Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly

Hongjie Li, Jasper Janssens, Maxime de Waegeneer, Sai Saroja Kolluru, Kristofer Davie, Vincent Gardeux, Wouter Saelens, Fabrice David, Maria Brbić, Katina Spanier, et al.

Fly Cell Atlas: a single-cell transcriptomic atlas of the adult fruit fly

Authors: Hongjie Li[1,2,*], Jasper Janssens[3,4,*], Maxime De Waegeneer[3,4], Sai Saroja Kolluru[5], Kristofer Davie[3], Vincent Gardeux[6], Wouter Saelens[6], Fabrice David[6], Maria Brbić[7], Katina Spanier[3,4], Jure Leskovec[7], Colleen N. McLaughlin[1], Qijing Xie[1], Robert C. Jones[5], Katja Brueckner[8], Jiwon Shim[9], Sudhir Gopal Tattikota[10], Frank Schnorrer[11], Katja Rust[12,13], Todd G. Nystul[13], Zita Carvalho-Santos[14], Carlos Ribeiro[14], Soumitra Pal[15], Sharvani Mahadevaraju[25], Teresa M. Przytycka[15], Aaron M. Allen[16], Stephen F. Goodwin[16], Cameron W. Berry[17], Margaret T. Fuller[17,18], Helen White-Cooper[19], Erika L. Matunis[20], Stephen DiNardo[21], Anthony Galenza[22], Lucy Erin O'Brien[22], Julian A. T. Dow[23], **FCA Consortium**[26], Heinrich Jasper[24], Brian Oliver[25], Norbert Perrimon[10]†, Bart Deplancke[6]†, Stephen R. Quake[5]†, Liqun Luo[1]†, and Stein Aerts[3,4]†

Affiliations:
[1] Department of Biology, Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA
[2] Huffington Center on Aging and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[3] VIB-KU Leuven Center for Brain & Disease Research, KU Leuven, Leuven 3000, Belgium
[4] Laboratory of Computational Biology, Department of Human Genetics, KU Leuven, Leuven 3000, Belgium
[5] Departments of Bioengineering and Applied Physics, Stanford University, Stanford CA USA, and Chan Zuckerberg Biohub, San Francisco CA USA
[6] Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland
[7] Department of Computer Science, Stanford University, Stanford, CA 94305, USA
[8] Department of Cell and Tissue Biology, University of California, San Francisco, CA 94143, USA
[9] Department of Life Science, College of Natural Science, Hanyang University, Seoul, Republic of Korea 04763
[10] Department of Genetics, Blavatnik Institute, Harvard Medical School, Harvard University, Boston, MA 02115; Howard Hughes Medical Institute, Boston, MA, USA
[11] Aix-Marseille University, CNRS, IBDM (UMR 7288), Turing Centre for Living systems, 13009 Marseille, France
[12] Institute of Physiology and Pathophysiology, Department of Molecular Cell Physiology, Philipps-University, Marburg, Germany
[13] Department of Anatomy, University of California, San Francisco, CA 94143
[14] Behavior and Metabolism Laboratory, Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal
[15] National Center of Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD 20894, USA
[16] Centre for Neural Circuits & Behaviour, University of Oxford, Tinsley Building, Mansfield road, Oxford, OX1 3SR
[17] Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
[18] Department of Genetics, Stanford University School of Medicine, Stanford CA 94305, USA
[19] Molecular Biosciences Division, Cardiff University, Cardiff, CF10 3AX UK
[20] Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, MD 21205 USA
[21] Perelman School of Medicine, The University of Pennsylvania, and The Penn Institute for Regenerative Medicine Philadelphia, PA 19104 USA
[22] Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA
[23] Institute of Molecular, Cell & Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK.
[24] Immunology Discovery, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA
[25] Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Kidney and Digestive Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[26] FCA Consortium Authors listed before Acknowledgements, and all author contributions and affiliations listed in the Supplemental Materials

* equal contribution
† corresponding authors: perrimon@genetics.med.harvard.edu (N.P.), bart.deplancke@epfl.ch (B.D.), steve@quake-lab.org (S.R.Q.), lluo@stanford.edu (L.L.), stein.aerts@kuleuven.be (S.A.)

Abstract
We present the Tabula *Drosophilae*, a single cell atlas of the adult fruit fly that includes 580k cells from 15 individually dissected sexed tissues as well as the entire head and body. Over 100 researchers from the fly community contributed annotations to >250 distinct cell types across all tissues. We provide an in-depth analysis of cell type-related gene signatures and transcription factor markers, as well as sexual dimorphism, across the whole animal. Analysis of common cell types that are shared between tissues, such as blood and muscle cells, allowed the discovery of rare cell types and tissue-specific subtypes. This atlas provides a valuable resource for the entire *Drosophila* community and serves as a comprehensive reference to study genetic perturbations and disease models at single-cell resolution.

One Sentence Summary: A single-cell transcriptomic map of the entire adult Drosophila melanogaster

Main Text

The fly *Drosophila melanogaster* has a fruitful history in biological research, dating back to experiments of Thomas Hunt Morgan a century ago (*1*). *Drosophila* has been at the basis of many key discoveries in genetics, development, neurobiology, circadian rhythms, stem cell biology, and aging (*2*). The highly collaborative nature of the *Drosophila* community contributed to many of these successes, and led to the development of essential research resources, including a high-quality genome (*3*), a large collection of mutant lines and a large variety of genetic and molecular tools, as well as important databases such as Flybase (*4*), FlyMine (*5*), FlyLight (*6*), VirtualFlyBrain (*7*) and ModERN (*8*). The fruit fly genome contains about 17,000 genes, including an estimated 13,968 protein-coding genes of which ~63% have human orthologues. Studies such as ModENCODE (*9*) and FlyAtlas (*10*) have explored their expression patterns in embryonic, larval, pupal, and adult tissues, but have lacked cell type resolution. Recent advances in single-cell technologies have enabled the transcriptomic profiling of thousands of cells at once, leading to greater insight in tissue composition and differential gene expression. Several studies have already applied single-cell RNA sequencing (scRNA-seq) on multiple *Drosophila* tissues and developmental stages (*11*, *12*). However, these data have been generated by different laboratories on different genetic backgrounds, with different dissociation protocols and sequencing platforms, hindering systematic comparison of gene expression across cells and tissues.

Here, we present a single cell transcriptomic atlas of the entire adult *Drosophila,* separately analyzing male vs female samples, using a uniform genotype and a unified single-nucleus RNA-seq (snRNA-seq) platform (*13*) with two sequencing strategies: droplet-based 10x Genomics and plate-based Smart-seq2. The resulting Tabula *Drosophilae*, the first comprehensive dataset within the Fly Cell Atlas consortium, contains over 580k cells, resulting in >250 distinct cell types annotated by >100 experts from 40 laboratories. This atlas reports cellular signatures for each tissue, providing a valuable resource for the entire *Drosophila* community as a reference for studies that probe the effects of genetic perturbations and disease models at single-cell resolution. All data and

annotations can be accessed through multiple visualization and analysis portals from https://flycellatlas.org, also summarized in fig. S1-S3.

Sampling single cells across the entire adult fly

We used a unified snRNA-seq platform for all samples, because it is difficult to isolate intact cells from many adult *Drosophila* cuticular tissues, such as the antenna, wing, leg, and haltere. In addition, snRNA-seq can be applied to large multinucleated cells (e.g., muscle) and facilitate (frozen) tissue collection from different laboratories. A recent study (*13*) revealed that 70–90% of transcriptomic information was preserved from snRNA-seq compared to scRNA-seq of the same fly cell types.

To achieve a comprehensive sampling of cell types, we used two complementary strategies. First, we dissected 12 individual tissues from both males and females, plus 3 sex-specific tissues (see Methods). snRNA-seq was performed using droplet-based 10x Genomics (*14*) and plate-based Smart-seq2 (*15*) (**Fig. 1A**). For tissues (fat body, oenocytes, and trachea) that are localized across the body and cannot be directly dissected, we used specific GAL4 lines driving nuclear-GFP to label nuclei and collected them using FACS (see Methods). In addition, two rare cell types were sequenced only with Smart-seq2: insulin-producing cells (IPCs) and corpora cardiaca cells (CCs). Second, we sorted nuclei from the entire head and body of sexed flies and profiled them on 24 10x Chromium lanes, aiming to detect cell types that may not be covered by the selected tissues. In total, we obtained 580k high-quality cells: 570k from 10x Genomics of 63 runs and 10k from Smart-seq2 of 35 384-well plates (**Fig. 1A)**.

To analyze the 10x Genomics data in a reproducible manner, we used an automated computational pipeline called VSN (see Methods), which takes the raw sequencing data as input, performs filtering, normalization, doublet removal, batch effect correction, dimensionality reduction, and clustering, and produces *LoomX* formatted files with expression data, embeddings and clusterings (**Fig. 1B** and **fig. S4**). As we observed a presumed artefactual cluster of cells showing expression of nearly all genes, we added an additional preprocessing and filtering step that models and then subtracts ambient RNA from the gene expression values (*16*), resulting in computational removal of this cluster. However, since adjusting the gene expression values per cell can also introduce other biases (e.g., overcorrection, removal of non-doublet cells), we retained both a *Relaxed* dataset of 570k cells, and a *Stringent* dataset of 510k cells (see Methods and **Fig. 1C**). In the analyses below, unless otherwise mentioned (e.g., Fig. 2C), we focus on the *Stringent* dataset.

Cells from 10x Genomics and Smart-seq2 were well aligned after batch correction using Harmony (*17*) (**Fig. 1D**). As expected, Smart-seq2 yielded a higher number of detected genes for most tissues (**Fig. 1E**) as cells were sequenced to a higher depth. We analyzed each tissue separately, combining the male and female runs, which yielded between 6.5k (haltere) and 100k (head) cells per tissue with a median of 16.5k cells per tissue for 10x and between 263 (male reproductive gland) and 1,349 (fat body) cells per tissue with a median of 534 cells for Smart-seq2 (**Fig. 1F**). We obtained a similar number of male and female cells for non-sex-specific tissues. We retrieved on average 1895 unique molecular identifiers (UMIs) and 828 genes per cell with variations between tissues (**fig. S5**), suggesting that different tissues have different numbers of active genes, similar to other atlas data such as Tabula Muris (*18*). Next, all cells were combined in a meta-analysis, showing tissue-specific clusters like the germline cells of the testis and ovary, and shared clusters of common cell types, such as neurons, glial cells, fat cells, epithelial cells, hemocytes, and muscle cells from different tissues (**Fig. 1G**; see fig. S24, 25).

Crowd-based cell type annotation by tissue experts

Experts from 40 laboratories collaborated on cell type annotation for 15 individual tissues, including 12 tissues for both males and females: antenna, body wall, fat body, haltere, heart, gut, leg, Malpighian tubule, oenocyte,

proboscis with maxillary palp, trachea, and wing; and 3 sex-specific tissues: male reproductive gland and testis, and female ovary (**Fig. 2A**). For this, we developed a consensus-voting strategy within the SCope web application (https://flycellatlas.org/scope) (*19*), allowing curators and users to access and annotate the snRNA-seq data simultaneously. Tissue experts worked with the SCope team in a series of virtual 'jamborees' to annotate clusters at multiple resolutions, with additional analysis performed in ASAP (https://flycellatlas.org/asap) (*20*). If the same cluster was annotated with two terms, the voting system allows users to vote for their preferred annotation. To ensure that cell type annotations are consistent with previous literature and databases and to allow *a posteriori* computational analyses at different anatomical resolutions, we linked the controlled vocabulary to the Flybase anatomy ontology (*21*).

Since some cell types are present at low clustering resolution, and others appear only at higher resolution, we used the VSN pipeline to calculate different Leiden resolutions ranging from 0.8 to 8 (**fig. S6A**). Annotations were then collapsed across resolutions and the annotation with the highest number of up-votes was selected per cell. All initial annotations were performed on the *Relaxed* dataset, and were then exported to the *Stringent* dataset, where field experts verified the accuracy of the annotation transfer (**Fig. 2 and fig. S6-S18**). Overall, we annotated 251 cell types in the *Stringent* dataset (262 cell types if combining *Relaxed* and *Stringent datasets,* Table S1), with a median of 15 cell types per tissue.

Our dataset provides the first single-cell transcriptomic profiling for several adult tissues, including the haltere, heart, leg, Malpighian tubule, proboscis, maxillary palp, trachea, and wing (**fig. S6–S18**). In these tissues, all major expected cell types were identified. For example, in the proboscis and maxillary palp (**fig. S7A, B**), we could annotate gustatory and olfactory receptor neurons, mechanosensory neurons and several glial clusters, and all 7 olfactory receptors expressed in the maxillary palp could be detected. In the wing (**fig. S8**), we could identify four different neuronal types – gustatory receptor neurons, pheromone-sensing neurons, nociceptive neurons and mechanosensory neurons, as well as three glial clusters. In the leg (**fig. S9**), we could distinguish gustatory receptor neurons from two clusters of mechanosensory neurons. In the heart (**fig. S10**), we found a large proportion of resident hemocytes and muscle cells, with the cardial cells marked by the genes *Hand* and *tinman* constituting a small proportion. In the Malpighian tubule (**fig. S11**), we were able to identify 15 cell types, including the different principal cells of the stellate and main segments. In the haltere (**fig. S13**), we identified two clusters of neurons, three clusters of glial cells and a large population of epithelial cells. In some tissues, one cell type formed a big cluster instead of being split into distinct populations. In these cases, we tried to identify genes or pathways that showed gradient or compartmentalized expression. For example, in the fat body (**fig. S14 and S19**), the main fat body cells formed one big cluster, but our metabolic pathway enrichment analysis performed through ASAP (*20*) revealed that fatty acid biosynthesis and degradation are in fact compartmentalized, highlighting possible fat body cell heterogeneity in terms of metabolic capacities.

Our crowd annotations with tissue experts also revealed many cell types not profiled previously, such as multinucleated muscle cells (**Fig. 2B**) and two distinct types of nuclei among main cells in the male accessory gland (**fig. S17**), a cell type previously thought to be uniform. The high number of nuclei analyzed allowed identification of rare cell types. For example, in the testis (**Fig. 2C**), one of the best characterized but also most complex tissues in the fly, our FCA data allowed identification of 25 unique cell types, covering all expected cell types, including very rare cells, such as germinal proliferation center hub cells (79 nuclei in the Relaxed version, out of 44,621 total testis nuclei). In addition to annotation of differentiated cell types, the snRNA-seq data from testis captured and displayed in the geography of the UMAP the dynamic sequence of gene expression changes in the resident germline and somatic cyst stem cell lineages, identifying many genes that alter expression at specific steps in the differentiation process (see trajectory analyses in **Fig. 6**). Clustering at higher resolution (Leiden 6) was especially

useful in these stem cell lineages as it highlighted as an emergent property the stepwise changes in gene expression as the development proceeds.

Next, we compared the distribution of cells between 10x and Smart-seq2 for all tissues profiled by both sequencing technologies. In almost all cases, cells from two technologies were well matched from co-clustering analysis (**fig. S20 and S21**). Since Smart-seq2 cells only account for a small fraction, our previous annotations focused on 10x cells. The cell-matched co-clustering analysis allowed us to directly transfer annotations from 10x to Smart-seq2 datasets (**fig. S20E**), validating the transfer using cluster-specific markers (**fig. S20F**). We also identified genes specifically detected using Smart-seq2, presumably because the deeper sequencing allowed a higher gene detection rate (**fig. S20G** and **Fig. 1E**). In summary, the 10x datasets, containing many more cells, provide an important resource for identifying cell types while the Smart-seq2 datasets facilitate the detection of low-expressed genes and enable future exploration of cell-specific isoform information.

Correspondence between dissected tissues and whole head and body

To generate a complete atlas of the fly, we performed snRNA-seq experiments on whole-head and whole-body samples, inspired by earlier work on whole-body single-cell experiments of less complex animals (*22*, *23*). Full head and body sequencing can provide a practical means to assess the impact of mutations or to track disease mechanisms, without having to focus on, or dissect, specific tissues. In addition, it could yield cell types that are not covered by any of the targeted tissue dissections.

First, we analyzed the head and body separately. In the head, we annotated 81 unique cell types with most cells being neurons (**Fig. 3A and S19**). We identified photoreceptors and cone cells based on *gl* and *lz* expression respectively, olfactory receptor neurons with *orco* expression, glial cells based on *repo* expression, head hemocytes based on *Hml* expression, and different types of muscle and epithelial cells. In the body, we annotated the top 33 most abundant cell classes, including epithelia, muscle, and ventral nerve cord and peripheral neurons, followed by fat body cells, oenocytes, germ line cells, glial cells, and tracheal cells (**Fig. 3B and S20**). Many of these cell classes can be further divided into cell types for further annotation, as from tissue sequencing (see **Fig. 2** and **fig. S6-S18**).

Next, we examined how well the body and head samples covered the cell types from the dissected tissues. Analyzing the head and body samples together with the tissue samples revealed that most of the selected tissues cluster together with the body sample, as expected. We also detected head and body enriched clusters (**Fig. 3C**) and even clusters without cells from the dissected tissues (**Fig. 3D**). One body-specific cluster was cuticle cells, likely from connective tissue. Others were relatively rare cell types in their respective tissues, such as adult stem cells. Conversely, most tissue clusters contained body cells, with only a small number being completely specific to dissected tissues. As tissue-specific clusters were mostly observed in tissues with high cell coverage, such as the testis and Malpighian tubule, we anticipate that these clusters would also be identified in the body upon sampling a larger number of cells.

For head, antenna and proboscis with maxillary palp were dissected for tissue sequencing. Cell types from those two tissues largely overlapped with head sequencing data. Many other cell types, such as central brain cells, including Kenyon cells (*ey*, *prt*) and lamina glia (*repo*, *Optix*), were only detected in the head sample.

To compare our data with existing dataset, we integrated our head snRNA-seq dataset ("head" hereafter) with previously published brain single-cell RNA-seq data ("brain" hereafter) (*19*, *24–26*) (**Fig. 3E**). Clusters unique to head made up 20% of the cells, including the antennae, photoreceptors, muscle, cone cells and cuticular cell types, while the other 80% were present in mixed clusters containing both head- and brain-derived cells covering the neuronal and glial cell types of the brain. This co-clustering across genotypes and protocols underscores the usefulness of our snRNA-seq data, as well as its quality compared to conventional scRNA-seq data. Next, we used

a neural network for optic lobe cell types and trained a support-vector machine on a whole brain sample to predict annotations per cluster that were manually checked in an additional Jamboree. Given the high number of neuron types, additional automatic subclustering was performed on each cluster, identifying different groups of peptidergic neurons based on *dimm* and *Pdf* expression and olfactory projection neurons based on *oaz*, *c15,* and *kn*. Finally, we identified many cell types in the optic lobe, including lamina (e.g. L1–L5), medulla (e.g. Mi1, Mi15), lobula (e.g. LC), and lobula plate (e.g. LPLC). Using *acj6* and *SoxN*, we identified the T4/T5 neurons of the optic lobe that split in T4/T5a-b and T4/T5c-d subtypes by subclustering. Lastly, we identified the major types of glia, leading to a total of 81 cell types in the head. It is worth noting that many neurons cluster together into a big clump that we were unable to annotate (**Fig. 3A**), indicating that our dataset does not contain a sufficient number of cells to resolve the complexity of the central brain, which may contain hundreds or thousands of neuron types.

Subclustering analysis performed on cells within larger groups was especially useful in identifying cell types in the combined dataset (**Fig. 3F**). This separated photoreceptors into inner and outer, with the inner photoreceptors further splitting into R7 and R8 subtypes, each with pale and yellow types based on *rhodopsin* expression. Additionally, subclustering identified dorsal rim area and ocellar photoreceptors and further split Kenyon cells into the three different KC subtypes: α/β, α'/β' and γ based on known marker genes (*19*). These case studies highlight the resolution in our dataset and the potential of using subclustering to discover rare cell types.

Cross-tissue analyses allow comparison of cell types by location

Using the whole body and head sequencing data, we created a compilation of the entire fly, assigning cells to major cell classes (e.g. epithelial cells, neurons, muscle cells, hemocytes) using a hierarchy of ontology terms and marker gene expression (**Fig. 4A-C, S22, S23**). Epithelial cells and muscle cells were the most abundant cell classes in the body, while neurons were the most abundant in the head. These assignments allowed us to compare common cell classes across tissues (**fig. S24, S25**).

First, we compared blood cells, termed hemocytes in the fly, across tissues by selecting all *Hml*-positive cells, a known marker for hemocytes (**Fig. 4D**). Hemocytes are generated in the embryo and larva and become resident cells in many tissues. Combining hemocytes across tissues revealed a major group of plasmatocytes, the most common hemocyte type (~56%), crystal cells (1.5%, PPO1, PPO2), and several unknown types (**fig. S26A, B**). Looking deeper into the plasmatocytes, we uncovered different gradients of gene expression that were not seen in the plasmatocyte clusters of individual tissues. This gradient is based on the expression of *Pxn*, *LysX*, *Tep4*, *trol* and *Nplp2* and can be linked to maturation of hemocytes and their plasticity with *Pxn* positive cells showing the highest *Hml* expression, while *Tep4*, *trol* and *Nplp2* are prohemocyte markers (*27–29*). Furthermore, different antimicrobial peptide (AMP) families such as the *Attacins* and *Cecropins* were expressed in different subgroups of plasmatocytes, indicating further specialization. Finally, expression of acetylcholine receptors was specific for a subset of hemocytes, which could be related to the cholinergic anti-inflammatory pathway as described in humans and mice (*30*, *31*). Lamellocytes were not observed in adults as previously suggested (*32*). On the contrary, an unknown hemocyte type expressed *Antp* and *kn* (43 cells, 0.5%) reminiscent of the posterior signaling center in the lymph gland, an organization center previously thought to be absent in the adult (*33*, *34*) (**Fig. S26B**). These findings highlight the value of performing a whole organism-level single cell analysis and constitute a foundation to investigate the fly immune system in greater detail.

Second, we compared the muscle cells of the different tissues (**Fig. 4E, S23C**). Muscle cells are syncytia—individual cells containing many nuclei, and to our knowledge have not been profiled by single-cell sequencing prior to our study. With snRNA-seq, we recovered all known muscle cell types, with specific enrichment in the body, body wall, and leg. The integrated analysis of all muscle cells provided a comprehensive view of the fly

muscular system, highlighting separation of visceral, skeletal, and indirect flight muscle based on the expression of different *troponins*. Specifically, we discovered previously unreported gradients of *dysf* and *fln* in the indirect flight muscle, which may indicate regional gradients of transcriptional activity in these very large cells containing more than 1000 nuclei per cell (**fig. S26D**). In visceral muscles, we noted a separation of types based on neuropeptide receptors being expressed. We identified 4 types of visceral muscle in the gut based on expression of the *AstC, Ms, Dh31* and *CCAP* neuropeptide receptors, indicating potential modulators for muscle contraction (*35*). Receptors for Ms and Dh31 have been described to be expressed in spatially restricted domains in the gut, suggesting that the other two neuropeptides might also exhibit spatial domains (*36*, *37*). Furthermore, the corresponding neuropeptides are released by neuroendocrine cells in similar spatially bound sections (*38*). All visceral muscle cells, including the ovarian sheath muscle, are enriched for the receptor of *Pdf*, a neuropeptide involved in circadian rhythms in the brain, which has also been found to regulate ureter contractions, pointing towards a dual role of circadian rhythm and muscle contraction (*39*).

Transcription factors and cell type specificity

Our data allow comparison of gene expression across the entire fly for 251 annotated cell types. Clustering cell types showed the germline cells as the most distinct group, followed by neurons (**fig. S27-S32**). Marker genes were calculated for every cell type using the whole FCA data as background, with 14,240 genes found as a marker for at least one cell type and a median of 638 markers per cell type [min: visceral muscle (94), max: spermatocyte (7736)]. Notably, markers specific for cell types in a tissue were not always specific in the whole body **(fig. S33)**.

To begin to understand what underlies cell-type-specific transcriptomes, we analyzed the expression of all predicted transcription factors (TFs) (*4*) encoded by the fly genome, and investigated whether certain TFs are specifically expressed in a single cell type (**Fig. 5A**). We used the *tau* score of tissue specificity (*40*) and identified 500 TFs with a score > 0.85, indicating a high specificity for one or very few cell types (**Table S2**). 127 of these TFs were "CG" numbers (computed genes), indicating that their functions are poorly studied. These patterns provide hypotheses about where these genes may function. We found that the male germline stands out in showing expression of a great number of cell type-specific TFs. This may be related to the broad activation of many genes in late spermatocytes discussed below.

Similar analysis across broad cell types **(Fig 5B, C)** identified 156 TFs with high *tau* scores, for example the previously known regulators *grh* for epithelial cells and *repo* for glia; as well as 24 uncharacterized genes. These factors can be visualized in a network, also taking semantic similarity into account for linking similar terms (see Methods). This led to the grouping of CNS neurons and sensory organ cells, including many sensory neurons, with shared pan-neuronal factors such as *onecut* and *scrt* but each cluster having a unique set of TFs, such as *ey*, *scro* and *dati* for CNS neurons and *lz* and *gl* for sensory neurons. Further, *esg*, *Sox21a* and *Sox100B* expression was mostly restricted to somatic precursor cells.

In addition to the specificity of TF expression, we predicted gene regulatory networks based on co-expression and motif enrichment using the SCENIC method (*41*). Because of the stochasticity of this network inference method, we ran SCENIC 100 times, enabling the ranking of predicted target genes by their recurrence. This approach selected 6112 "regulons" for 583 unique TFs across all tissues, whereby each regulon consists of the TF, its enriched motif, and the set of predicted target genes. In fat cells, our analysis predicted a regulon for *sugarbabe* (*sug*), a sugar-sensitive TF necessary for the induction of lipogenesis (*42*). In photoreceptors, the analysis identified a *glass* (*gl*) regulon, with key photoreceptor markers such as *Arr1*, *eya* and multiple rhodopsins as predicted target genes **(Fig. 5D, E)**. This is in line with previous studies identifying *gl* as a key photoreceptor

regulator (*43*, *44*). The SCENIC predictions for all cell types can be easily investigated via the SCope browser (https://flycellatlas.org/scope).

Comparative analysis of genes across broad cell types or all tissues (**Fig. 5F, fig. S34**) allowed identification of common genes and specifically expressed genes, such as a shared set of 555 housekeeping genes that are expressed in all tissues. There was a wide spectrum between tissues, with the testis having the highest number of uniquely expressed genes consistent with previous reports (*45–47*), followed by the Malpighian tubule and male reproductive glands (**fig. S34**). These tissue-specific genes seemed to be evolutionarily "younger" based on GenTree age compared to the set of commonly expressed genes that are all present in the common ancestor. This suggests that natural selection works on the tissue specialization level, with the strongest selection on testis, male reproductive tract, and Malgpighian tubules (*48*). In addition, this analysis allowed an estimation of transcriptomic similarity or difference measured by the number of shared unique genes. For example, the two flight appendages, the haltere and wing, share a set of 16 uniquely expressed genes including *vestigial*, a key factor for wing and haltere development (*49*), reflecting the evolutionary origin of halteres as a modified wing (*50*) (**fig. S34**).

In conclusion, TF expression analysis reveals many known regulators, with specific expression across the adult, as well as dozens of previously uncharacterized genes with specific expression that constitute promising candidates for downstream functional characterization.

Analysis of sex-biased expression and sex-specialized tissues

Much of the genome showed sex-biased expression (*51*), presumably to support the function of highly sex-specialized gonads and reproductive tract tissues. Sequencing male and female samples separately allowed us to study sex-related differences in tissues common to both sexes. Comparing male- versus female-derived nuclei for all common tissues (**fig. S35**), we found the top male- or female-specific genes were known male or female markers (*roX1/2*, *Yp1/2/3*). Notably, a large fraction of genes with male-enriched expression were CGs (*52*). The primary sex determination pathway in somatic cells is activated by a two-X-chromosome karyotype followed by an alternative splicing cascade resulting in sex-specific splicing of *doublesex* (*dsx*) to encode female- or male-specific TFs (*53*) (**Fig. 6A**). Consistent with this, we found *dsx* expression in a largely non-sex-specific pattern, while many other genes showed sex-biased expression (**Fig. 6B**).

As many cell types common to both sexes showed sex-biased expression, we performed differential expression between sexes for all cell types. Notably, cell types tended to show either high female- or male-bias, not both (**Fig. 6B-C**). We found strong female-bias in the excretory system, including the principal and stellate cells of the Malpighian tubule (MT) and in the pericardial nephrocytes (**Fig. 6C**). Examples of genes expressed with female-bias include *eEF1A1*, *eEF2*, *irk3*, and *Rack1*, which are also differentially expressed when Cl^- channels are disrupted in a *Drosophila* cystic fibrosis model (*54*). Other female-biased genes (i.e., *Ics* and *whe*) were differentially expressed under high salt conditions (*55*), suggesting sex-bias in nephric ion transport. Across cell types, sex-biased expression strongly correlated with *dsx* expression (**Fig. 6D**) (*56*, *57*), consistent with the role of Dsx as a key regulator.

For analysis of sex-specialized tissues, we focused on the testis plus seminal vesicle as a case study, as this dissected sample had one of the most complex cell type compositions, the highest number of 'unique' marker genes **(Fig. 5F)**, and two dynamic differentiation trajectories. Among all tissues in the adult fly, those best characterized that have ongoing cellular differentiation are the gut, ovaries, and testis. Trajectory analysis has been performed on gut and ovary stem cell lineages in previous studies (*58–60*), and our FCA data on gut and ovary accurately co-clustered with these published datasets (**fig. S36, S37**).

The testis has two populations of stem cells, the somatic cyst stem cells (CySCs) that produce cell types

with supporting roles essential to spermatogenesis, and the germline stem cells (GSCs) that produce haploid sperm (**Fig. 2C**). Trajectory inference on germ cells from the testis (**Fig. 6E-F**) revealed transitions from GSCs and proliferating spermatogonia, to cells expressing early spermatocyte markers, followed by progressively maturing spermatocytes, early spermatids, and finally mid-to-late elongation stage spermatids (**Fig. 2C**). As expected, the spermatocyte stage featured a robust increase in the number of genes being transcribed (high UMI) (**Fig. 6E**), leading to a substantial rise in RNA complexity. Many of the genes that are strongly upregulated in spermatocytes (*kmg, Rbp4, fzo, can, sa*, and, for later spermatocytes, Y-linked fertility factors *kl-3* and *kl-5*) are not substantially expressed in most other cell types. An unexpected finding in the FCA data was that late spermatocytes showed expression of many genes that were normally thought of as markers of somatic cells (*Upd1* and *eya*, for example). Notable otherwise cell type-specific examples were marker genes from **Fig. 4B**, *grh* (epithelial cells), *Mhc* (muscle), and *Hml* (**Fig. 5A**), although the level of expression of these genes in spermatocytes was considerably lower than in their cognizant marked tissue. This late spermatocyte expression may contribute to the expression of many transcription factors in male germ cells reported in Figure 5A. Early spermatids, as expected from their transcriptional quiescence, had a very low number of nuclear transcripts (**Fig. 6F**, low UMI). Later, elongating spermatids turn on a coordinated burst of new transcription as previously identified (*61*). Our snRNA-seq study identified many new such genes (**Fig. 6F**). Again, the comprehensive Tabula *Drosophilae* snRNA-seq effort identified many new genes to be investigated for roles in differentiation of specific cell types.

The early stages of the somatic cyst cell lineage also emerged as a clear trajectory, starting from CySCs expressing the cell cycle marker *string*, transitioning into post-mitotic (no *string* expression) early cyst cells, then branching into two related clusters of cyst cells likely associated with spermatocytes (**Fig. 6G)**. Together, the FCA data from the progression of germ cell types and early somatic cyst cell types identify many additional genes differentially up or down regulated at specific steps in the respective stem cell lineages, presenting a unique opportunity for functional tests to identify new players in the transitions from one cell state to the next.

Discussion

Recent technological development has enabled single-cell transcriptomic atlases of *C. elegans* (*23*) and selected tissues in mice and humans (*18*, *62–65*). Here, we provide the first single-cell transcriptomic map of the entire adult *Drosophila melanogaster*, a premier model organism for studies of fundamental and evolutionarily conserved biological mechanisms. The Fly Cell Atlas provides a valuable resource for the *Drosophila* community as a reference for studies of gene function at single-cell resolution. The ability to annotate cell types by whole-head and whole-body sequencing facilitate the use of *Drosophila* to study biological processes and to model human diseases at a whole-organism level with cell-type resolution.

A key challenge in large-scale cell atlas projects is the definition of cell types. We addressed this using a consensus-based voting system across multiple resolutions. An FCA cell type is thus defined as a cluster detected at any clustering resolution that could be separated by the expression of known marker genes from other clusters. Further, all annotations were manually curated by tissue experts, leading to a high-confidence dataset with over 250 annotated cell types. We note differences in annotation depth for different cell groups, with some cell types only linked to broad classes (e.g. epithelial cell, CNS neuron), in contrast to more detailed cell types (e.g., different types of olfactory receptor neurons). We also note that while many marker genes are useful in identifying cell types, some marker gene expression was not congruent with cluster expression. This can be caused by discrepancies between mRNA expression and protein or by mistakes made in literature. These examples highlight the need and the opportunities presented by Tabula *Drosophilae* to serve as the basis for future validation.

Comparisons of our snRNA-seq dataset with previously published scRNA-seq datasets from three tissues—

brain (**Fig. 3E**), gut (**fig. S36**) and ovary (**fig. S37**)—not only validated consistency of these methods, but also revealed additional cell types not detected in previous studies. For example, the FCA dataset has increased our knowledge of late-stage follicle cells and allowed identification of new marker genes that were validated using *in vivo* genetic reporters (**fig. S37**). In addition, snRNA-seq offers a more immediate picture of changes in transcript expression as cells progress, which may be important for dynamic cell types differentiating in adult stem cell lineages.

The list of cell type-specific marker genes, in particular TFs, provides a blueprint for generating cell type-specific reporters and expression drivers for studying new aspects of cellular function. However, cell type-specific markers in one tissue may be expressed in more cell types in other tissues. The FCA dataset presents the first opportunity to distinguish marker genes on different levels of specificity.

Our analysis also presents several technical novelties, including the use of Nextflow pipelines improving reproducibility (VSN, https://github.com/vib-singlecell-nf), the availability of raw and processed datasets for users to explore, and the development of a crowd-annotation platform with voting, comments and references via SCope (https://flycellatlas.org/scope). These elements may inspire future atlassing projects. Indeed, FCA data together with atlases from other species will allow cross-species studies. Furthermore, Tabula *Drosophilae* is fully linked to existing *Drosophila* databases by a common vocabulary, benefitting its use and integration in future projects. Finally, all the FCA data are freely available for further analysis via multiple portals and can be downloaded for custom analysis using other single cell tools (**fig. S1**) (all links can be found on https://www.flycellatlas.org).

References and Notes

1. T. H. Morgan, SEX LIMITED INHERITANCE IN DROSOPHILA. *Science*. 32, 120–122 (1910).
2. H. J. Bellen, C. Tong, H. Tsuda, 100 years of Drosophila research and its impact on vertebrate neuroscience: a history lesson for the future. *Nat. Rev. Neurosci.* 11, 514–522 (2010).
3. M. D. Adams, S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, J. C. Venter, The genome sequence of Drosophila melanogaster. *Science*. 287, 2185–2195 (2000).
4. A. Larkin, S. J. Marygold, G. Antonazzo, H. Attrill, G. Dos Santos, P. V. Garapati, J. L. Goodman, L. S. Gramates, G. Millburn, V. B. Strelets, C. J. Tabone, J. Thurmond, FlyBase Consortium, FlyBase: updates to the Drosophila melanogaster knowledge base. *Nucleic Acids Res.* 49, D899–D907 (2021).
5. R. Lyne, R. Smith, K. Rutherford, M. Wakeling, A. Varley, F. Guillier, H. Janssens, W. Ji, P. Mclaren, P. North, D. Rana, T. Riley, J. Sullivan, X. Watkins, M. Woodbridge, K. Lilley, S. Russell, M. Ashburner, K. Mizuguchi, G. Micklem, FlyMine: an integrated database for Drosophila and Anopheles genomics. *Genome Biol.* 8, R129 (2007).
6. A. Jenett, G. M. Rubin, T.-T. B. Ngo, D. Shepherd, C. Murphy, H. Dionne, B. D. Pfeiffer, A. Cavallaro, D. Hall, J. Jeter, N. Iyer, D. Fetter, J. H. Hausenfluck, H. Peng, E. T. Trautman, R. R. Svirskas, E. W. Myers, Z. R. Iwinski, Y. Aso, G. M. DePasquale, C. T. Zugates, A GAL4-driver line resource for Drosophila neurobiology. *Cell Rep.* 2, 991–1001 (2012).
7. N. Milyaev, D. Osumi-Sutherland, S. Reeve, N. Burton, R. A. Baldock, J. D. Armstrong, The Virtual Fly Brain browser and query interface. *Bioinformatics*. 28, 411–415 (2012).
8. M. M. Kudron, A. Victorsen, L. Gevirtzman, L. W. Hillier, W. W. Fisher, D. Vafeados, M. Kirkey, A. S. Hammonds, J. Gersch, H. Ammouri, M. L. Wall, J. Moran, D. Steffen, M. Szynkarek, S. Seabrook-Sturgis, N. Jameel, M. Kadaba, J. Patton, R. Terrell, M. Corson, R. H. Waterston, The ModERN Resource: Genome-Wide Binding Profiles for Hundreds of Drosophila and Caenorhabditis elegans Transcription Factors. *Genetics*. 208, 937–949 (2018).
9. modENCODE Consortium, S. Roy, J. Ernst, P. V. Kharchenko, P. Kheradpour, N. Negre, M. L. Eaton, J. M. Landolin, C. A. Bristow, L. Ma, M. F. Lin, S. Washietl, B. I. Arshinoff, F. Ay, P. E. Meyer, N. Robine, N. L. Washington, L. Di Stefano, E. Berezikov, C. D. Brown, M. Kellis, Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science*. 330, 1787–1797 (2010).
10. V. R. Chintapalli, J. Wang, J. A. T. Dow, Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat. Genet.* 39, 715–720 (2007).
11. H. Li, Single-cell RNA sequencing in Drosophila: Technologies and applications. *Wiley Interdiscip. Rev. Dev. Biol.* 10, e396 (2021).
12. S. E. Mohr, S. G. Tattikota, J. Xu, J. Zirin, Y. Hu, N. Perrimon, Methods and tools for spatial mapping of single-cell RNAseq clusters in Drosophila. *Genetics*. 217 (2021), doi:10.1093/genetics/iyab019.
13. C. N. McLaughlin, M. Brbić, Q. Xie, T. Li, F. Horns, S. S. Kolluru, J. M. Kebschull, D. Vacek, A. Xie, J. Li, R. C. Jones, J. Leskovec, S. R. Quake, L. Luo, H. Li, Single-cell transcriptomes of developing and adult olfactory receptor neurons in Drosophila. *eLife*. 10 (2021), doi:10.7554/eLife.63856.
14. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049 (2017).
15. S. Picelli, Å. K. Björklund, O. R. Faridani, S. Sagasser, G. Winberg, R. Sandberg, Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*. 10, 1096–1098 (2013).
16. S. Yang, S. E. Corbett, Y. Koga, Z. Wang, W. E. Johnson, M. Yajima, J. D. Campbell, Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol.* 21, 57 (2020).
17. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods*. 16, 1289–1296 (2019).

18. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, Principal investigators, Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*. 562, 367–372 (2018).
19. K. Davie, J. Janssens, D. Koldere, M. De Waegeneer, U. Pech, Ł. Kreft, S. Aibar, S. Makhzami, V. Christiaens, C. Bravo González-Blas, S. Poovathingal, G. Hulselmans, K. I. Spanier, T. Moerman, B. Vanspauwen, S. Geurs, T. Voet, J. Lammertyn, B. Thienpont, S. Liu, S. Aerts, A Single-Cell Transcriptome Atlas of the Aging *Drosophila* Brain. *Cell*. 174, 982-998.e20 (2018).
20. F. P. A. David, M. Litovchenko, B. Deplancke, V. Gardeux, ASAP 2020 update: an open, scalable and interactive web-based portal for (single-cell) omics analyses. *Nucleic Acids Res.* 48, W403–W414 (2020).
21. M. Costa, S. Reeve, G. Grumbling, D. Osumi-Sutherland, The Drosophila anatomy ontology. *J. Biomed. Semantics*. 4, 32 (2013).
22. S. Levy, A. Elek, X. Grau-Bové, S. Menéndez-Bravo, M. Iglesias, A. Tanay, T. Mass, A. Sebé-Pedrós, A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. *Cell*. 184, 2973-2987.e18 (2021).
23. J. Cao, J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S. N. Furlan, F. J. Steemers, A. Adey, R. H. Waterston, C. Trapnell, J. Shendure, Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. 357, 661–667 (2017).
24. M. N. Özel, F. Simon, S. Jafari, I. Holguera, Y.-C. Chen, N. Benhra, R. N. El-Danaf, K. Kapuralin, J. A. Malin, N. Konstantinides, C. Desplan, Neuronal diversity and convergence in a visual system developmental atlas. *Nature*. 589, 88–95 (2021).
25. H. Li, F. Horns, B. Wu, Q. Xie, J. Li, T. Li, D. J. Luginbuhl, S. R. Quake, L. Luo, Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing. *Cell*. 171, 1206-1220.e22 (2017).
26. Y. Z. Kurmangaliyev, J. Yoo, J. Valdes-Aleman, P. Sanfilippo, S. L. Zipursky, Transcriptional programs of circuit assembly in the drosophila visual system. *Neuron*. 108, 1045-1057.e6 (2020).
27. B. Cho, S.-H. Yoon, D. Lee, F. Koranteng, S. G. Tattikota, N. Cha, M. Shin, H. Do, Y. Hu, S. Y. Oh, D. Lee, A. Vipin Menon, S. J. Moon, N. Perrimon, J.-W. Nam, J. Shim, Single-cell transcriptome maps of myeloid blood cell lineages in Drosophila. *Nat. Commun.* 11, 4483 (2020).
28. M. Grigorian, T. Liu, U. Banerjee, V. Hartenstein, The proteoglycan Trol controls the architecture of the extracellular matrix and balances proliferation and differentiation of blood progenitors in the Drosophila lymph gland. *Dev. Biol.* 384, 301–312 (2013).
29. R. Makki, M. Meister, D. Pennetier, J.-M. Ubeda, A. Braun, V. Daburon, J. Krzemień, H.-M. Bourbon, R. Zhou, A. Vincent, M. Crozatier, A short receptor downregulates JAK/STAT signalling to control the Drosophila cellular immune response. *PLoS Biol.* 8, e1000441 (2010).
30. M. Gallowitsch-Puerta, K. J. Tracey, Immunologic role of the cholinergic anti-inflammatory pathway and the nicotinic acetylcholine alpha 7 receptor. *Ann. N. Y. Acad. Sci.* 1062, 209–219 (2005).
31. V. A. Pavlov, K. J. Tracey, The cholinergic anti-inflammatory pathway. *Brain Behav. Immun.* 19, 493–499 (2005).
32. P. Sanchez Bosch, K. Makhijani, L. Herboso, K. S. Gold, R. Baginsky, K. J. Woodcock, B. Alexander, K. Kukar, S. Corcoran, T. Jacobs, D. Ouyang, C. Wong, E. J. V. Ramond, C. Rhiner, E. Moreno, B. Lemaitre, F. Geissmann, K. Brückner, Adult drosophila lack hematopoiesis but rely on a blood cell reservoir at the respiratory epithelia to relay infection signals to surrounding tissues. *Dev. Cell*. 51, 787-803.e5 (2019).
33. J. Krzemień, L. Dubois, R. Makki, M. Meister, A. Vincent, M. Crozatier, Control of blood cell homeostasis in Drosophila larvae by the posterior signalling centre. *Nature*. 446, 325–328 (2007).
34. L. Mandal, J. A. Martinez-Agosto, C. J. Evans, V. Hartenstein, U. Banerjee, A Hedgehog- and Antennapedia-dependent niche maintains Drosophila haematopoietic precursors. *Nature*. 446, 320–324 (2007).
35. R. J. Siviter, G. M. Coast, A. M. Winther, R. J. Nachman, C. A. Taylor, A. D. Shirras, D. Coates, R. E. Isaac, D. R. Nässel, Expression and functional characterization of a Drosophila neuropeptide precursor with homology to mammalian preprotachykinin A. *J. Biol. Chem.* 275, 23273–23280 (2000).
36. D. R. LaJeunesse, B. Johnson, J. S. Presnell, K. K. Catignas, G. Zapotoczny, Peristalsis in the junction

region of the Drosophila larval midgut is modulated by DH31 expressing enteroendocrine cells. *BMC Physiol.* 10, 14 (2010).
37. D. Hadjieconomou, G. King, P. Gaspar, A. Mineo, L. Blackie, T. Ameku, C. Studd, A. de Mendoza, F. Diao, B. H. White, A. E. X. Brown, P.-Y. Plaçais, T. Préat, I. Miguel-Aliaga, Enteric neurons increase maternal food intake during reproduction. *Nature*. 587, 455–459 (2020).
38. D. R. Nässel, Substrates for neuronal cotransmission with neuropeptides and small molecule neurotransmitters in drosophila. *Front. Cell. Neurosci.* 12, 83 (2018).
39. A. D. Talsma, C. P. Christov, A. Terriente-Felix, G. A. Linneweber, D. Perea, M. Wayland, O. T. Shafer, I. Miguel-Aliaga, Remote control of renal physiology by the intestinal neuropeptide pigment-dispersing factor in Drosophila. *Proc Natl Acad Sci USA*. 109, 12177–12182 (2012).
40. I. Yanai, H. Benjamin, M. Shmoish, V. Chalifa-Caspi, M. Shklar, R. Ophir, A. Bar-Even, S. Horn-Saban, M. Safran, E. Domany, D. Lancet, O. Shmueli, Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*. 21, 650–659 (2005).
41. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods*. 14, 1083–1086 (2017).
42. J. Mattila, V. Hietakangas, Regulation of Carbohydrate Energy Metabolism in Drosophila melanogaster. *Genetics*. 207, 1231–1253 (2017).
43. F. J. Bernardo-Garcia, C. Fritsch, S. G. Sprecher, The transcription factor Glass links eye field specification with photoreceptor differentiation in Drosophila. *Development*. 143, 1413–1423 (2016).
44. K. Moses, M. C. Ellis, G. M. Rubin, The glass gene encodes a zinc-finger protein required by Drosophila photoreceptor cells. *Nature*. 340, 531–536 (1989).
45. H. Kaessmann, Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20, 1313–1326 (2010).
46. E. Witt, S. Benjamin, N. Svetec, L. Zhao, Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in Drosophila. *eLife*. 8 (2019), doi:10.7554/eLife.47138.
47. S. Kondo, J. Vedanayagam, J. Mohammed, S. Eizadshenass, L. Kan, N. Pang, R. Aradhya, A. Siepel, J. Steinhauer, E. C. Lai, New genes often acquire male-specific functions but rarely become essential in Drosophila. *Genes Dev.* 31, 1841–1846 (2017).
48. Y. Shao, C. Chen, H. Shen, B. Z. He, D. Yu, S. Jiang, S. Zhao, Z. Gao, Z. Zhu, X. Chen, Y. Fu, H. Chen, G. Gao, M. Long, Y. E. Zhang, GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Res.* 29, 682–696 (2019).
49. J. A. Williams, J. B. Bell, S. B. Carroll, Control of Drosophila wing and haltere development by the nuclear vestigial gene product. *Genes Dev.* 5, 2481–2495 (1991).
50. E. B. Lewis, A gene complex controlling segmentation in Drosophila. *Nature*. 276, 565–570 (1978).
51. M. Parisi, R. Nuttall, P. Edwards, J. Minor, D. Naiman, J. Lü, M. Doctolero, M. Vainer, C. Chan, J. Malley, S. Eastman, B. Oliver, A survey of ovary-, testis-, and soma-biased gene expression in Drosophila melanogaster adults. *Genome Biol.* 5, R40 (2004).
52. J. Andrews, G. G. Bouffard, C. Cheadle, J. Lü, K. G. Becker, B. Oliver, Gene discovery using computational and microarray analysis of transcription in the Drosophila melanogaster testis. *Genome Res.* 10, 2030–2043 (2000).
53. H. K. Salz, J. W. Erickson, Sex determination in Drosophila: The view from the top. *Fly (Austin)*. 4, 60–70 (2010).
54. K. Kim, E. A. Lane, A. Saftien, H. Wang, Y. Xu, F. Wirtz-Peitz, N. Perrimon, Drosophila as a model for studying cystic fibrosis pathophysiology of the gastrointestinal system. *Proc Natl Acad Sci USA*. 117, 10357–10367 (2020).
55. K. Kim, R.-J. Hung, N. Perrimon, miR-263a Regulates ENaC to Maintain Osmotic and Intestinal Stem Cell Homeostasis in Drosophila. *Dev. Cell*. 40, 23–36 (2017).
56. E. Clough, E. Jimenez, Y.-A. Kim, C. Whitworth, M. C. Neville, L. U. Hempel, H. J. Pavlou, Z.-X. Chen, D. Sturgill, R. K. Dale, H. E. Smith, T. M. Przytycka, S. F. Goodwin, M. Van Doren, B. Oliver, Sex- and tissue-specific functions of Drosophila doublesex transcription factor target genes. *Dev. Cell*. 31, 761–773

(2014).
57. M. N. Arbeitman, F. N. New, J. M. Fear, T. S. Howard, J. E. Dalton, R. M. Graze, Sex Differences in Drosophila Somatic Gene Expression: Variation and Regulation by doublesex. *G3 (Bethesda)*. 6, 1799–1808 (2016).
58. R.-J. Hung, Y. Hu, R. Kirchner, Y. Liu, C. Xu, A. Comjean, S. G. Tattikota, F. Li, W. Song, S. Ho Sui, N. Perrimon, A cell atlas of the adult Drosophila midgut. *Proc Natl Acad Sci USA*. 117, 1514–1523 (2020).
59. K. Rust, L. E. Byrnes, K. S. Yu, J. S. Park, J. B. Sneddon, A. D. Tward, T. G. Nystul, A single-cell atlas and lineage analysis of the adult Drosophila ovary. *Nat. Commun.* 11, 5628 (2020).
60. A. Jevitt, D. Chatterjee, G. Xie, X.-F. Wang, T. Otwell, Y.-C. Huang, W.-M. Deng, A single-cell atlas of adult Drosophila ovary identifies transcriptional programs and somatic cell lineage regulating oogenesis. *PLoS Biol.* 18, e3000538 (2020).
61. C. Barreau, E. Benson, E. Gudmannsdottir, F. Newton, H. White-Cooper, Post-meiotic transcription in Drosophila testes. *Development*. 135, 1897–1902 (2008).
62. X. Han, R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, A. Saadatpour, Z. Zhou, H. Chen, F. Ye, D. Huang, Y. Xu, W. Huang, M. Jiang, X. Jiang, J. Mao, Y. Chen, C. Lu, J. Xie, Q. Fang, G. Guo, Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*. 173, 1307 (2018).
63. J. Cao, D. R. O'Day, H. A. Pliner, P. D. Kingsley, M. Deng, R. M. Daza, M. A. Zager, K. A. Aldinger, R. Blecher-Gonen, F. Zhang, M. Spielmann, J. Palis, D. Doherty, F. J. Steemers, I. A. Glass, C. Trapnell, J. Shendure, A human cell atlas of fetal gene expression. *Science*. 370 (2020), doi:10.1126/science.aba7721.
64. X. Han, Z. Zhou, L. Fei, H. Sun, R. Wang, Y. Chen, H. Chen, J. Wang, H. Tang, W. Ge, Y. Zhou, F. Ye, M. Jiang, J. Wu, Y. Xiao, X. Jia, T. Zhang, X. Ma, Q. Zhang, X. Bai, G. Guo, Construction of a human cell landscape at single-cell level. *Nature*. 581, 303–309 (2020).
65. Tabula Muris Consortium, A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*. 583, 590–595 (2020).
66. G. L. Henry, F. P. Davis, S. Picard, S. R. Eddy, Cell type-specific genomics of Drosophila neurons. *Nucleic Acids Res.* 40, 9691–9704 (2012).
67. K. Polański, M. D. Young, Z. Miao, K. B. Meyer, S. A. Teichmann, J.-E. Park, BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics*. 36, 964–965 (2020).
68. B. Hie, B. Bryson, B. Berger, Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* 37, 685–691 (2019).
69. X. Li, K. Wang, Y. Lyu, H. Pan, J. Zhang, D. Stambolian, K. Susztak, M. P. Reilly, G. Hu, M. Li, Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis. *Nat. Commun.* 11, 2338 (2020).
70. M. Brbić, M. Zitnik, S. Wang, A. O. Pisco, R. B. Altman, S. Darmanis, J. Leskovec, MARS: discovering novel cell types across heterogeneous single-cell experiments. *Nat. Methods*. 17, 1200–1206 (2020).
71. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 20, 296 (2019).
72. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell*. 177, 1888-1902.e21 (2019).
73. M. Sarov, C. Barz, H. Jambor, M. Y. Hein, C. Schmied, D. Suchold, B. Stender, S. Janosch, V. V. K J, R. T. Krishnan, A. Krishnamoorthy, I. R. S. Ferreira, R. K. Ejsmont, K. Finkl, S. Hasse, P. Kämpfer, N. Plewka, E. Vinis, S. Schloissnig, E. Knust, F. Schnorrer, A genome-wide resource for the analysis of protein localisation in Drosophila. *eLife*. 5, e12068 (2016).
74. C. Schönbauer, J. Distler, N. Jährling, M. Radolf, H.-U. Dodt, M. Frasch, F. Schnorrer, Spalt mediates an evolutionarily conserved switch to fibrillar muscle fate in insects. *Nature*. 479, 406–409 (2011).
75. M. B. Chechenova, S. Maes, S. T. Oas, C. Nelson, K. G. Kiani, A. L. Bryantsev, R. M. Cripps, Functional redundancy and nonredundancy between two Troponin C isoforms in Drosophila adult muscles. *Mol. Biol. Cell*. 28, 760–770 (2017).
76. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502 (2015).
77. M. Slaidina, T. U. Banisch, S. Gupta, R. Lehmann, A single-cell atlas of the developing Drosophila ovary

identifies follicle stem cell progenitors. *Genes Dev.* 34, 239–249 (2020).
78. L. Bai, A. L. Goldman, J. R. Carlson, Positive and negative regulation of odor receptor gene choice in Drosophila by acj6. *J. Neurosci.* 29, 12940–12947 (2009).
79. J. C. Lucchesi, M. I. Kuroda, Dosage compensation in Drosophila. *Cold Spring Harb. Perspect. Biol.* 7 (2015), doi:10.1101/cshperspect.a019398.
80. B. R. Graveley, A. N. Brooks, J. W. Carlson, M. O. Duff, J. M. Landolin, L. Yang, C. G. Artieri, M. J. van Baren, N. Boley, B. W. Booth, J. B. Brown, L. Cherbas, C. A. Davis, A. Dobin, R. Li, W. Lin, J. H. Malone, N. R. Mattiuzzo, D. Miller, D. Sturgill, S. E. Celniker, The developmental transcriptome of Drosophila melanogaster. *Nature*. 471, 473–479 (2011).

FCA Consortium authors (last name, A–Z):

Stein Aerts, Devika Agarwal, Yasir Ahmed-Braimah, Aaron M. Allen, Michelle Arbeitman, Majd M. Ariss, Jordan Augsburger, Kumar Ayush, Catherine C. Baker, Torsten Banisch, Cameron W. Berry, Katja Birker, Rolf Bodmer, Benjamin Bolival, Susanna E. Brantley, Maria Brbic, Julie A. Brill, Nora C. Brown, Katja Brueckner, Norene A. Buehner, Xiaoyu Tracy Cai, Rita Cardoso-Figueiredo, Zita Carvalho-Santos, Fernando Casares, Amy Chang, Thomas R. Clandinin, Sheela Crasta, Fabrice PA David, Kristofer Davie, Bart Deplancke, Claude Desplan, Angela M. Detweiler, Darshan B. Dhakan, Stephen DiNardo, Erika Donà, Julian A. T. Dow, Stefanie Engert, Swann Floc'hlay, Margaret T. Fuller, Anthony Galenza, Vincent Gardeux, Nancy George, Amanda Gonzalez, Stephen F Goodwin, Andrew K. Groves, Samantha Gumbin, Yanmeng Guo, Devon E. Harris, Yael Heifetz , Stephen L. Holtz, Felix Horns, Bruno Hudry, Ruei-Jiun Hung, Yuh Nung Jan, Jasper Janssens, Heinrich Jasper, Jacob S. Jaszczak, Gregory S.X.E. Jefferis, Robert C. Jones, Jim Karkanias, Timothy L. Karr, Nadja Sandra Katheder, James Kezos, Anna A. Kim, Seung K. Kim, Lutz Kockel, Sai Saroja Kolluru, Nikolaos Konstantinides, Thomas B Kornberg, Henry M. Krause, Andrew Thomas Labott, Meghan Laturney, Ruth Lehmann, Sarah Leinwand, Jure Leskovec, Hongjie Li, Jiefu Li, Joshua Shing Shun Li, Kai Li, Ke Li, Liying Li, Tun Li, Maria Litovchenko, Han-Hsuan Liu, Yifang Liu, Tzu-Chiao Lu, Liqun Luo, Sharvani Mahadevaraju, Jonathan Manning, Anjeli Mase, Mikaela Matera-Vatnick, Neuza Reis Matias, Erika L. Matunis, Caitlin E. McDonough-Goldstein, Aaron McGeever, Alex D. McLachlan, Colleen N McLaughlin, Paola Moreno-Roman, Norma Neff, Megan Neville, Sang Ngo, Tanja Nielsen, Todd G. Nystul, Caitlin E. O'Brien, Lucy Erin O'Brien, Brian Oliver, David Osumi-Sutherland, Mehmet Neset Özel, Soumitra Pal, Irene Papatheodorou, Norbert Perrimon, Maja Petkovic, Clare Pilgrim, Angela Oliveira Pisco, Teresa M Przytycka, Stephen R. Quake, Carolina Reisenman, Carlos Ribeiro, Katja Rust, Wouter Saelens, Erin Nicole Sanders, Gilberto dos Santos, Frank Schnorrer, Kristin Scott, Aparna Sherlekar, Jiwon Shim, Philip Shiu, David Sims, Rene V. Sit, Maija Slaidina, Harold E. Smith, Katina Spanier, Gabriella Sterne, Yu-Han Su, Daniel Sutton, Marco Tamayo, Michelle Tan, Ibrahim Tastekin, Sudhir Gopal Tattikota, Christoph Treiber, David Vacek, Georg Vogler, Scott Waddell, Maxime De Waegeneer, Wanpeng Wang, Helen White-Cooper, Rachel I. Wilson, Mariana F. Wolfner, Yiu-Cheung E. Wong, Anthony Xie, Qijing Xie, Jun Xu, Shinya Yamamoto, Jia Yan, Zepeng Yao, Kazuki Yoda, Ruijun Zhu, Robert P Zinzen

ACKNOWLEDGMENTS:

We thank the whole fly community for the enthusiastic support for this project, Bill Burkholder for organizing the FCA meeting at Biohub, Cathryn Murphy for helping organize the monthly FCA Zoom meeting, and Kathleen Vogelaers for coordinating all Jamborees.

Funding: The sequencing experimental work was supported by the Chan Zuckerberg Biohub (S. Quake), Genentech Inc (H. Jasper), National Institutes of Health (B. Oliver), and Howard Hughes Medical Institute and a National Institutes of Health grant (L. Luo). Computational work was supported by the KU Leuven and the Flemish Supercomputer Center (VSC) (S. Aerts) and EPFL (B. Deplancke). Due to the space limit, we have provided FCA Consortium Funding in the Supplemental Materials.

Author contributions: See FCA Consortium Contributions in Supplementary Materials.
Competing interests: The authors declare no competing interests.
Data and materials availability
All data are available for user-friendly querying via https://flycellatlas.org/scope and for additional custom analyses at https://flycellatlas.org/asap . For each tissue, a CellxGene portal is also available (links can be found on www.flycellatlas.org). Raw sequencing data and count matrices can be downloaded from ArrayExpress (accession number E-MTAB-10519 for 10x dataset, and E-MTAB-10628 for Smart-seq2 dataset). Files with expression data, clustering, embeddings, and annotation can be downloaded for each tissue, or all data combined, in h5ad and loomX formats from www.flycellatlas.org. We have included three Supplemental Figures on how to access and explore the portals for FCA data: fig. S1 for summary of Data Availability, fig. S2 for how to use SCope and fig. S3 for how to use ASAP. In addition, we have made a video tutorial for using Scope (https://www.youtube.com/watch?v=yNETQVaSJYM&t=349s). Analysis codes are at Github (https://github.com/flycellatlas).

Supplementary Materials

FCA Consortium Contributions
FCA Consortium Author Affiliations
Materials and Methods
FCA Consortium Funding
FCA Consortium Author Affiliations
Figures S1 to S37
Tables S1 to S5, provided as Excel files
Captions for Table S1 to S4

Other Supplementary Materials for this manuscript include the following:

Table S1 [excel]: Supplementary Table for Fig2, cell type annotation
Table S2 [excel]: Supplementary Table for Fig5, cell type-specific transcription factors
Table S3 [excel]: Supplementary Table for Fig6: sex-differences, mannually removed cells
Table S4 [excel]: Supplementary Table for Fig6: sex-differences cell level data
Table S5 [excel]: Supplementary Table for Fig6: sex-differences cluster level data

FIGURE LEGENDS:

Figure 1. Overview of the Fly Cell Atlas
(A) Experimental platform of snRNA-seq using 10x Genomics and Smart-seq2 (SS2).
(B) Data analysis pipeline and data visualization using SCope (*19*) and ASAP (*20*).
(C) Two versions of 10x datasets have been generated: *Relaxed* and *Stringent*. tSNE colors based on gene expression: *grh* (epithelia, red), *Mhc* (muscle, green) and *Syt1* (neuron, blue). Red arrow denotes an artefactual cluster with co-expression of all three markers in the *Relaxed* dataset.
(D) tSNE visualization of cells from the *Stringent* 10x dataset and Smart-seq2 (SS2) cells. 10x cells are from individual tissues. Integrated data is colored by tissue (left) and platform (right).
(E) Tissue-level comparison of the number of detected genes between 10x and Smart-seq2 platforms (see Methods).

(F) Number of cells for each tissue by 10x and Smart-seq2. Male and female cells are indicated. Mixed cells are from pilot experiments where flies were not sexed. Different batches are separated by vertical white lines.
(G) All 10x cells from the *Stringent* dataset clustered together; cells are colored by tissue type. Tissue names and colors indexed in F.

Figure 2: Cell type annotation for dissected tissues

(A) Illustration of 15 individual tissues. 12 sequenced separately from males and females, 3 sex-specific. Fat body, oenocyte, and tracheal nuclei were labeled using a tissue-specific GAL4 driving UAS-nuclearGFP.
(B) tSNE plot with annotations for body wall from the *Stringent* 10x dataset. *1, epidermal cells of the abdominal posterior compartment. *2, epidermal cells specialized in antimicrobial response.
(C) UMAP plot with annotations for the testis from the *Relaxed* 10x dataset.
(D) tSNE plots of the other 13 tissues from the *Stringent* 10x dataset. Detailed annotations are in fig. S6–S18.
(E) Number of unique annotations for each tissue. Fractions of annotated cells over all analyzed cells from the *Relaxed* dataset are indicated in red.

Figure 3: Whole-head and whole-body sequencing leads to full coverage of the entire fly

(A) tSNE of the whole-head sample with 81 annotated clusters. See fig. S22 for full cell types. Many cells in the middle (gray) are unannotated, most of which are central brain neurons.

(B) tSNE of the whole-body sample with 33 annotated clusters, many of which can be further divided into sub-clusters. Cells in gray are unannotated. See fig. S23 for full cell types.

(C) (left) tSNE of the entire dataset colored by standardized tissue enrichment, leading to the identification of head- and body-specific clusters. (right) Bar plots showing tissue composition (head, body, or dissected tissues) for different clusters at Leiden resolution 50.

(D) Examples of head- and body-specific clusters.

(E) Integration of a brain scRNA-seq dataset with the head snRNA-seq for label transfer. Outlined are example clusters revealed by the head snRNA-seq dataset but not by the brain scRNA-seq datasets, including epithelial cells (EPI), photoreceptors (PRs), olfactory receptor neurons (ORNs), and muscle cells (MUS).

(F) Subclustering analysis reveals types of photoreceptors, including inner and outer photoreceptors, with the inner photoreceptors further splitting into R7 and R8 types, and mushroom body Kenyon cells, forming three distinct KC types: α/β, α'/β' and γ.

Figure 4: Cross-tissue analyses of common cell classes

(A) Overview of the different main cell classes identified throughout the fly cell atlas. Som. pre., somatic precursor cells; male repr., male reproductive system; fem. repr., female reproductive system; male germ., male germline cells; fem. germ., female germline cells.
(B) tSNE plots showing marker expression four common cell classess: *grh* for epithelial cells, *Syt1* for neurons, *Mhc* for muscle cells, and *Hml* for hemocytes.
(C) Composition of whole head and body samples, showing a shift from neurons to epithelial and muscle cells. Composition of the entire fly cell atlas shows enrichment for rarer cell classes compared to the whole-body sample. The bar plot shows the number of cells.
(D) Cross-tissue analysis of hemocytes reveals different cell states of plasmatocytes. Annotations marked as blue are hemocytes containing markers of different cell types, including lymph gland posterior signaling center (LGP), muscle (MUS), antenna (ANT), neurons (NEU), photoreceptor (PR), male accessory glands (MAG), glia (G), male testis and spermatocyte (MS), olfactory-binding proteins (OBP), and heat-shock proteins (Hsp). Other abbreviations show top marker gene(s) in red. Plasmatocytes and crystal cells are indicated. On the right are genes showing compartmentalized expression patterns within the plasmatocyte cluster.

(E) Cross-tissue analysis of muscle cells reveals subdivision of the visceral muscle cells based on neuropeptide receptors. Annotations marked as blue are muscle cells containing markers of different cell types, including neuron (NEU) and male testis and spermatocyte (MS). Muscle cells from three body parts are indicated: head muscle (HEAD), body muscle (BODY), and testis muscle (TESTIS). Other annotated muscle types include indirect flight muscle (IFM), ovarian sheath muscle (OSM), abdominal visceral muscle (ABD), *dpy* expressing muscle (DPY), visceral muscle of the midgut *AstC-R2* (VMM-A), visceral muscle of the crop *MsR1* (VMC-M), visceral muscle of the midgut *Dh31-R* (VMM-D), and visceral muscle *CCAP-R* (VM-C). *Pdfr* expressed in all visceral muscle cells, including the ovarian sheath muscle; other four receptor genes (*AstC-R2, MsR1, Dh31-R, CCAP-R*) expressed in different gut visceral muscle types.

Figure 5: Transcription factor (TF) pleiotropy versus cell-type specificity

(A) Heatmap showing the expression of key marker genes and unique TF profiles for each of the annotated cell types. TFs were selected based on *tau* score. Cell types were grouped based on hierarchical terms: CNS neurons (N), sensory organ cells (S), epithelial cells (E), muscle cells (M), glia (G), fat cells (F), oenocytes (O), hemocytes (H), (fe)male reproductive system and germline (MR, MG, FR, FG), excretory system (X), tracheal cell (T), gland (L), cardial cell (C), somatic precursor cell (P).

(B) A network analysis of TFs and cell classes based on similarity of ontology terms, reveals unique and shared TFs across the individual tissues.

(C) Heatmap showing the expression of unique TFs per cell class. Factors from the literature are highlighted.

(D) *Glass* is uniquely expressed in photoreceptors and cone cells in the head.

(E) Overview of the *Glass* regulon of 444 target genes, highlighting known photoreceptor marker genes.

(F) Gene expression comparison across broad cell types. Only sets with more than 10 genes are shown. The left bar graph shows the number of uniquely expressed genes for each tissue. The top bar graph shows the gene age in branches, ranging from the common ancestor to *Drosophila melanogaster*-specific genes (http://gentree.ioz.ac.cn). See fig. S34 for tissue-based comparison.

Figure 6. Sex-biased expression and trajectory analysis of testis cell lineages

(A) Simplified sex determination pathway. Sex chromosome karyotype (XX) activates Sex-lethal (Sxl) which regulates transformer (Tra), resulting in a female Dsx isoform (Dsx^F). In XY (or X0) flies, Sxl and Tra are inactive (light gray) and the male-specific Dsx^M is produced.

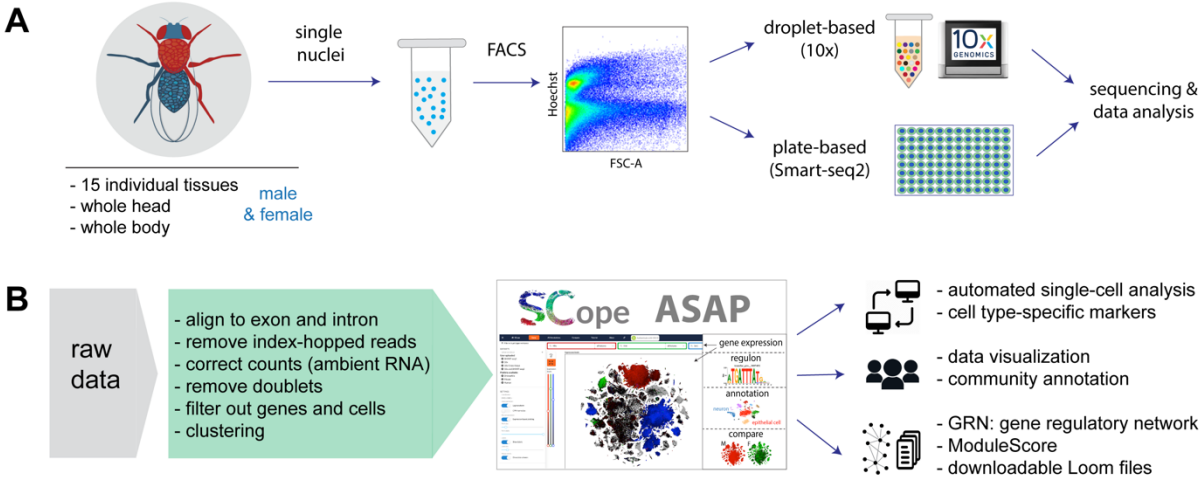
(B) Dsx expression and female- and male-biased expression projected on tSNE plots of all female (left column) and male (right column) cells except reproductive tissue cells (Table S4). The top two plots showing *Dsx* expression. The bottom two plots showing female- and male-biased expression, measured as the percentage of genes in the cluster showing biased expression in favor of the respective sex (Table S5). These percentage values were computed for each annotated cluster and those cluster-level values were projected on the individual cells in the corresponding clusters. For all four tSNE plots, values outside the scale in the heatmap key are represented by the closest extreme color (denoted by > and < signs in the scale).

(C) Scatter plot of female- and male-bias values across non-reproductive cell clusters defined as % sex-biased genes (at least 2-fold change with FDR < 0.05 on Wilcoxon test and BH correction) in the cluster (Table S5). Data point size shows the number of cells per cluster (key). Selected clusters are labeled, with those from excretory cells highlighted (brown). MT, Malpighian tubule.

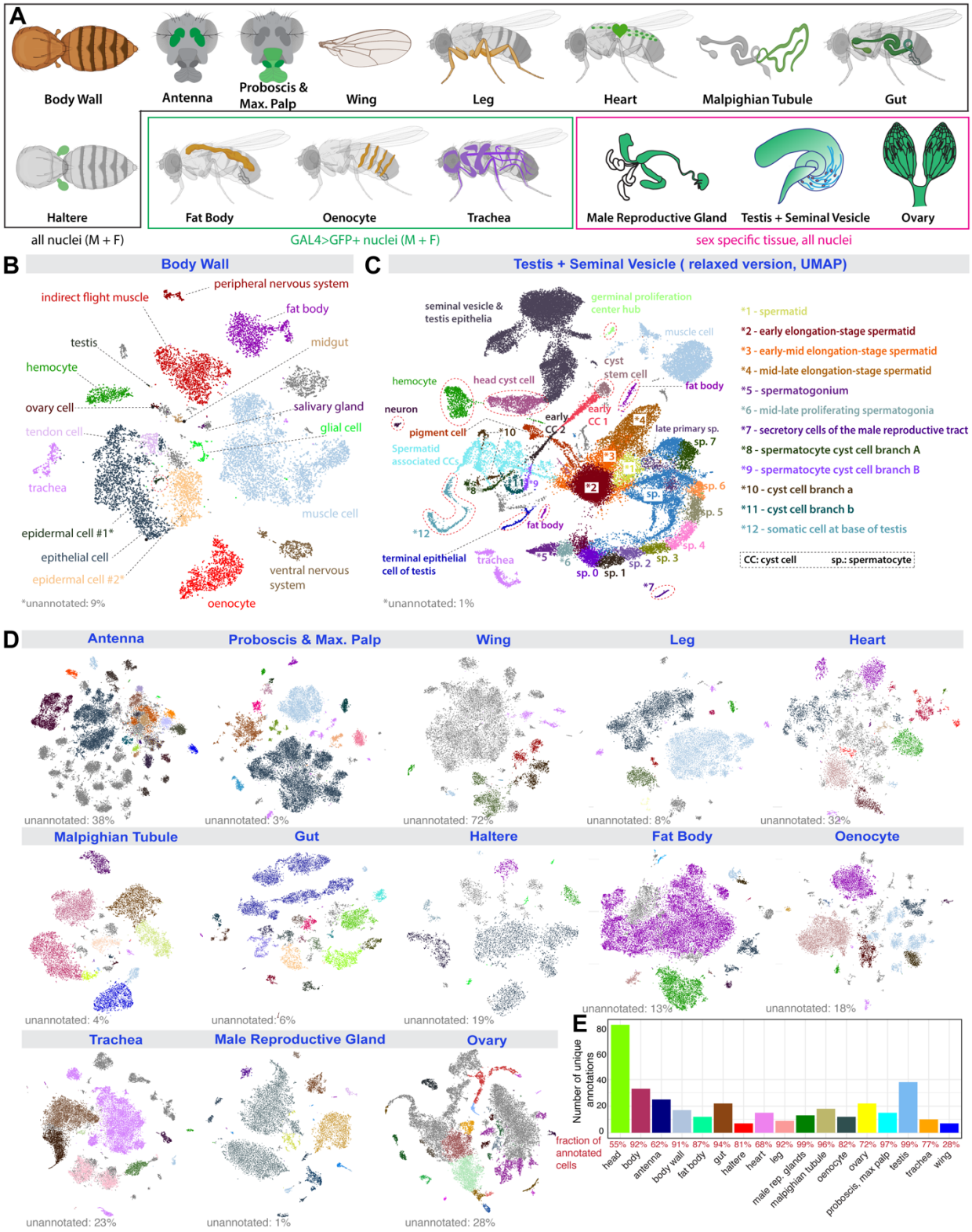
(D) Box plots showing the relationship between *dsx* gene expression and sex-biased expression (Table S5). Clusters (B) were partitioned into the set of clusters with Dsx expression (dsx^+) or not (no/low) using *dsx* expression in germ cells as an expression cut-off. Sex-, female-, and male-biased expression in the two groups were plotted. Each box shows hinges at first and third quartiles and median in the middle. The upper whisker extends from the upper hinge to the largest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance

between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Outliers are not shown. Wilcoxon test p values shown.
(E–G) Trajectory of testis subsets. We used slingshot to infer a possibly branching trajectory for spermatogonia-spermatocytes (E), spermatids (F), and early cyst cells (G). Shown are the trajectories on a UMAP (top) and the expression patterns of the strongest differentially expressed genes, together with the smoothed proportions of annotated cells and average number of unique molecular identifiers (UMIs) along the trajectory (bottom).

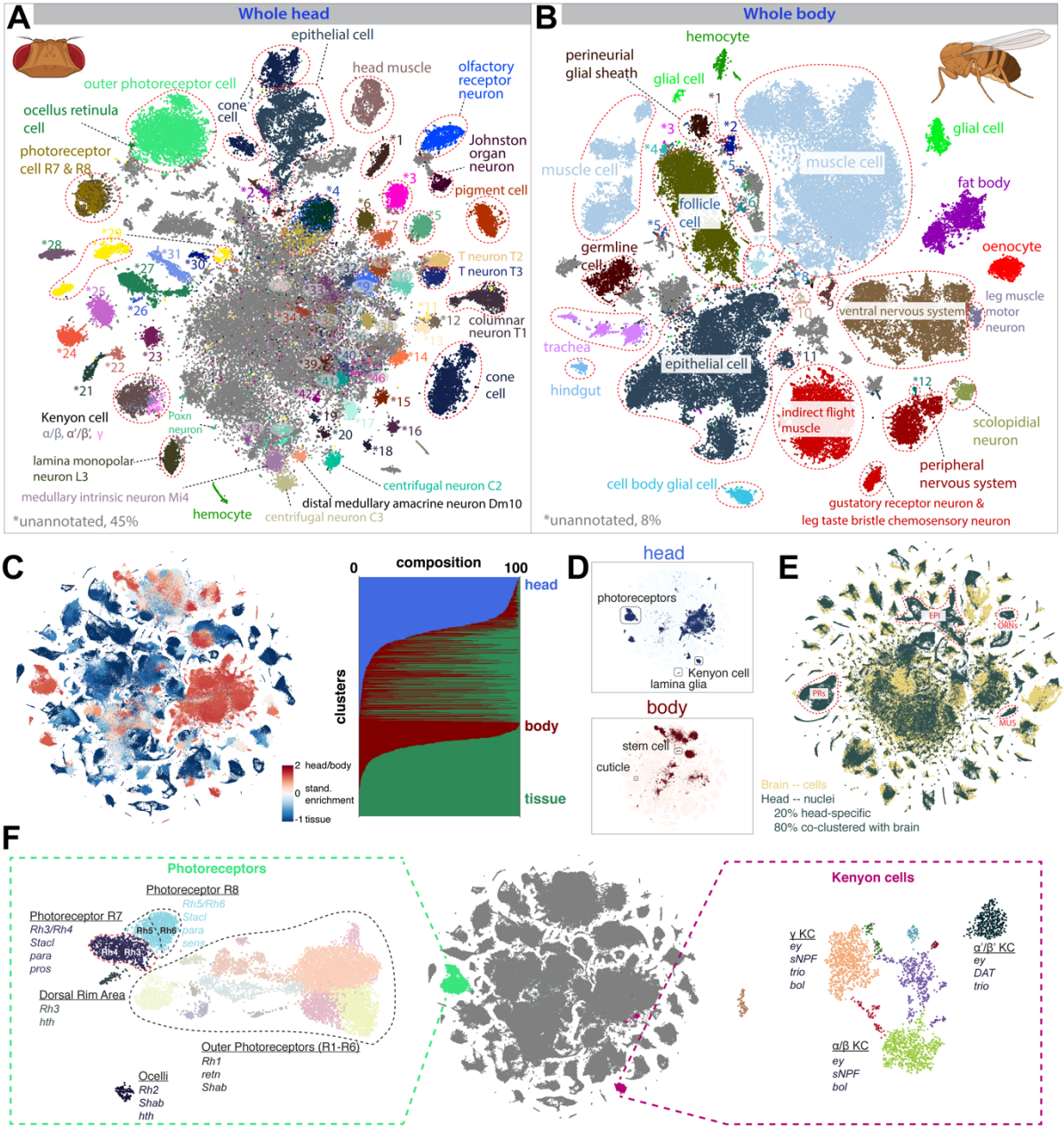
FCA Figure 1

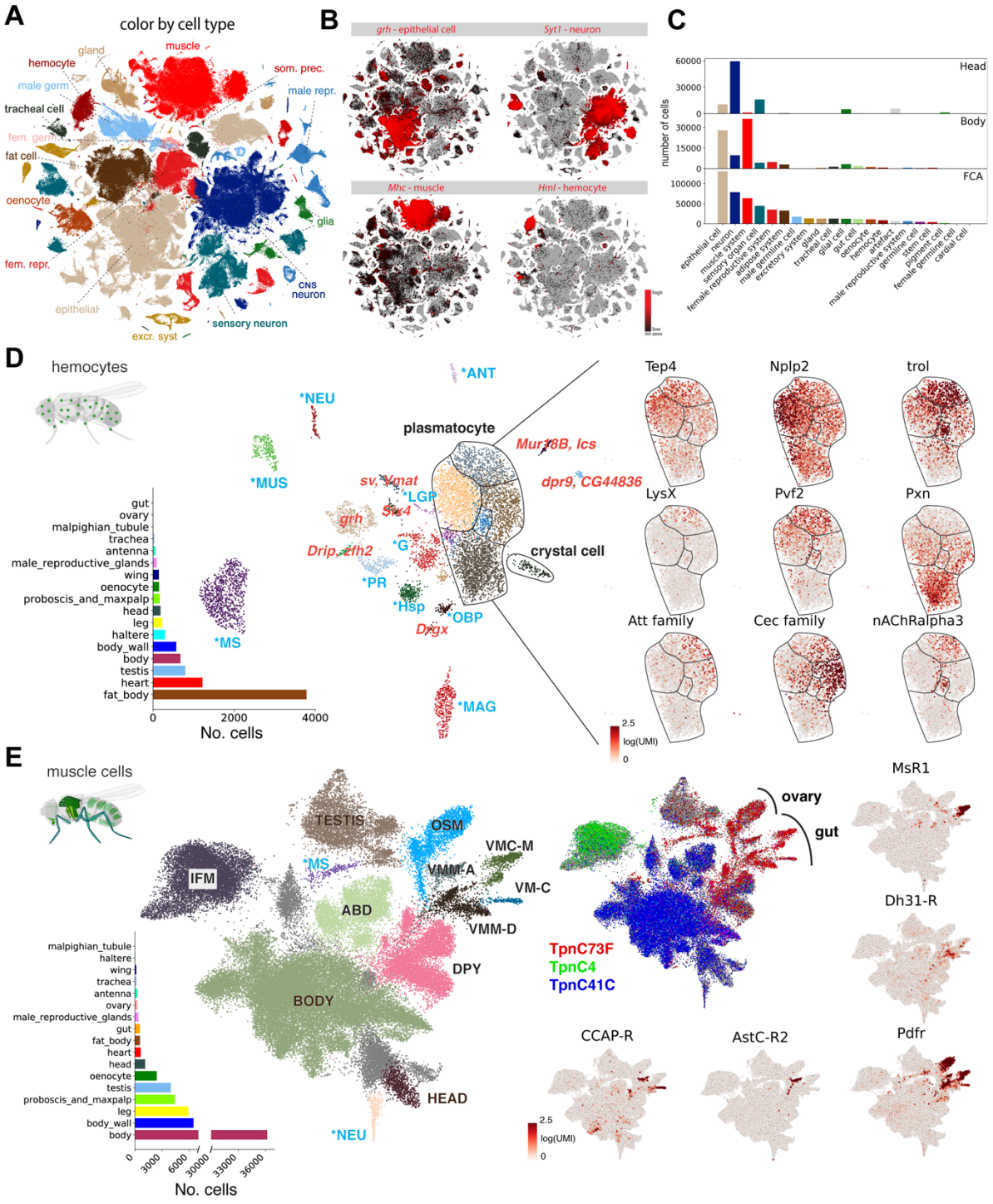

A
single nuclei
FACS
Hoechst
FSC-A
droplet-based (10x)
plate-based (Smart-seq2)
sequencing & data analysis
- 15 individual tissues
- whole head
- whole body
male & female
B
raw data
- align to exon and intron
- remove index-hopped reads
- correct counts (ambient RNA)
- remove doublets
- filter out genes and cells
- clustering
SCope ASAP
gene expression
regulon
annotation
compare
- automated single-cell analysis
- cell type-specific markers
- data visualization
- community annotation
- GRN: gene regulatory network
- ModuleScore
- downloadable Loom files
C 10x: relaxed version
stringent version
grh
Mhc
Syt1
D 10x + SS2
10x
SS2
E gene detection per tissue
Number of genes
12000
10000
8000
6000
antenna
body wall
fat body
gut
haltere
heart
leg
male rep. gland
malpighian tubule
oenocyte
ovary
prob. max. palps
testis
trachea
wing
F
10x
SS2
tissue
head
body
testis
antenna
ovary
fat body
trachea
proboscis & max palp
body wall
wing
oenocyte
leg
malpighian tubule
male reproductive gland
gut
heart
haltere
insulin producing cell (IPC)
corpora cardiaca (CC)
male (47%)
female (50%)
mix (3%)
100k
75k
50k
25k
0
500
1k
number of cells
G
color by tissue

FCA Figure 2

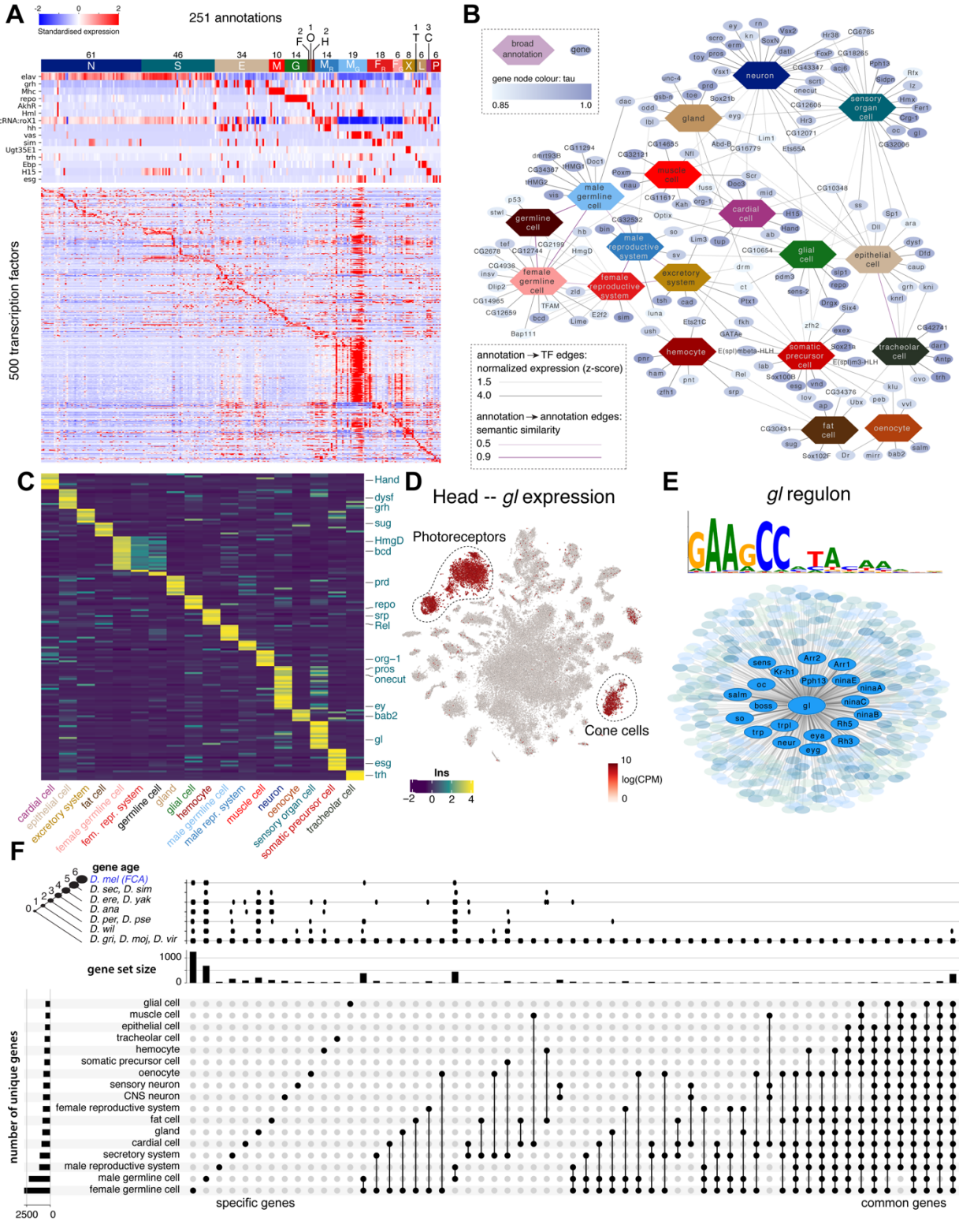

A
Body Wall
Antenna
Proboscis & Max. Palp
Wing
Leg
Heart
Malpighian Tubule
Gut
Haltere
Fat Body
Oenocyte
Trachea
Male Reproductive Gland
Testis + Seminal Vesicle
Ovary
all nuclei (M + F)
GAL4>GFP+ nuclei (M + F)
sex specific tissue, all nuclei
B
Body Wall
indirect flight muscle
peripheral nervous system
fat body
testis
midgut
hemocyte
salivary gland
ovary cell
glial cell
tendon cell
trachea
muscle cell
epidermal cell #1*
epithelial cell
epidermal cell #2*
ventral nervous system
oenocyte
*unannotated: 9%
C
Testis + Seminal Vesicle (relaxed version, UMAP)
seminal vesicle & testis epithelia
germinal proliferation center hub
muscle cell
cyst stem cell
hemocyte
head cyst cell
fat body
neuron
early CC 1
early CC 2
pigment cell
late primary sp.
Spermatid associated CCs
sp. 7
sp. 6
sp. 5
sp. 4
sp. 3
sp. 2
sp. 1
sp. 0
fat body
terminal epithelial cell of testis
trachea
*1 - spermatid
*2 - early elongation-stage spermatid
*3 - early-mid elongation-stage spermatid
*4 - mid-late elongation-stage spermatid
*5 - spermatogonium
*6 - mid-late proliferating spermatogonia
*7 - secretory cells of the male reproductive tract
*8 - spermatocyte cyst cell branch A
*9 - spermatocyte cyst cell branch B
*10 - cyst cell branch a
*11 - cyst cell branch b
*12 - somatic cell at base of testis
CC: cyst cell
sp.: spermatocyte
*unannotated: 1%
D
Antenna
unannotated: 38%
Proboscis & Max. Palp
unannotated: 3%
Wing
unannotated: 72%
Leg
unannotated: 8%
Heart
unannotated: 32%
Malpighian Tubule
unannotated: 4%
Gut
unannotated: 6%
Haltere
unannotated: 19%
Fat Body
unannotated: 13%
Oenocyte
unannotated: 18%
Trachea
unannotated: 23%
Male Reproductive Gland
unannotated: 1%
Ovary
unannotated: 28%
E
Number of unique annotations
0
20
40
60
80
fraction of annotated cells
55% 92% 62% 91% 87% 94% 81% 68% 92% 99% 96% 82% 72% 97% 99% 77% 28%
head
body
antenna
body wall
fat body
gut
haltere
heart
leg
male rep. glands
malpighian tubule
oenocyte
ovary
proboscis, max palp
testis
trachea
wing

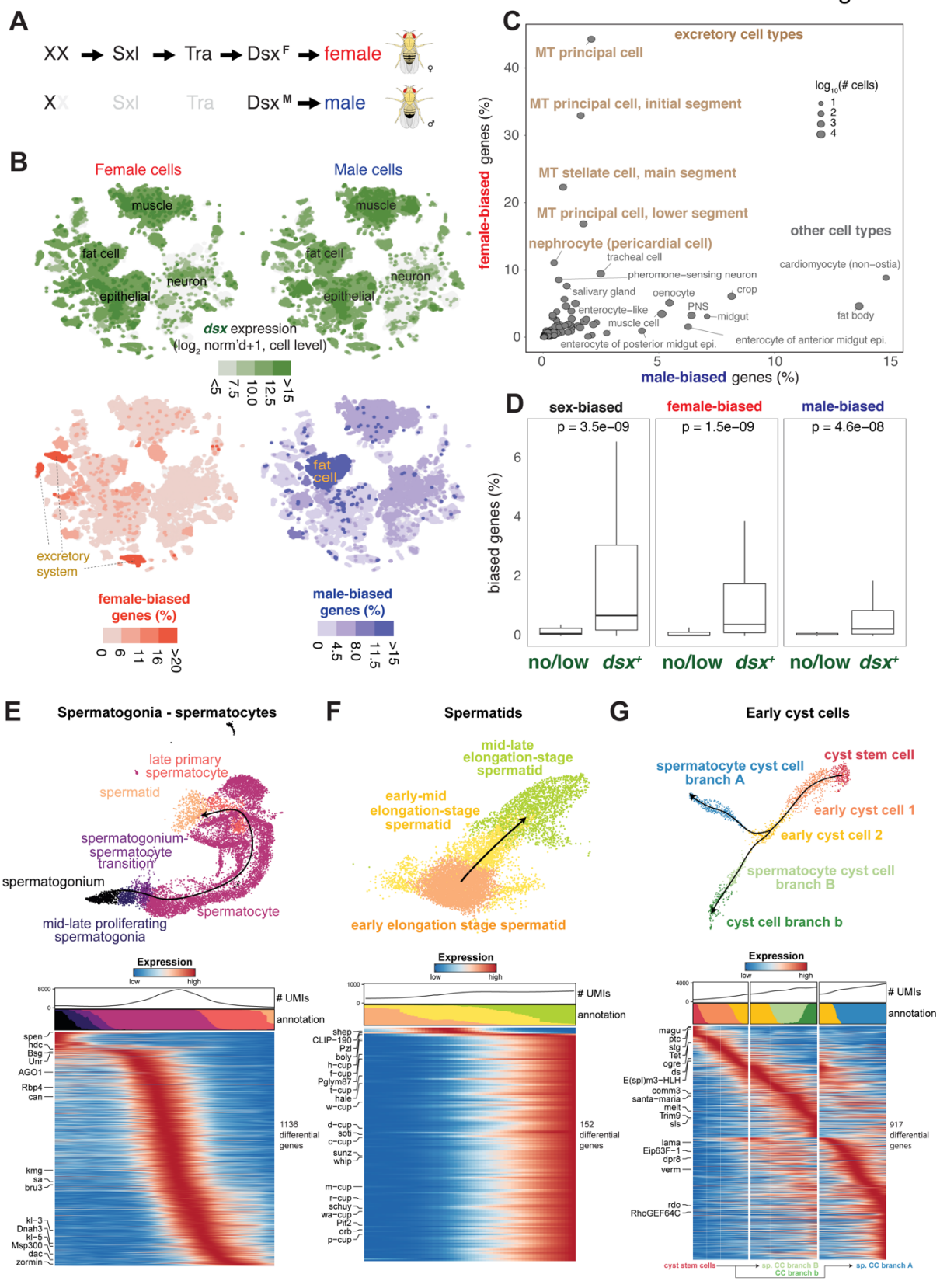
FCA Figure 3


A
Whole head
epithelial cell
head muscle
olfactory receptor neuron
outer photoreceptor cell
cone cell
ocellus retinula cell
photoreceptor cell R7 & R8
Johnston organ neuron
pigment cell
T neuron T2
T neuron T3
columnar neuron T1
cone cell
Kenyon cell α/β, α′/β′, γ
Poxn neuron
lamina monopolar neuron L3
medullary intrinsic neuron Mi4
hemocyte
centrifugal neuron C3
distal medullary amacrine neuron Dm10
centrifugal neuron C2
*unannotated, 45%
B
Whole body
hemocyte
perineurial glial sheath
glial cell
muscle cell
follicle cell
germline cell
fat body
oenocyte
leg muscle motor neuron
ventral nervous system
trachea
epithelial cell
hindgut
indirect flight muscle
scolopidial neuron
peripheral nervous system
cell body glial cell
gustatory receptor neuron & leg taste bristle chemosensory neuron
*unannotated, 8%
C
composition
0
100
head
body
tissue
clusters
2 head/body
0 stand. enrichment
-1 tissue
D
head
photoreceptors
Kenyon cell
lamina glia
body
stem cell
cuticle
E
EPI
ORNs
PRs
MUS
Brain -- cells
Head -- nuclei
20% head-specific
80% co-clustered with brain
F
Photoreceptors
Photoreceptor R8
Rh5/Rh6
Photoreceptor R7
Rh3/Rh4
Stacl
para
pros
Dorsal Rim Area
Rh3
hth
Outer Photoreceptors (R1-R6)
Rh1
retn
Shab
Ocelli
Rh2
Shab
hth
Kenyon cells
γ KC
ey
sNPF
trio
bol
α′/β′ KC
ey
DAT
trio
α/β KC
ey
sNPF
bol


A
color by cell type
gland
muscle
hemocyte
som. prec.
male germ
male repr.
tracheal cell
fem. germ
fat cell
oenocyte
glia
fem. repr.
CNS neuron
epithelial
sensory neuron
excr. syst
B
grh - epithelial cell
Syt1 - neuron
Mhc - muscle
Hml - hemocyte
high
low zero
C
Head
Body
FCA
number of cells
epithelial cell
neuron
muscle system
sensory organ cell
female reproductive system
adipose system
male germline cell
excretory system
gland
tracheal cell
glial cell
gut cell
oenocyte
hemocyte
antefact
male reproductive system
germline cell
stem cell
pigment cell
female germline cell
cardial cell
D
hemocytes
*NEU
*ANT
plasmatocyte
Mur18B, lcs
*MUS
sv, Vmat
dpr9, CG44836
*LGR
grh
Sw4
*G
Drip, elh2
crystal cell
*PR
*Hsp
*OBP
Dngx
*MS
*MAG
gut
ovary
malpighian_tubule
trachea
antenna
male_reproductive_glands
wing
oenocyte
proboscis_and_maxpalp
head
leg
haltere
body_wall
body
testis
heart
fat_body
No. cells
Tep4
Nplp2
trol
LysX
Pvf2
Pxn
Att family
Cec family
nAChRalpha3
log(UMI)
E
muscle cells
TESTIS
OSM
VMC-M
*MS
VMM-A
IFM
VM-C
ABD
VMM-D
DPY
BODY
HEAD
*NEU
malpighian_tubule
haltere
wing
trachea
antenna
ovary
male_reproductive_glands
gut
fat_body
heart
head
oenocyte
testis
proboscis_and_maxpalp
leg
body_wall
body
No. cells
ovary
gut
TpnC73F
TpnC4
TpnC41C
MsR1
Dh31-R
CCAP-R
AstC-R2
Pdfr
log(UMI)

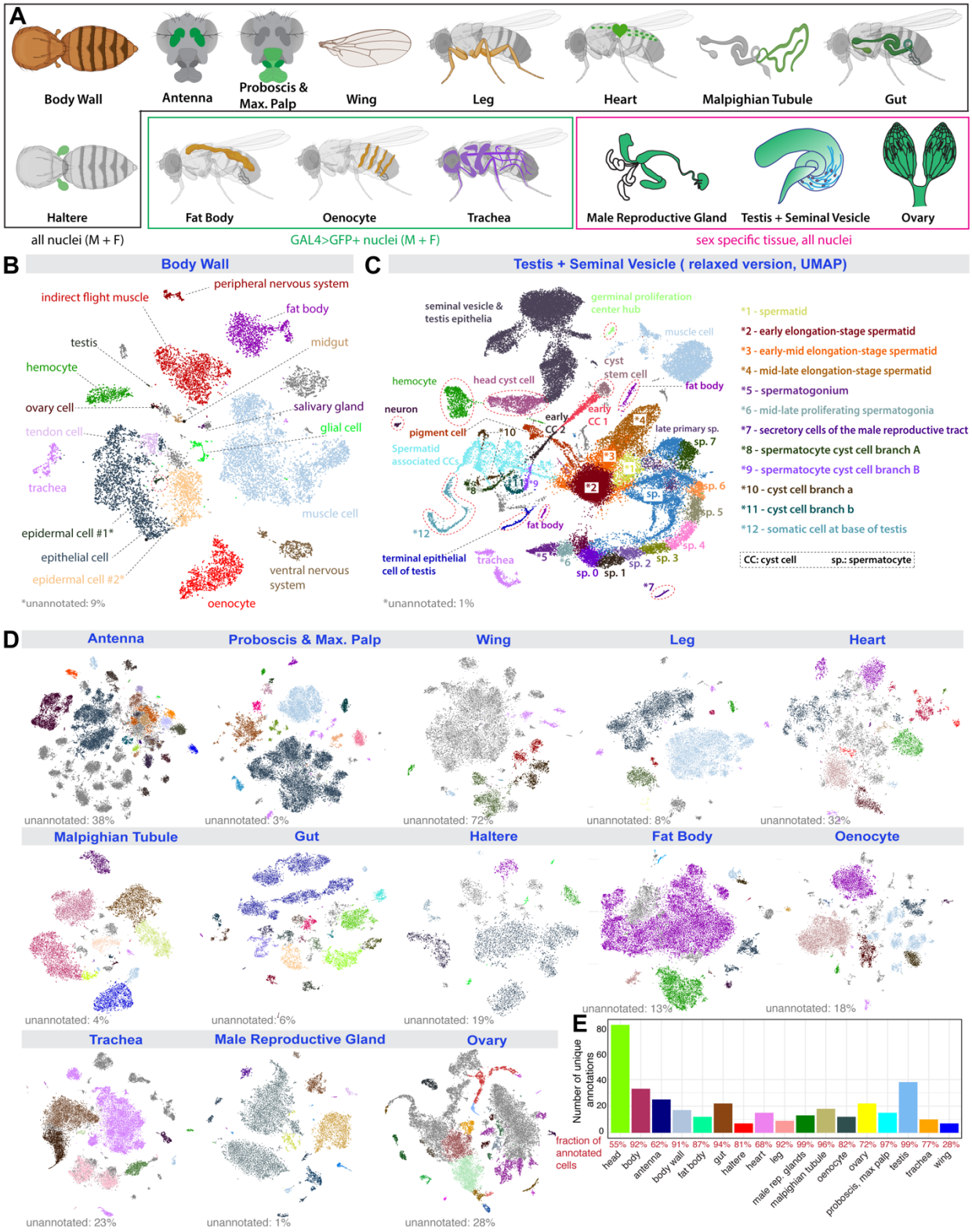
FCA Figure 5

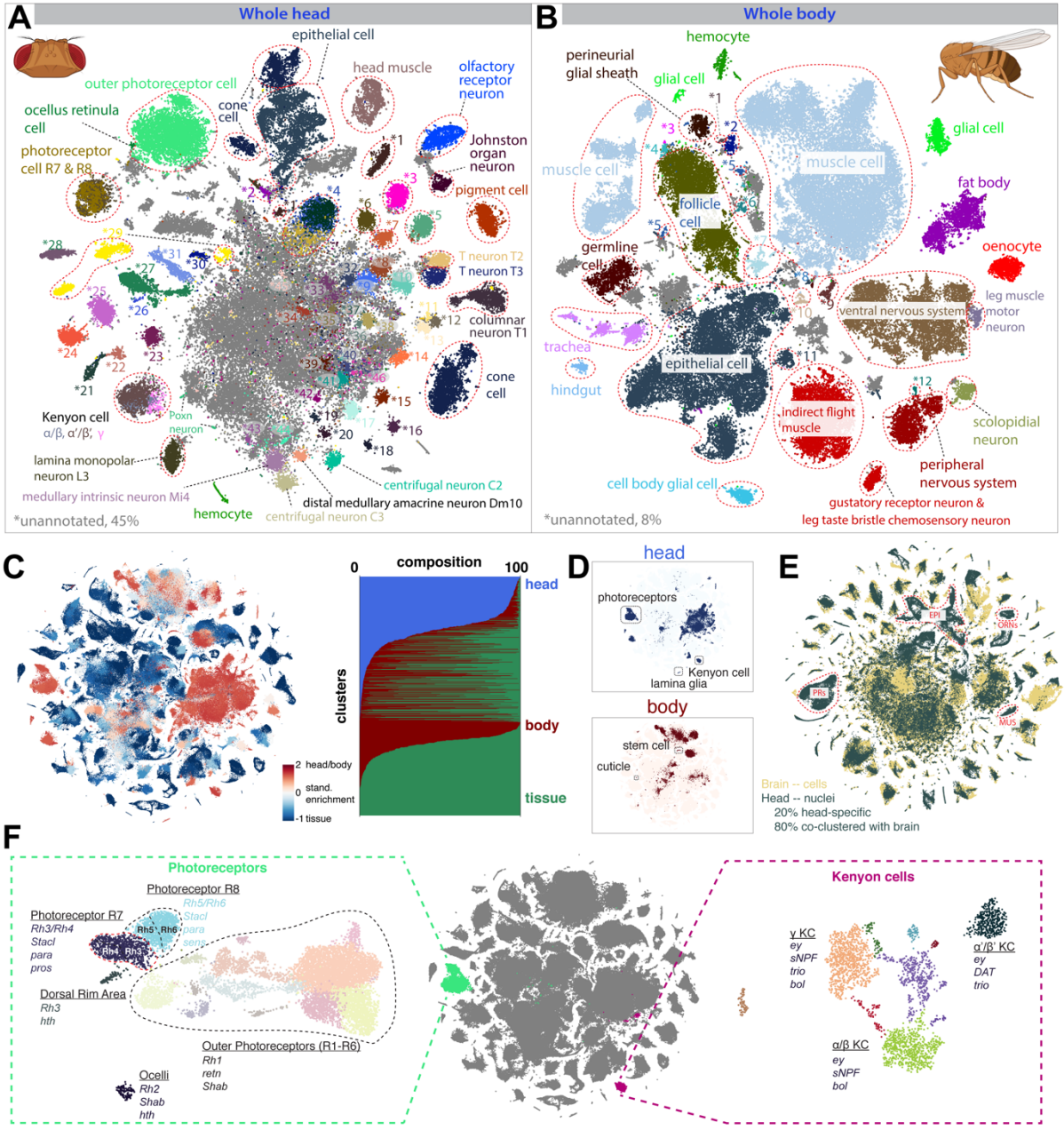

A
251 annotations
Standardised expression
500 transcription factors
B
broad annotation
gene
gene node colour: tau
0.85
1.0
neuron
gland
muscle cell
male germline cell
germline cell
male reproductive system
female germline cell
female reproductive system
excretory system
cardial cell
glial cell
epithelial cell
sensory organ cell
hemocyte
somatic precursor cell
tracheolar cell
fat cell
oenocyte
annotation → TF edges:
normalized expression (z-score)
1.5
4.0
annotation → annotation edges:
semantic similarity
0.5
0.9
C
Hand
dysf
grh
sug
HmgD
bcd
prd
repo
srp
Rel
org−1
pros
onecut
ey
bab2
gl
esg
trh
cardial cell
epithelial cell
excretory system
fat cell
female germline cell
fem. repr. system
germline cell
gland
glial cell
hemocyte
male germline cell
male repr. system
muscle cell
neuron
oenocyte
sensory organ cell
somatic precursor cell
tracheolar cell
D
Head -- gl expression
Photoreceptors
Cone cells
lns
log(CPM)
E
gl regulon
sens
Arr2
Kr-h1
Arr1
oc
Pph13
ninaE
ninaA
salm
boss
gl
ninaC
so
ninaB
trp
trpl
Rh5
neur
eya
Rh3
eyg
F
gene age
D. mel (FCA)
D. sec, D. sim
D. ere, D. yak
D. ana
D. per, D. pse
D. wil
D. gri, D. moj, D. vir
gene set size
1000
0
number of unique genes
glial cell
muscle cell
epithelial cell
tracheolar cell
hemocyte
somatic precursor cell
oenocyte
sensory neuron
CNS neuron
female reproductive system
fat cell
gland
cardial cell
secretory system
male reproductive system
male germline cell
female germline cell
2500
0
specific genes
common genes


A
XX → Sxl → Tra → Dsx F → female
X Sxl Tra Dsx M → male
B
Female cells
Male cells
muscle
fat cell
neuron
epithelial
dsx expression
(log2 norm'd+1, cell level)
<5 7.5 10.0 12.5 >15
excretory system
female-biased genes (%)
0 6 11 16 >20
male-biased genes (%)
0 4.5 8.0 11.5 >15
C
excretory cell types
MT principal cell
MT principal cell, initial segment
MT stellate cell, main segment
MT principal cell, lower segment
nephrocyte (pericardial cell)
other cell types
tracheal cell
pheromone-sensing neuron
salivary gland
oenocyte
crop
enterocyte-like
PNS
midgut
muscle cell
fat body
cardiomyocyte (non-ostia)
enterocyte of posterior midgut epi.
enterocyte of anterior midgut epi.
log10(# cells)
female-biased genes (%)
male-biased genes (%)
D
sex-biased
female-biased
male-biased
p = 3.5e-09
p = 1.5e-09
p = 4.6e-08
biased genes (%)
no/low dsx+
E
Spermatogonia - spermatocytes
late primary spermatocyte
spermatid
spermatogonium-spermatocyte transition
spermatogonium
spermatocyte
mid-late proliferating spermatogonia
F
Spermatids
mid-late elongation-stage spermatid
early-mid elongation-stage spermatid
early elongation stage spermatid
G
Early cyst cells
cyst stem cell
spermatocyte cyst cell branch A
early cyst cell 1
early cyst cell 2
spermatocyte cyst cell branch B
cyst cell branch b
Expression
low high
UMIs
annotation
1136 differential genes
152 differential genes
917 differential genes
cyst stem cells
sp. CC branch B
CC branch b
sp. CC branch A

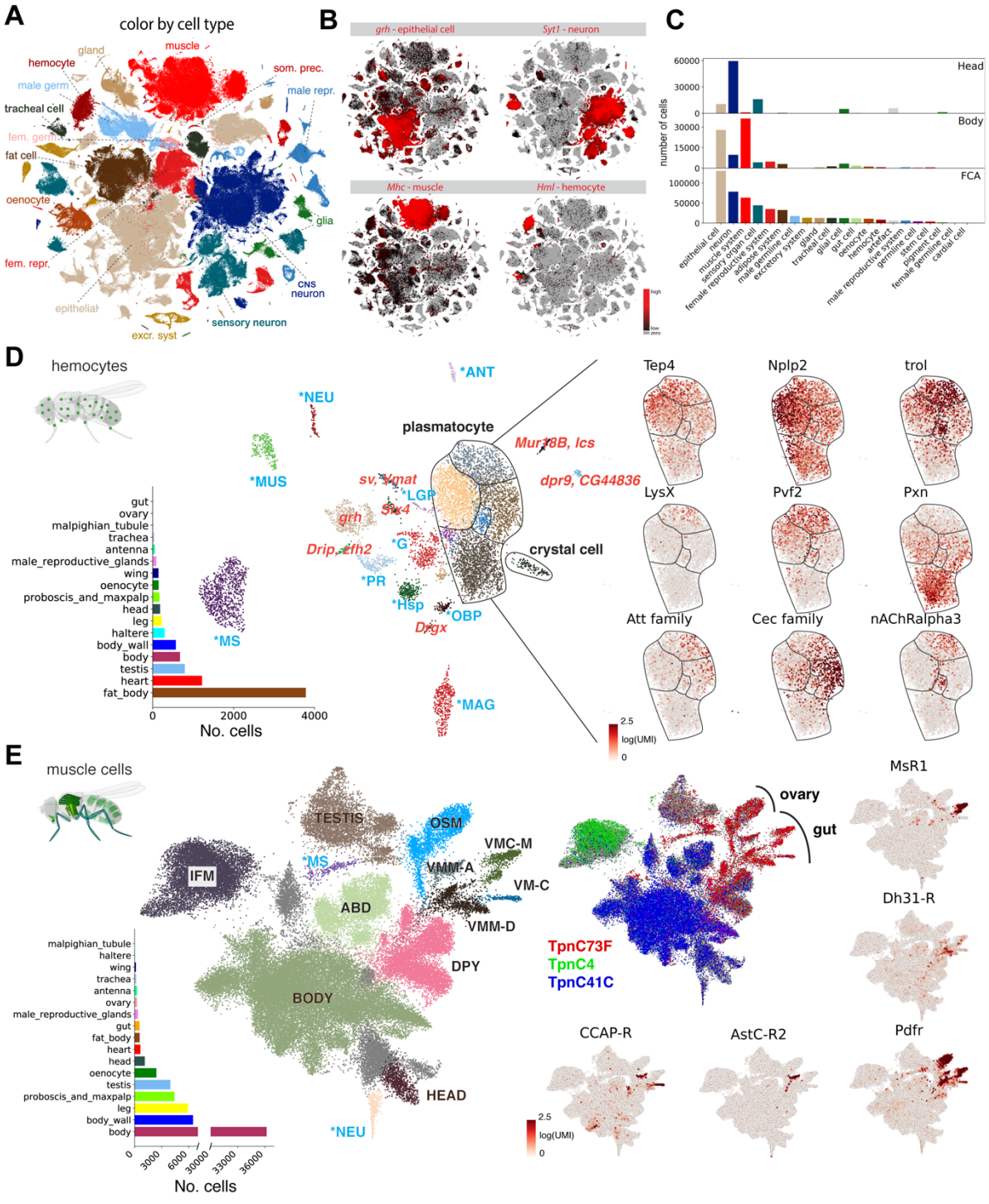
There is a repeat for figure 2-6, please ignore the later fig 2-6. sorry.

FCA Figure 2

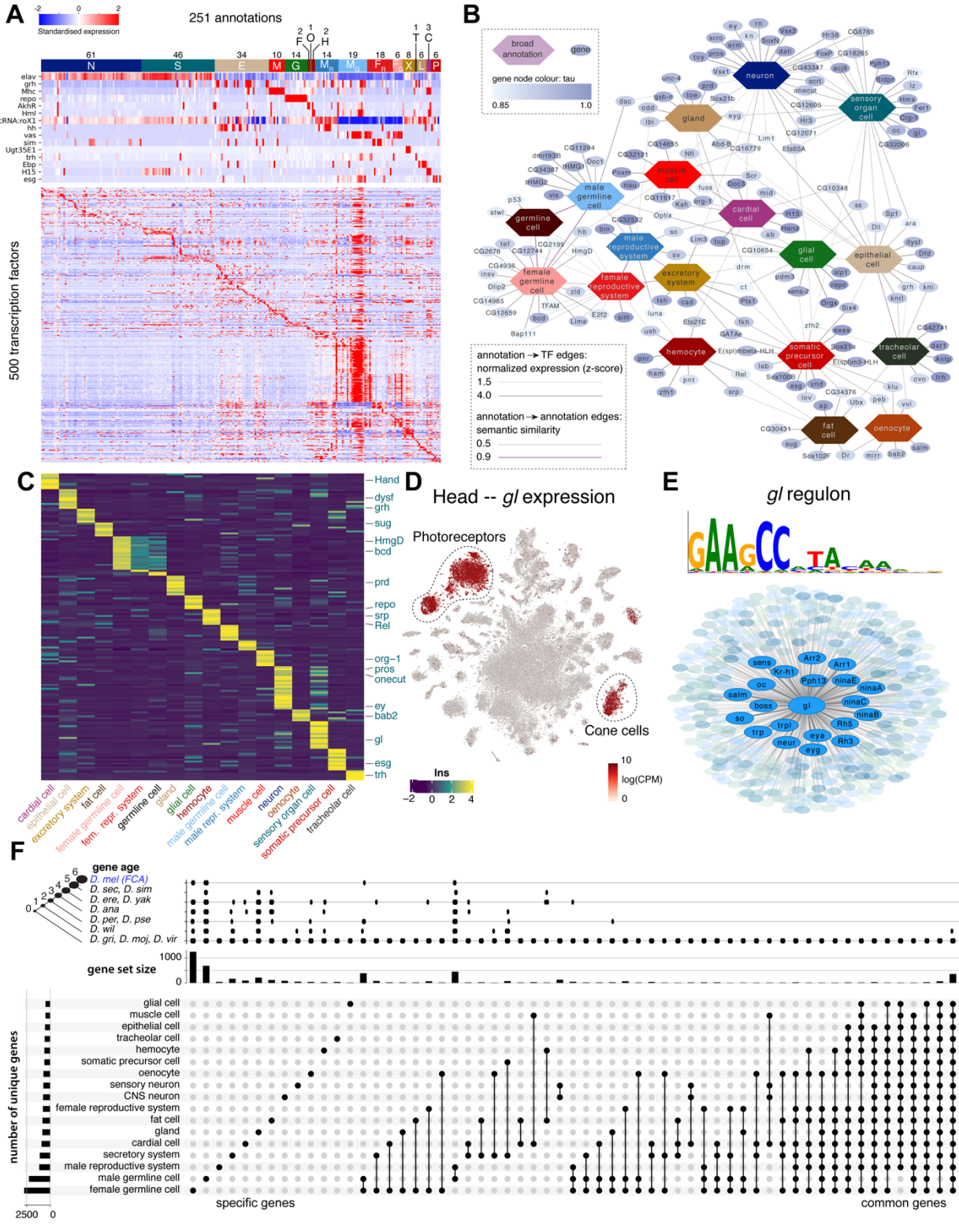


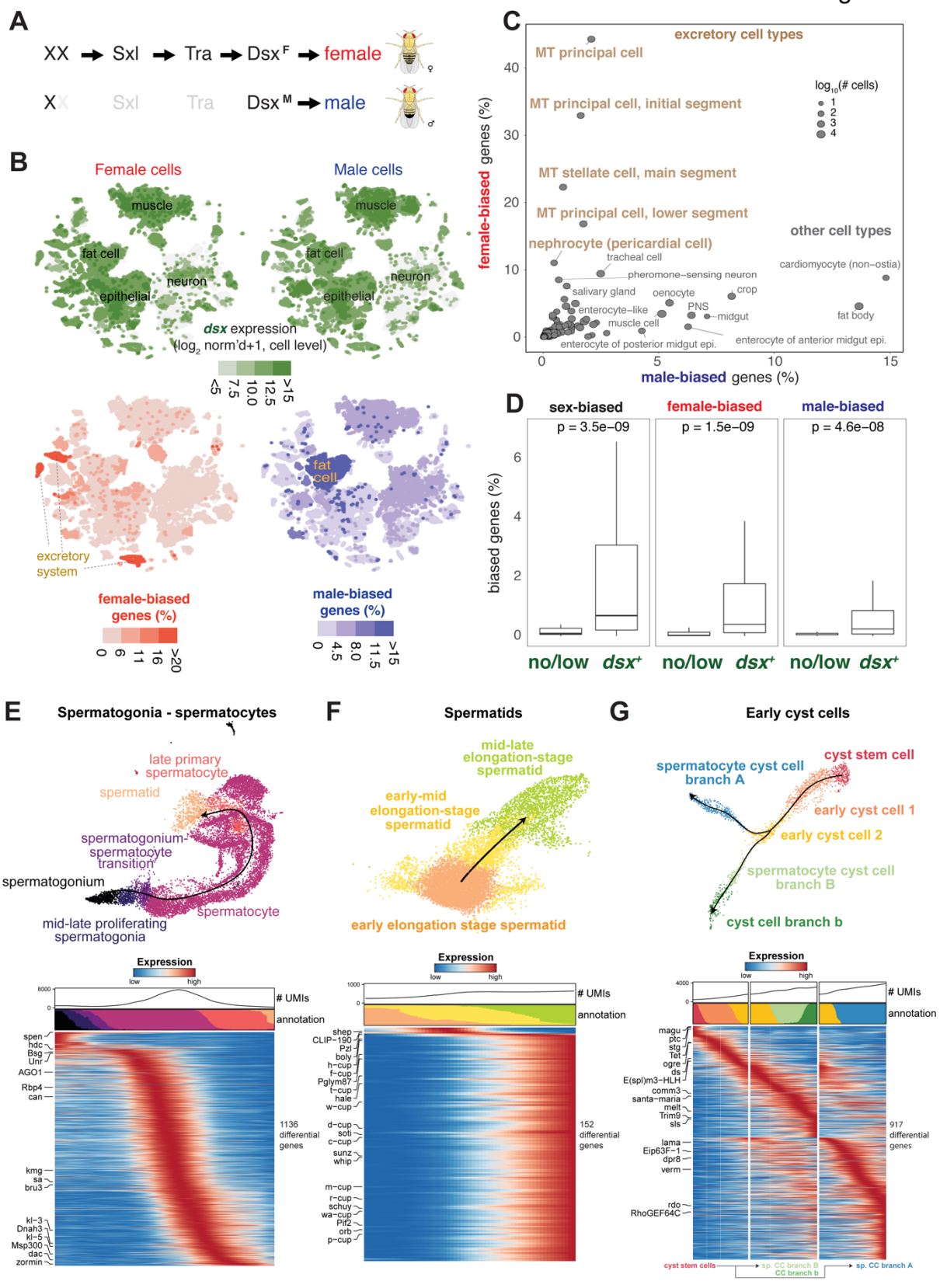

A
Whole head
epithelial cell
head muscle
olfactory receptor neuron
outer photoreceptor cell
cone cell
ocellus retinula cell
photoreceptor cell R7 & R8
Johnston organ neuron
pigment cell
T neuron T2
T neuron T3
columnar neuron T1
cone cell
Kenyon cell α/β, α′/β′, γ
Poxn neuron
lamina monopolar neuron L3
medullary intrinsic neuron Mi4
hemocyte
centrifugal neuron C3
centrifugal neuron C2
distal medullary amacrine neuron Dm10
*unannotated, 45%
B
Whole body
hemocyte
perineurial glial sheath
glial cell
muscle cell
follicle cell
germline cell
fat body
oenocyte
leg muscle motor neuron
ventral nervous system
trachea
hindgut
epithelial cell
indirect flight muscle
scolopidial neuron
peripheral nervous system
cell body glial cell
gustatory receptor neuron & leg taste bristle chemosensory neuron
*unannotated, 8%
C
composition
0
100
head
body
tissue
clusters
2 head/body
0 stand. enrichment
-1 tissue
D
head
photoreceptors
Kenyon cell
lamina glia
body
stem cell
cuticle
E
EPI
ORNs
PRs
MUS
Brain -- cells
Head -- nuclei
20% head-specific
80% co-clustered with brain
F
Photoreceptors
Photoreceptor R8
Rh5/Rh6
Photoreceptor R7
Rh3/Rh4
Stacl
para
pros
Dorsal Rim Area
Rh3
hth
Outer Photoreceptors (R1-R6)
Rh1
retn
Shab
Ocelli
Rh2
Shab
hth
Kenyon cells
γ KC
ey
sNPF
trio
bol
α′/β′ KC
ey
DAT
trio
α/β KC
ey
sNPF
bol

FCA Figure 4


A
color by cell type
gland
muscle
hemocyte
som. prec.
male germ
male repr.
tracheal cell
fem. germ
fat cell
oenocyte
glia
fem. repr.
CNS neuron
epithelial
sensory neuron
excr. syst
B
grh - epithelial cell
Syt1 - neuron
Mhc - muscle
Hml - hemocyte
high
low zero
C
Head
Body
FCA
number of cells
epithelial cell
neuron
muscle system
sensory organ cell
female reproductive system
adipose system
male germline cell
excretory system
gland
tracheal cell
glial cell
gut cell
oenocyte
hemocyte
antefact
male reproductive system
germline cell
stem cell
pigment cell
female germline cell
cardial cell
D
hemocytes
*NEU
*ANT
plasmatocyte
*MUS
Mur18B, lcs
sv, Vmat
*LGR
dpr9, CG44836
grh
Sw4
*G
crystal cell
Drip, elh2
*PR
*Hsp
*OBP
Dngx
*MS
*MAG
gut
ovary
malpighian_tubule
trachea
antenna
male_reproductive_glands
wing
oenocyte
proboscis_and_maxpalp
head
leg
haltere
body_wall
body
testis
heart
fat_body
No. cells
Tep4
Nplp2
trol
LysX
Pvf2
Pxn
Att family
Cec family
nAChRalpha3
log(UMI)
E
muscle cells
TESTIS
OSM
VMC-M
*MS
VMM-A
IFM
VM-C
ABD
VMM-D
DPY
BODY
HEAD
*NEU
malpighian_tubule
haltere
wing
trachea
antenna
ovary
male_reproductive_glands
gut
fat_body
heart
head
oenocyte
testis
proboscis_and_maxpalp
leg
body_wall
body
No. cells
ovary
gut
TpnC73F
TpnC4
TpnC41C
MsR1
Dh31-R
CCAP-R
AstC-R2
Pdfr
log(UMI)

FCA Figure 5


A
251 annotations
Standardised expression
500 transcription factors
B
broad annotation
gene
gene node colour: tau
annotation → TF edges:
normalized expression (z-score)
annotation → annotation edges:
semantic similarity
C
D
Head -- gl expression
Photoreceptors
Cone cells
log(CPM)
E
gl regulon
F
gene age
D. mel (FCA)
D. sec, D. sim
D. ere, D. yak
D. ana
D. per, D. pse
D. wil
D. gri, D. moj, D. vir
gene set size
number of unique genes
specific genes
common genes


A
XX → Sxl → Tra → Dsx F → female
X Sxl Tra Dsx M → male
B
Female cells
Male cells
muscle
fat cell
neuron
epithelial
dsx expression
(log2 norm'd+1, cell level)
<5 7.5 10.0 12.5 >15
excretory system
female-biased genes (%)
0 6 11 16 >20
male-biased genes (%)
0 4.5 8.0 11.5 >15
C
excretory cell types
MT principal cell
MT principal cell, initial segment
MT stellate cell, main segment
MT principal cell, lower segment
nephrocyte (pericardial cell)
other cell types
tracheal cell
pheromone-sensing neuron
salivary gland
oenocyte
crop
enterocyte-like
PNS
midgut
muscle cell
fat body
cardiomyocyte (non-ostia)
enterocyte of posterior midgut epi.
enterocyte of anterior midgut epi.
log10(# cells)
female-biased genes (%)
male-biased genes (%)
D
sex-biased
female-biased
male-biased
p = 3.5e-09
p = 1.5e-09
p = 4.6e-08
biased genes (%)
no/low dsx+
E
Spermatogonia - spermatocytes
late primary spermatocyte
spermatid
spermatogonium-spermatocyte transition
spermatogonium
spermatocyte
mid-late proliferating spermatogonia
F
Spermatids
mid-late elongation-stage spermatid
early-mid elongation-stage spermatid
early elongation stage spermatid
G
Early cyst cells
cyst stem cell
spermatocyte cyst cell branch A
early cyst cell 1
early cyst cell 2
spermatocyte cyst cell branch B
cyst cell branch b
Expression
low high
UMIs
annotation
1136 differential genes
152 differential genes
917 differential genes
cyst stem cells
sp. CC branch B
CC branch b
sp. CC branch A