examined, some of which were also markers of cluster 7, *hemicentin*, *netrin*, *Irx4* and the *Pax2.1* [g1094] paralog, were expressed along the proximo-distal axis of the developing limbs (S8C Fig) [29]. Given one of the functions of *Pax2.1* in spiders, cluster 6 may represents neurons that innervate sensory organs in limb appendages [29]. These two clusters therefore likely represent two classes of PNS neurons, one involved in motor movement and the other in sensory detection.

New resolution of SAZ regionalisation and maturation of segments during posterior development

Our analysis of Hox gene expression in the scRNA-seq data identified cell clusters likely related to opisthosomal regions (Fig 2A, S3C). In particular, posterior Hox genes were most highly expressed in clusters 3 and 5 (2A, 2D, S3C Fig). Given the wealth of genes previously known to be expressed in the SAZ and involved in posterior segmentation, we first examined whether they were markers of these two clusters. We found that *Wnt8* [g19404], *Wnt11b-2* [aug3.g1356], *hh*, *even-skipped* (*eve*) [g21109], *runt* [g18815], *noto1* [g13049], *bcl11* [g25539], were all markers of cluster 3 (Fig 5A, 5B) [38, 41, 42, 44, 50, 79]. We also repeated ISH for markers *AP2* [aug3.g23531] and g30822 [aug3.g27670] and corroborated their expression in the SAZ (Fig 5C, S9 Fig) [44]. Furthermore, we analysed four new markers of cluster 3, *RNF220* [g27156], *dentin-like*/*DSPP* [g3028], *big-brother* [g8835] and *band4.1* [g22446] (S9 Fig). Like *Wnt8*, *AP2* and g30822 [41, 44], the expression of *big-brother* and *band4.1* was restricted to a broad domain in the posterior region of the SAZ (Fig 5C and S9 Fig). Whereas, like *hh*, *eve* and *runt* [38, 42], *RNF220* and *DSPP* had dynamic expression in stripes originating from the posterior SAZ and subsequently in forming segments (Fig 5C and S9 Fig). Therefore, these cluster 3 markers show two fundamentally different expression patterns in the SAZ despite being clustered together.

Cluster 5 was marked by fewer known spider segmentation genes than cluster 3, but included *msx1* [g11815] (Fig 5B) [44]. We further analysed two additional cluster 3 markers *gooseberry* (*gsb*/*prd2*) [g10589], and *notum* [g17362], which encodes a Wnt signalling repressor (Fig 5B and S9 Fig) [80]. Note that *gsb* expression up to stage 8 was reported previously [44]. While these three genes are expressed in many regions of the embryo, they have similar expression in the developing opisthosoma in a region anterior to the SAZ, like both *Antp* genes at stage 9 (Fig 5C, S9 Fig) [22].

We also identified another cluster that relates to segmentation. Cluster 2 likely corresponds to cells in the posterior compartment of formed segments because it was marked

by *en* [g24362] and *hh* (Fig 5B) [15, 42, 81, 82]. This cluster was also marked by *frizzled 2* [g16651] [83] and *noto1* [44] and we analysed three other markers *hemicentin* [g1871], *basonuclin* [g29744] and *Irx4* [g29290] and confirmed that they were all similarly expressed in the posterior compartment of all segments (Fig 5B, S9 Fig).

Since clusters 2, 3 and 5 appear to represent different points during the generation of posterior segments, we assessed these cells separately from the other 20 clusters. We combined cluster 2 cells expressing posterior Hox genes (to remove prosomal cells) with cluster 3 and 5 cells (Fig 5D). The abundance of cells per cluster per stage showed that proportionally, filtered cluster 2 (posterior compartment) and cluster 5 (notum-related) cells increase from stage 7 to 9, while cluster 3 (SAZ) cells decrease (Fig 5E). This was consistent with the increase in the number of mature opisthosomal segments as development progresses.

Re-clustering was then performed to be able to test any further sub-populations and helped to identify a clean lineage path and pseudotime (Fig 5F) (see Materials and Methods). Four transcriptional modules were identified along the pseudotime path (Fig 5H, 5G). Two modules (2 and 4) corresponded to the original SAZ cluster 3: module 2 encompassed the very posterior SAZ genes (*Wnt8*, g30822 and *AP2*), which have a single continuous domain (Fig 5I), and module 4 included genes such as *gsb*, *eve*, *hh*, *Irx4, RNF220* and *DPSS*, which all have striped expression progressing from the SAZ (Fig 5I). These differences further evidence that two regions of the SAZ can be detected in the scRNA-seq data. Module 3 contained *notum*, as well as *sog* and *VitK-C*, which were top markers for the ventral midline cluster 4 (Fig 5I, S5, S9 Fig). This suggests that as cells progress anteriorly from the SAZ, ventral determinants are expressed as segments begin to mature. Following this, module 1 relates to the posterior compartment cells, expressing *en*, *hemicentin* and *gas* [g1575], as well as an *en* paralog *engrailed-2* [aug3.g15983], which is also expressed in segmental stripes in posterior compartment cells [23, 44, 84].

These four transcriptional modules identified along the pseudotime path may correspond to spatially distinct regions during posterior segmentation. To assess this further,

we used double fluorescence ISH combinations of genes from within and between modules. This showed that genes from module 2 (*Wnt8*, *AP2* and g30822) all had overlapping expression restricted to the very posterior of the SAZ, and that *AP2* and *g30822* expression was mutually exclusive from *eve*, collectively supporting that there is a distinct zone at the very posterior of the SAZ (Fig 5C). Anterior to this, genes from module 4 (*eve*, *DSPP* and *RNF220*) displayed striped expression, however, expression of *DSPP* and *RNF220* showed different phasing of expression relative to *eve* (Fig 5C). This suggests that module 4 may capture cells across (a) whole segment(s), with different gene expression profiles representing their sub-segmental position across the A-P axis of each segment (Fig 5C). Modules 2 and 4 were spatially distinct from module 3 (Fig 5C) consistent with neither *AP2* nor *eve* expression overlapping with the module 3 marker *notum*. Interestingly, *notum* expression overlaps with *gsb*, therefore module 3 likely represents cells of a transitioning zone where nascent segments have left the SAZ and are maturing (Fig 5C).

During assessment of SAZ markers we also observed that several markers of cluster 3, including *eve*, *DSPP* and *RNF220* were also cluster 1 markers (Fig 5B and S9 Fig). Cluster 1 was also marked by genes, not present in SAZ cluster 3, such as *SoxD2* [g19045] [43], *LotoB* [g25432] [85] and *Cux2* [g31597] [23], which are known to be expressed in the mesoderm of opisthosomal segments (Fig 5B and S9 Fig). We also found that two other cluster 1 markers, g13175 and *ZNF-like* [g18071], which were not previously analysed, were expressed in a somewhat similar region of the forming SAZ and also more anteriorly (S9 Fig). This suggests that cluster 1 represents SAZ cells that are internalised posteriorly and contribute to the mesoderm anteriorly as segments are added. Consistent with this, the expression of many cluster 1 markers overlap anterior to the SAZ where *twist* is expressed [26, 27, 41]. This offers an intriguing new insight into how the mesoderm may be specified during the addition of opisthosomal segments.

Discussion

The application of ACME and SPLiT-seq methods is a powerful combination for cellular resolution analysis of development

We used a combination of ACME dissociation [3] and SPLiT-seq [59] for the first time in arthropod embryos. ACME dissociation was an essential approach to fix samples early in the dissociation process, allowing us to collect multiple stages and samples that could be processed and stored prior to the single-cell transcriptomic experiments. This is highly desirable because embryogenesis progresses at a fast pace in many arthropods, and capturing distinct stages using lengthy (~hours) dissociation processes might alter the single-cell expression patterns. Furthermore, cells taken out of their developmental context for hours are prone to stress, which alters their gene expression patterns. We first fixed the embryos in ACME solution, which stopped biological activity, to then rupture and dissociate the embryos ACME therefore conveniently overcomes these hurdles, making the approach highly suited for our study of embryonic development. We then used a variation of SPLiT-seq which allowed us to obtain tens of thousands of single-cell transcriptomic profiles from multiple stages. The combined approach using both ACME and SPLiT-seq is therefore a powerful method to robustly profile developmental model systems.

A single-cell atlas of spider developmental

Our analysis of the scRNA-seq data from three embryonic stages of *P. tepidariorum* has enabled a new understanding of spider embryogenesis at cellular resolution. We were able to relate all 23 clusters to specific biological processes (Fig 6A), capturing known as well as novel cell types involved in multiple aspects of spider embryogenesis. We found multiple previously characterised genes in the clusters, which allowed us to establish their identities. Most cell clusters were also marked by genes that had not been previously studied and therefore offer new insights into these cell types and the regulation of developmental processes. Collectively, our dataset constitutes the first cell type atlas of spider embryonic stages 7 to 9.

While there was considerable diversity in cluster identities, there were fundamental patterning aspects relating to Hox patterning of the A-P axis (Fig 6B), the formation of the D-V axis (Fig 6C) and to the germ layer cell types (Fig 6D). These aspects reveal that the genetic components that regulate these core geometries also regionalise cells described by our scRNAseq data. The Hox genes were clearly segregated into three Hox positive groups, relating to pedipalpal, leg bearing and opisthosoma segments (Fig 6B). Yet within each of these regions, ohnologs appeared to show differential divergences that might relate to sub- and/or neo-functionalisation. The ectodermal versus the mesodermal and endodermal cell layers showed strikingly strong separation in the data (Fig 6D). This potentially reflects the early determination of these cell types in spider development and that they are transcriptional very distinct at stages 7 to 9.

Identification of new domains found in the process of precheliceral patterning

During prosomal development, *otd1* helps determine the precheliceral region [39], where structures/organs such as the brain, eyes and stomodeum develop . We identified four clusters that add further insights into the patterning of this region. At stage 9.1 the precheliceral region displays point-like depressions that correspond to the neural precursor cell internalisation and the first morphological indicator of neurogenesis [19]. *tll* encodes a nuclear receptor and is a marker of clusters 8, 10, 16 and 19, while *ham* is only a marker of cluster 8 (Fig 3, S2 Fig). Both *tll* and *ham* regulate neural stem cell fate and their broad expression in the head suggests specification and differentiation of neural stem cells is already underway at this early stage of head development [64, 65, 86, 87]. Markers of clusters 10, 16 and 19 have more specific spatial expression consistent with demarcation of different regions in the developing head. For example, expression of both *six3* paralogs remains at the anterior rim and suggests conserved roles for these genes in forebrain development [67]. In contrast, expression that demarcates clusters 10 and 19 are initially specified in the anterior but subsequently shift posteriorly to different extents and separate along the D-V axis. Markers of the ventral posterior head lobe

cluster, *lim1a* and *Pax6.2*, and the dorsal posterior head lobe, *SoxE2* and *Pax6.1*, are known for their essential roles in head and neural development in other animals/arthropods [88-91]. Overall, our scRNA-seq reveals three cell clusters present from stage 7 that likely prefigure different regions of the brain, as well as an additional cluster that more broadly contributes to the precheliceral region (Fig 3). Exploration of other markers of these clusters is likely to provide further new insights into conserved and derived aspects of spider brain and head development.

Developmental system drift in *P. tepidariorum* DV patterning?

The D-V axis of *P. tepidariorum* is specified by cumulus formation and migration of these cells to the periphery of the germ disc, where the dorsal field opens and ultimately surrounds the germ band in dorsal identity [92]. Like other arthropods, *P. tepidariorum* D-V axis formation and patterning is regulated by ventral *sog* and dorsal *dpp* signals [21, 25, 92]. While genes like *Ets-4*, *hh* and *fgf8* are also known to control cumulus migration, there is still much to be learned about the gene regulatory network that governs spider D-V patterning [20, 45, 49].

We observed *noggin* expression in the cumulus, dorsal field, and dorsal region of the germ band, which do not express *sog* (S5 Fig). Curiously, *noggin* is expressed dorsally in *Xenopus*, with *chordin*/*sog* [93], and *Xenopus noggin* is able to ventralise *Drosophila* embryos, revealing its conserved function as a BMP signalling inhibitor [93]. It is therefore surprising that *P. tepidariorum* exhibits dorsal expression of *noggin* where *dpp* signalling is active and necessary for dorsalisation. It may suggest that while *sog* and *dpp* play conserved roles in D-V specification in *P. tepidariorum,* other genes like *noggin* have diverged in function and therefore indicating potential developmental system drift in D-V regulation in this spider.

Contribution of extra-embryonic cells to the gut, hematopoiesis and the immune system

The germ disc and germ band of spiders has been largely considered to be exclusive from the extra embryonic region and yolk. However, the gut genes *serpent* and *hnf-4* are expressed in the extra embryonic region, suggesting these cells contribute to endoderm development in spiders [28]. These two genes were present in a single cluster, along with other previously unknown markers that were expressed in the extra-embryonic region (Fig 5). However, expression of marker genes from two other clusters in the extra-embryonic region suggests a need for reinterpretation of this tissue beyond just the contribution to the gut.

Insights from our scRNA-seq data revealed two clusters (17 and 20) with markers known to regulate heart and blood/immune system development in other animals, respectively (Fig 5). Heart development in the spider *C. salei* originates in the dorsal opisthosoma with the expression of the heart marker, *tinman* [94]. This was the only genetic marker characterised previously for the circulatory system in spiders. The combination of *Ush* and *GATA* genes, markers for cluster 17, have previously been reported to coordinate heart development, hematopoiesis, immune response and dorsal closure in *Drosophila* [95-97]. Like *tinman*, we found that they were expressed at the very dorsal periphery of the germ band. However, expression of other cluster 17 markers like *Tbx20*, *CAD-like*, *platelet glycoprotein* and *cytochrome p450* extends from this dorsal region across the extra-embryonic region, suggesting that the extra-embryonic region contributes cells to heart development (Fig 5).

Previously little was known about the specification of peripheral circulatory cells in spiders. We found that cluster 20 had combined expression of *Mef2* paralogs, which are later expressed in the heart (Fig 5C), and *hemocyanins*, strongly suggesting that this cluster may relate to the hemocyte cells. These cells initially surround the dorsal region of the precheliceral region, but subsequently migrate into the extra embryonic region. Interestingly, in *Drosophila*, embryonic hemocytes also originate from the head mesoderm indicating that there is perhaps a conserved origin of these cells between chelicerates and insects [71-74]. Our data offer an

exciting opportunity to investigate blood/immune system evolution and function further. In *Drosophila* and *Tribolium*, the extra-embryonic tissue is involved in immune response, and our results of hematopoiesis markers suggest for the first time that cells in the extra-embryonic region of spider could also contribute to this function [95, 98].

Taken together these new insights into the roles of extra-embryonic cells further highlight the utility of our scRNA-seq data in the detection of known cell types, but also its' power to detect previous unknown cell diversity and functions in understudied spider tissues.

Appendage diversity and their innervation by the peripheral nervous system

We identified several clusters relating to developing spider appendages (S7 Fig). Our data clearly distinguished the pedipalps from the four walking legs. However, no cluster related specifically to the cheliceral segment. This is possibly due to its smaller size, and therefore reduced contribution to cell numbers. Our clustering also appeared to differentiate between L1-2 and L3-4, presumably related to the differences in *Dfd* and *Scr* expression, along with markers like *Notch2* and *hb* (Fig 2, S7 Fig).

Spiders use appendages to build sophisticated webs for courtship, communication and capturing prey. They require a fine-tuned nervous system and various mechanoreceptive sensilla to detect vibrations, cuticle deflections as well as air flow [99]. However, the genetic regulators of the development of appendage sense organs have not been extensively investigated [100-102]. The only developmental study of appendage sensilla in *P. tepidariorum* revealed that *Pax2.1* is conserved in sensilla development in arthropods [29]. Interestingly, we also identified *Pax2.1* as a marker of appendage related neuron cells (S8 Fig). However, we identified addintional CNS and PNS clusters, along with many appendage-related clusters (S5, S7, S8 Fig). These highlight the diversity in spider appendage and PNS identity that likely corresponds to their function in various complex behaviours. Further analyses of these cells and their marker genes may help to map the development of the complex set of appendage and sensory systems of spiders.

Posterior segmentation can be regionalised based on genetic signatures of segment formation and maturation

How the addition of posterior segments from a SAZ is regulated has been studied extensively in spiders and other arthropods [27, 38, 41, 98, 103-109]. Although these studies have identified the involvement of some key genes and their interactions, we still have a poor understanding of how a SAZ works. Previous studies have found evidence that SAZs are likely genetically sub-structured, representing different regions that cells have to progress through to form new segments [38, 110]. The power of scRNA-seq has facilitated the identification of more robust genetic signatures of SAZ sub-structure in the clustering and at the same time revealed new genes involved in regulating segmentation. This allows us to propose an extended model for the structure of the SAZ and production of new segments.

Our cluster, ISH, trajectory, and expression module analysis suggest that opisthosomal segmentation can be divided into four regions representing different phases of segment formation (Fig 5H, 5I, 6E): region one relates to the most posterior cells in the SAZ, which is marked by genes like *Wnt8, AP2* and g30822 [44]. As previously suggested, this first region probably represents a pool of undifferentiated cells that continuously contributes to building new segments, perhaps analogous to the caudal region of the vertebrate presomitic mesoderm [41, 111].

Anteriorly in the SAZ, region two is marked by phased expression of pair-rule gene orthologs, like *eve* and *runt* [38], and new genes identified in our analysis including *RNF220* and *DSPP* (Fig 5C, 6E, S9 Fig). These phased domains appear to broadly relate to anterior and posterior regions of the forming segment (Fig 5C, 6E, S9 Fig). Anteriorly expressed genes include *caudal*, *eve*, *runt*, *DSPP* and *msx1* and posteriorly expressed genes *RNF220*, *hh* and *h* [38, 42, 44, 60]. Interestingly, *RNF220* enhances canonical Wnt signalling in other animals, and therefore might modulate the Wnt8 activity in the SAZ of spiders [112]. Region two

therefore likely represents forming segments anteriorly from the SAZ, but still lacking expression of the segment polarity gene *en* (Fig 5C, 6E) [15].

Region three is marked by expression of *notum,* which is also expressed in a similar pattern in another spider, *Cupiennius salei* [113]. *notum* is expressed in the posterior region of segments in this region, and like the overlapping expression of *gsb*, is mutually exclusive of *eve* expression (Fig 5C). However, it is unknown how this Wnt suppressor functions in the process of spider segmentation, although it may act on *Wnt8* activity to facilitate segment maturation [41, 80]. Due to this transition from the SAZ to a more differentiated region we describe this region as the segment maturation zone (SMZ) (Fig 6E).

Anteriorly, the formed segments then express segment polarity genes like *en*, which marks the fourth and final region (Fig 5H, 5I) [15]. This region has the clearest genetic signature of a mature segment and is strongly associated with posterior compartment cells. Interestingly, one marker of region four, *hemicentin*, is an extracellular matrix protein expressed in developing somites of zebrafish [114], and associated with the adhesion in the basement membrane, with knockdowns displaying detachment phenotypes between cell types. In this fourth region, ventral PNS genes (S8 Fig), like *netrin* are expressed. *netrin* is found ventrally in other arthropods, hemichordates and chordates and has conserved roles in axon guidance [115-118]. The combined regulation of *hemicentin* and *netrin* is involved in the coordination of invasions of anchor cells in *C. elegans* [118]. The known roles of these genes in other animals indicate that in cells of posterior segments in region four there could be complex interactions between cell types as the segment differentiates.

The SAZ cluster 3 is marked by several genes that are also expressed in cluster 1 (Fig 5B). Cluster 1 appears to correspond to mesodermal cells, given the expression of *SoxD2*, *LotoB*/*Toll* and *Cux2*, which are also all expressed in the mesoderm of other animals [23, 43, 85, 119, 120]. Both *LotoB* and *SoxD2* are expressed in region one to three, but later, along with *Cux2* are expressed in mesodermal metameric blocks in the opisthosoma and prosoma [23, 43, 85]. This is highly reminiscent of the *SoxD2* vertebrate ortholog, *Sox5*, which is expressed in the presomitic mesoderm and then later within each newly formed somite of

salmon [120]. RNAi knockdown of *Loto* genes in *P. tepidariorum* causes widening of the germ band due to perturbed cell intercalation and convergent extension [85]. Furthermore, disruption of the SAZ by RNAi against *Wnt8* or *Dl* results in ectopic expression of the mesoderm gene *twist* in the SAZ [26, 27, 41]. Given the role of *Dl* in specifying ectodermal versus mesodermal fate, this strongly suggests that there is dynamic specification and sorting of ectodermal and mesodermal cells in the SAZ throughout opisthosomal segmentation. One other marker of cluster 3, *Ubx-A*, has been shown to suppress *twist* in somatic myogenesis, which raises the possibility that this Hox gene may also function in a similar manner in the SAZ [121].

Conclusion

We captured gene expression profiles of single-cells at stages 7, 8 and 9, which encompass key processes of embryonic body plan establishment and the onset of the differentiation of several body structures. Our scRNA-seq cell type atlas of spider development corroborates previous findings and also provides novel insights into important processes during spider embryogenesis, including head patterning, posterior segmentation and hematopoiesis. Given that only three embryonic stages have been examined here, there is still a wealth of information awaiting capture from scRNA-seq of other time points in spider development. Future work to compare spider and other chelicerate cell atlases to other arthropods will also more broadly uncover evolutionary characteristics of cell types. Furthermore, comparisons between arachnids could prove fruitful for assessing the evolution of cell types after WGD events.

Materials and Methods

Dissociation of *P. tepidariorum* embryos for single-cell sequencing

Staging of embryos [19] and dissociation of whole embryos was performed as previously described [3]. Stage 7, 8.1 or 9.1 cocoons were selected and weighed without silk to determine the sample size. Embryos were dechorionated with bleach (sodium hypochlorite, 5% active chlorine, Arcos) and tap water (1:1). Embryos were washed several times with ultrapure water (UltraPure™ DNase/RNase-Free Distilled Water, Invitrogen) to remove bleach traces. Unfertilized embryos were removed, and embryos were immersed in 10 ml ACME solution (3:1:2:14 of methanol, glacial acetic acid, glycerol, and ultrapure water). To break open the vitelline membranes and allow complete dissociation, embryos were treated with a few pulses of polytron homogenisation. Embryos in ACME solution were incubated for 1 hour at room temperature (RT) on a rocking platform (Stuart SSL4) at 70 oscillations per minute. The cell suspension was filtered through a 50 µm filter (Sysmex/Partec CellTrics, Wolflabs) to separate remaining cell clumps and debris (vitelline membranes). Dissociated cells were pelleted at 1500 rpm for 5 minutes at 4°C. The supernatant was discarded, and the cell pellet washed with 7 ml PBS/1% Bovine Serum Albumin (BSA) (BSA Microbiological Grade Powder, Fisher BioReagents). The cells were pelleted at 1500 rpm for 5 minutes at 4°C, the supernatant was discarded, and the cell pellet resuspended in 1 ml 1x PBS-1% BSA and DMSO (9:1) and stored at -20°C.

Flow cytometry and cell dilution

ACME-dissociated cells from stage 7, 8 and 9 embryos were thawed on ice. Thawed cells were centrifuged twice at 1500 rpm for 5 minutes at 4ºC, to remove the DMSO, and resuspended in 400 µl of fresh 1x PBS-1% BSA. Samples were then filtered through a 50 µm filter (Sysmex/Partec CellTrics, Wolflabs) and collected into new 1.5 ml Eppendorf tubes, on ice. 50 µl of filtered cells were added to 100 µl of 1x PBS-1% BSA. The remaining undiluted samples were kept on ice, in the fridge, for the rest of the analysis.

Dilutions were stained with 0.4 µl of DRAQ5 (5 mM stock solution, Bioscience) and 0.8 µl of Concanavalin-A conjugated with AlexaFluor 488 (1 mg/ml stock solution, Invitrogen), and incubated in the dark for 25 min at room temperature. DRAQ5 was used as nuclear dye, while Concanavalin-A (Con-A) was used as cytoplasmic dye. We visualized and counted our cells using a CytoFlex S Flow Cytometer (Beckman Coulter). For each stained dilution, we made three measurements of 10 µl and registered the average number of total events (total ungated population). From this we calculated the number of total events per µl in our undiluted samples.

When multiple samples were available, we selected those with the highest percentage of singlets (DRAQ5-positive & Concanavalin-A-positive single-cells). To obtain this percentage of singlets, we used the following gating strategy: FSC-H vs FSC-A, where we selected only well-correlated events (first filter to remove aggregates); Con-A vs FSC-A, where we selected Con-A positive events (events with cytoplasm); DRAQ5 vs FSC-A, where we selected DRAQ5 positive events (events with nucleus); DRAQ5-A vs DRAQ5-H, where we selected only well-correlated events (second filter to remove aggregates); and DRAQ5 vs Con-A, where we obtained the final number of singlets.

To prepare the cells for the SPLiT-seq protocol, we unstained the undiluted/unstained samples. Samples were diluted in fresh 0.5x PBS buffer to a final concentration of 625 events/µl and kept on ice.

Re-annotation of *P. tepidariorum* genome for mapping SPLiT-seq data

SPLiT-seq has a bias towards capturing the 3' region of transcripts. To ensure capture of the signal form the SPLiT-seq data we re-annotated the genome (GCA000365465.2) of *P. tepidariorum*. Bulk RNA-seq data [57] was combined with multiple paired end libraries from a range of other embryonic stages. All data was quality trimmed with Trimmomatic v0.39 [122] and then mapped to the genome using Star v2.7.9a [123] using the 2-pass method for better detection of splice junctions. Alignment information was used as evidence for Braker v2 [124]

annotation, with ten rounds of optimisation, UTR training and considering CRF models. This annotation was combined with the previous genome annotation [22] by first merging the gene coordinates with bamtools merge. Gene models with a new annotation were replaced. Those with multiple new annotations to one previous annotation were rejected and the previous annotation was retained. New annotations that compounded multiple previous annotations were retained. This final annotation contained 33,413 gene models compared to the 27,950 in the Schwager et al. (2017) annotation. The majority (18,544) of these genes show a 1:1 relationship between annotations, as well as additional annotations captured by the new version, and the fusion of split genes by merging annotations. Old annotations are given as aug3.g* whereas new annotations are given as g*. The reannotation GTF and amino acid fasta files have been uploaded to figshare (https://doi.org/10.6084/m9.figshare.c.6032888.v2).

Mitochondrial genome assembly of *P. tepidariorum*

Mitochondrial expression in single-cell sequence data can be indicative of cell stress and therefore a useful metric to measure. We assembled a partial, but near complete, version of the mitochondrial genome to include in the mapping steps. DNA-seq data (SRR891587) from the genome assembly was trimmed with Trimmomatic v0.39 [122] and assembled with Spades v3.13.1 [125] using kmer sizes 21, 33, 55, 77 with the *P. tepidariorum* CO1 sequence (DQ029215.1) as a trusted contig. The spades contig matching the CO1 sequence was extracted and then extended with NOVOPlasty v4.2 [126] to achieve a final assembly of length 14,427 bp, though it was not circularisable. MiToS v2 [127] identified all expected features. The sequence was added to the genome assembly and a feature spanning the full length was added to the GTF gene coordinates file for mapping. The mitochondrial genome assembly has been uploaded to figshare (https://doi.org/10.6084/m9.figshare.c.6032888.v2).

SPLiT-seq, filtering, pre-processing, and clustering analysis

The SPLiT-seq protocol was performed as previously described with some modifications (S1 Text) [3]. Libraries were sequenced with 150 bp paired-end Illumina NovaSeq 6000 S4 flow cell, provided commercially by Novogene. Raw reads have been uploaded to the ENA with BioProject PRJEB53350.

Total sequencing output was 103.9 + 40.7 Gb, constituting a total of 963,482,454 raw reads, with >99.98% clean reads and a Q20 >93.94%. All data and samples passed FastQC inspection. Adapters and low-quality bases were trimmed with Cutadapt v1.18 [128] and properly paired reads were combined with Picard FastqToSam v2.20.5. All sequence runs were combined with Picard MergeSamFiles to attain paired reads for downstream expression analysis. To generate reference files, first Picard CreateSequenceDictionary was used to generate a dictionary from the genome plus mitochondrial sequence and re-annotations. Then converted to a RefFlat, a reduced GTF and intervals with DropSeq v2.4.0 [129] tools ConvertToRefFlat, ReduceGtf and CreateIntervalsFiles, respectively. For mapping the data, a Star v2.7.9a [123] genome index was generated with sjdbOverhang 99. These reference files and genome index were used as inputs for the Split-seq_pipeline.sh [129]. An expression matrix was generated with dropseq DigitalExpression, including reads mapping to introns, with a barcode edit distance of one, and outputting cells that had at least 100 genes. Cells from each stage were extracted from this matrix using the 16 sequences from cell barcode one.

Each library per stage was first processed for doublet removal. The expression matrix for each stage was loaded into Seurat v4 [63] and subset to contain cells where genes are expressed in at least 20 cells; that have genes numbers between 400 and 1800; UMI counts between minimums of 650 for stage 7, 700 for stage 8 and 500 for stage 9 and maximum of 4500; and no more than 1% mitochondrial expression. This initial dataset contained cells for stages 7 (1967 and 3111), 8 (2058 and 3029) and 9 (3866 and 5459), for libraries one and two respectively. Each sample was normalised with SCTransform with the glmGamPoi method and variable features threshold of 1.4 and regressing the mitochondrial expression, UMI counts and gene counts. Forty PCs and neighbours were computed using k. param 100,

and clusters were identified at a resolution of 1 for stage 7 and 8 and 1.2 for stage 9. Using doubletFinder v3 [130], 5% doublets we removed, identifying an appropriate pK with an initial parameter sweep, and retained singlets were extracted

Doublet filtered stage specific matrices were then processed for integration in Seurat, normalising with SCTransform using the glmGamPoi method and a variable feature threshold of 1.3 and regressing the mitochondrial expression, UMI counts and gene counts. The three stages were integrated by selecting 450 integration features, using the reciprocal PCA method (50 PCs) and 45 anchors. Fifty PCs were computed on the integrated data and used for UMAPs with the umap-learn method. Clusters were initially determined using 200 neighbors and a resolution of 2. The clustering resolutions were guided by ChooseR [131] and clustree [132] analysis. FindAllMarkers was used to extract markers for each cluster using the Wilcoxon method and including genes that were expressed in at least 40% of the cells in their respective cluster and a return threshold of 1e-5. Marker genes were annotated initially with the NCBI nr database using diamond v2.0.8.146 [133] and refined for existing genes already characterised in *P. tepidariorum*.

Monocle pseudotime and module analysis

For the analysis of segmentation dynamics, clusters 2, 3 and 5 were subset from the data and reprocessed. Only cluster 2 cells that expressed posterior Hox (*ftz* to *AbdB*) genes were used. Each stage and library were merged and processed as per the integration process, except the clustering parameters used 20 neighbors and a resolution of 0.6 to attain a finer scale resolution of clusters in these cell populations. Using this clustering and UMAP, a trajectory graph and pseudotime ordering were estimated with Monocle v3 [9], using the cluster that represented the most posterior cells (inferred by ISH) as the root population. Modules were then identified along the pseudotime projection, using cells with an expression correlation q value of 1e-30 and module detect resolution of 6e-1.

Gene cloning and expression analysis

For gene expression characterisation in *P. tepidariorum* embryos we performed colorimetric in situ hybridisation (ISH) [134], fastred [29] and double fluorescent ISH (dFISH) [43] as previously described with minor modifications (S1 Text).

Data Availability

Raw reads for the scRNAseq have been uploaded to the ENA with BioProject PRJEB53350. The re-annotation of the genome and the mitochondrial genome assembly of *P. tepidariorum* have been uploaded to figshare (https://doi.org/10.6084/m9.figshare.c.6032888.v2).

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been partially funded by grants from the Leverhulme Trust (RPG-2016-234) and NERC (NE/T006854/2) to APM; the Nigel Groome studentship at Oxford Brookes University to GB and HGC; a BBSRC DTP studentship to AH; the Rutherford Discovery Fellowship (MFP-UOO2109) from the Royal Society of New Zealand to NJK; the MRC grant (MR/S007849/1) and the BBSRC grant (BB/V014447/1) to JS; the DFG grants PE 2075/1-2 to RW and PE 2075/3-1 to NS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

Conceptualization: APM, AS, DJL, JS

Data curation: AS, DJL, HGC, NJK, AH, GB, LBG, MP, NT

Formal analysis: DJL, NJK

Funding acquisition: APM, JS, NJK, MP

Investigation: AS, DJL, APM, AH, GB, HGC, LBG, MP, NT, RW, NS, AGN, VSPK, NJK

Project administration: APM, DJL, AS, JS

Resources: APM, JS, MP, NT

Software: DJL, NJK

Supervision: APM, AS, DJL, JS, MP, NT

Visualization: DJL, GB, AH, LBG, MP, NT

Writing – original draft: DJL, APM, AS, HGC, AH

Writing – review & editing: DJL, APM, AS, MP, NT, NJK, JS, AH, GB, RW, LBG

Acknowledgements

We thank Lauren Sumney-Rooney and Fritz Vollrath for discussions. We thank Helen Ferry and Liam Hardy at the Experimental Medicine Division Flow Cytometry Facility at the Nuffield Department of Clinical Medicine (University of Oxford), Michal Maj, and Robert Hedley at the Flow Cytometry Facility at the Dunn School of Pathology (University of Oxford) for their help and advice with flow cytometry. We thank Vincent Mason for technical help. We thank the Center for Advanced Light Microscopy at the LMU Munich and Gregor Bucher for providing the microscope set-up for imaging whole mount ISH.

References

1. Carroll SB, Grenier JK, Weatherbee SD. From DNA to diversity: Molecular genetics and the evolution of animal design. Malden, Mass: Blackwell Science; 2001.
2. Wang J, Sun H, Jiang M, Li J, Zhang P, Chen H, et al. Tracing cell-type evolution by cross-species comparison of cell atlases. Cell Rep. 2021;34(9):108803.
3. Garcia-Castro H, Kenny NJ, Iglesias M, Alvarez-Campos P, Mason V, Elek A, et al. ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. Genome Biol. 2021;22(1):89. Epub 20210408.
4. Plass M, Solana J, Wolf FA, Ayoub S, Misios A, Glazar P, et al. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. Science. 2018;360(6391). Epub 20180419.
5. Tanay A, Sebe-Pedros A. Evolutionary cell type mapping with single-cell genomics. Trends Genet. 2021;37(10):919-32. Epub 20210518.
6. Briggs JA, Weinreb C, Wagner DE, Megason S, Peshkin L, Kirschner MW, et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science. 2018;360(6392). Epub 20180426.
7. Cao C, Lemaire LA, Wang W, Yoon PH, Choi YA, Parsons LR, et al. Comprehensive single-cell transcriptome lineages of a proto-vertebrate. Nature. 2019;571(7765):349-54. Epub 20190710.
8. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science. 2017;357(6352):661-7.
9. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature. 2019;566(7745):496-502. Epub 20190220.
10. Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science. 2018;360(6392). Epub 20180426.
11. Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, et al. The Drosophila embryo at single-cell transcriptome resolution. Science. 2017;358(6360):194-9. Epub 20170831.
12. Packer JS, Zhu Q, Huynh C, Sivaramakrishnan P, Preston E, Dueck H, et al. A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution. Science. 2019;365(6459). Epub 20190905.
13. Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science. 2018;360(6392):981-7. Epub 20180426.
14. Giribet G, Edgecombe GD. The Phylogeny and Evolutionary History of Arthropods. Curr Biol. 2019;29(12):R592-R602.
15. Schwager EE, Schöenauer A, Leite DJ, Sharma PP, McGregor AP. Chelicerata. In: Wanninger A, editor. Evolutionary Developmental Biology of Invertebrates 3: Ecdysozoa I: Non-Tetraconata.: Springer-Verlag; 2015.
16. Oda H, Akiyama-Oda Y. The common house spider Parasteatoda tepidariorum. Evodevo. 2020;11:6. Epub 20200320.
17. McGregor AP, Hilbrant M, Pechmann M, Schwager EE, Prpic NM, Damen WG. Cupiennius salei and Achaearanea tepidariorum: Spider models for investigating evolution and development. Bioessays. 2008;30(5):487-98.
18. Hilbrant M, Damen WG, McGregor AP. Evolutionary crossroads in developmental biology: the spider Parasteatoda tepidariorum. Development. 2012;139(15):2655-62.
19. 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(4):189-216. Epub 20120509.

20. Pechmann M, Benton MA, Kenny NJ, Posnien N, Roth S. A novel role for Ets4 in axis specification and cell migration in the spider Parasteatoda tepidariorum. Elife. 2017;6. Epub 20170829.
21. Akiyama-Oda Y, Oda H. Axis specification in the spider embryo: dpp is required for radial-to-axial symmetry transformation and sog for ventral patterning. Development. 2006;133(12):2347-57.
22. Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, et al. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biol. 2017;15(1):62. Epub 20170731.
23. Leite DJ, Baudouin-Gonzalez L, Iwasaki-Yokozawa S, Lozano-Fernandez J, Turetzek N, Akiyama-Oda Y, et al. Homeobox Gene Duplication and Divergence in Arachnids. Mol Biol Evol. 2018;35(9):2240-53.
24. Khadjeh S, Turetzek N, Pechmann M, Schwager EE, Wimmer EA, Damen WG, et al. Divergent role of the Hox gene Antennapedia in spiders is responsible for the convergent evolution of abdominal limb repression. Proc Natl Acad Sci U S A. 2012;109(13):4921-6. Epub 20120315.
25. Akiyama-Oda Y, Oda H. Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces Dpp signals received by germ disc epithelial cells. Development. 2003;130(9):1735-47.
26. Yamazaki K, Akiyama-Oda Y, Oda H. Expression patterns of a twist-related gene in embryos of the spider Achaearanea tepidariorum reveal divergent aspects of mesoderm development in the fly and spider. Zoolog Sci. 2005;22(2):177-85.
27. Oda H, Nishimura O, Hirao Y, Tarui H, Agata K, Akiyama-Oda Y. Progressive activation of Delta-Notch signaling from around the blastopore is required to set up a functional caudal lobe in the spider Achaearanea tepidariorum. Development. 2007;134(12):2195-205. Epub 20070516.
28. Feitosa NM, Pechmann M, Schwager EE, Tobias-Santos V, McGregor AP, Damen WGM, et al. Molecular control of gut formation in the spider Parasteatoda tepidariorum. Genesis. 2017;55(5). Epub 20170422.
29. Janeschik M, Schacht MI, Platten F, Turetzek N. It takes Two: Discovery of Spider Pax2 Duplicates Indicates Prominent Role in Chelicerate Central Nervous System, Eye, as Well as External Sense Organ Precursor Formation and Diversification After Neo- and Subfunctionalization Frontiers in Ecology and Evolution. 2022;10.
30. Setton EVW, Sharma PP. Cooption of an appendage-patterning gene cassette in the head segmentation of arachnids. Proc Natl Acad Sci U S A. 2018;115(15):E3491-E500. Epub 20180326.
31. Schomburg C, Turetzek N, Schacht MI, Schneider J, Kirfel P, Prpic NM, et al. Molecular characterization and embryonic origin of the eyes in the common house spider Parasteatoda tepidariorum. Evodevo. 2015;6:15. Epub 20150428.
32. Baudouin-Gonzalez L, Harper A, McGregor AP, Sumner-Rooney L. Regulation of Eye Determination and Regionalization in the Spider Parasteatoda tepidariorum. Cells. 2022;11(4). Epub 20220211.
33. Turetzek N, Pechmann M, Schomburg C, Schneider J, Prpic NM. Neofunctionalization of a Duplicate dachshund Gene Underlies the Evolution of a Novel Leg Segment in Arachnids. Mol Biol Evol. 2016;33(1):109-21. Epub 20151006.
34. Turetzek N, Khadjeh S, Schomburg C, Prpic NM. Rapid diversification of homothorax expression patterns after gene duplication in spiders. BMC Evol Biol. 2017;17(1):168. Epub 20170714.
35. Pechmann M, Schwager EE, Turetzek N, Prpic NM. Regressive evolution of the arthropod tritocerebral segment linked to functional divergence of the Hox gene labial. Proc Biol Sci. 2015;282(1814).
36. Heingard M, Turetzek N, Prpic NM, Janssen R. FoxB, a new and highly conserved key factor in arthropod dorsal-ventral (DV) limb patterning. Evodevo. 2019;10:28. Epub 20191108.

37. Schwager EE, Meng Y, Extavour CG. vasa and piwi are required for mitotic integrity in early embryogenesis in the spider Parasteatoda tepidariorum. Dev Biol. 2015;402(2):276-90. Epub 20140923.
38. Schonauer A, Paese CL, Hilbrant M, Leite DJ, Schwager EE, Feitosa NM, et al. The Wnt and Delta-Notch signalling pathways interact to direct pair-rule gene expression via caudal during segment addition in the spider Parasteatoda tepidariorum. Development. 2016;143(13):2455-63. Epub 20160610.
39. Pechmann M, Khadjeh S, Turetzek N, McGregor AP, Damen WG, Prpic NM. Novel function of Distal-less as a gap gene during spider segmentation. PLoS Genet. 2011;7(10):e1002342. Epub 20111020.
40. Paese CLB, Schoenauer A, Leite DJ, Russell S, McGregor AP. A SoxB gene acts as an anterior gap gene and regulates posterior segment addition in a spider. Elife. 2018;7. Epub 20180821.
41. McGregor AP, Pechmann M, Schwager EE, Feitosa NM, Kruck S, Aranda M, et al. Wnt8 is required for growth-zone establishment and development of opisthosomal segments in a spider. Curr Biol. 2008;18(20):1619-23.
42. Kanayama M, Akiyama-Oda Y, Nishimura O, Tarui H, Agata K, Oda H. Travelling and splitting of a wave of hedgehog expression involved in spider-head segmentation. Nat Commun. 2011;2:500. Epub 20111011.
43. Baudouin-Gonzalez L, Schoenauer A, Harper A, Blakeley G, Seiter M, Arif S, et al. The Evolution of Sox Gene Repertoires and Regulation of Segmentation in Arachnids. Mol Biol Evol. 2021;38(8):3153-69.
44. Akiyama-Oda Y, Oda H. Hedgehog signaling controls segmentation dynamics and diversity via msx1 in a spider embryo. Sci Adv. 2020;6(37). Epub 20200909.
45. Akiyama-Oda Y, Oda H. Cell migration that orients the dorsoventral axis is coordinated with anteroposterior patterning mediated by Hedgehog signaling in the early spider embryo. Development. 2010;137(8):1263-73.
46. Pechmann M. Formation of the germ-disc in spider embryos by a condensation-like mechanism. Front Zool. 2016;13:35. Epub 20160811.
47. Hemmi N, Akiyama-Oda Y, Fujimoto K, Oda H. A quantitative study of the diversity of stripe-forming processes in an arthropod cell-based field undergoing axis formation and growth. Dev Biol. 2018;437(2):84-104. Epub 20180316.
48. Oda H, Iwasaki-Yokozawa S, Usui T, Akiyama-Oda Y. Experimental duplication of bilaterian body axes in spider embryos: Holm's organizer and self-regulation of embryonic fields. Dev Genes Evol. 2020;230(2):49-63. Epub 20190410.
49. Wang R, Karadas L, Schiffer P, Pechmann M. FGF signalling is involved in cumulus migration in the common house spider Parasteatoda tepidariorum. bioRxiv. 2021.
50. Oda H, Akiyama-Oda Y. Dataset on gene expressions affected by simultaneous knockdown of Hedgehog and Dpp signaling components in embryos of the spider Parasteatoda tepidariorum. Data Brief. 2020;28:105088. Epub 20200103.
51. Leite DJ, Ninova M, Hilbrant M, Arif S, Griffiths-Jones S, Ronshaugen M, et al. Pervasive microRNA Duplication in Chelicerates: Insights from the Embryonic microRNA Repertoire of the Spider Parasteatoda tepidariorum. Genome Biol Evol. 2016;8(7):2133-44. Epub 20160803.
52. Harper A, Baudouin Gonzalez L, Schonauer A, Janssen R, Seiter M, Holzem M, et al. Widespread retention of ohnologs in key developmental gene families following whole-genome duplication in arachnopulmonates. G3 (Bethesda). 2021;11(12).
53. Ohno S. Evolution by Gene Duplication. Berlin, Heidelberg: Springer; 1970.
54. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, et al. The amphioxus genome and the evolution of the chordate karyotype. Nature. 2008;453(7198):1064-71.
55. Marletaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, et al. Amphioxus functional genomics and the origins of vertebrate gene regulation. Nature. 2018;564(7734):64-70. Epub 20181121.

56. Iwasaki-Yokozawa S, Akiyama-Oda Y, Oda H. Genome-scale embryonic developmental profile of gene expression in the common house spider Parasteatoda tepidariorum. Data Brief. 2018;19:865-7. Epub 20180524.
57. Posnien N, Zeng V, Schwager EE, Pechmann M, Hilbrant M, Keefe JD, et al. A comprehensive reference transcriptome resource for the common house spider Parasteatoda tepidariorum. PLoS One. 2014;9(8):e104885. Epub 20140813.
58. Stuart T, Satija R. Integrative single-cell analysis. Nat Rev Genet. 2019;20(5):257-72.
59. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science. 2018;360(6385):176-82. Epub 20180315.
60. Schwager EE, Pechmann M, Feitosa NM, McGregor AP, Damen WG. hunchback functions as a segmentation gene in the spider Achaearanea tepidariorum. Curr Biol. 2009;19(16):1333-40. Epub 20090723.
61. Pechmann M, McGregor AP, Schwager EE, Feitosa NM, Damen WG. Dynamic gene expression is required for anterior regionalization in a spider. Proc Natl Acad Sci U S A. 2009;106(5):1468-72. Epub 20090115.
62. Kanayama M, Akiyama-Oda Y, Oda H. Early embryonic development in the spider Achaearanea tepidariorum: Microinjection verifies that cellularization is complete before the blastoderm stage. Arthropod Struct Dev. 2010;39(6):436-45.
63. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-87 e29. Epub 20210531.
64. Leszczynski P, Smiech M, Parvanov E, Watanabe C, Mizutani KI, Taniguchi H. Emerging Roles of PRDM Factors in Stem Cells and Neuronal System: Cofactor Dependent Regulation of PRDM3/16 and FOG1/2 (Novel PRDM Factors). Cells. 2020;9(12). Epub 20201204.
65. Hohenauer T, Moore AW. The Prdm family: expanding roles in stem cells and development. Development. 2012;139(13):2267-82.
66. Schacht MI, Schomburg C, Bucher G. six3 acts upstream of foxQ2 in labrum and neural development in the spider Parasteatoda tepidariorum. Dev Genes Evol. 2020;230(2):95-104. Epub 20200210.
67. Steinmetz PR, Urbach R, Posnien N, Eriksson J, Kostyuchenko RP, Brena C, et al. Six3 demarcates the anterior-most developing brain region in bilaterian animals. Evodevo. 2010;1(1):14. Epub 20101229.
68. Stollewerk A, Tautz D, Weller M. Neurogenesis in the spider: new insights from comparative analysis of morphological processes and gene expression patterns. Arthropod Struct Dev. 2003;32(1):5-16.
69. Ranganayakulu G, Zhao B, Dokidis A, Molkentin JD, Olson EN, Schulz RA. A series of mutations in the D-MEF2 transcription factor reveal multiple functions in larval and adult myogenesis in Drosophila. Dev Biol. 1995;171(1):169-81.
70. Crittenden JR, Skoulakis EMC, Goldstein ES, Davis RL. Drosophila mef2 is essential for normal mushroom body and wing development. Biol Open. 2018;7(9). Epub 20180907.
71. Spahn P, Huelsmann S, Rehorn KP, Mischke S, Mayer M, Casali A, et al. Multiple regulatory safeguards confine the expression of the GATA factor Serpent to the hemocyte primordium within the Drosophila mesoderm. Dev Biol. 2014;386(1):272-9. Epub 20131217.
72. de Velasco B, Mandal L, Mkrtchyan M, Hartenstein V. Subdivision and developmental fate of the head mesoderm in Drosophila melanogaster. Dev Genes Evol. 2006;216(1):39-51. Epub 20051025.
73. Holz A, Bossinger B, Strasser T, Janning W, Klapper R. The two origins of hemocytes in Drosophila. Development. 2003;130(20):4955-62. Epub 20030820.
74. Tepass U, Fessler LI, Aziz A, Hartenstein V. Embryonic origin of hemocytes and their relationship to cell death in Drosophila. Development. 1994;120(7):1829-37.
75. Konigsmann T, Turetzek N, Pechmann M, Prpic NM. Expression and function of the zinc finger transcription factor Sp6-9 in the spider Parasteatoda tepidariorum. Dev Genes Evol. 2017;227(6):389-400. Epub 20171107.

76. Sharma PP. Chelicerates and the Conquest of Land: A View of Arachnid Origins Through an Evo-Devo Spyglass. Integr Comp Biol. 2017;57(3):510-22.
77. Curtiss J, Heilig JS. Establishment of Drosophila imaginal precursor cells is controlled by the Arrowhead gene. Development. 1995;121(11):3819-28.
78. Sagasti A, Hobert O, Troemel ER, Ruvkun G, Bargmann CI. Alternative olfactory neuron fates are specified by the LIM homeobox gene lim-4. Genes Dev. 1999;13(14):1794-806.
79. Janssen R, Pechmann M, Turetzek N. A chelicerate Wnt gene expression atlas: novel insights into the complexity of arthropod Wnt-patterning. Evodevo. 2021;12(1):12. Epub 20211109.
80. Giraldez AJ, Copley RR, Cohen SM. HSPG modification by the secreted enzyme Notum shapes the Wingless morphogen gradient. Dev Cell. 2002;2(5):667-76.
81. Tabata T, Eaton S, Kornberg TB. The Drosophila hedgehog gene is expressed specifically in posterior compartment cells and is a target of engrailed regulation. Genes Dev. 1992;6(12B):2635-45.
82. Morata G, Lawrence PA. Control of compartment development by the engrailed gene in Drosophila. Nature. 1975;255(5510):614-7.
83. Janssen R, Schonauer A, Weber M, Turetzek N, Hogvall M, Goss G, et al. The evolution and expression of panarthropod frizzled genes. Frontiers in Ecology and Evolution 2015;3.
84. Janssen R, Turetzek N, Pechmann M. Lack of evidence for conserved parasegmental grooves in arthropods. Dev Genes Evol. 2022;232(1):27-37. Epub 20220117.
85. Benton MA, Pechmann M, Frey N, Stappert D, Conrads KH, Chen YT, et al. Toll Genes Have an Ancestral Role in Axis Elongation. Curr Biol. 2016;26(12):1609-15. Epub 20160519.
86. Liu HK, Wang Y, Belz T, Bock D, Takacs A, Radlwimmer B, et al. The nuclear receptor tailless induces long-term neural stem cell expansion and brain tumor initiation. Genes Dev. 2010;24(7):683-95.
87. Gui H, Li ML, Tsai CC. A tale of tailless. Dev Neurosci. 2011;33(1):1-13. Epub 20101202.
88. Shawlot W, Behringer RR. Requirement for Lim1 in head-organizer function. Nature. 1995;374(6521):425-30.
89. Lilly B, O'Keefe DD, Thomas JB, Botas J. The LIM homeodomain protein dLim1 defines a subclass of neurons within the embryonic ventral nerve cord of Drosophila. Mech Dev. 1999;88(2):195-205.
90. Zhu J, Palliyil S, Ran C, Kumar JP. Drosophila Pax6 promotes development of the entire eye-antennal disc, thereby ensuring proper adult head formation. Proc Natl Acad Sci U S A. 2017;114(23):5846-53.
91. Stolt CC, Wegner M. SoxE function in vertebrate nervous system development. Int J Biochem Cell Biol. 2010;42(3):437-40. Epub 20090730.
92. Ferguson EL. Conservation of dorsal-ventral patterning in arthropods and chordates. Curr Opin Genet Dev. 1996;6(4):424-31.
93. Holley SA, Neul JL, Attisano L, Wrana JL, Sasai Y, O'Connor MB, et al. The Xenopus dorsalizing factor noggin ventralizes Drosophila embryos by preventing DPP from activating its receptor. Cell. 1996;86(4):607-17.
94. Janssen R, Damen WG. Diverged and conserved aspects of heart formation in a spider. Evol Dev. 2008;10(2):155-65.
95. Tingvall TO, Roos E, Engstrom Y. The GATA factor Serpent is required for the onset of the humoral immune response in Drosophila embryos. Proc Natl Acad Sci U S A. 2001;98(7):3884-8. Epub 20010306.
96. Lenz J, Liefke R, Funk J, Shoup S, Nist A, Stiewe T, et al. Ush regulates hemocyte-specific gene expression, fatty acid metabolism and cell cycle progression and cooperates with dNuRD to orchestrate hematopoiesis. PLoS Genet. 2021;17(2):e1009318. Epub 20210218.

97. Lada K, Gorfinkiel N, Martinez Arias A. Interactions between the amnioserosa and the epidermis revealed by the function of the u-shaped gene. Biol Open. 2012;1(4):353-61. Epub 20120216.
98. Egger B, Lapraz F, Tomiczek B, Muller S, Dessimoz C, Girstmair J, et al. A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. Curr Biol. 2015;25(10):1347-53. Epub 2015/04/14.
99. Foelix RF. Biology of Spiders. Oxford: Oxford University Press; 1996.
100. Gold K, Cotton JA, Stollewerk A. The role of Notch signalling and numb function in mechanosensory organ formation in the spider Cupiennius salei. Dev Biol. 2009;327(1):121-31. Epub 20081216.
101. Stollewerk A, Weller M, Tautz D. Neurogenesis in the spider Cupiennius salei. Development. 2001;128(14):2673-88.
102. Stollewerk A, Seyfarth EA. Evolutionary changes in sensory precursor formation in arthropods: embryonic development of leg sensilla in the spider Cupiennius salei. Dev Biol. 2008;313(2):659-73. Epub 20071117.
103. Stollewerk A, Schoppmeier M, Damen WG. Involvement of Notch and Delta genes in spider segmentation. Nature. 2003;423(6942):863-5.
104. Pueyo JI, Lanfear R, Couso JP. Ancestral Notch-mediated segmentation revealed in the cockroach Periplaneta americana. Proc Natl Acad Sci U S A. 2008;105(43):16614-9. Epub 20081016.
105. Chesebro JE, Pueyo JI, Couso JP. Interplay between a Wnt-dependent organiser and the Notch segmentation clock regulates posterior development in Periplaneta americana. Biol Open. 2013;2(2):227-37. Epub 20121219.
106. Brena C, Akam M. An analysis of segmentation dynamics throughout embryogenesis in the centipede Strigamia maritima. BMC Biol. 2013;11:112. Epub 20131129.
107. Green J, Akam M. Evolution of the pair rule gene network: Insights from a centipede. Dev Biol. 2013;382(1):235-45. Epub 20130626.
108. Clark E, Peel AD. Evidence for the temporal regulation of insect segmentation by a conserved sequence of transcription factors. Development. 2018. Epub 20180503.
109. Janssen R. Segment polarity gene expression in a myriapod reveals conserved and diverged aspects of early head patterning in arthropods. Dev Genes Evol. 2012;222(5):299-309. Epub 20120818.
110. Auman T, Vreede BMI, Weiss A, Hester SD, Williams TA, Nagy LM, et al. Dynamics of growth zone patterning in the milkweed bug Oncopeltus fasciatus. Development. 2017;144(10):1896-905. Epub 20170421.
111. Shimizu T, Bae YK, Muraoka O, Hibi M. Interaction of Wnt and caudal-related genes in zebrafish posterior body formation. Dev Biol. 2005;279(1):125-41.
112. Ma P, Yang X, Kong Q, Li C, Yang S, Li Y, et al. The ubiquitin ligase RNF220 enhances canonical Wnt signaling through USP7-mediated deubiquitination of beta-catenin. Mol Cell Biol. 2014;34(23):4355-66. Epub 20140929.
113. Prpic NM, Damen WG. A homolog of the hydrolase Notum is expressed during segmentation and appendage formation in the Central American hunting spider Cupiennius salei. Naturwissenschaften. 2005;92(5):246-9. Epub 20050416.
114. Feitosa NM, Zhang J, Carney TJ, Metzger M, Korzh V, Bloch W, et al. Hemicentin 2 and Fibulin 1 are required for epidermal-dermal junction formation and fin mesenchymal cell migration during zebrafish development. Dev Biol. 2012;369(2):235-48. Epub 20120706.
115. Nagel AN, Marshak S, Manitt C, Santos RA, Piercy MA, Mortero SD, et al. Netrin-1 directs dendritic growth and connectivity of vertebrate central neurons in vivo. Neural Dev. 2015;10:14. Epub 20150610.
116. Lowe CJ, Terasaki M, Wu M, Freeman RM, Jr., Runft L, Kwan K, et al. Dorsoventral patterning in hemichordates: insights into early chordate evolution. PLoS Biol. 2006;4(9):e291.
117. Hiramoto M, Hiromi Y, Giniger E, Hotta Y. The Drosophila Netrin receptor Frazzled guides axons by controlling Netrin distribution. Nature. 2000;406(6798):886-9.

118. Hagedorn EJ, Yashiro H, Ziel JW, Ihara S, Wang Z, Sherwood DR. Integrin acts upstream of netrin signaling to regulate formation of the anchor cell's invasive membrane in C. elegans. Dev Cell. 2009;17(2):187-98.
119. Tavares AT, Tsukui T, Izpisua Belmonte JC. Evidence that members of the Cut/Cux/CDP family may be involved in AER positioning and polarizing activity during chick limb development. Development. 2000;127(23):5133-44.
120. Rescan PY, Ralliere C. A Sox5 gene is expressed in the myogenic lineage during trout embryonic development. Int J Dev Biol. 2010;54(5):913-8.
121. Domsch K, Schroder J, Janeschik M, Schaub C, Lohmann I. The Hox Transcription Factor Ubx Ensures Somatic Myogenesis by Suppressing the Mesodermal Master Regulator Twist. Cell Rep. 2021;34(1):108577.
122. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. Epub 2014/04/04.
123. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21. Epub 20121025.
124. Bruna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 2021;3(1):lqaa108. Epub 2021/02/13.
125. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77. Epub 20120416.
126. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45(4):e18.
127. Donath A, Juhling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, et al. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 2019;47(20):10543-52.
128. Kechin A, Boyarskikh U, Kel A, Filipenko M. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. J Comput Biol. 2017;24(11):1138-43. Epub 20170717.
129. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015;161(5):1202-14.
130. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 2019;8(4):329-37 e4. Epub 20190403.
131. Patterson-Cross RB, Levine AJ, Menon V. Selecting single cell clustering parameter values using subsampling-based robustness metrics. BMC Bioinformatics. 2021;22(1):39. Epub 20210201.
132. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. Gigascience. 2018;7(7).
133. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59-60. Epub 20141117.
134. Prpic NM, Schoppmeier M, Damen WG. Whole-mount in situ hybridization of spider embryos. CSH Protoc. 2008;2008:pdb prot5068. Epub 20081001.

Fig 1. scRNAseq of stage 7, 8 and 9 *P. tepidariorum* embryos.

(A) Three stages of *P. tepidariorum* embryos used for scRNAseq. Stage 7 showing the A – anterior and P- posterior of the germ band. The SAZ – segment addition zone is located in the posterior. The Pc – precheliceral region, Ch – chelicerae, Pp – pedipalps, and four walking legs L1 to L4 are indicated at stage 9. (B) Schematic overview of the single-cell sequencing protocol. Stage specific dissociations were independently barcoded, and cells were processed using SPLiTseq. Each stage was pre-processed before integration and cluster prediction. (C) Metrics of raw data, showing the count per cell, genes per cell and percentage of mitochondrial reads per cell. Stage 7, 8 and 9 with libraries 1 and 2 for each stage. (D) Stage 7, 8 and 9 projections after doublet removal for two libraries show a general increase in data structure through developmental time. Clustering is based on per sample clustering and does not relate to integrated clusters. (E) Integrated data UMAP showing 23 identified clusters. (F) Integrated data UMAP showing stages/libraries. (G) Proportion of cells in each cluster from each stage/library, sub-setting 1800 random cells per stage/library. Sample colours relate to legend in (F). (H) Comparison of the number of cells per cluster from all stages/libraries versus number of marker genes detected per cluster. Cluster colours relate to legend in (E).

Fig 2. Hox expression in scRNAseq data from duplicated Hox cluster A and B.

(A) All 13 Hox markers of clusters hierarchically ordered, showing three main cluster types that relate to pedipalpal, leg bearing and opisthosomal regional identity. (B) Temporal expression of 13 Hox markers across the three developmental stages. (C) Pearson's correlation coefficients of SCT normalised average Hox expression across all 23 clusters. (D) Schematics of Hox gene expression taken from ISH for stage 7, 8 and 9, with a corresponding UMAP using the integrated data. Numbers above gene names are the Pearson's correlations coefficients from (C) with green greater than 0.8, orange between 0.8 and 0.4 and red less than 0.4.

Fig 3. Four cell clusters contributing to the spider precheliceral region.

(A) Cells highlighted for clusters 8, 10, 16 and 19 relating to precheliceral region patterning. (B) Dotplot of markers with ISH data from clusters 8, 10, 16 and 19. (C) Expression of marker genes

Fig 4. Novel cell types in the dorsal head and extra-embryonic regions.

(A) Cells highlighted for clusters 17 and 20 relating to extra embryonic expression. (B) Dotplot of markers with ISH data from clusters 17 and 20. (C) ISH expression of marker genes.

Fig 5. Segmentation clusters, markers and modules.

(A) Cells highlighted for clusters 1, 2, 3, and 5 that relate to segmentation. (B) Dotplot of markers with ISH data from clusters 1, 2, 3, and 5. (C) Double FISH expression of marker genes showing the spatial relationship between markers. (D) UMAP of clusters 2, 3 and 5, with only posterior Hox gene expressing cluster 2 cells. (E) Percentage of cells per clusters, as indicated in (D) for each stage and library. (F) Finer scaled re-clustering of UMAP in (D). (G) Path through clusters and pseudotime estimation of re-clustered data. (H) Four modules detected and their association with clusters from the re-clustering. (I) Association of modules through pseudotime on the UMAP.

Fig 6. Summary of scRNAseq data clustering, annotation, core patterns and segmentation regions.

(A) Annotation of clusters based on ISH data. (B) Hox expression can compartmentalise data into posterior opisthosoma (purple), leg (green), pedipalp (blue), and a region void of Hox expression in the precheliceral region (red). (C) Ventral (red) and dorsal (blue) clusters are separated across the data. (D) Clear segregation of ectoderm clusters (green) from mesoderm clusters (blue) and the endoderm cluster (red). (E) Summary of the four regions of posterior segmentation. Initially region one (red) contains genes that are expressed in the caudal lobe/SAZ, with region three (green) marking the anterior boundary. As segments are added the SAZ forms two distinct regions, with a region two (yellow) that represented by phased

gene expression. Region three maintains a posterior boundary with the SAZ. Region four (purple) represents the demarcation of the posterior compartment (PC) cells that mark mature segments, which occur with region three but also persist and differentiate in formed segments.

S1 Fig. Elbow plots of all samples indicate informative PCAs for downstream use.

Significant PCs are highlighted in red. Transition from red to blue marks the point where the percent change in variation between the consecutive PCs is less than 0.1%.

S2 Fig. *hamlet* (*ham*) expression in scRNAseq data and ISH.

(A) Cells highlighted for cluster 0. (B) Dotplot of *hamlet* expression, which was a marker for clusters 0, 4, 7, 8 and 22. (C) Expression of *hamlet* in the head, ventral midline, SAZ and appendages.

S3 Fig. Hox expression across the A-P and clusters associated with Hox expression.

(A) Summary of Hox gene expression domains during *P. tepidariorum* embryogenesis based on in situ hybridization from Schwager et al. (2017) supplemented with additional *Hox3-A* ISH data (B). Columns represent segments from anterior to posterior. Bars represent the extent of a gene's expression domain with respect to the segments. Opacity of the colours indicates strength of expression. *ftz* and *Antp-A* are expressed dynamically (i.e. budding off stripes) in the SAZ, and *AbdB-B* is continuously expressed in the SAZ until the end of segmentation. Yellow square outlines in what is otherwise the O12 segment depict the SAZ expression patterns. Abbreviations: Ch, cheliceral segment; Pp, Pedipalpal segment; SAZ, segment addition zone. (B) Updated expression of *Hox3-A* in *P. tepidariorum*, with improved detection than previously published [22].

S4 Fig. Double fluorescence ISH for precheliceral region patterning genes.

(A – A''') Expression of *hh* and *Pax6.2*. Expression of *hh* marks the posterior region of *Pax6.2* (A). Both genes expression appears to migrate posteriorly (A' – A'''). The broad domain of

Pax6.2 splits and forms two domains in the ventral posterior region of the precheliceral region, with *hh* at the very posterior (A'''). (B – B''') Expression of *otd1* and *Pax6.1* with the former overlapping with the posterior part of the latter. (C – C''') Expression of Pax6 paralogs starts at the anterior rim and migrates posteriorly. Expression of *Pax6.1* splits at stage 8.1 whereas *Pax6.2* is still broadly expressed at stage 8.1 (C'). By 9.1 both paralogs have split into two domains in the precheliceral region. *Pax6.1* expression is more dorsal and slightly more anterior to *Pax6.2* expression, with only some minor overlap.

S5 Fig. D-V patterning genes in two clusters.

(A) Cells highlighted for clusters 4 and 11 relating to D-V patterning. (B) Dotplot of markers with ISH data from clusters 4 and 11. (C) ISH expression of marker genes.

S6 Fig. Endodermal gut cell markers are found in a single cluster.

(A) Cells highlighted for cluster 22 relating to endodermal gut cells. (B) Dotplot of markers with ISH data from cluster 22. (C) ISH expression of marker genes.

S7 Fig. Prosomal and opisthosomal appendage related clusters.

(A) Cells highlighted for clusters 9, 12 ,13 ,14 ,15 ,18, 21 relating to prosomal and opisthosomal appendages. (B) Dotplot of markers with ISH data from clusters 9, 12 ,13 ,14 ,15 ,18, 21. (C) ISH expression of marker genes.

S8 Fig. Peripheral nervous system is represented by many cells across the embryo.

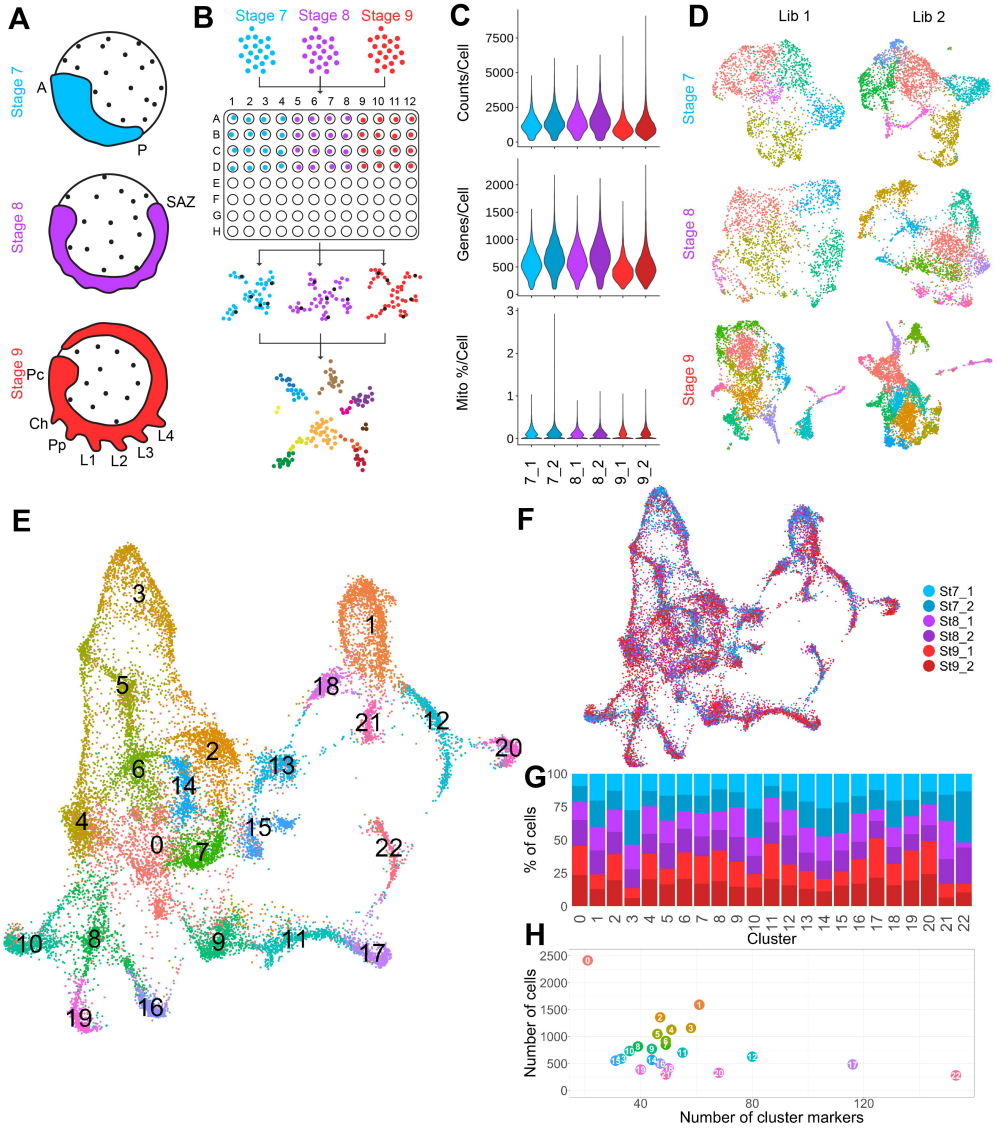
(A) Cells highlighted for clusters 6 and 7 relating to the PNS. (B) Dotplot of markers with ISH data from clusters 6 and 7. (C) ISH expression of marker genes.

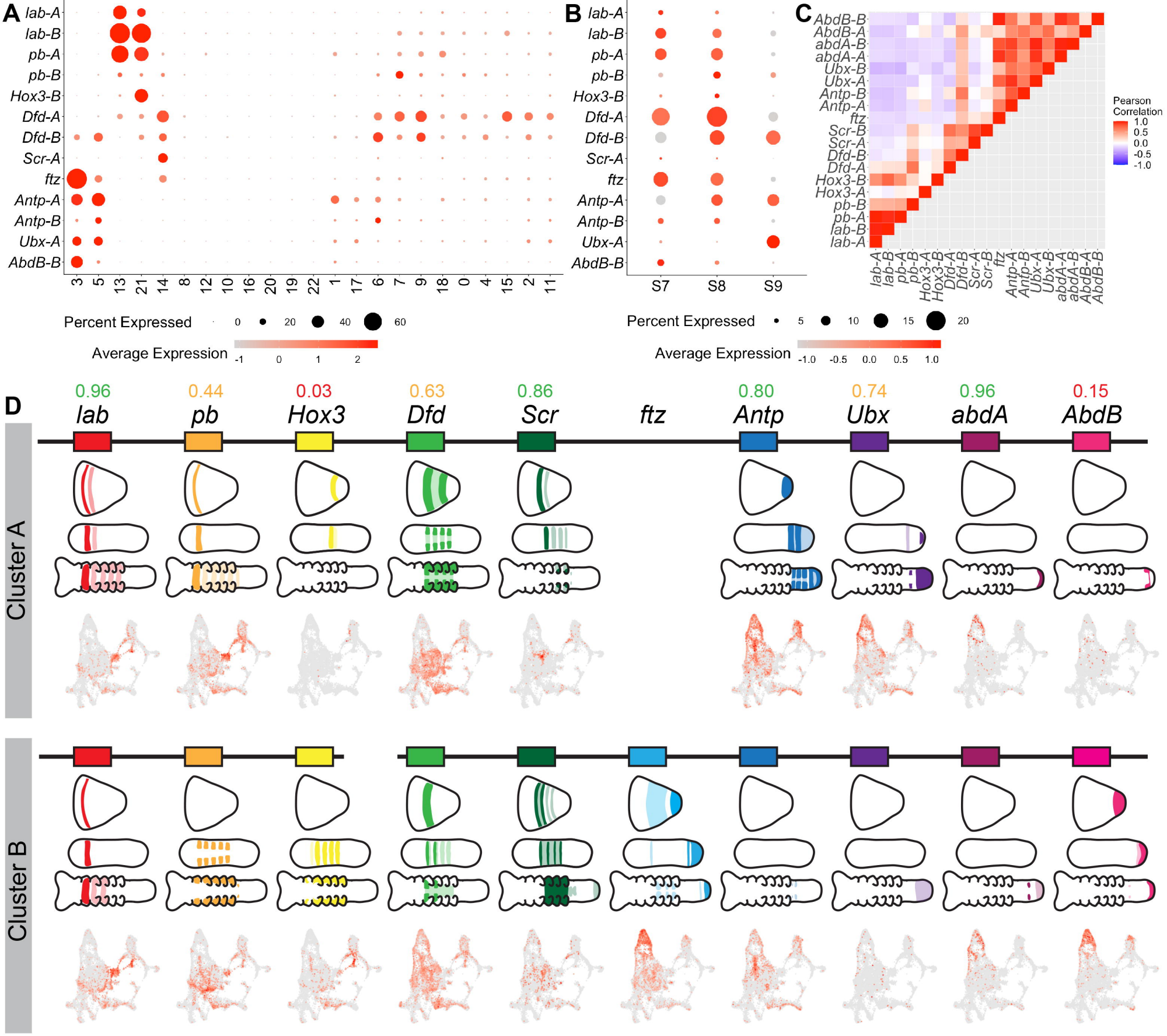
S9 Fig. Single ISH expression of segmentation genes for cluster 1, 2, 3 and 5.

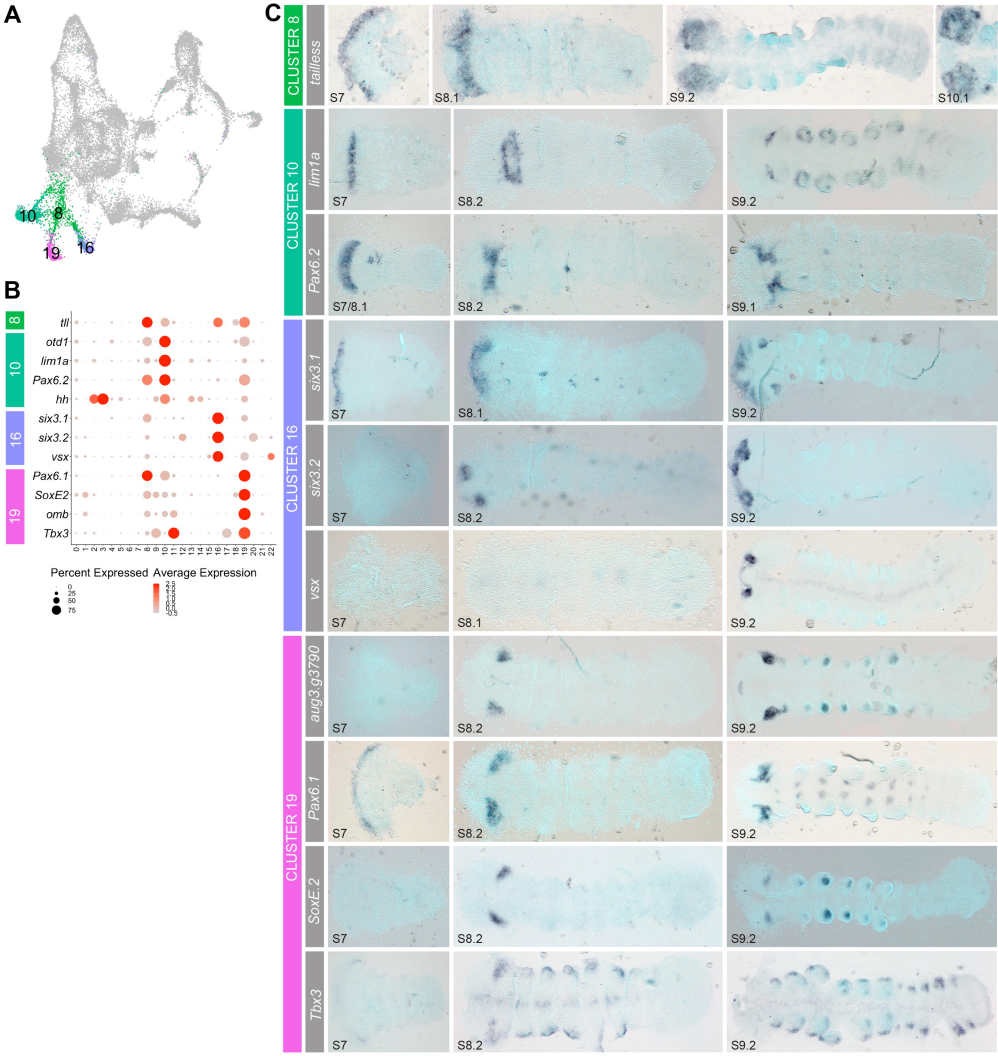
ISH of all markers analysed for clusters 1, 2, 3, and 5, including further information for genes shown in Fig 5C.

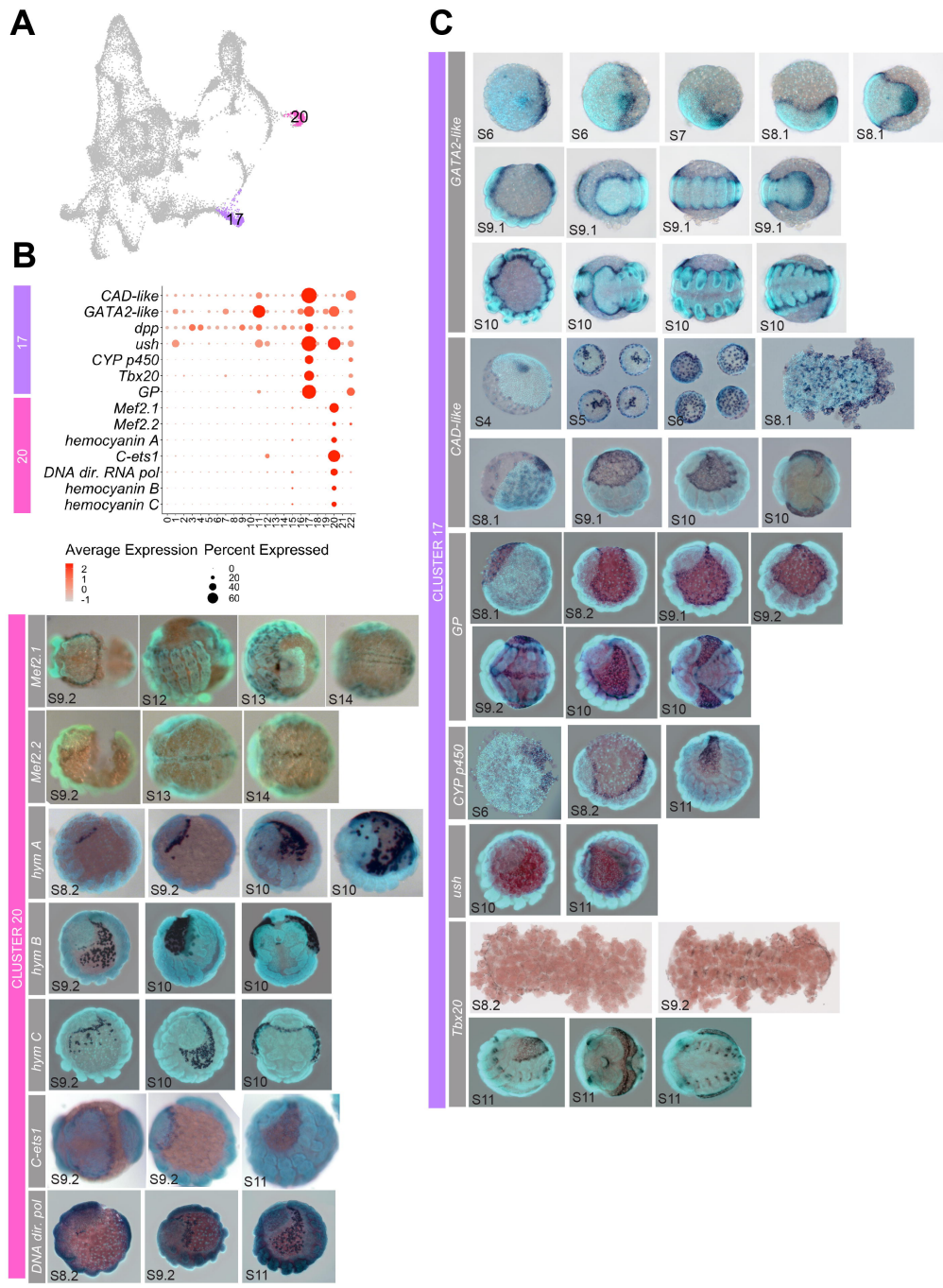
S1 Text. Supplementary methods for SPLiT-seq and in situ hybridisations.

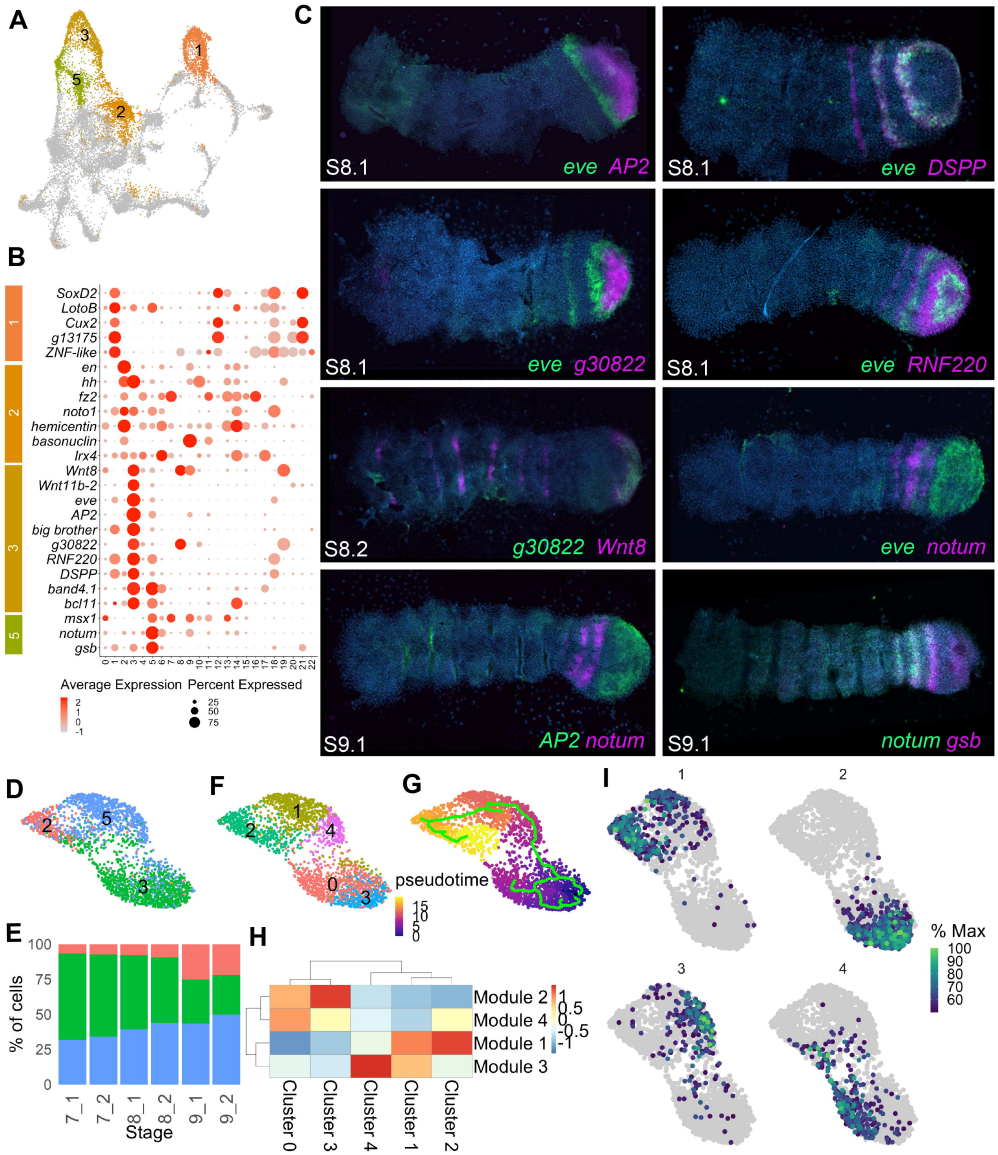
S1 Table. Primer list for in situ hybridisations.

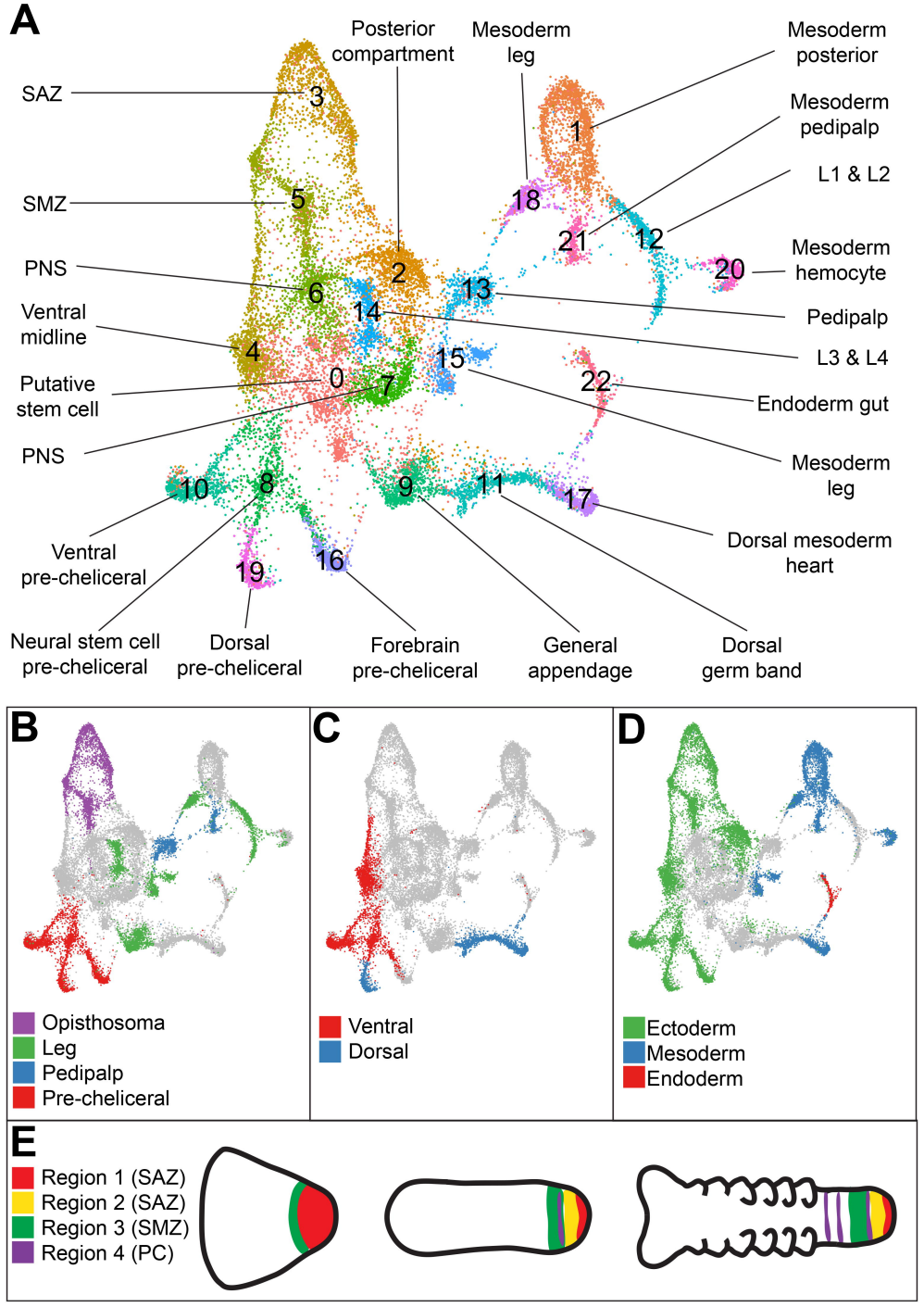

A
Stage 7
A
P
Stage 8
SAZ
Stage 9
Pc
Ch
Pp
L1
L2
L3
L4
B
Stage 7
Stage 8
Stage 9
1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H
C
Counts/Cell
7500
5000
2500
0
Genes/Cell
2000
1500
1000
500
0
Mito %/Cell
3
2
1
0
7_1
7_2
8_1
8_2
9_1
9_2
D
Lib 1
Lib 2
Stage 7
Stage 8
Stage 9
E
3
1
18
5
21
12
20
2
6
13
14
4
15
0
7
22
10
8
9
11
17
19
16
F
St7_1
St7_2
St8_1
St8_2
St9_1
St9_2
G
% of cells
100
75
50
25
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Cluster
H
Number of cells
2500
2000
1500
1000
500
0
40
80
120
Number of cluster markers


A
lab-A
lab-B
pb-A
pb-B
Hox3-B
Dfd-A
Dfd-B
Scr-A
ftz
Antp-A
Antp-B
Ubx-A
AbdB-B
3 5 13 21 14 8 12 10 16 20 19 22 1 17 6 7 9 18 0 4 15 2 11
Percent Expressed 0 20 40 60
Average Expression -1 0 1 2
B
S7 S8 S9
Percent Expressed 5 10 15 20
Average Expression -1.0 -0.5 0.0 0.5 1.0
C
Pearson Correlation
1.0
0.5
0.0
-0.5
-1.0
D
0.96 0.44 0.03 0.63 0.86 0.80 0.74 0.96 0.15
lab pb Hox3 Dfd Scr ftz Antp Ubx abdA AbdB
Cluster A
Cluster B


A
10
8
19
16
B
8
10
16
19
tll
otd1
lim1a
Pax6.2
hh
six3.1
six3.2
vsx
Pax6.1
SoxE2
omb
Tbx3
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Percent Expressed
Average Expression
0
25
50
75
C
CLUSTER 8
tailless
S7
S8.1
S9.2
S10.1
CLUSTER 10
lim1a
S7
S8.2
S9.2
Pax6.2
S7/8.1
S8.2
S9.1
CLUSTER 16
six3.1
S7
S8.1
S9.2
six3.2
S7
S8.2
S9.2
vsx
S7
S8.1
S9.2
CLUSTER 19
aug3.g3790
S7
S8.2
S9.2
Pax6.1
S7
S8.2
S9.2
SoxE.2
S7
S8.2
S9.2
Tbx3
S7
S8.2
S9.2


A
20
17
B
17
20
CAD-like
GATA2-like
dpp
ush
CYP p450
Tbx20
GP
Mef2.1
Mef2.2
hemocyanin A
C-ets1
DNA dir. RNA pol
hemocyanin B
hemocyanin C
0 1 3 4 5 6 8 9 11 12 14 16 17 19 20 21 22
Average Expression
2
1
0
-1
Percent Expressed
0
20
40
60
CLUSTER 20
Mef2.1
S9.2 S12 S13 S14
Mef2.2
S9.2 S13 S14
hym A
S8.2 S9.2 S10 S10
hym B
S9.2 S10 S10
hym C
S9.2 S10 S10
C-ets1
S9.2 S9.2 S11
DNA dir. pol
S8.2 S9.2 S11
C
CLUSTER 17
GATA2-like
S6 S6 S7 S8.1 S8.1
S9.1 S9.1 S9.1 S9.1
S10 S10 S10 S10
CAD-like
S4 S5 S6 S8.1
S8.1 S9.1 S10 S10
GP
S8.1 S8.2 S9.1 S9.2
S9.2 S10 S10
CYP p450
S6 S8.2 S11
ush
S10 S11
Tbx20
S8.2 S9.2
S11 S11 S11


A
3
1
5
2
B
1
2
3
5
SoxD2
LotoB
Cux2
g13175
ZNF-like
en
hh
fz2
noto1
hemicentin
basonuclin
lrx4
Wnt8
Wnt11b-2
eve
AP2
big brother
g30822
RNF220
DSPP
band4.1
bcl11
msx1
notum
gsb
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Average Expression
2
1
0
-1
Percent Expressed
25
50
75
C
S8.1
eve AP2
S8.1
eve DSPP
S8.1
eve g30822
S8.1
eve RNF220
S8.2
g30822 Wnt8
eve notum
S9.1
AP2 notum
S9.1
notum gsb
D
2
5
3
F
1
2
4
0
3
G
pseudotime
15
10
5
0
I
1
2
3
4
% Max
100
90
80
70
60
E
100
75
50
25
0
% of cells
7_1
7_2
8_1
8_2
9_1
9_2
Stage
H
Module 2
Module 4
Module 1
Module 3
1
0.5
0
-0.5
-1
Cluster 0
Cluster 3
Cluster 4
Cluster 1
Cluster 2


A
Posterior compartment
Mesoderm leg
Mesoderm posterior
Mesoderm pedipalp
L1 & L2
Mesoderm hemocyte
Pedipalp
L3 & L4
Endoderm gut
Mesoderm leg
Dorsal mesoderm heart
SAZ
SMZ
PNS
Ventral midline
Putative stem cell
PNS
Ventral pre-cheliceral
Neural stem cell pre-cheliceral
Dorsal pre-cheliceral
Forebrain pre-cheliceral
General appendage
Dorsal germ band
3
1
18
21
12
20
5
2
13
6
14
4
15
0
7
22
10
8
9
11
17
19
16
B
Opisthosoma
Leg
Pedipalp
Pre-cheliceral
C
Ventral
Dorsal
D
Ectoderm
Mesoderm
Endoderm
E
Region 1 (SAZ)
Region 2 (SAZ)
Region 3 (SMZ)
Region 4 (PC)