Fly Cell Atlas: a single-nucleus 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[24], Teresa M. Przytycka[15], Aaron M. Allen[16], Stephen F. Goodwin[16], Cameron W. Berry[17], Margaret T. Fuller[17], Helen White-Cooper[18], Erika L. Matunis[19], Stephen DiNardo[20], Anthony Galenza[21], Lucy Erin O'Brien[21], Julian A. T. Dow[22], **FCA Consortium**[25], Heinrich Jasper[23], Brian Oliver[24], 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, and Chan Zuckerberg Biohub, San Francisco CA, 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, USA
[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, UK
[17] Department of Developmental Biology and Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
[18] Molecular Biosciences Division, Cardiff University, Cardiff, CF10 3AX UK
[19] Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[20] Perelman School of Medicine, The University of Pennsylvania, and The Penn Institute for Regenerative Medicine Philadelphia, PA 19104, USA
[21] Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305, USA
[22] Institute of Molecular, Cell & Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK.
[23] Immunology Discovery, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA

[24] Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Kidney and Digestive Diseases, National Institutes of Health, Bethesda, MD 20892, USA
[25] FCA Consortium: All authors listed before Acknowledgements, and all contributions and affiliations listed in the Supplementary 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
For over 100 years, the fruit fly *Drosophila melanogaster* has been one of the most studied model organisms. Here we present a single cell atlas of the adult fly, Tabula *Drosophilae*, that includes 580k nuclei from 15 individually dissected sexed tissues as well as the entire head and body, annotated to >250 distinct cell types. 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 between tissues, such as blood and muscle cells, reveals rare cell types and tissue-specific subtypes. This atlas provides a valuable resource for the entire *Drosophila* community and serves as a reference to study genetic perturbations and disease models at single-cell resolution.

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

Main Text

Drosophila melanogaster has a fruitful history in biological research, dating back to experiments of Thomas Hunt Morgan a century ago (*1*) and has been at the basis of many key biological discoveries. 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 (*2*), a large collection of genetic and molecular tools, and important databases such as Flybase (*3*), FlyMine (*4*), FlyLight (*5*), VirtualFlyBrain (*6*) and ModERN (*7*). The fly genome contains about 17,000 genes, including 13,968 protein-coding genes of which ~63% have human orthologues. Studies such as ModENCODE (*8*) and FlyAtlas (*9*) explored expression patterns in different tissues, but lacked cell type resolution. Recent advances in single-cell technologies have enabled the transcriptomic profiling of thousands of cells at once, facilitating the creation of tissue-wide atlases. Several studies have already applied single-cell RNA sequencing (scRNA-seq) to multiple *Drosophila* tissues and developmental stages (*10*). However, these data were 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 (*11*) with two sequencing strategies: droplet-based 10x Genomics (*12*) and plate-based Smart-seq2 (*13*). The resulting Tabula *Drosophilae*, the first dataset within the Fly Cell Atlas consortium (FCA), 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 the entire *Drosophila* community 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 (**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* tissues, especially cuticular ones (e.g., antenna, wing) and adipocyte-enriched ones (e.g., fat body). In addition, snRNA-seq can be applied to large multinucleated cells (e.g., muscle) and facilitates (frozen) tissue collection from different laboratories. Finally, 70–90% of transcriptomic information is preserved from snRNA-seq compared to scRNA-seq of the same fly cell types (*11*).

To achieve a comprehensive sampling, we used two complementary strategies. First, we dissected 12 individual tissues from both males and females, plus 3 sex-specific tissues (**Fig. 1A**). For tissues that are localized across the body (fat body, oenocytes, and trachea) and cannot be directly dissected, we used specific GAL4 lines driving nuclear-GFP to label and collect nuclei using FACS. In addition, two rare cell types were sequenced only with Smart-seq2: insulin-producing cells (IPCs) and corpora cardiaca cells (CCs). Second, we sorted and profiled nuclei from the entire head and body, aiming to detect cell types not covered by the selected tissues. In total, we obtained 580k high-quality nuclei: 570k from 10x Genomics and 10k from Smart-seq2 (**Fig. 1A)**.

To analyze the 10x Genomics data in a reproducible manner, we used the automated VSN pipeline (*14*) (Methods, **Table S1**), which takes the raw sequencing data as input and performs preprocessing (e.g., normalization, doublet removal, batch effect correction) to produce *LoomX* formatted files with expression data, embeddings and clusterings (**Fig. 1B** and **fig. S4**). A presumed artifactual cluster showed expression of nearly all genes, so we added an additional preprocessing step that models and subtracts ambient RNA signals (*15*) to remove this cluster, resulting in a *Stringent* dataset of 510k cells (see Methods and **Fig. 1C**). However, since adjusting the gene expression values per cell can introduce other biases (e.g., overcorrection, removal of non-doublet cells), we also retained the original *Relaxed* dataset of 570k cells. In the analyses below, unless mentioned otherwise (e.g., Fig. 2C), the *Stringent* dataset was used.

Cells from 10x Genomics and Smart-seq2 were well integrated after batch correction using Harmony (*16*) (**Fig. 1D**). 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 and a median of 16.5k cells per tissue for 10x and between 263 (male reproductive gland) and 1,349 (fat body) cells and a median of 534 cells per tissue for Smart-seq2 (**Fig. 1F**). We obtained similar numbers of male and female cells for non-sex-specific tissues with on average 1895 unique molecular identifiers (UMIs) and 828 genes per cell (**fig. S5**). 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 (**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 sexes: 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, testis, and ovary (**Fig. 2A**). We developed a consensus-voting strategy within the SCope web application (https://flycellatlas.org/scope) (*17*), where curators annotated clusters at multiple resolutions (ranging from 0.8 to 8, **fig. S6A**), with additional analysis performed in ASAP (https://flycellatlas.org/asap) (*18*). 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 used Flybase anatomy ontology terms (*19*).

Since some cell types are annotated at low, and others at high resolutions, we collapsed all annotations across resolutions and retained the annotation with the highest number of up-votes. 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. 2A-E** and **fig. S6-S18**). Overall, we annotated 251 cell types in the *Stringent* dataset (262 cell types if combining *Relaxed* and *Stringent datasets,* **Table S2**), with a median of 15 cell types per tissue.

Our dataset provides a single-cell transcriptomic profiling for several adult tissues not profiled previously, 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. In the proboscis and maxillary palp (**fig. S7A, B**), we could annotate gustatory and olfactory receptor neurons, mechanosensory neurons, and several glial clusters. All 7 olfactory receptors expressed in the maxillary palp were detected. In the wing (**fig. S8**), we could identify four different neuronal types – gustatory receptor neurons, pheromone-sensing neurons, nociceptive neurons, 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**), 15 cell types were identified, 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, cell types formed a big cluster instead of being split into distinct populations. In these cases, we identified genes or pathways that showed a 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 (*18*) revealed that fatty acid biosynthesis and degradation are in fact compartmentalized, highlighting possible fat body cell heterogeneity in metabolic capacities.

Our crowd annotations with tissue experts also revealed cell types that had not been profiled previously, such as multinucleated muscle cells (**Fig. 2B**) and two distinct types of nuclei among the main cells in the male accessory gland (**fig. S17**), a cell type that was previously thought to be uniform. The high number of nuclei analyzed allowed identification of rare cell types. For example, in the testis (**Fig. 2C**), we identified 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).

Next, we compared the distribution of cells between 10x and Smart-seq2, finding a good match based on a 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 transfer annotations from 10x to Smart-seq2 datasets (**fig. S20E**), using cluster-specific markers as validation (**fig. S20F**). We also identified genes that were specifically detected using Smart-seq2 thanks to its higher gene detection rate (**fig. S20G** and **Fig. 1E**). In summary, the high-throughput 10x datasets form the basis for identifying cell types while the Smart-seq2 datasets facilitate the detection of lowly 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 next performed snRNA-seq experiments on whole-head and whole-body samples. Whole-body single-cell experiments were previously performed on less complex animals (*20*, *21*). Full head and body sequencing provides a practical means to assess the impact of mutations or to track disease

mechanisms, without having to focus on specific tissues. In addition, it could yield cell types that are not covered by any of the targeted tissue dissections.

In the head, we annotated 81 mostly neuronal cell types (**Fig. 3A and S22**). 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 cells, oenocytes, germ line cells, glia, and tracheal cells (**Fig. 3B and S23**). Many of these cell classes can be further divided into cell types for further annotation (see **Fig. 2** and **fig. S6-S18**).

Next, we examined how well the head and body samples covered the cell types from the dissected tissues. We analyzed head, body, and tissue samples together, with most of the selected tissues clustering together with the body. We also detected head and body enriched clusters (**Fig. 3C**). One body-specific cluster contained cuticle cells, likely from connective tissue (**Fig. 3D)**. 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 the head, antenna and proboscis with maxillary palp were dissected for tissue sequencing. Cell types from those two tissues largely overlapped with head cells. 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 datasets, we integrated our head snRNA-seq dataset ("head" hereafter) with published brain single-cell RNA-seq data ("brain" hereafter) (*17*, *22–24*) (**Fig. 3E**). Head unique clusters made up 20% of the cells, including the antennae, photoreceptors, muscle, cone cells and cuticular cell types, whereas the other 80% were present in 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 quality and utility of our snRNA-seq data compared to scRNA-seq data. Next, we used machine learning models to predict annotations per cluster, followed by manual curation (*22*). Given the high number of neuron types, additional subclustering was performed on each cluster, identifying subtypes of peptidergic neurons (*dimm*, *Pdf)* 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. A big clump of neurons remained unannotated (**Fig. 3A**), indicating that our dataset cannot resolve the complexity of the central brain, which may contain hundreds to thousands of neuron types.

Subclustering in the combined dataset separated inner and outer photoreceptors from dorsal rim area and ocellar photoreceptors, with the inner photoreceptors further splitting into R7 and R8 types, each with *pale* and *yellow* types based on *rhodopsin* expression (**Fig. 3F**). Additionally, Kenyon cells were split into three types: α/β, α'/β' and γ (*17*). These cases 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 assigned cells to major cell classes (e.g., epithelial cells, neurons, muscle cells, hemocytes), allowing us to compare common classes across tissues (**Fig. 4A-C and fig. S24, S25**). First, we compared blood cells across tissues by selecting all *Hml*-positive cells, a known marker for hemocytes (**Fig. 4D**). 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 gradients based on the expression of *Pxn*, *LysX*, *Tep4*, *trol*

and *Nplp2* that can be linked to maturation and plasticity with *Pxn* positive cells showing the highest *Hml* expression, while *Tep4*, *trol* and *Nplp2* are prohemocyte markers (*25*). Furthermore, different antimicrobial peptide (AMP) families such as the *Attacins* and *Cecropins* were expressed in different subgroups indicating specialization. Finally, expression of acetylcholine receptors was specific for a subset of hemocytes, relating to the cholinergic anti-inflammatory pathway as described in humans and mice (*26*). Lamellocytes were not observed in adults as previously suggested (*27*). 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 (*28*, *29*) (**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** and **fig. S26C, D**). 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. This comprehensive view of the fly muscular system highlights a separation of visceral, skeletal, and indirect flight muscle based on the expression of different *troponins*. Specifically, we discovered gradients of *dysf* and *fln* in the indirect flight muscle, which may indicate regional differences in these very large cells (>1000 nuclei) (**fig. S26E**). We identified four 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 (*30*). *Ms* and *Dh31* have been described to function in spatially restricted domains (30, 31, 32), suggesting similar domains for *AstC* and *CCAP*. All visceral muscle cells are enriched for the receptor of *Pdf*, a neuropeptide involved in circadian rhythms, pointing towards a function in muscle contraction as well (33).

Transcription factors and cell type specificity

Our data allow the comparison of gene expression across the entire fly. Clustering cell types showed the germline cells as the most distinct group, followed by neurons (**fig. S27-S32**). We calculated marker genes 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)**.

Next, we calculated the *tau* score of tissue specificity (*34*) for all predicted transcription factors (TFs) (*3*), identifying 500 TFs with a score > 0.85, indicating a high specificity for one or very few cell types (**Fig. 5A, Table S3**). 127 of these TFs were "CGs" (computed genes), indicating that their functions are poorly studied. 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, as discussed below.

Similar analysis across broad cell types **(Fig. 5B, C)** identified 156 TFs with high *tau* scores, for example the known regulators *grh* for epithelial cells and *repo* for glia, as well as 24 uncharacterized genes. Network visualization shows 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.

In addition to the specificity of TF expression, we predicted gene regulatory networks based on co-expression and motif enrichment using SCENIC (*31*). Because of the stochasticity of this network inference method, we ran SCENIC 100 times, ranking 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 target genes that are predicted in at least 5/100 runs. In fat cells, our analysis predicted a regulon for *sugarbabe* (*sug*), a sugar-sensitive TF necessary for the induction of lipogenesis (*32*). 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)(*33*)**. The SCENIC predictions for all cell types are available via SCope (https://flycellatlas.org/scope).

Comparative analysis of genes across broad cell types or tissues (**Fig. 5F, fig. S34**) identified common genes and specifically expressed genes, such as a shared set of 555 housekeeping genes that are expressed in all tissues. The testis has the highest number of uniquely expressed genes consistent with previous reports (*34*), 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 Malpighian tubules (*35*). 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, reflecting the evolutionary origin of halteres as a modified wing (*36*) (**fig. S34**).

Analysis of sex-biased expression and sex-specialized tissues

To study sex-related differences, we compared male- versus female-derived nuclei for all common tissues (**fig. S35**), finding *roX1/2* and *Yp1/2/3* as the top male- and female-specific genes, respectively. Notably, a large fraction of genes with male-enriched expression were uncharacterized (*37*). The primary sex determination pathway in somatic cells leads to sex-specific splicing of *doublesex* (*dsx*) to encode female- or male-specific TFs (*38*) (**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**).

Next, 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**). Female-biased genes (i.e., *Ics* and *whe*) were differentially expressed under high salt conditions, suggesting sex-bias in nephric ion transport. Across cell types, sex-biased expression strongly correlated with *dsx* expression (**Fig. 6D**) (*39*), consistent with the role of Dsx as a key regulator.

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 the gut and ovary stem cell lineages in previous studies (*40–42*), and our FCA data on gut and ovary accurately co-clustered with these published datasets (**fig. S36, S37**). Therefore, we focused on the testis plus seminal vesicle as a case study. 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**). The main testis analysis (**Fig. 2C**) revealed transitions from GSCs and proliferating spermatogonia, spermatocytes, to maturing spermatids, and finally late elongation stage spermatids.

We further performed trajectory inference on spermatocytes and spermatids separately (**Fig. 6E-F**). As expected, the spermatocyte stage featured a continuous increase in the number of genes being transcribed (**Fig. 6E**), with many of the strongly upregulated genes (*kmg, Rbp4, fzo, can, sa*, and, for later spermatocytes, Y-linked fertility factors *kl-3* and *kl-5*) not substantially expressed in any other cell type. Late spermatocytes, however, showed expression of marker genes from many other cell types like somatic cells (*Upd1*, *eya*)), epithelial cells (*grh*), muscle (*Mhc*) or hemocytes (*Hml*) (**Fig. 5A**), although their expression level was lower than in their marked cell type. Early spermatids are in transcriptional quiescence, as can be seen by a very low number of nuclear transcripts (**Fig. 6F,** low UMI), followed by a burst of new transcription in elongating spermatids including many *cup* genes (48).

In the somatic cyst cell lineage, we found CySCs expressing the cell cycle marker *string*, transitioning into post-mitotic (no *string* expression) early cyst cells, and branching into two related clusters of cyst cells likely associated with spermatocytes (**Fig. 6G)**.

Discussion

Recent technological development has enabled single-cell transcriptomic atlases of *C. elegans* (*21*) and selected tissues in mice and humans (*43–46*). Here, we provide a single-cell transcriptomic map of the entire adult *Drosophila melanogaster*, a premier model organism for studies of fundamental and evolutionarily conserved biological mechanisms. The FCA provides a resource for the *Drosophila* community as a reference for studies of gene function at single-cell 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 transcriptomic 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), in contrast to other, more detailed cell types (e.g., different ORNs). 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 and expression or by mistakes that were made in the literature. These examples highlight the need and the opportunities presented by Tabula *Drosophilae* to serve as the basis for future validation.

We have generated lists of marker genes per cell type with different levels of specificity, ranging from tissue-wide to animal-wide. This unique level of precision presents a blueprint for future integration with other data modalities such as single-cell ATAC-seq (*47*) and spatial omics, and for generating cell-type reporter lines to study new cellular functions. Furthermore, the large number of uncharacterized genes that show cell-type specific, sex-biased or trajectory-dependent expression provides the foundation for many follow-up studies. Our analysis also presents several technical novelties, including the use of reproducible Nextflow pipelines (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), linked to an online analysis platform in ASAP (https://asap.epfl.ch/fca). These elements may inspire future atlas projects. Given the work in other model organisms, we also envision a use for the FCA data in 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 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**; links available on https://www.flycellatlas.org).

References and Notes

1. T. H. Morgan, SEX LIMITED INHERITANCE IN DROSOPHILA. *Science*. **32**, 120–122 (1910).

2. M. D. Adams et al., The genome sequence of Drosophila melanogaster. *Science*. **287**, 2185–2195 (2000).

3. A. Larkin et al., FlyBase Consortium, FlyBase: updates to the Drosophila melanogaster knowledge base. *Nucleic Acids Res.* **49**, D899–D907 (2021).

4. R. Lyne et al., FlyMine: an integrated database for Drosophila and Anopheles genomics. *Genome Biol.* **8**, R129 (2007).

5. A. Jenett et al., A GAL4-driver line resource for Drosophila neurobiology. *Cell Rep.* **2**, 991–1001 (2012).

6. N. Milyaev et al., The Virtual Fly Brain browser and query interface. *Bioinformatics*. **28**, 411–415 (2012).

7. M. M. Kudron et al., The ModERN Resource: Genome-Wide Binding Profiles for Hundreds of Drosophila and Caenorhabditis elegans Transcription Factors. *Genetics*. **208**, 937–949 (2018).

8. modENCODE Consortium et al., Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science*. **330**, 1787–1797 (2010).

9. 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).

10. H. Li, Single-cell RNA sequencing in Drosophila: Technologies and applications. *Wiley Interdiscip. Rev. Dev. Biol.* **10**, e396 (2021).

11. C. N. McLaughlin et al., Single-cell transcriptomes of developing and adult olfactory receptor neurons in Drosophila. *eLife*. **10** (2021), doi:10.7554/eLife.63856.

12. G. X. Y. Zheng et al., Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

13. S. Picelli et al., Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*. **10**, 1096–1098 (2013).

14. M. De Waegeneer, C. C. Flerin, K. Davie, G. Hulselmans, vib-singlecell-nf/vsn-pipelines: v0.26.0 (v0.26.0). Zenodo. https://doi.org/10.5281/zenodo.5055627. *Zenodo* (2021).

15. S. Yang et al., Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol.* **21**, 57 (2020).

16. I. Korsunsky et al., Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods*. **16**, 1289–1296 (2019).

17. K. Davie et al., A Single-Cell Transcriptome Atlas of the Aging *Drosophila* Brain. *Cell*. **174**, 982-998.e20 (2018).

18. 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).

19. M. Costa, S. Reeve, G. Grumbling, D. Osumi-Sutherland, The Drosophila anatomy ontology. *J. Biomed. Semantics*. **4**, 32 (2013).

20. S. Levy et al., A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. *Cell*. **184**, 2973-2987.e18 (2021).

21. J. Cao et al., Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. **357**, 661–667 (2017).

22. M. N. Özel et al., Neuronal diversity and convergence in a visual system developmental atlas. *Nature*. **589**, 88–95 (2021).

23. H. Li et al., Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing. *Cell*. **171**, 1206-1220.e22 (2017).

24. Y. Z. Kurmangaliyev et al., Transcriptional programs of circuit assembly in the drosophila visual system. *Neuron*. **108**, 1045-1057.e6 (2020).

25. B. Cho et al., Single-cell transcriptome maps of myeloid blood cell lineages in Drosophila. *Nat. Commun.*

11, 4483 (2020).
26. V. A. Pavlov, K. J. Tracey, The cholinergic anti-inflammatory pathway. *Brain Behav. Immun.* **19**, 493–499 (2005).
27. P. Sanchez Bosch et al., 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).
28. J. Krzemień, L et al., Control of blood cell homeostasis in Drosophila larvae by the posterior signalling centre. *Nature*. **446**, 325–328 (2007).
29. L. Mandal et al., A Hedgehog- and Antennapedia-dependent niche maintains Drosophila haematopoietic precursors. *Nature*. **446**, 320–324 (2007).
30. R. J. Siviter et al., Expression and functional characterization of a Drosophila neuropeptide precursor with homology to mammalian preprotachykinin A. *J. Biol. Chem.* **275**, 23273–23280 (2000).
31. S. Aibar et al., SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods*. **14**, 1083–1086 (2017).
32. J. Mattila, V. Hietakangas, Regulation of Carbohydrate Energy Metabolism in Drosophila melanogaster. *Genetics*. **207**, 1231–1253 (2017).
33. 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).
34. H. Kaessmann, Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**, 1313–1326 (2010).
35. Y. Shao et al., GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Res.* **29**, 682–696 (2019).
36. E. B. Lewis, A gene complex controlling segmentation in Drosophila. *Nature*. **276**, 565–570 (1978).
37. J. Andrews et al., Gene Discovery Using Computational and Microarray Analysis of Transcription in the *Drosophila melanogaster* Testis. *Genome Res.* **10**, 2030–2043 (2000).
38. H. K. Salz, J. W. Erickson, Sex determination in Drosophila: The view from the top. *Fly (Austin)*. **4**, 60–70 (2010).
39. E. Clough et al., Sex- and tissue-specific functions of Drosophila doublesex transcription factor target genes. *Dev. Cell*. **31**, 761–773 (2014).
40. R.-J. Hung et al., A cell atlas of the adult Drosophila midgut. *Proc Natl Acad Sci USA*. **117**, 1514–1523 (2020).
41. K. Rust et al., A single-cell atlas and lineage analysis of the adult Drosophila ovary. *Nat. Commun.* **11**, 5628 (2020).
42. A. Jevitt et al., A single-cell atlas of adult Drosophila ovary identifies transcriptional programs and somatic cell lineage regulating oogenesis. *PLoS Biol.* **18**, e3000538 (2020).
43. Tabula Muris Consortium et al., Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*. **562**, 367–372 (2018).
44. X. Han et al., Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*. **173**, 1307 (2018).
45. J. Cao et al., A human cell atlas of fetal gene expression. *Science*. **370** (2020), doi:10.1126/science.aba7721.
46. X. Han et al., Construction of a human cell landscape at single-cell level. *Nature*. **581**, 303–309 (2020).
47. J. Janssens et al., Decoding gene regulation in the fly brain. *Nature* (2022), doi:10.1038/s41586-021-04262-z.

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 J González-Segarra, 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 entire fly community for the enthusiastic support for this project, Bill Burkholder, Cathryn Murphy, and Kathleen Vogelaers for coordinating FCA and all Jamboree meetings. **Funding:** The sequencing 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). FCA Consortium Funding in the Supplemental Materials. **Author contributions:** See FCA Consortium Contributions in Supplementary Materials. **Competing interests:** H. Jasper, N.S. Katheder and X.T. Cai are employees of Genentech, Inc. Other authors declare no competing interests. **Data and materials availability**: All data are available for user-friendly querying via https://flycellatlas.org/scope and for custom analyses at https://flycellatlas.org/asap. For each tissue, a CellxGene portal is also available (www.flycellatlas.org). Raw data and count matrices can be downloaded from ArrayExpress (accession number E-MTAB-10519 for 10x, and E-MTAB-10628 for Smart-seq2; the same accession numbers are available at EBI Single Cell Expression Atlas https://www.ebi.ac.uk/gxa/sc). 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. Three Supplemental Figures describe how to access and explore FCA data: fig. S1 for summary of Data Availability, fig. S2 and S3 for how to use SCope and ASAP. We also include a video tutorial for using Scope (https://www.youtube.com/watch?v=yNETQVaSJYM&t=349s). Analysis codes are at Github

(https://github.com/flycellatlas). Dataset access: GSE107451 (scRNA-seq adult fly brain), GSE120537 (scRNA-seq adult fly gut), GSE136162, GSE146040 and GSE131971 (scRNA-seq adult ovary). The neural network from (*22*) (Appendix 1).

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 S6
References (*48-58*)

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 (*17*) and ASAP (*18*).
(C) Two versions of 10x datasets: *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.
(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 are 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 comprising three distinct types: α/β, α'/β' and γ.

Figure 4: Cross-tissue analyses of common cell classes

(A) Overview of main cell classes identified throughout the fly cell atlas. Som. pre., somatic precursor cells; male repr. and fem. repr., male and female reproductive system; male germ. and fem. germ., male and female germline cells.

(B) tSNE plots showing expression of four markers in four common cell classes.

(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.

(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* is expressed in all visceral muscle cells, including the ovarian sheath muscle; other four receptor genes (*AstC-R2, MsR1, Dh31-R, CCAP-R*) are 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), cardiac cell (C), somatic precursor cell (P).

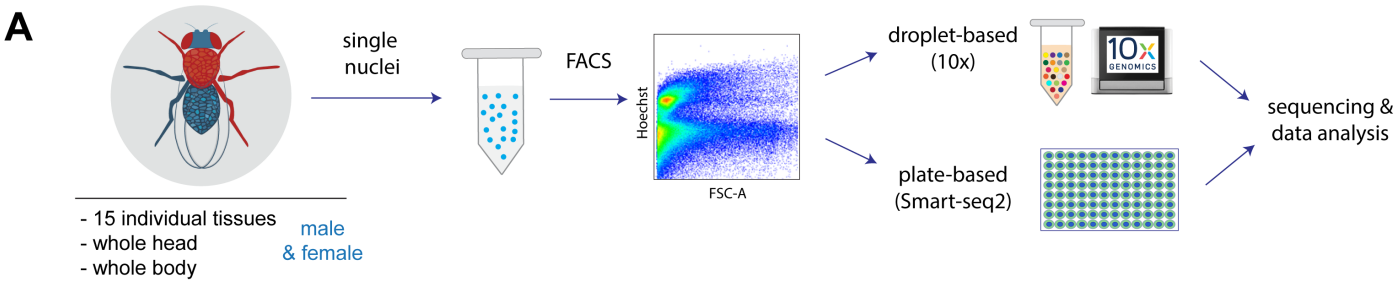
(B) A network analysis of TFs and cell classes based on similarity of ontology terms, reveals unique and shared TFs across the individual tissues.

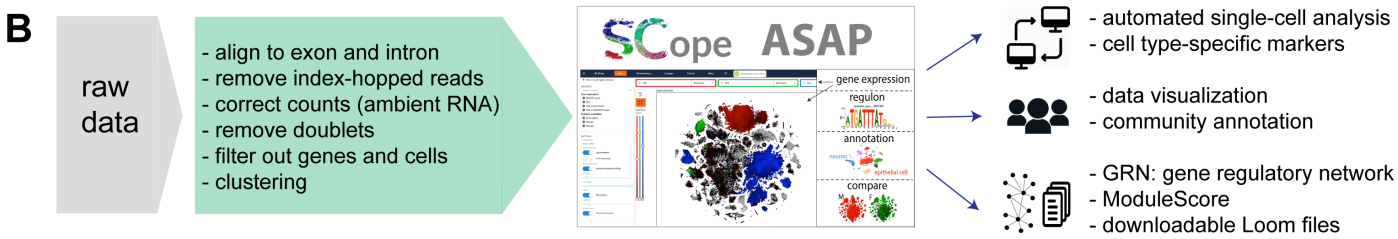
(C) Heatmap showing the expression of unique TFs per cell class. Factors from the literature are highlighted.

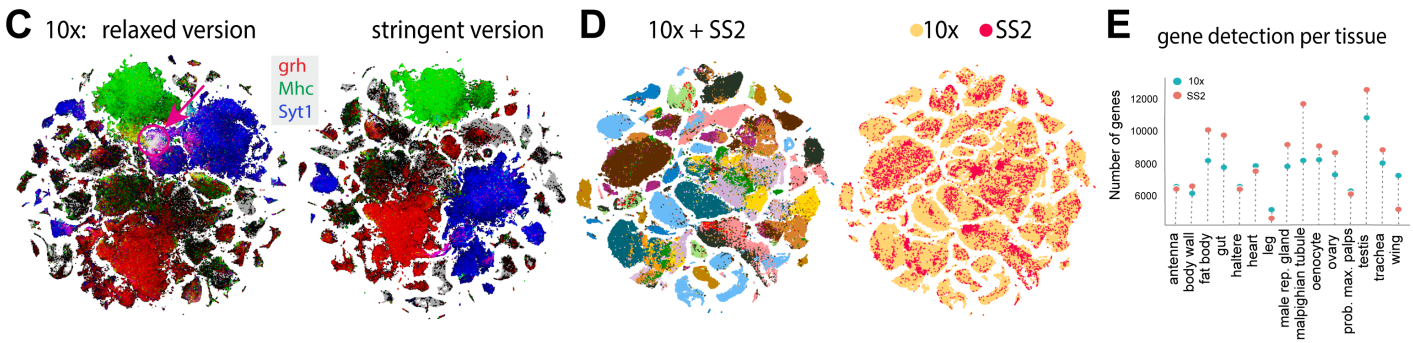
(D) *Glass* is uniquely expressed in photoreceptors and cone cells in the head.
(E) Overview of the *Glass* regulon of 444 target genes, highlighting known photoreceptor marker genes.
(F) Gene expression comparison across broad cell types. Only sets with more than 10 genes are shown. The left bar graph 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.

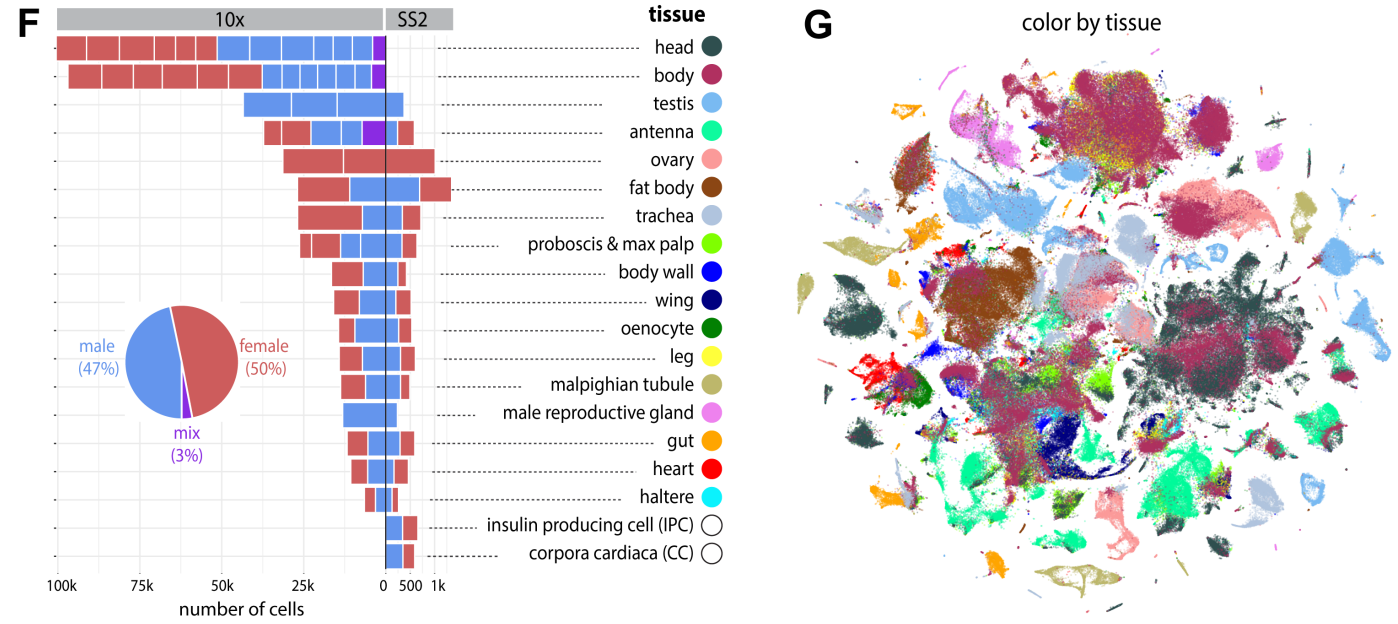
Figure 6. Sex-biased expression and trajectory analysis of testis cell lineages
(A) Simplified sex determination pathway. Sex chromosome karyotype (XX) activates Sex-lethal (Sxl) which regulates transformer (Tra), resulting in a female Dsx isoform (Dsx^F). In XY (or X0) flies, Sxl and Tra are inactive (light gray) and the male-specific Dsx^M is produced.
(B) Top, Dsx expression and female- and male-biased expression projected onto tSNE plots of all female (left column) and male (right column) cells except reproductive tissue cells (**Table S4 and S5**). female- and male-biased expression measured as the percentage of genes in the cluster showing biased expression in favor of the respective sex (**Table S6**). These percentage values were computed for each annotated cluster and those cluster-level values were projected onto 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 (> 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 S6**). Data point size indicatess cell numbers 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. 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. p-values are based on Wilcoxon test.
(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

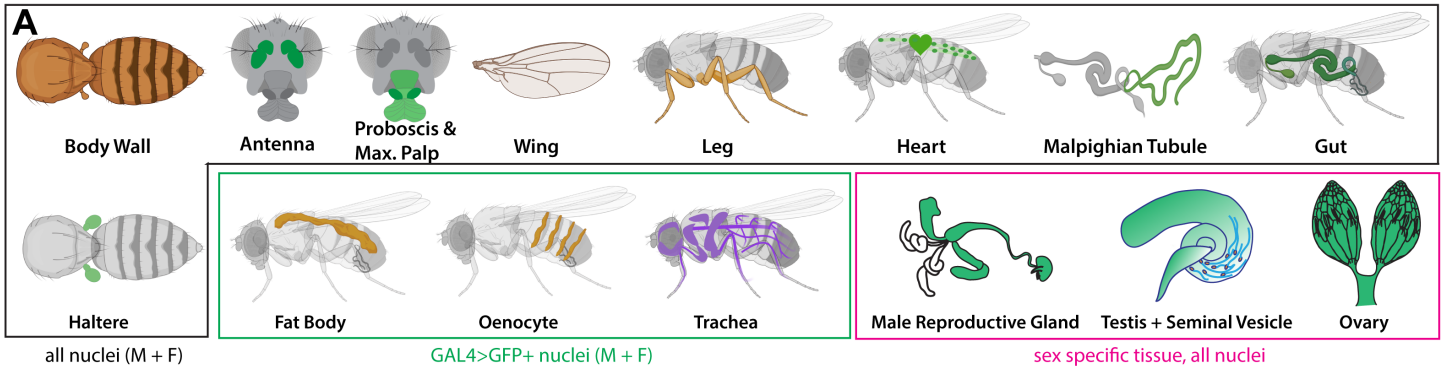

A
single nuclei
FACS
Hoechst
FSC-A
droplet-based (10x)
plate-based (Smart-seq2)
10x GENOMICS
sequencing & data analysis
- 15 individual tissues
- whole head
- whole body
male & female

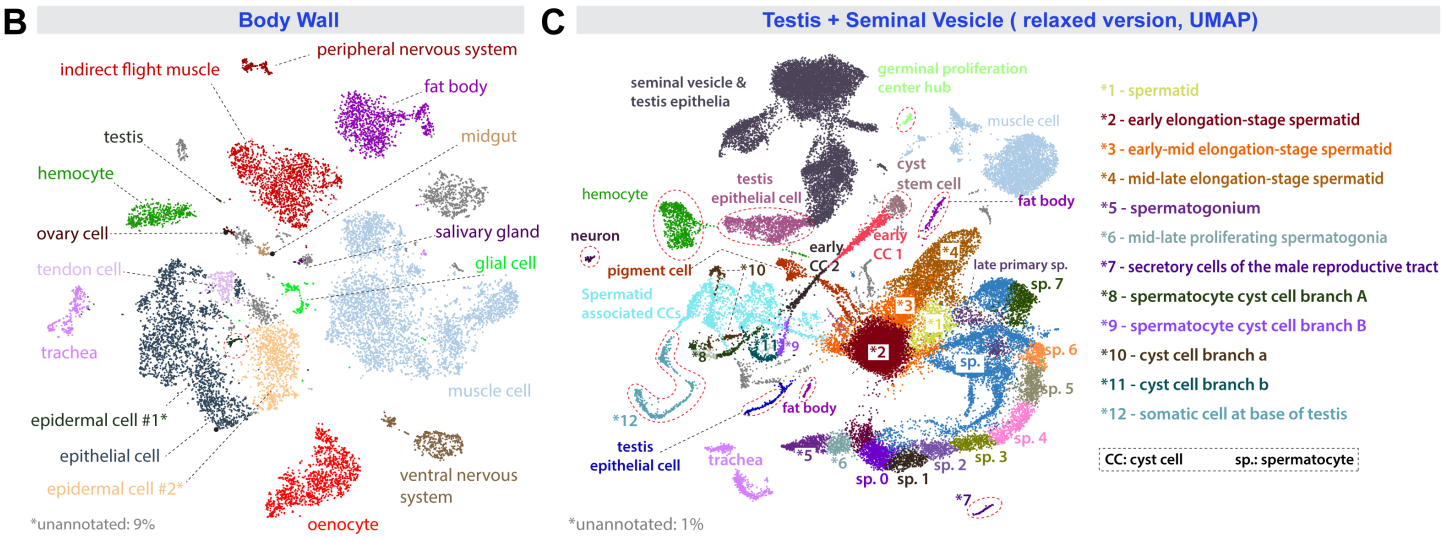

B
raw data
- align to exon and intron
- remove index-hopped reads
- correct counts (ambient RNA)
- remove doublets
- filter out genes and cells
- clustering
SCope ASAP
gene expression
regulon
annotation
compare
- automated single-cell analysis
- cell type-specific markers
- data visualization
- community annotation
- GRN: gene regulatory network
- ModuleScore
- downloadable Loom files

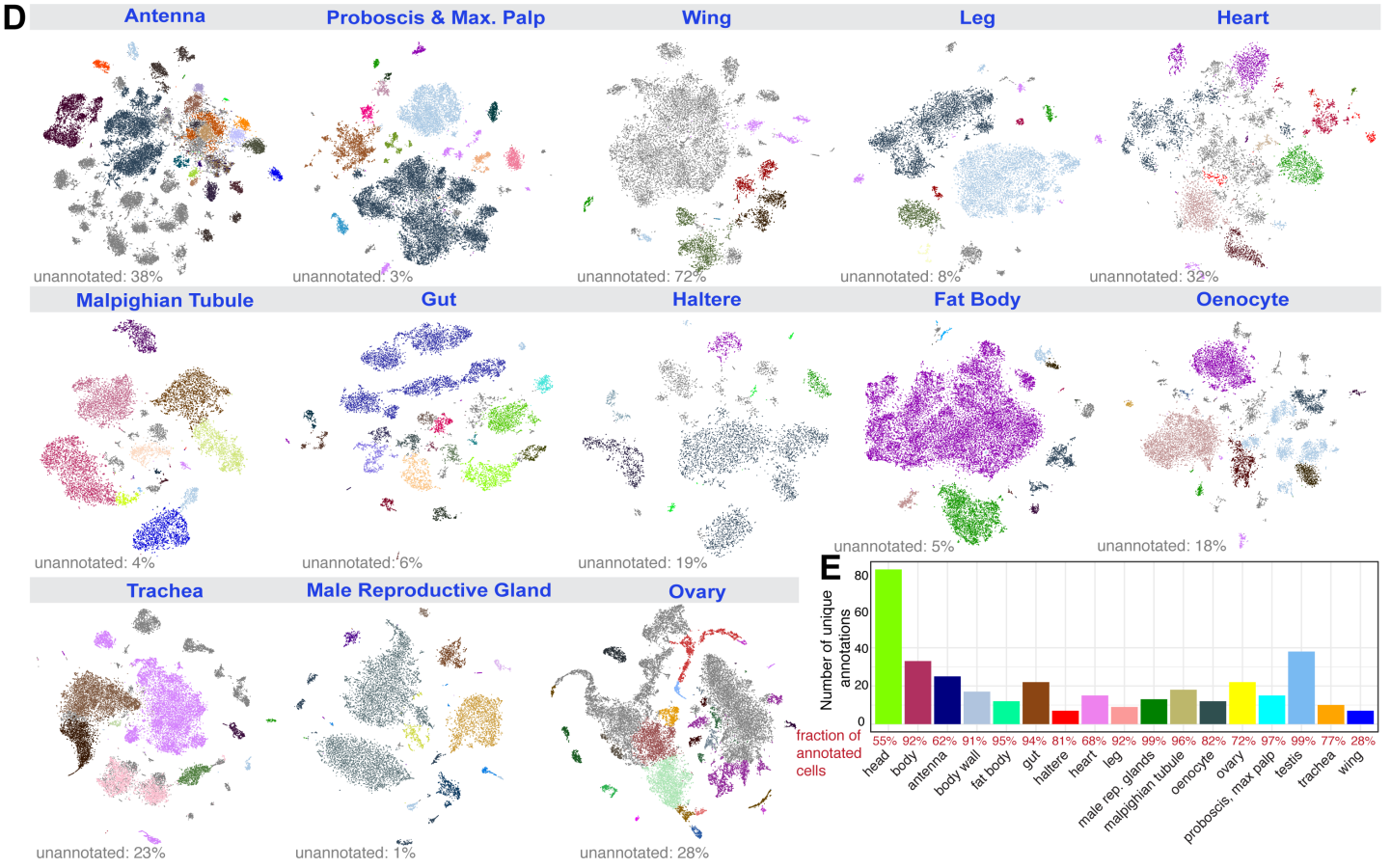

C
10x: relaxed version
stringent version
grh
Mhc
Syt1
D
10x + SS2
10x
SS2
E
gene detection per tissue
Number of genes
12000
10000
8000
6000
antenna
body wall
fat body
gut
haltere
heart
leg
male rep. gland
malpighian tubule
oenocyte
ovary
prob. max. palps
testis
trachea
wing


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

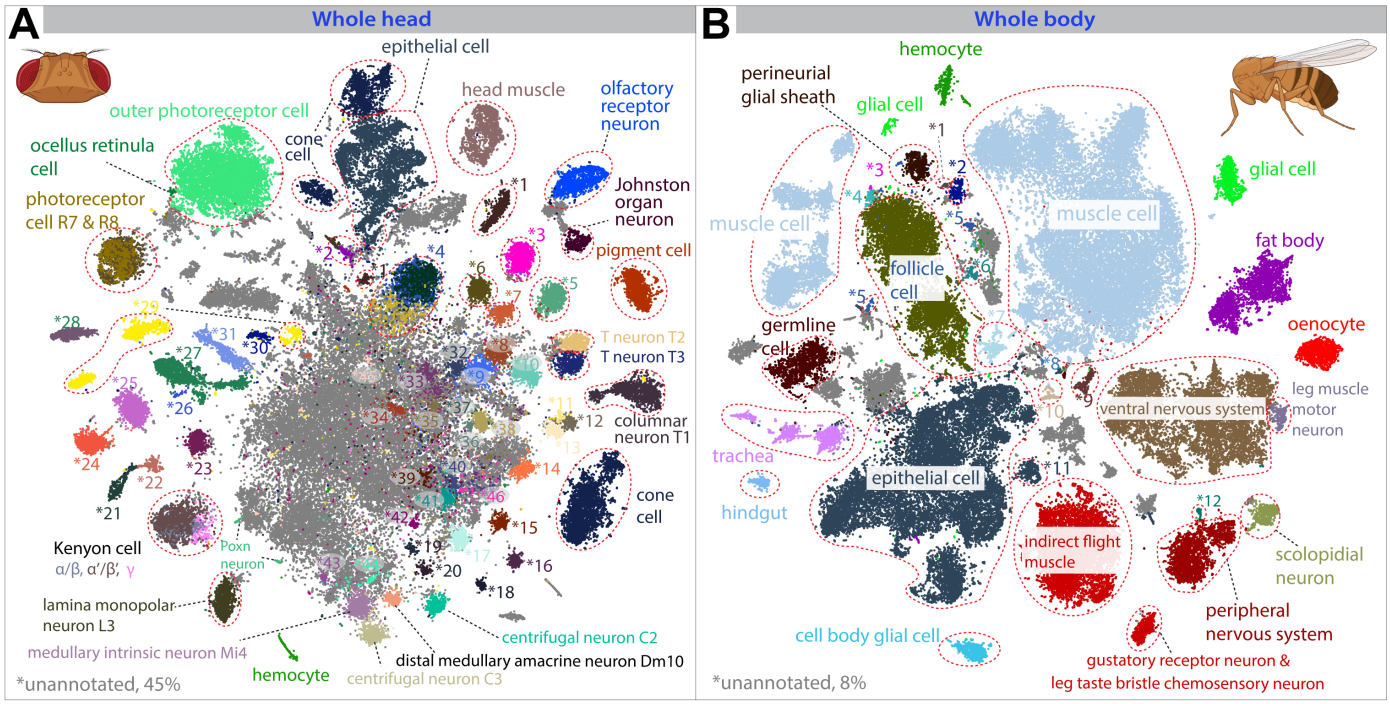
FCA Figure 2

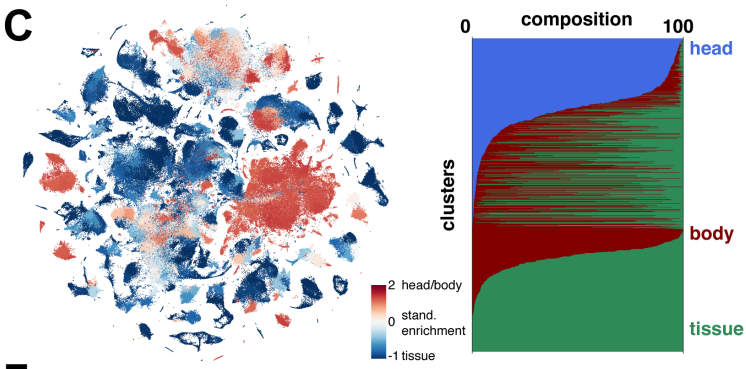

A
Body Wall
Antenna
Proboscis & Max. Palp
Wing
Leg
Heart
Malpighian Tubule
Gut
Haltere
Fat Body
Oenocyte
Trachea
Male Reproductive Gland
Testis + Seminal Vesicle
Ovary
all nuclei (M + F)
GAL4>GFP+ nuclei (M + F)
sex specific tissue, all nuclei

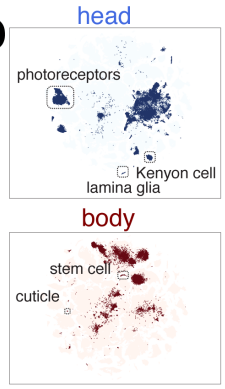

B
Body Wall
indirect flight muscle
peripheral nervous system
fat body
testis
midgut
hemocyte
ovary cell
salivary gland
tendon cell
glial cell
trachea
muscle cell
epidermal cell #1*
epithelial cell
epidermal cell #2*
ventral nervous system
oenocyte
*unannotated: 9%
C
Testis + Seminal Vesicle (relaxed version, UMAP)
seminal vesicle & testis epithelia
germinal proliferation center hub
muscle cell
cyst stem cell
testis epithelial cell
hemocyte
fat body
neuron
early CC 1
early CC 2
pigment cell
late primary sp.
sp. 7
Spermatid associated CCs
sp. 6
sp. 5
fat body
sp. 4
testis epithelial cell
trachea
sp. 3
sp. 2
sp. 0
sp. 1
*unannotated: 1%
*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

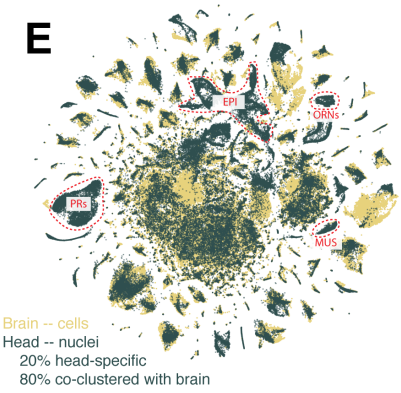

D
Antenna
unannotated: 38%
Proboscis & Max. Palp
unannotated: 3%
Wing
unannotated: 72%
Leg
unannotated: 8%
Heart
unannotated: 32%
Malpighian Tubule
unannotated: 4%
Gut
unannotated: 6%
Haltere
unannotated: 19%
Fat Body
unannotated: 5%
Oenocyte
unannotated: 18%
Trachea
unannotated: 23%
Male Reproductive Gland
unannotated: 1%
Ovary
unannotated: 28%
E
Number of unique annotations
fraction of annotated cells
55% 92% 62% 91% 95% 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

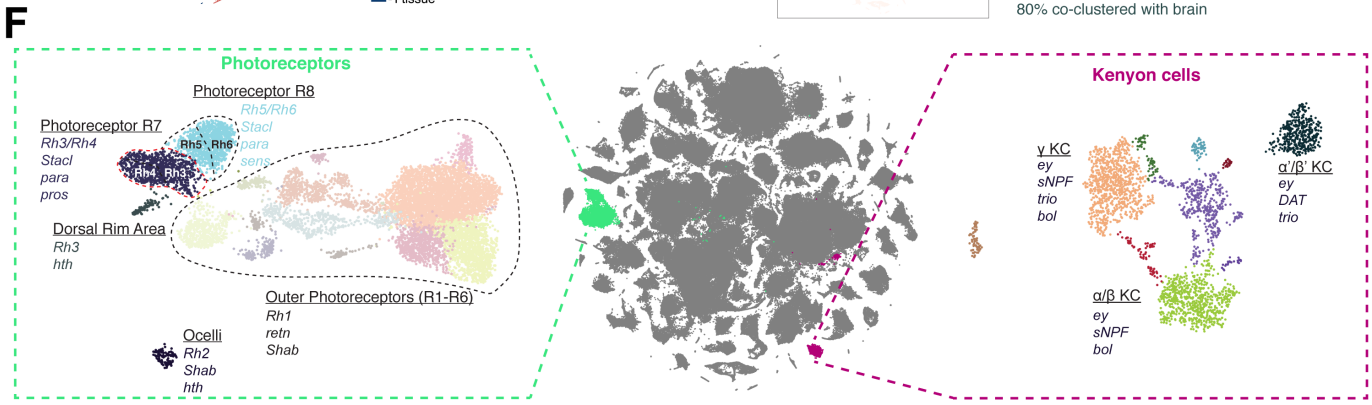
FCA Figure 3

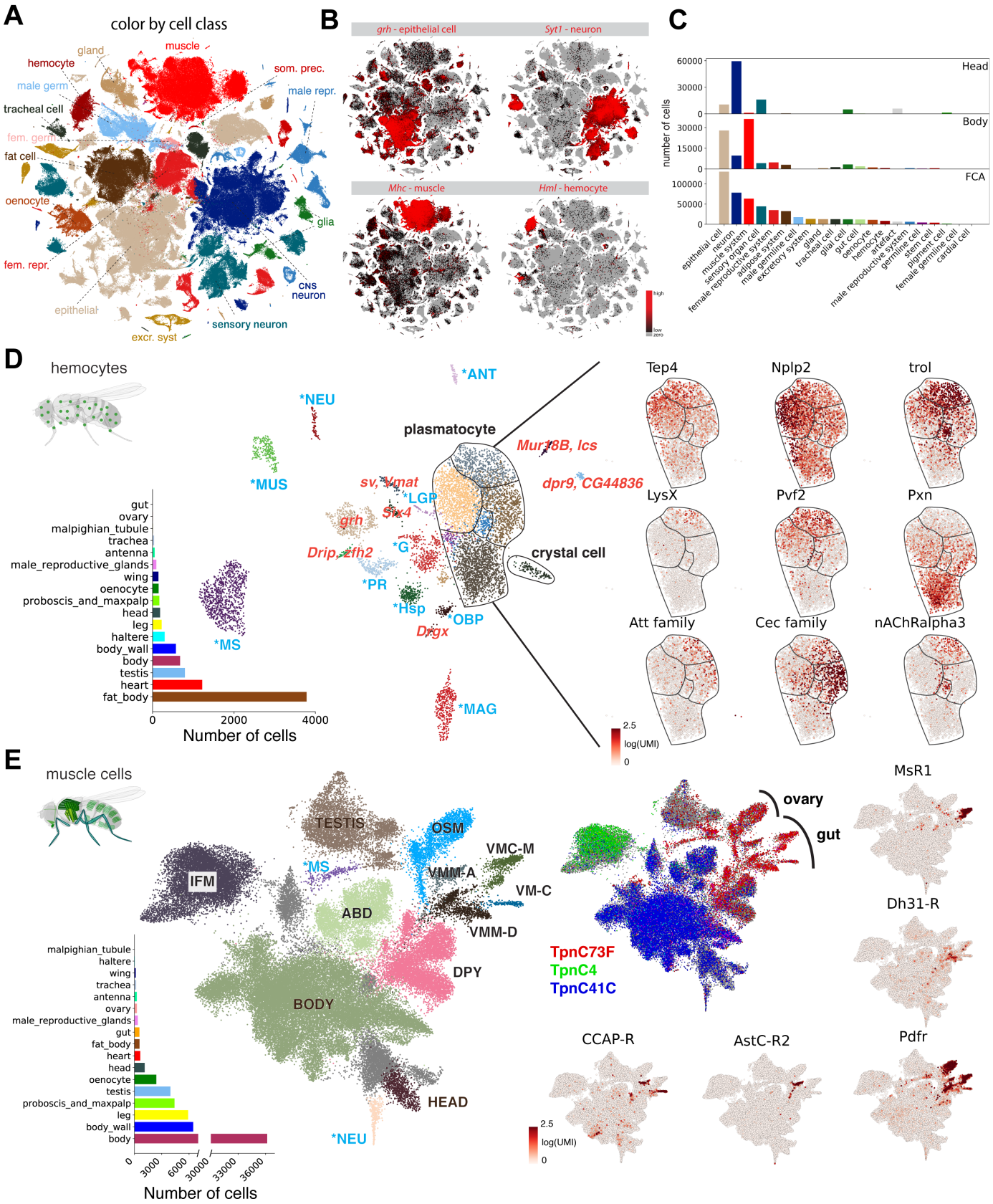

A
Whole head
epithelial cell
head muscle
olfactory receptor neuron
outer photoreceptor cell
cone cell
ocellus retinula cell
photoreceptor cell R7 & R8
Johnston organ neuron
pigment cell
T neuron T2
T neuron T3
columnar neuron T1
cone cell
Kenyon cell
α/β, α′/β′, γ
Poxn neuron
lamina monopolar neuron L3
medullary intrinsic neuron Mi4
centrifugal neuron C2
distal medullary amacrine neuron Dm10
hemocyte
centrifugal neuron C3
*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

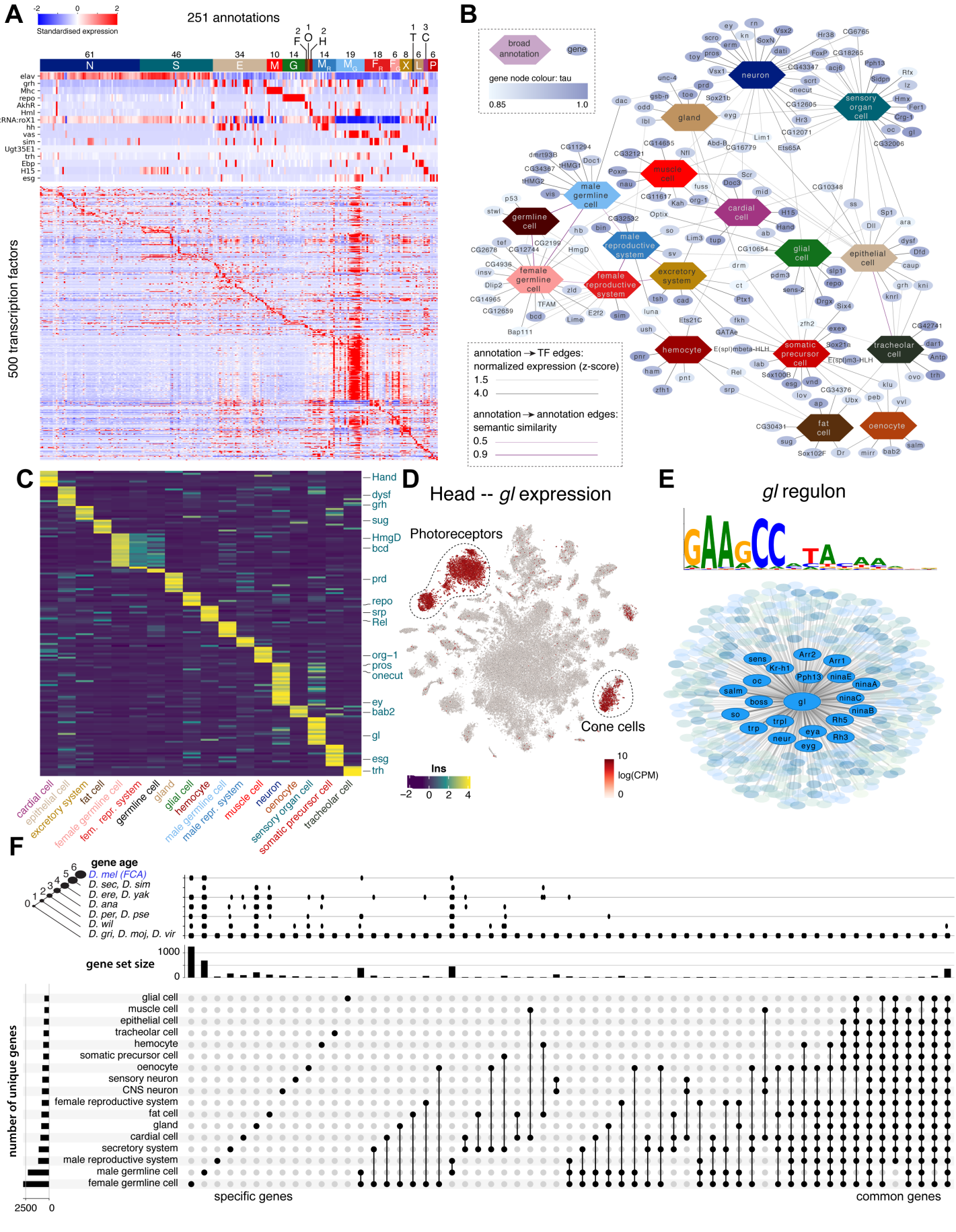

D
head
photoreceptors
Kenyon cell
lamina glia
body
stem cell
cuticle

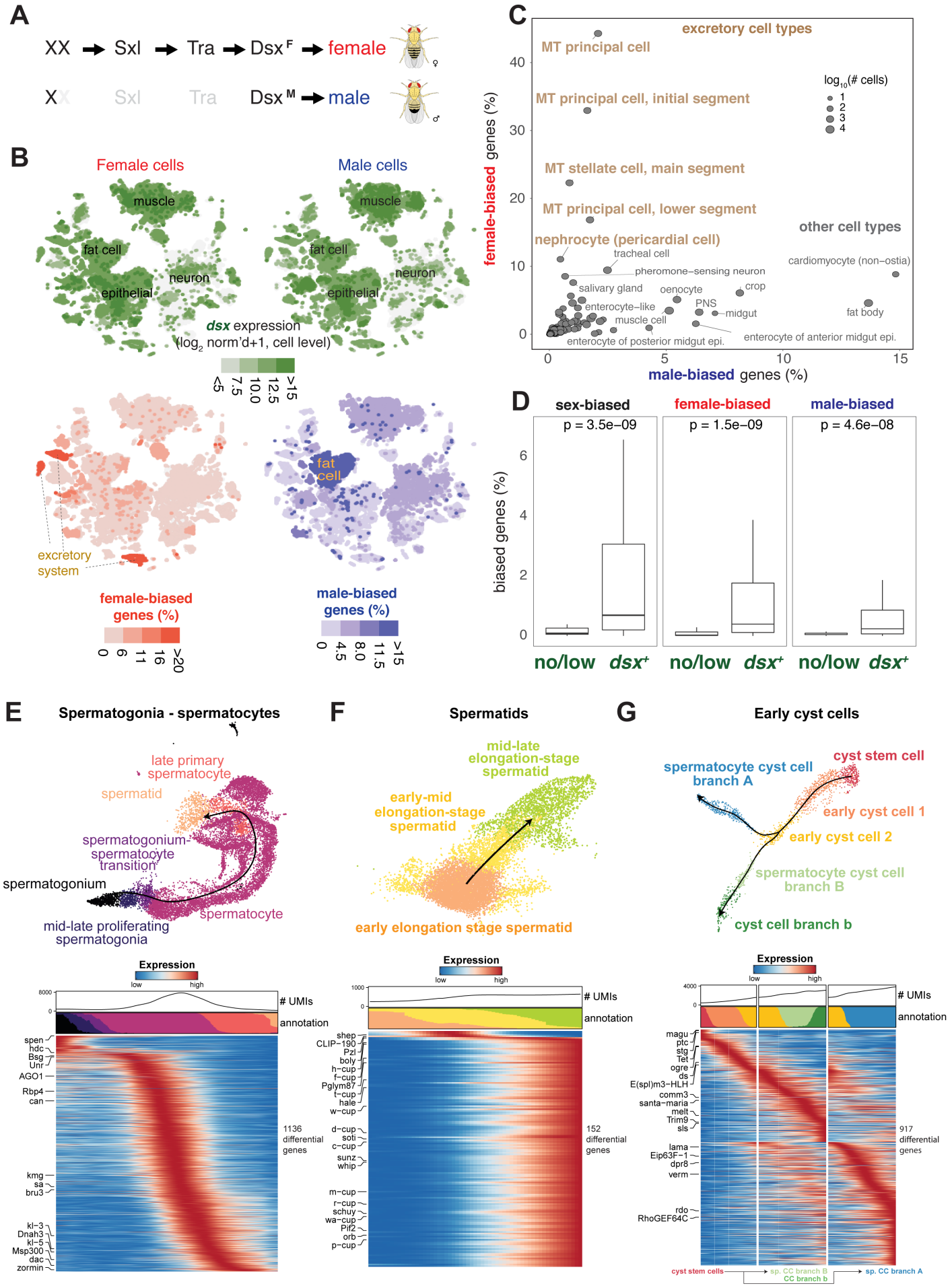

E
EPI
ORNs
PRs
MUS
Brain -- cells
Head -- nuclei
20% head-specific
80% co-clustered with brain


F
Photoreceptors
Photoreceptor R8
Rh5/Rh6
Stacl
para
sens
Photoreceptor R7
Rh3/Rh4
Stacl
para
pros
Rh5
Rh6
Rh4
Rh3
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 class
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
glia cell
gut cell
oenocyte
hemocyte
artefact
male reproductive system
germline cell
stem cell
pigment cell
female germline cell
cardial cell
D
hemocytes
*NEU
*ANT
plasmatocyte
Mur18B, Ics
*MUS
sv, Vmat
*LGP
dpr9, CG44836
Six4
grh
*G
crystal cell
Drip, zfh2
*PR
*Hsp
*OBP
Dvgx
*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
Number of cells
Tep4
Nplp2
trol
LysX
Pvf2
Pxn
Att family
Cec family
nAChRalpha3
2.5
log(UMI)
0
E
muscle cells
TESTIS
OSM
VMC-M
IFM
*MS
VMM-A
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
Number of cells
ovary
gut
TpnC73F
TpnC4
TpnC41C
MsR1
Dh31-R
CCAP-R
AstC-R2
Pdfr
2.5
log(UMI)
0

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
XX 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)
1 2 3 4
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
spen hdc Bsg Unr AGO1 Rbp4 can kmg sa bru3 kl-3 Dnah3 kl-5 Msp300 dac zormin
1136 differential genes
shep CLIP-190 Pzl boly h-cup f-cup Pglym87 t-cup hale w-cup d-cup soti c-cup sunz whip m-cup r-cup schuy wa-cup Pif2 orb p-cup
152 differential genes
magu ptc stg Tet ogre ds E(spl)m3-HLH comm3 santa-maria melt Trim9 sls lama Eip63F-1 dpr8 verm rdo RhoGEF64C
917 differential genes
cyst stem cells
sp. CC branch B
CC branch b
sp. CC branch A

Science
AAAS

Supplementary Materials for

Fly Cell Atlas: a single-nucleus 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[24], Teresa M. Przytycka[15], Aaron M. Allen[16], Stephen F. Goodwin[16], Cameron W. Berry[17], Margaret T. Fuller[17], Helen White-Cooper[18], Erika L. Matunis[19], Stephen DiNardo[20], Anthony Galenza[21], Lucy Erin O'Brien[21], Julian A. T. Dow[22], **FCA Consortium**[25], Heinrich Jasper[23], Brian Oliver[24], Norbert Perrimon[10]†, Bart Deplancke[6]†, Stephen R. Quake[5]†, Liqun Luo[1]†, and Stein Aerts[3,4]†

* 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.)

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Table S1 [excel]: for Fig1, analysis parameters
Table S2 [excel]: for Fig2, cell type annotation
Table S3 [excel]: for Fig5, cell type-specific transcription factors
Table S4 [excel]: for Fig6: sex-differences, manually removed cells
Table S5 [excel]: for Fig6: sex-differences cell level data
Table S6 [excel]: for Fig6: sex-differences cluster level data

FCA Consortium Contributions

Overall coordination

Hongjie Li, Jasper Janssens, Norbert Perrimon, Bart Deplancke, Stephen R. Quake, Liqun Luo, Stein Aerts

Logistical coordination

Hongjie Li, Jasper Janssens, Maxime De Waegeneer, Sai Saroja Kolluru, Robert C. Jones, Norbert Perrimon, Bart Deplancke, Stephen R. Quake, Liqun Luo, Stein Aerts, Aaron McGeever, Angela Oliveira Pisco, Jim Karkanias, Sheela Crasta, Tzu-Chiao Lu, Gil dos Santos, Clare Pilgrim, Alex McLachlan, David Osumi-Sutherland, Irene Papatheodorou, Nancy George, Jonathan Manning, Robert P Zinzen

Tissue dissection

Liqun Luo lab (head, body, antenna, haltere): Liqun Luo, Hongjie Li, Jiefu Li, David Vacek, Anthony Xie
Lucy O'Brien lab (gut): Lucy Erin O'Brien, Yu-Han Su, Erin Nicole Sanders, Samantha Gumbin, Paola Moreno-Roman, Aparna Sherlekar, Andrew Thomas Labott, Sang Ngo
Norbert Perrimon lab (Malpighian tubule): Norbert Perrimon, Ruei-Jiun Hung, Jun Xu
Yuh-Nung Jan lab (body wall): Yuh Nung Jan, Jacob S. Jaszczak, Ruijun Zhu, Ke Li, Yanmeng Guo, Kai Li, Liying Li, Tun Li, Han-Hsuan Liu, Caitlin E. O'Brien, Wanpeng Wang, Maja Petkovic
Rolf Bodmer lab (heart): Rolf Bodmer, Georg Vogler, Marco Tamayo, James Kezos, Katja Birker, Tanja Nielsen
Mariana Wolfner lab (male reproductive glands): Mariana F. Wolfner, Norene A. Buehner
Kristin Scott lab (leg, wing, proboscis & max palp), Kristin Scott, Amy Chang, Stefanie Engert, Amanda J González-Segarra, Meghan Laturney, Sarah Leinwand, Carolina Reisenman, Philip Shiu, Gabriella Sterne, Zepeng Yao
Heinrich Jasper lab (fat body, oenocyte): Heinrich Jasper, Xiaoyu Tracy Cai, Nadja Sandra Katheder
Tom Kornberg lab (trachea): Thomas B Kornberg, Wanpeng Wang
Margaret Fuller lab (testis): Margaret T. Fuller, Neuza Reis Matias, Cameron W. Berry, Susanna E. Brantley, Catherine C. Baker, Devon E. Harris, Yiu-Cheung E. Wong, Benjamin Bolival
Todd Nystul lab (ovary): Todd G. Nystul, Katja Rust
Katja Brueckner lab (hemocyte): Katja Brueckner, Jordan Augsburger, Anjeli Mase
Seung Kim lab (insulin-producing cell, corpora cardiaca cell): Seung K. Kim, Lutz Kockel

Library preparation and sequencing

Liqun Luo: Hongjie Li, Colleen N. McLaughlin, Qijing Xie,
Stephen Quake lab: Sai Saroja Kolluru, Robert C. Jones, Felix Horns
Biohub: Angela M. Detweiler, Jia Yan, Michelle Tan, Norma Neff, Rene V. Sit
NIH group: Harold E. Smith, Brian Oliver

Main Data analysis

Stein Aerts lab: Jasper Janssens, Maxime De Waegeneer, Kristofer Davie, Swann Floc'hlay, Katina Spanier
Bart Deplancke lab: Vincent Gardeux, Wouter Saelens, Fabrice David, Maria Litovchenko,
Jure Leskovec lab: Maria Brbić, Kumar Ayush
Liqun Luo lab: Hongjie Li

Case study analysis/writing

Common cell - hemocyte: Katja Brueckner, Jiwon Shim, Sudhir Tattikota, Jasper Janssens, Hongjie Li
Common cell - muscle: Frank Schnorrer, Jasper Janssens
Gut data integration: Wouter Saelens
Metabolic pathway: Carlos Ribeiro, Zita Carvalho-Santos, Darshan B. Dhakan, Rita Cardoso-Figueiredo,
Ovary data integration: Katja Rust , Todd G. Nystul
Sex differences: Brian Oliver, Soumitra Pal, Teresa Przytycka, Aaron M. Allen, Devika Agarwal, Stephen F Goodwin, Julian A. T. Dow

Testis annotation & trajectory: Margaret T. Fuller, Cameron W. Berry, Erika L. Matunis, Stephen DiNardo, Helen White-Cooper, Brian Oliver, Sharvani Mahadevaraju, Julie A. Brill, Henry M. Krause, Wouter Saelens, Bart Deplancke

Cell type annotation

SCope team: Stein Aerts, Jasper Janssens, Maxime De Waegeneer, Kristofer Davie

ASAP team: Bart Deplancke, Vincent Gardeux, Wouter Saelens, Fabrice David

Gut: Lucy Erin O'Brien, Anthony Galenza, Aparna Sherlekar, Erin Nicole Sanders, Yu-Han Su, Anna A. Kim, Kazuki Yoda, Norbert Perrimon, Joshua Shing Shun Li

Malpighian tubule: Norbert Perrimon, Julian A. T. Dow, Jun Xu, Jasper Janssens, Yifang Liu

Body wall: Yuh Nung Jan, Jacob S. Jaszczak, Ruijun Zhu, Ke Li, Yanmeng Guo, Liying Li, Hongjie Li

Heart: Rolf Bodmer, Georg Vogler, Hongjie Li, Jasper Janssens

Muscle: Frank Schnorrer, Jasper Janssens

Male reproductive glands: Mariana F. Wolfner, Nora C. Brown, Yasir Ahmed-Braimah, Helen White-Cooper, Mikaela Matera-Vatnick, Timothy L. Karr

Leg, wing, prob. and max palp: Kristin Scott, Zepeng Yao, Carlos Ribeiro, Ibrahim Tastekin

Antenna and haltere, Liqun Luo, Hongjie Li, Colleen N. McLaughlin, Andrew K. Groves, Shinya Yamamoto, Daniel Sutton, Rachel I. Wilson, Stephen L. Holtz

Fat body and oenocyte: Heinrich Jasper, Xiaoyu Tracy Cai, Nadja Sandra Katheder, Sudhir Gopal Tattikota, Carlos Ribeiro, Zita Carvalho-Santos, Rita Cardoso-Figueiredo, Hongjie Li, Jasper Janssens

Trachea: Thomas B Kornberg, Wanpeng Wang, Hongjie Li

Testis (see case study)

Ovary: Todd G. Nystul, Katja Rust, Ruth Lehman, Maija Slaidina, Torsten Banisch, Zita Carvalho-Santos, Mariana F. Wolfner, Wanpeng Wang, Brian Oliver, Sharvani Mahadevaraju

Hemocyte, Katja Brueckner, Jiwon Shim, Sudhir Gopal Tattikota, Jasper Janssens, Hongjie Li

Insulin-producing cell (IPC) and corpora cardiaca cell (CC): Seung K. Kim, Lutz Kockel, Maria Brbić

Head and body: Aaron M. Allen, Bruno Hudry, Caitlin E. McDonough-Goldstein, Christoph Treiber, Clare Pilgrim, Claude Desplan, David Sims, Devika Agarwal, Erika Donà, Fernando Casares, Gregory S.X.E. Jefferis, Majd M. Ariss, Megan Neville, Michelle Arbeitman, Mehmet Neset Özel, Nikolaos Konstantinides, Scott Waddell, Stephen F Goodwin, Thomas R. Clandinin, Jasper Janssens, Hongjie Li, Stein Aerts

Writing group

Hongjie Li, Jasper Janssens, Norbert Perrimon, Bart Deplancke, Stephen R. Quake, Liqun Luo, Stein Aerts

Principal investigators (A-Z):

Stein Aerts, Yasir Ahmed-Braimah, Rolf Bodmer, Julie A. Brill, Katja Brueckner, Fernando Casares, Thomas R. Clandinin, Bart Deplancke, Claude Desplan, Stephen DiNardo, Julian A. T. Dow, Margaret T. Fuller, Stephen F Goodwin, Andrew K. Groves, Bruno Hudry, Yuh Nung Jan, Heinrich Jasper, Gregory S.X.E. Jefferis, Timothy L. Karr, Seung K. Kim, Nikolaos Konstantinides, Thomas B Kornberg, Henry M. Krause, Jure Leskovec, Hongjie Li, Liqun Luo, Erika L. Matunis, Todd G. Nystul, Lucy Erin O'Brien, Brian Oliver, Norbert Perrimon, Teresa M Przytycka, Stephen R. Quake, Carlos Ribeiro, Katja Rust, Frank Schnorrer, Kristin Scott, Jiwon Shim, Scott Waddell, Helen White-Cooper, Rachel I. Wilson, Mariana F. Wolfner, Shinya Yamamoto, Robert P Zinzen

FCA Consortium Author Affiliations

Names (158; last name A-Z)	email	Affiliation
Stein Aerts	stein.aerts@kuleuven.be	VIB-KU Leuven Center for Brain & Disease Research, KU Leuven, Leuven 3000, Belgium; Laboratory of Computational Biology, Department of Human Genetics, KU Leuven, Leuven 3000, Belgium
Devika Agarwal	devika.agarwal@imm.ox.ac.uk	MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK
Yasir Ahmed-Braimah	yahmed@syr.edu	Department of Biology, Syracuse University, Syracuse NY 13244 USA
Aaron M. Allen	aaron.allen@cncb.ox.ac.uk	Centre for Neural Circuits & Behaviour, University of Oxford, Tinsley Building, Mansfield road, Oxford, OX1 3SR, UK
Michelle Arbeitman	michelle.arbeitman@med.fsu.edu	Biomedical Sciences Department, Florida State University, Tallahassee, Florida, USA
Majd M. Ariss	majdariss@gmail.com	Department of Genetics, Harvard Medical School, Boston, MA, USA
Jordan Augsburger	Jordan.Augsburger@ucsf.edu	Department of Cell and Tissue Biology, University of California, San Francisco, CA 94143, USA
Kumar Ayush	kayush@stanford.edu	Department of Computer Science, Stanford University, Stanford, CA 94305, USA
Catherine C. Baker	ccb@stanford.edu	Department of Developmental Biology, Stanford University School of Medicine, Stanford CA 94305 USA
Torsten Banisch	Torsten.Banisch@nyulangone.org	Skirball Institute and HHMI, New York University Langone Medical Center, New York City, NY 10016, USA
Cameron W. Berry	cwberry@stanford.edu	Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
Katja Birker	kbirker@sbpdiscovery.org	Development, Aging and Regeneration Program, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037 USA
Rolf Bodmer	rolf@SBPdiscovery.org	Development, Aging and Regeneration Program, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037 USA
Benjamin Bolival	bbolival@stanford.edu	Department of Developmental Biology, Stanford University School of Medicine, Stanford CA 94305 USA
Susanna E. Brantley	susanna.brantley@duke.edu	Department of Developmental Biology, Stanford University School of Medicine, Stanford CA 94305 USA
Maria Brbic	mbrbic@cs.stanford.edu	Department of Computer Science, Stanford University, Stanford, CA 94305, USA
Julie A. Brill	julie.brill@sickkids.ca	Cell Biology Program, The Hospital for Sick Children (SickKids), Toronto, ON, M5G 0A4, Canada; Department of Molecular Genetics, University of Toronto, Toronto, ON, M5S 1A8, Canada
Nora C. Brown	nb579@cornell.edu	Department of Molecular Biology and Genetics, Cornell University, Ithaca NY 14853, USA
Katja Brueckner (deceased)	NA	Department of Cell and Tissue Biology, University of California, San Francisco, CA 94143, USA

Norene A. Buehner	nab28@cornell.edu	Department of Molecular Biology and Genetics, Cornell University, Ithaca NY 14853, USA
Xiaoyu Tracy Cai	caix15@gene.com	Immunology Discovery, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA
Rita Cardoso-Figueiredo	rita.figueiredo@neuro.fchampalimaud.org	Behavior and Metabolism Laboratory, Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal
Zita Carvalho-Santos	zita.santos@neuro.fchampalimaud.org	Behavior and Metabolism Laboratory, Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal
Fernando Casares	fcasfer@upo.es	CABD (Andalusian Centre for Developmental Biology), CSIC-UPO-JA. Seville 41013, SPAIN
Amy Chang	amykhinechang@berkeley.edu	Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
Thomas R. Clandinin	trc@stanford.edu	Department of Neurobiology, Stanford University, Stanford, CA 94305, USA
Sheela Crasta	sheela.crasta@czbiohub.org	Departments of Bioengineering and Applied Physics, Stanford University, Stanford CA USA, and Chan Zuckerberg Biohub, San Francisco CA USA
Fabrice PA David	fabrice.david@epfl.ch	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
Kristofer Davie	kristofer.davie@kuleuven.vib.be	VIB-KU Leuven Center for Brain & Disease Research, KU Leuven, Leuven 3000, Belgium
Bart Deplancke	bart.deplancke@epfl.ch	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
Claude Desplan	cd38@nyu.edu	Department of Biology, New York University, New York, New York 10003, USA
Angela M. Detweiler	angela.detweiler@czbiohub.org	Chan Zuckerberg Biohub, San Francisco CA USA
Darshan B. Dhakan	darshan.dhakan@research.fchampalimaud.org	Behavior and Metabolism Laboratory, Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal
Stephen DiNardo	sdinardo@pennmedicine.upenn.edu	Perelman School of Medicine, The University of Pennsylvania, and The Penn Institute for Regenerative Medicine Philadelphia, PA 19104 USA
Erika Donà	edona@mrc-lmb.cam.ac.uk	Neurobiology Division, MRC Laboratory of Molecular Biology, Cambridge, CB2 0QH, UK
Julian A. T. Dow	Julian.dow@glasgow.ac.uk	Institute of Molecular, Cell & Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK.
Stefanie Engert	stefanie.engert@berkeley.edu	Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
Swann Floc'hlay	swann.flochlay@kuleuven.be	VIB-KU Leuven Center for Brain & Disease Research, KU Leuven, Leuven 3000, Belgium; Laboratory of Computational Biology, Department of Human Genetics, KU Leuven, Leuven 3000, Belgium
Margaret T. Fuller	mtfuller@stanford.edu	Department of Developmental Biology and Department of Genetics, Stanford University School of Medicine, Stanford CA 94305, USA

Anthony Galenza	agalenza@stanford.edu	Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA
Vincent Gardeux	vincent.gardeux@epfl.ch	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
Nancy George	ngeorge@ebi.ac.uk	European Molecular Biology Laboratory, European Bioinformatics Institute, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK.
Amanda J González-Segarra	amandagonzalez@berkeley.edu	Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
Stephen F Goodwin	stephen.goodwin@dpag.ox.ac.uk	Centre for Neural Circuits & Behaviour, University of Oxford, Tinsley Building, Mansfield road, Oxford, OX1 3SR, UK
Andrew K. Groves	akgroves@bcm.edu	Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA
Samantha Gumbin	sgumbin@stanford.edu	Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA
Yanmeng Guo	yanmeng.guo@ucsf.edu	Department of Physiology, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA; Howard Hughes Medical Institute, San Francisco, California, USA
Devon E. Harris	deharris@stanford.edu	Department of Developmental Biology, Stanford University School of Medicine, Stanford CA 94305 USA
Yael Heifetz	yael.heifetz@mail.huji.ac.il	The Hebrew University of Jerusalem, Jerusalem, Israel
Stephen L. Holtz	holtz@fas.harvard.edu	Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA
Felix Horns	rfhorns@gmail.com	Department of Bioengineering and Biophysics Graduate Program, Stanford University, Stanford, CA 94305 USA
Bruno Hudry	Bruno.HUDRY@univ-cotedazur.fr	Université Côte d'Azur, CNRS, INSERM, iBV, France
Ruei-Jiun Hung	rjhung@hotmail.com	Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts, USA
Yuh Nung Jan	yuhnung.jan@ucsf.edu	Department of Physiology, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA; Howard Hughes Medical Institute, San Francisco, California, USA
Jasper Janssens	jasper.janssens@kuleuven.vib.be	VIB-KU Leuven Center for Brain & Disease Research, KU Leuven, Leuven 3000, Belgium; Laboratory of Computational Biology, Department of Human Genetics, KU Leuven, Leuven 3000, Belgium
Heinrich Jasper	jasper.heinrich@gene.com	Immunology Discovery, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA
Jacob S. Jaszczak	jacob.jaszczak@ucsf.edu	Department of Physiology, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA; Howard Hughes Medical Institute, San Francisco, California, USA
Gregory S.X.E. Jefferis	jefferis@mrc-lmb.cam.ac.uk	Neurobiology Division, MRC Laboratory of Molecular Biology, Cambridge, CB2 0QH, UK

Robert C. Jones	jonesbob@stanford.edu	Departments of Bioengineering and Applied Physics, Stanford University, Stanford CA USA, and Chan Zuckerberg Biohub, San Francisco CA USA
Jim Karkanias	jim.karkanias@czbiohub.org	Chan Zuckerberg Biohub, San Francisco CA USA
Timothy L. Karr	tkarr@asu.edu	Biodesign Institute, Arizona State University, Tempe AZ 85281, USA
Nadja Sandra Katheder	katheden@gene.com	Immunology Discovery, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA
James Kezos	jkezos@sbpdiscovery.org	Development, Aging and Regeneration Program, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037 USA
Anna A. Kim	anna.kim@angstrom.uu.se	Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA; University of California, Santa Barbara, CA 93106, USA; Uppsala University, Sweden
Seung K. Kim	seungkim@stanford.edu	Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305; Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA
Lutz Kockel	lkockel@stanford.edu	Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
Sai Saroja Kolluru	saisarojakolluru@gmail.com	Departments of Bioengineering and Applied Physics, Stanford University, Stanford CA USA, and Chan Zuckerberg Biohub, San Francisco CA USA
Nikolaos Konstantinides	nikos.konstantinides@ijm.fr	Institut Jacques Monod, Centre National de la Recherche Scientifique-UMR 7592, Université Paris Diderot, Paris, France
Thomas B Kornberg	Tom.Kornberg@ucsf.edu	Cardiovascular Research Institute, University of California, San Francisco, CA 94143, USA
Henry M. Krause	h.krause@utoronto.ca	Donnelly Centre for Cellular and Biomolecular Research, Department of Molecular Genetics, University of Toronto, Toronto, ON, M5S 3E1, Canada
Andrew Thomas Labott	alabott@stanford.edu	Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA
Meghan Laturney	meghan.laturney@berkeley.edu	Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
Ruth Lehmann	lehmann@wi.mit.edu	Skirball Institute, Department of Cell Biology, and HHMI, New York University Langone Medical Center, New York City, NY 10016; present affiliation Whitehead Institute and Department of Biology, MIT, Boston USA
Sarah Leinwand	sleinwand@berkeley.edu	Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
Jure Leskovec	jure@cs.stanford.edu	Department of Computer Science, Stanford University, Stanford, CA 94305, USA and Chan Zuckerberg Biohub, San Francisco CA USA
Hongjie Li	hongjie.li@bcm.edu	Department of Biology, Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA; Huffington Center on Aging, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
Jiefu Li	jiefuli@stanford.edu	Department of Biology, Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA

Joshua Shing Shun Li	joshua_li@hms.harvard.edu	Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts, USA
Kai Li	kai.li2@ucsf.edu	Department of Physiology, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA; Howard Hughes Medical Institute, San Francisco, California, USA
Ke Li	ke.li@ucsf.edu	Department of Physiology, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA; Howard Hughes Medical Institute, San Francisco, California, USA
Liying Li	liying.li@ucsf.edu	Department of Physiology, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA; Howard Hughes Medical Institute, San Francisco, California, USA
Tun Li	tun.li@ucsf.edu	Department of Physiology, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA; Howard Hughes Medical Institute, San Francisco, California, USA
Maria Litovchenko	maria.litovchenko@gmail.com	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
Han-Hsuan Liu	han-hsuan.liu@ucsf.edu	Department of Physiology, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA; Howard Hughes Medical Institute, San Francisco, California, USA
Yifang Liu	yifang_liu@hms.harvard.edu	Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts, USA
Tzu-Chiao Lu	Tzu-Chiao.Lu@bcm.edu	Huffington Center on Aging, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
Liqun Luo	lluo@stanford.edu	Department of Biology, Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA
Sharvani Mahadevaraju	sharvani.mahadevaraju@nih.gov	Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Kidney and Digestive Diseases, National Institutes of Health, Bethesda, MD, 20892, USA
Jonathan Manning	jmanning@ebi.ac.uk	European Molecular Biology Laboratory, European Bioinformatics Institute, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK
Anjeli Mase	Anjeli.mase@ucsf.edu	Department of Cell and Tissue Biology, University of California, San Francisco, CA 94143, USA
Mikaela Matera-Vatnick	mm2568@cornell.edu	Department of Molecular Biology and Genetics, Cornell University, Ithaca NY 14853, USA
Neuza Reis Matias	nmatias@stanford.edu	Department of Developmental Biology, Stanford University School of Medicine, Stanford CA 94305 USA
Erika L. Matunis	ematuni1@jhmi.edu	Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, MD 21205 USA
Caitlin E. McDonough-Goldstein	mcdonouce@gmail.com	Department of Biology, Syracuse University, Syracuse, NY, USA; Department of Evolutionary Biology, University of Vienna, Vienna, Austria
Aaron McGeever	aaron.mcgeever@czbiohub.org	Chan Zuckerberg Biohub, San Francisco CA USA

Alex D. McLachlan	adm71@cam.ac.uk	Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK
Colleen N McLaughlin	cnm@stanford.edu	Department of Biology, Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA
Paola Moreno-Roman	paola.moreno.roman@gmail.com	Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA
Norma Neff	norma.neff@czbiohub.org	Chan Zuckerberg Biohub, San Francisco CA USA
Megan Neville	megan.goodwin@dpag.ox.ac.uk	Centre for Neural Circuits & Behaviour, University of Oxford, Tinsley Building, Mansfield road, Oxford, OX1 3SR, UK
Sang Ngo	sangn@alumni.stanford.edu	Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA
Tanja Nielsen	tnielsen@sbpdiscovery.org	Development, Aging and Regeneration Program, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037 USA
Todd G. Nystul	Todd.Nystul@ucsf.edu	Department of Anatomy, University of California, San Francisco, CA 94143, USA
Caitlin E. O'Brien	caitlin.obrien@ucsf.edu	Department of Physiology, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA; Howard Hughes Medical Institute, San Francisco, California, USA
Lucy Erin O'Brien	lucye@stanford.edu	Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA
Brian Oliver	brian.oliver@nih.gov	Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Kidney and Digestive Diseases, National Institutes of Health, Bethesda, MD 20892, USA
David Osumi-Sutherland	davidos@ebi.ac.uk	European Bioinformatics Institute (EMBL/EBI), Wellcome Trust Genome Campus, Cambridge, UK
Mehmet Neset Özel	no24@nyu.edu	Department of Biology, New York University, New York, NY 10003, USA
Soumitra Pal	soumitra.pal@nih.gov	National Center of Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD 20894, USA
Irene Papatheodorou	irenep@ebi.ac.uk	European Molecular Biology Laboratory, European Bioinformatics Institute, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK.
Norbert Perrimon	perrimon@genetics.med.harvard.edu	Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts, USA; Howard Hughes Medical Institute, Boston, Massachusetts, USA
Maja Petkovic	maja.petkovic@ucsf.edu	Department of Physiology, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA; Howard Hughes Medical Institute, San Francisco, California, USA
Clare Pilgrim	cp390@cam.ac.uk	Department of Physiology, Development and Neuroscience, University of Cambridge, CB2 3DY, UK
Angela Oliveira Pisco	angela.pisco@czbiohub.org	Chan Zuckerberg Biohub, San Francisco CA USA
Teresa M Przytycka	przytyck@nih.gov	National Center of Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD 20894, USA

Stephen R. Quake	steve@quake-lab.org	Departments of Bioengineering and Applied Physics, Stanford University, Stanford CA USA, and Chan Zuckerberg Biohub, San Francisco CA USA
Carolina Reisenman	creisenman@berkeley.edu	Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
Carlos Ribeiro	carlos.ribeiro@neuro.fchampalimaud.org	Behavior and Metabolism Laboratory, Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal
Katja Rust	katja.rust@uni-marburg.de	Institute of Physiology and Pathophysiology, Department of Molecular Cell Physiology, Philipps-University, Marburg, Germany.;Department of Anatomy, University of California, San Francisco, CA 94143, USA
Wouter Saelens	w.saelens@epfl.ch	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
Erin Nicole Sanders	ensand12@stanford.edu	Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA
Gilberto dos Santos	dossantos@morgan.harvard.edu	The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
Frank Schnorrer	frank.schnorrer@univ-amu.fr	Aix-Marseille University, CNRS, IBDM (UMR 7288), Turing Centre for Living systems, 13009 Marseille, France
Kristin Scott	kscott@berkeley.edu	Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
Aparna Sherlekar	aparnasb@stanford.edu	Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA
Jiwon Shim	jshim@hanyang.ac.kr	Department of Life Science, College of Natural Science, Hanyang University, Seoul, Republic of Korea 04763
Philip Shiu	philshiu@berkeley.edu	Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
David Sims	david.sims@imm.ox.ac.uk	MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK
Rene V. Sit	renevsit@gmail.com	Chan Zuckerberg Biohub, San Francisco CA USA
Maija Slaidina	maija.slaidina@nyulangone.org	Skirball Institute, Faculty of Medicine, New York University, New York, NY 10016; current affiliation - University of Bergen, Faculty of Medicine, Bergen, Norway
Harold E. Smith	harold.smith@nih.gov	Genomics Core, National Institute of Diabetes and Digestive and Kidney Diseases, US National Institutes of Health, Bethesda, MD, USA
Katina Spanier	katina.spanier@kuleuven.be	VIB-KU Leuven Center for Brain & Disease Research, KU Leuven, Leuven 3000, Belgium; Laboratory of Computational Biology, Department of Human Genetics, KU Leuven, Leuven 3000, Belgium
Gabriella Sterne	sternegr@berkeley.edu	Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
Yu-Han Su	elsasu@stanford.edu	Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA

Daniel Sutton	daniel.sutton@bcm.edu	Graduate Program in Genetics and Genomics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, 77030 USA
Marco Tamayo	mtamayo@sbpdiscovery.org	Development, Aging and Regeneration Program, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037 USA
Michelle Tan	themichelletan@gmail.com	Chan Zuckerberg Biohub, San Francisco CA USA
Ibrahim Tastekin	ibrahim.tastekin@research.fchampalimaud.org	Behavior and Metabolism Laboratory, Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal
Sudhir Gopal Tattikota	sudhir_gt@hms.harvard.edu	Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts, USA
Christoph Treiber	christoph.treiber@cncb.ox.ac.uk	Centre for Neural Circuits & Behaviour, University of Oxford, Tinsley Building, Mansfield Road, Oxford, OX1 3TA, UK
David Vacek	dvacek@stanford.edu	Department of Biology, Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA
Georg Vogler	gvogler@sbpdiscovery.org	Development, Aging and Regeneration Program, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037 USA
Scott Waddell	scott.waddell@cncb.ox.ac.uk	Centre for Neural Circuits & Behaviour, University of Oxford, Tinsley Building, Mansfield Road, Oxford, OX1 3TA, UK
Maxime De Waegeneer	maxime.dewaegeneer@kuleuven.vib.be	VIB-KU Leuven Center for Brain & Disease Research, KU Leuven, Leuven 3000, Belgium; Laboratory of Computational Biology, Department of Human Genetics, KU Leuven, Leuven 3000, Belgium
Wanpeng Wang	wanpeng.wang@ucsf.edu	Cardiovascular Research Institute, University of California, San Francisco, CA 94143, USA
Helen White-Cooper	white-cooperh@cardiff.ac.uk	School of Biosciences, Cardiff University, Cardiff, CF10 3AX UK
Rachel I. Wilson	rachel_wilson@hms.harvard.edu	Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA
Mariana F. Wolfner	mariana.wolfner@cornell.edu	Department of Molecular Biology and Genetics, Cornell University, Ithaca NY 14853, USA
Yiu-Cheung E. Wong	yiuwong@stanford.edu	Department of Developmental Biology, Stanford University School of Medicine, Stanford CA 94305 USA
Anthony Xie	anthonyx@stanford.edu	Department of Biology, Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA
Qijing Xie	cqxie@stanford.edu	Department of Biology, Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA
Jun Xu	Jun_Xu@hms.harvard.edu	Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts, USA
Shinya Yamamoto	yamamoto@bcm.edu	Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX 77030, USA
Jia Yan	rose.yan@czbiohub.org	Chan Zuckerberg Biohub, San Francisco CA USA
Zepeng Yao	zepengyao@berkeley.edu	Department of Molecular and Cell Biology, University of California, Berkeley,

		Berkeley, CA 94720, USA
Kazuki Yoda	yodak@stanford.edu	Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford CA 94305 USA
Ruijun Zhu	ruijun.zhu@ucsf.edu	Department of Physiology, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA; Howard Hughes Medical Institute, San Francisco, California, USA
Robert P Zinzen	Robert.Zinzen@mdc-berlin.de	Laboratory for Systems Biology of Neural Tissue Differentiation, Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrueck Centre for Molecular Medicine (MDC) in the Helmholtz Association, Robert-Roessle-Strasse 12, 13125 Berlin, Germany

Materials and Methods

Fly sample information

				Number of cells after filtering				
tissue	**genotype**	**age**	**Dissection lab**	**10x male**	**10x female**	**10x mix**	**SS2 male**	**SS2 female**
head	w^{1118}	5d	Liqun Luo	33292	59275	4359	-	-
body	w^{1118}	5d	Liqun Luo	47409	49105	4013	-	-
antenna	w^{1118}	5d	Liqun Luo	15586	14446	7222	242	344
haltere	w^{1118}	5d	Liqun Luo	3148	3379	-	128	135
proboscis & maxillary palp	w^{1118}	5d	Kristin Scott	13765	12536	-	333	307
wing	w^{1118}	5d	Kristin Scott	8053	7836	-	207	313
leg	w^{1118}	5d	Kristin Scott	7120	7077	-	302	305
gut	w^{1118}	5d	Lucy O'Brien	5450	6338	-	292	302
body wall	w^{1118}	5d	Yuh-Nung Jan	6851	9700	-	249	176
heart	w^{1118}	5d	Rolf Bodmer	5515	5171	-	169	295
Male reproductive gland	w^{1118}	2-3d*	Mariana Wolfner	13143	-	-	238	-
testis	w^{1118}	1d*	Margaret Fuller	43454	-	-	374	-
ovary	w^{1118}	5d	Todd Nystul	-	31401	-	-	1011
Malpighian tubule	*drip-GAL4 > UAS-nlsGFP* (all nuclei,** see note)	5d	Norbert Perrimon	6185	7589	-	303	191
fat body	*Cg-GAL4 > UAS-lamGFP* (GFP enriched nuclei)	5d	Heinrich Jasper	10983	15943	-	693	656
oenocyte	*PromE800-GAL4 > UAS-unc84GFP* (GFP enriched nuclei)	5d	Heinrich Jasper	9420	4990	-	264	270
trachea	*Btl-GAL4 > UAS-lamGFP* (***GFP enriched nuclei)	5d	Tom Kornberg	7112	19794	-	339	380
insulin-producing cell (IPC)	*dilp2-GAL4 > UAS-unc84GFP* (GFP enriched nuclei)	5d	Seung Kim	-	-	-	345	313
corpora cardiaca cell (CC)	*Akh-GAL4 > UAS-unc84GFP* (GFP enriched nuclei)	5d	Seung Kim	-	-	-	351	241

For most samples, 5-day adults were used with two exceptions (see below). Male and female flies were collected on day 1 and kept together for mating, and on day 5, flies were sexed and dissected. *UAS-lamGFP* was from Bloomington Drosophila Stock Center (BDSC# 7378). *UAS-unc84GFP* was from (*48*).

*For the male reproductive gland, we used 2–3 day old virgin males in order to detect active transcripts which may be not detectable in 5 day old virgin flies. For the testis, we chose 0–1 day old males to be consistent with the phenotypic studies in the field. If virgin males are held away from females until they are 5 days old, the seminal vesicle fills up with huge numbers of mature sperm, and their condensed, inactive haploid nuclei would be a problem if they displaced nuclei from earlier germ cell stages.

**For the Malpighian tubule, we used *Drip-GAL4>UAS-nlsGFP* to label stellate cells and sequenced all nuclei from dissected tissues. During clustering analysis, we found that the stellate cells can be readily annotated based on the marker genes, *tsh, SecCl,* and *Drip*.

***For the trachea, flies cannot survive to adulthood using *btl-GAL4* driving *UAS-UNC84GFP* or *UAS-lamGFP*. In order to label and collect adult trachea cells, we crossed *btl-GAL4*, *tub-Gal80ts* flies with *UAS-lamGFP* at 18℃ and transferred the young adult flies to 29℃ for 5 days before dissection.

How many tissues to use for each sample?

During sample preparation, we estimated the required tissue number based on three factors: total cell/nucleus number in the tissue, targeted cell number for 10x Genomics and Smart-seq2, and the FACS recovery rate. Our FACS recovery rate is about 5%. For example, if a tissue contains 10,000 nuclei, we were able to collect 500 nuclei after FACS. So if we want to collect 10,000 nuclei after FACS, we will dissect 20 tissues. Please note that since we don't have the exact number of cells for many tissues, our dissection labs have prepared more samples than estimated. Here we list the numbers of tissues we prepared. If not specified, the number is used for each sex to target one 10x run, which is about 10,000 cells:

Head, 100 heads for a total 6 10x runs; body, 50 bodies for a total 6 10x runs; antenna, 300; haltere, 500; proboscis with maxillary palp, 100; wing, 200; leg, 120; gut, 200; body wall, 40; heart, 250; male reproductive gland, 200; testis, 600; ovary, 30; malpighian tubule, 300; fat body, 250; oenocyte, 250; trachea, 100; insulin-producing cell (IPC), 250 for one 384-well plate; corpora cardiaca cell (CC), 250 for one 384-well plate.

Single-nucleus RNA-seq

Fly dissection and single-nucleus suspension: Fly tissues were dissected by different dissection labs, flash-frozen using liquid nitrogen, stored at –80℃ , and shipped to Stanford University for processing in the Luo and Quake labs. When using nuclear GFP to label tissues, we compared *UAS-nlsGFP*, *UAS-UNC84GFP* and *UAS-lamGFP*, and found that there was no GFP signal from *UAS-nlsGFP* after nucleus isolation, while the other two gave good fluorescent signal of the nuclear labeling. Single-nucleus suspensions were prepared as detailed below, largely adapted from our recently published protocol (*11*).

1. Prepare w^{1118} flies or GAL4 driving UAS-nuclear-GFP flies.
2. Dissect tissue in cold Schneider's medium, and use P20 pipette (coat the tip with fly body fat) or forceps to transfer them into 100 µl Schneider's medium in a nuclease-free 1.5ml EP tube on ice. Label the tube clearly using permanent marker. Note: for tissues that float in the medium (e.g., adult antennae), before dissection, prepare three clean dishes: 1st with 100% ethanol, 2nd and 3rd with Schneider's medium. Rinse the fly in the 1st dish with 100% ethanol for 5 seconds, then rinse the fly in the 2nd dish, and dissect in the 3rd dish.

3. After dissection, spin down samples in 100 µl Schneider's medium using a bench top spinner.
4. Fresh: The sample can be processed for extraction of nuclei immediately following dissection. Frozen: Alternatively, the sample can be flash-frozen for long-term storage. Seal the 1.5 ml EP tube with parafilm and put into liquid nitrogen for >30s. Immediately store the sample at -80ºC freezer for long-term storage (several months).
5. Prepare fresh homogenization buffer (see details below) and keep on ice.
6. Thaw samples from -80ºC on ice if using frozen samples. Spin down samples in 100 µl Schneider's medium using the bench top spinner, discard medium as much as possible, and add 100 µl Homogenization butter.
7. Optional: if sample pieces are too big, e.g. whole body or whole head, use a pestle motor (Kimble 6HAZ6) to grind the sample for 30s–60s on ice.
8. Add 900 µl homogenization buffer, and transfer 1000 µl homogenized sample into the 1ml dounce (Wheaton 357538). Dounce sets should be autoclaved at 200ºC >5h or overnight.
9. Release nuclei by 20 strokes of loose dounce pestle and 40 of tight dounce pestle. Keep on ice. Avoid bubbles.
10. Filter 1000 µl sample through 5 ml cell strainer (35 µm), and then filter sample using 40 µm Flowmi (BelArt, H13680-0040) into 1.5 ml EP tube.
11. Centrifuge for 10 min at 1000g at 4ºC. Discard the supernatant. Do not disturb the pellet.
12. Re-suspend the nuclei using the desired amount (we normally use 500-1000 µl) of 1xPBS/0.5%BSA with RNase inhibitor (9.5 ml 1x PBS, 0.5 ml 10% BSA, 50 µl RNasin Plus). Pipet more than 20 times to completely re-suspend the nuclei. Filter sample using 40 µm Flowmi into a new 5 ml FACS tube and keep the tube on ice. Now the single-nucleus suspensions are ready for FACS.

According to our experience, the nuclei are stickier than whole cells. For users making single-nucleus suspension for the first time, we suggest taking 10 µl of the single-nucleus suspension, stain with Hoechst (Invitrogen 33342), and check on a cell counter slide to confirm if they are mostly individual nuclei. If nuclei are not sufficiently dissociated, adjust above steps (e.g., increase the number of strokes of the tight pestle when releasing nuclei).

	Amount	Storage	Item (add in this order)	Final concentration
1	10 ml	RT	H2O (nuclease free)	
2	0.856 g	RT	Sucrose (nuclease free)	250 mM
3	100 µl	4ºC	1M Tris PH 8.0	10 mM
4	250 µl	4ºC	1M KCl	25 mM
5	50 µl	4ºC	1M MgCl2	5 mM
6	100 µl	4ºC	10% Triton-x 100	0.1%
7	50 µl	-20ºC	RNasin Plus (Promega, N2615)	0.5%
8	200 µl	-20ºC aliquots	50x protease inhibitor (Promega, G6521)	1x
9	50 µl	-20ºC	20mM DTT	0.1 mM

FACS: We used the SONY SH800 FACS sorter for collecting nuclei. Nuclei were stained by Hoechst-33342 (1:1000; >5min). For wildtype tissues, Hoechst+ nuclei were collected; for collecting GFP+ nuclei, we first gated on Hoechst+ events and then chose the GFP+ population.

Since polyploidy is common for many fly tissues, we observed different populations of nuclei according to DNA content (Hoechst signal). Conversely, many haploid nuclei are present in testis. After FACS of nuclei from the testis plus seminal vesicle by level of Hoeschst and nuclear size, we observed 6 different populations with different Hoechst signal intensity, and confirmed that the Hoechst signal intensity correlated well with nuclear size. For the fly gut, we observed 5 different populations with different Hoechst signals. In order to include all cell populations with different nuclear sizes, we have included all nuclear populations from the FACS in samples for 10x sequencing, except for testis.

In the testis, 64 haploid spermatids are eventually produced for each germ line stem cell division. To avoid overrepresentation of small haploid spermatids in the testis sample, we used the following strategy to collect testis nuclei. For the first of the three total 10x runs for testis plus seminal vesicle, we collected nuclei from all 6 of the populations discerned by FACS for Hoechst x size. For the two subsequent 10x runs, we collected nuclei without the population of the smallest size (haploid spermatids)

Individual nuclei were collected either to 384-well plates for smart-seq2 or to one tube for 10x Genomics. For 10x Genomics, nuclei were collected into a 15ml tube with 500ul 1x PBS with 0.5% BSA as the receiving buffer (RNase inhibitor added). For each 10x Genomics run, 100k–400k nuclei were collected. Nuclei were spinned for 10min at 1000g at 4℃, and then resuspended using 40ul or desired amount of 1x PBS with 0.5% BSA (RNase inhibitor added). 2ul nucleus suspension was used for counting the nuclei with hemocytometers to calculate the concentration. When loading to the 10x controller, we always target at 10k nuclei for each channel. We observed that loading 1.5 folds more nuclei as recommended by the protocol allowed us to recover about 10k cells after sequencing. For example, if the concentration is 1500 nuclei per ul for one sample, we treat it as 1000 nuclei per ul when loading to the 10x controller.

Library preparation and sequencing: Smart-seq2 sequencing libraries were prepared following the protocol we previously described (*11*). Sequencing was performed using the Novaseq 6000 Sequencing system (Illumina) with 100 paired-end reads and 2x12 bp index reads.

10x Genomics sequencing libraries were prepared following the standard protocol from 10x Genomics 3' v3.1 kit with following settings. All PCR reactions were performed using the Biorad C1000 Touch Thermal cycler with 96-Deep Well Reaction Module. 13 cycles were used for cDNA amplification and 16 cycles were used for sample index PCR. As per 10x protocol, 1:10 dilutions of amplified cDNA and final libraries were evaluated on a bioanalyzer. Each library was diluted to 4 nM, and equal volumes of 18 libraries were pooled for each NovaSeq S4 sequencing run. Pools were sequenced using 100 cycle run kits and the Single Index configuration. Read 1, Index 1 (i7), and Read 2 are 28 bp, 8 bp and 91 bp respectively. A PhiX control library was spiked in at 0.2 to 1% concentration. Libraries were sequenced on the NovaSeq 6000 Sequencing System (Illumina).

Sequencing read alignment

Prior to read alignment, the raw FASTQ files were processed with the index-hopping-filter software from 10x Genomics (version 1.0.1) to remove index-hopped reads. More information about this software is available at https://support.10xgenomics.com/docs/index-hopping-filter.

A Cell Ranger (version 3.1.0) index was built from a pre-mRNA GTF which was derived from the Flybase version r6.31 GTF. A complete recipe on how to build this custom pre-mRNA GTF is available here: https://github.com/FlyCellAtlas/genome_references/tree/master/flybase/r6.31

For 10x, the filtered reads (index-hopped reads removed) were processed all the way up to the gene expression matrix using Cell Ranger (version 3.1.0). For Smart-seq2, reads were aligned to the *Drosophila* melanogaster genome (r6.31), the same as for 10x read alignment, using STAR (2.5.4). Gene counts were produced using HTseq (0.11.2) with default settings except '-m intersection-strict'. Gene counts were generated using the same GTF file as for 10x, covering both exonic and intronic regions. Low-quality nuclei having fewer than 10,000 uniquely mapped reads were removed.

Cell filtering and clustering: *Relaxed* version

To verify the accuracy and robustness of the data processing steps, we examined marker gene expression and compared multiple methods and parameter settings for the various preprocessing steps, including index hopping filtering, genome annotation and counting intronic reads, doublet detection (scrublet), read decontamination (DecontX), batch effect removal (harmony), dimensionality reduction (automated PC determination), and clustering (Leiden). At each step QC plots are generated (fig. S4) and the final loom files can be visualised in our SCope and ASAP analysis and visualization platforms, or be downloaded for custom analysis.

Two versions of the processed data were generated: a *Relaxed* and a *Stringent* version. Here we focus on describing the *Relaxed* version of the data. To know more about the difference between the two versions please read the next section.

The Scrublet software was chosen for doublet removal. Doublet scores were calculated from the raw expression matrix generated by Cell Ranger and using Scrublet (version 0.2.1, Docker image: vibsinglecellnf/scrublet:0.1.4). The strategy taken here to remove the doublets from each sample relies on the multiplet rate one can expect from running a Single Cell 3' 10x Genomics experiment. This number depends on the number of recovered cells. In order to estimate this rate as a function of the number of recovered cells, a linear regression was performed on the multiplet rate table (see Chromium Single Cell 3' Reagent Kits v2 User Guide • Rev F) in order to determine the slope and the bias terms. Those numbers were 0.008 and 0.0527 respectively. Given this model, for each sample the top N cells, ranked by the doublet score, were considered as doublet hence removed.

The data was further processed using the Python package Scanpy (version 1.4.4.post1, through Docker image: vibsinglecellnf/scanpy:0.5.0).

Two additional filters, cellwise and genewise, were applied. The cell filter is based on hard thresholds applied on some of the quality metrics (QC). All cells expressing less than 200 genes were filtered out. Moreover, cells exceeding a 15% mitochondrial content were removed. Regarding the gene filter, all genes not expressed in at least 3 cells were filtered out.

For the different analysis runs with VSN-Pipelines, the samples were concatenated using the anndata.concatenate (join=outer). Consequently, the combined matrix was normalized (scanpy.pp.normalize_per_cell, with counts_per_cell_after=10000) and log transformed (scanpy.pp.log1p). Highly variable genes were selected using sc.pp.highly_variable_genes (min_mean=0.0125, max_mean=3, min_disp=0.5). The data was further scaled so that each gene had unit variance and values exceeding a standard deviance of 10 were clipped. In order to determine the number

of principal components (PC) to select, a cross-validation approach was performed using the pcacv module (available in the VSN-Pipelines). The scaled matrix was projected to a principal component analysis (PCA) space using the scanpy.tl.pca function (svd_solver=arpack). Batch correction was applied using the Harmony software (Docker image: vibsinglecellnf/pcacv:0.2.0) with default parameters and using sample as batch variable. The neighborhood graph was calculated from the corrected PCA space with scanpy.pp.neighbors and default parameters except for the number of PCs (see aforementioned). For visualization purposes, two non-linear dimensionality reduction methods were used: t-SNE and UMAP. The t-SNE embeddings were generated using scanpy.tl.tsne while the UMAP embeddings with sc.tl.umap. Default parameters were used for both methods except for the number of PCs (see aforementioned). Clustering was performed using the Leiden algorithm via scanpy.tl.leiden (default parameters) except for the resolution parameter where a range of values were selected. A default clustering was selected using a custom method available in the directs module from VSN-Pipelines (vibsinglecellnf/directs:0.1.0). The default clustering is selected as follows: for a range of min_cluster_size and min_samples, a density-based clustering is performed using HDBSCAN on the t-SNE embedding; an adjusted rand score is computed between this clustering and the previously generated Leiden clusterings; a clustering is assigned to each pair of parameters which maximizes the score; the final selected clustering is the one that maximizes the most the score over all pairs. Cluster markers for each of the generated clusterings were computed from the log normalized expression matrix by means of scanpy.tl.rank_genes_groups (method=wilcoxon, n_genes=0).

To ensure reproducibility of the 10x Genomics data processing, all the analyses from raw counts to final processed files (.loom and .h5ad) were performed using the VSN-Pipelines (https://githhub.com/vib-singlecell-nf/vsn-pipelines).

Cell filtering, decontamination and clustering: *Stringent* version

In the previous section, we described how the *Relaxed* version of the data was generated. The main reason to generate a *Stringent* version was that we identified a significant number of cells, we called "black hole" cells, which are expressing multiple general cell type markers e.g.: grh (epithelial cell), Mhc (muscle cell), onecut (neuron). These cells likely originated from droplets that were contaminated by ambient RNA.

DecontX was chosen in order to correct for this bias. Practically, the raw counts generated from the Cell Ranger pipeline were corrected using this algorithm, available in the celda R package (version 1.4.5, Docker image: vibsinglecellnf/celda:1.4.5). The corrected counts were then rounded using the R base::round function and newly generated empty cells were removed.

Additionally, we applied a more stringent filter on the cells. This cell filter is based on the median absolute deviation (MAD) from the median of the following quality control (QC) metrics across all cells: n_counts i.e.: number of counts per cell, n_genes i.e. number of genes per cell. A value is considered an outlier if it is greater than 3 MADs away (both directions) from the median of these two metrics. This filter strategy is applied in the log space of these QC metrics. Moreover, minima of 200 genes and 500 counts per cell are required for a cell to be considered in the downstream analysis. All cells exceeding a 5% mitochondrial content were removed.

Finally, after the highly variable gene selection, the number of UMIs per cell and the percentage of mitochondrial genes were regressed out using scanpy.pp.regress_out. All other steps described previously remained the same.

Relaxed versus Stringent dataset

As a more stringent filter was applied to generate the Stringent data, they were considered to be higher quality. For most analyses, we focused on the *Stringent* dataset, which should be used as a default for new users.

However, we noticed that in the testis data, the important hub cell cluster was filtered out by the stringent algorithms, likely due to the expression of many "somatic" transcripts including *Upd1* in late spermatocytes. Thus, the UMAP for testis presented in Figure 2C is plotted using Relaxed dataset.

SCope features and applications

SCope platform crowd annotation system was used to gather all annotations that were added during the Jamborees by all tissue experts. All tissue analysis results across all clustering resolutions were used as a basement to annotate the cells of the atlas. The system allows for tracking the author of the annotations as well as their confidence through a like/dislike feature. The latter feature was crucial for building a consensus annotation. Annotated Loom files can be directly downloaded from SCope.

ASAP features and applications

The ASAP platform (*18*) was used to perform more detailed analyses on the datasets. In particular, ASAP was used to perform sub-clustering and additional differential expression / marker gene discovery. Since the platform also allows annotating (e.g., a color gradient) cells according to the expression of a particular gene set (rather than a single gene), ASAP was also used to study the activity state of KEGG pathways. Finally, because ASAP allows users to share a project (or its copy) privately with a group, all FCA-related projects have been made public so that researchers can share/clone them freely, and annotated Loom files can then be directly downloaded in ASAP.

Jamboree annotation

We have tried two strategies to make sure our "jamboree" annotations are accurate. First, all tissue jamborees were led by Drosophila experts of corresponding tissues. Most annotations were based on ground truth knowledge, and for some uncertain annotations, we allowed experts to include notes besides the annotation. Second, we have implemented an upvote system, allowing experts to vote for clusters that have been assigned to different cell types. We also want to point out that there is no unified standard as a cutoff for FCA annotation, because some cell types can be specified by a single marker (for example, one type of olfactory receptor neuron can be determined by a single olfactory receptor gene), while some other cell types are specified by several different markers. To best document our jamboree annotation records, we have now included a supplemental table to show key information, including names, markers, references if any, and notes (Table S2).

For future annotation, please check flycellatlas.org website, where we provide information to users about how to provide new cell type annotations and where we will post our plan for updating these annotations.

Integration of 10x Genomics and Smart-Seq2 data (Fig. 1D)

We integrated 10x Genomics and Smart-Seq2 data using Harmony (*16*). To facilitate and improve integration, we first selected most relevant genes by performing differential gene expression on annotated 10x Genomics data (t-test; Benjamini-Hochberg corrected p-value < 0.1). To integrate individual tissues, we used genes differentially expressed between cell types of a given tissue. To integrate entire cell atlases,

we used genes differentially expressed between tissues. After selecting differentially expressed genes, we batch-corrected datasets using Harmony to remove the influence of the sequencing technology. We included an additional step of batch-correction for those tissues in which gender specific clusters were present. In total, we performed additional batch-correction based on the gender for 12/15 tissues.

We next systematically validated integration in the following way. If the marker genes were known, we visually inspected whether marker genes are expressed at the same tSNE location for 10x Genomics and Smart-Seq2 data after integration. In case marker genes were not known, we found differentially expressed genes based on annotations of 10x Genomics data and selected 3-5 genes that are as cluster specific as possible. Finally, for each tissue and cell type we came up with the list of genes and validated whether these genes are expressed at the same location in the tSNE space for 10x Genomics and Smart-Seq2 data. Besides Harmony, we considered three other integration approaches and finally decided to use Harmony based on our validation procedure.

Annotating Smart-Seq2 data (fig. S20, S21)

After integrating 10x Genomics and Smart-Seq2 data, we next aimed at transferring cell type annotations from 10x Genomics to unannotated Smart-Seq2 data. To develop an approach and quantitatively compare performance of different classification methods, we used Smart-Seq2 cells from olfactory receptor neurons (ORN) (*11*) annotated using MARS (*49*) and manually validated based on the marker-gene expressions. We integrated this dataset with ORN antenna 10x Genomics dataset that was annotated on the same granularity level. The classification accuracy was high (0.88) with a linear logistic regression model and did not improve by using non-linear models. Therefore, we decided to use logistic regression as the base classifier due to its simplicity and interpretability. Finally, for each tissue we used a 10x Genomics dataset as the train set and trained a logistic regression classifier to distinguish different cell types of annotated 10x Genomics data. We then applied the classifier on the Smart-Seq2 dataset to obtain cell type annotations. To confirm that Smart-Seq2 annotations are indeed correct, we checked expressions of known marker genes and validated if they agree with the predicted Smart-Seq2 annotations.

Comparison of 10x Genomics and Smart-Seq2 data (Fig. 1E)

To compare the number of detected genes between 10x Genomics and Smart-Seq2 data (Fig. 1E), we considered a gene detected if a single read maps to it. In particular, for 10x Genomics data we used UMI greater or equal to 1 as a threshold, while for Smart-Seq2 data we used log2(CPM+1) greater or equal to 1 as a threshold. Given these thresholds, we counted the number of genes detected in at least 1% of cells. To obtain examples of genes that are detected using Smart-Seq2, but not using 10x Genomics (fig. S20G) we obtained a list of genes that are expressed in less than 20 cells in 10x Genomics data, and two to four times more cells in Smart-Seq2 data depending on a tissue.

Brain-Head integration (Fig. 3E)

Single cell RNA-seq dataset from Davie et al. was downloaded from GEO (GSE107451) and processed using VSN. Next, the data was integrated with the single nucleus data from the FCA using Seurat. Data was normalized using SCT normalization (*50*) and batch correction was performed as described (*51*). 150 components were selected for clustering and UMAP/tSNE visualization. Annotations were added using computational approaches. First, we transferred annotation from annotated cells to clusters that contained at least 25% of annotated cells. Next, we used a classifier from Ozel et al. (*22*) to annotate optic lobe cell types. Finally, we trained an SVM classifier on the Davie et al. data, using scikit-learn in Python, following

10 fold cross-validation, optimizing C, kernel and gamma parameters. All computed annotations were then manually curated in jamborees.

Common cell type analysis – Hemocyte (Fig. 4D)

With the cross-tissue analysis, we extracted 8,391 Hml^{+} cells from most body parts, including the fat body, heart, body wall, oenocytes, legs, the Malpighian tubule, tissues in the head, and reproductive organs. Harmony was used to remove batch effects with different parameters to control against overcorrection. We tested lambda (ridge regression penalty parameter) and theta (diversity clustering penalty parameter) in a grid with each parameter ranging from 0 to 3. In the end, the T1L2 combination was found to preserve cell types while not overtly separating cells in batches (based on visual exploration). To explore hemocytes in adults, we annotated cell clusters according to the expression of previously published markers in larval hemocytes and identified twenty-one clusters of Hml^{+} cells (Figure 4D). Crystal cells are readily segregated by high PPO1, PPO2, and lozenge expressions, while plasmatocytes are largely combined as a population most akin to embryonic and larval hemocytes. Plasmatocytes are categorized into five clusters based on their gene expressions and we named the clusters with representative marker genes: Pxn^{High}, $Nplp2/Tep4^{High}$, $Cecropin^{High}$, $LysX/trol/Pvf2^{High}$, and $nAChRalpha3^{High}$. $LysX/trol/Pvf2^{High}$ plasmatocytes exhibit lower Hml compared to other plasmatocytes whereas Pxn^{High} shows the highest Hml with phagocytosis markers including crq, Sr-CI, and NimC1. $Nplp2/Tep4^{High}$ plasmatocytes show a prohemocyte marker, Tep4, and an intermediate prohemocyte marker, Nplp2, along with phagocytosis markers. $Cecropin^{High}$ plasmatocytes display immune- or stress responsive genes such as upd3, Mmp1, Mmp2, and puc. Further, we observed Antp and collier expressing Hml^{+} cells reminiscent of the posterior signaling center in the lymph gland . Yet, lamellocytes are not observed in adults as previously suggested (Bosch et al., 2019) (Figure 4D). In addition to Hml^{+} cells with classical hemocyte gene expressions, we noticed Hml^{+} cells originating from a single tissue, including the testis and antenna, constitute independent clusters significantly enriched with resident tissue marker genes. Overall, single-cell transcriptome profiles of adult hemocytes provide ample resources for understanding adult immunity, hematopoiesis and repertoires of tissue-resident hemocytes.

Hemocytes in adults are largely resident and the majority is found in the thorax or head while a small fraction circulates the hemolymph (*27*). Thus, cross-tissue dissection of adult hemocytes categorized represent hemocytes in adults. Although *Hml* is a well-known marker for plasmatocytes in embryonic- and larval stages, the expression of *Hml* is heterogenous during development which could hinder labeling the entire population of adult hemocytes.

Common cell type analysis - Muscle (Fig. 4E)

Muscle cell clusters were identified by their expression of common sarcomeric gene products, including *Mhc, sls, bt* and *Unc-89* (*52*). With the cross-tissue analysis, we extracted 63,441 muscle cells from most body parts. Harmony was used to remove batch effects with different parameters to control against overcorrection. We tested lambda (ridge regression penalty parameter) and theta (diversity clustering penalty parameter) in a grid with each parameter ranging from 0 to 3. In the end, the T1L2 combination was found to preserve cell types while not overtly separating cells in batches (based on visual exploration). The abundant indirect flight muscle nuclei cluster was uniquely identified by expression of flight muscle-specific markers *TpnC4, Act88F* and *fln* (*53*). Furthermore, the here identified specific expression of different troponinC gene isoforms (*TpnC4, TpnC73F, TpnC41C, TpnC47D, TpnC25D*) was used to further annotate the different muscle clusters taking into account their body part of origin (*54*).

Transcription factors and cell type specificity analysis (Fig. 5A-D)

Cell-type specific TFs were identified using the tau factor. First, calculated average expression profiles in log2CPM space for every cell type and subsequently Z-normalised per gene. We then calculated the tau value for each TF (Flybase r6.36 TF list) using the tspex Python package (version 0.6.2), leading to the identification of 500 TFs with a score higher than 0.85. These factors were plotted in Fig. 5A and are available in **Table S3**. Since the tau factor is calculated on average expression profiles, single-cell information regarding the number of cells where expression is detected is lost. To take this into account, we have split the high tau genes into three categories depending on the percentage of cells in the cell type where at least 1 UMI for the TF was detected (>50%, 50%<x<5% and <5%).

Regarding the analysis showing the TF specificity heatmap (Fig 5C), we used the log normalized gene by cell expression matrix. The cells' expressions were averaged by the broad annotations. We then calculated the tau value for each TF (Flybase r6.36 TF list) using the tspex Python package (version 0.6.2). The values shown in the heatmap are the feature-scaled values using the zscore function available in the package. Only genes passing the thresholds of tau greater than 0.85, log normalized expression greater than or equal to 0.5 and log normalized scaled expression greater than or equal to 1 were retained. The plotting was performed using the ComplexHeatmap R package.

The network shown in Fig 5B is based on 5 different .sif files (network files). The first one is the network based on TFs and broad annotations where links passing the aforementioned thresholds were kept. The second one is the network based on TFs and narrow annotations where 0.85, 2 and 5 thresholds were applied respectively. The other networks represent the narrow-to-narrow, broad-to-narrow and broad-to-broad annotation associations. Since the annotations were mainly driven by the EBI OLS system, most of them are associated with a curated FBbt term. We leverage this graph-based ontology structure in order to compute a semantic similarity between annotations using the ontologySimilarity R package (version 2.5). For broad-to-narrow and broad-to-broad annotation associations, the ones with a semantic similarity below 0.4 were removed except for a few of the broad-to-narrow associations that resulted in a loss of broad terms (fat cell to muscle cell, cardial cell to multidendritic neuron and neuron to multidendritic neuron). For narrow-to-narrow connections, after the expression-based filter, only terms that had a TF assigned were kept and moreover we selected for each term the two most connected terms.

Those 5 processed networks were used as input in the Cytoscape software (version 3.8.0) to build the visual network depicted in Fig 5B. The width of the edges represents the log normalized scaled expression (z-score) while the tau values are represented by the colour intensity of the gene nodes.

UpSetPlot (Fig. 5F)

Average expression profiles for tissues and broad cell types were calculated in log2CPM space. Next, all genes with log2CPM>1 were selected as being highly expressed in the tissue/broad cell type. For every gene the evolutionary age was determined using the GenTree database (http://gentree.ioz.ac.cn/). Finally, sets of genes with their evolutionary age were then plotted in an upset plot using Python (UpSetPlot version 0.4.4).

Sex-differences analysis (Fig. 6A-6D)

For the sex bias analysis, we CPM normalised the gene expression matrix. Next, we filtered out the cells from sex specific samples, i.e., testis, ovary and male reproductive glands and also the cells which were marked to be of ‘mix’ed sex or marked ‘artefact’ or ‘unannotated’ in the annotation. Since ‘body’ samples

also contain sex specific tissues, we further removed the cells annotated as germ cells or cells assigned to other sex specific clusters.

To be as stringent as possible, we further removed cells that were either (i) co-clustered in the t-SNE with these sex-specific cell types or organs, or (ii) might be improperly annotated as evidenced by co-expression of mutually exclusive cell-type specific markers. These cells were identified using the SCope web interface and "lasso" tool and removed. The list of cells removed by this manual procedure is included as in **Table S4**.

At the end of the filtering, 270,486 cells from 176 annotated clusters remained and were used for our analysis. **Table S5** gives the details of these cells and annotations. These were grouped by annotation and for each gene in each annotated cluster, we computed i) it's sex bias B (B = log2((male_avg+1e-9) / (female_avg+1e-9))) where male_avg and female_avg denote the average expression (computed from the normalized expression matrix) of the male and female cells, respectively, in the cluster and p-value for the bias (multiple tests corrected) using Wilcoxon test (scanpy function sc.tl.rank_genes_groups() with default parameters) of the difference in male and female means, and ii) average dsx expression (normalized).

For each annotated cluster and gene, we denoted the gene to be male-biased in this cluster, if sex-biased B was > 1 (i.e., 2-fold change in favor of male) with FDR < 0.05. Similarly, female-biased if sex-biased B was < -1 (i.e., 2-fold change in favor of female) with FDR < 0.05. A gene was considered sex biased if it was either male or female biased. Using this definition, we obtained the list of 9179 genes which are sex-biased in at least one annotated cluster. Next, for each cluster we computed what percentage of these sex biased genes were male (respectively female) biased in the given cluster. We define these fractions as male-bias (respectively female-bias) of the cluster. This information is kept in the data file **Table S6**.

Data used for the SNE visualizations on panel B is kept in the data file **Table S5**. This table also includes the dsx expression for each cell, extracted from the normalized expression matrix. The *dsx* expression level displayed on panel B uses log scale: (log2(dsx+1)). The cells with zero *dsx* expression are shown in gray and remaining using the color scale shown in the legend. For the bottom two subpanels each projected cell is colored according to the sex bias (as defined above) of the cluster it belongs to. For comparison with the top panels, we show the female-bias for female cells only on the left and male-bias for the male cells only on the right. For all four subpanels, if the displayed value is outside the scale, we use the closest extreme color (using sign < or >).

For the cut-off for *dsx* presence in panel D, we used 0.1 which is equal to maximum of average dsx expression (rounded to single decimal digit) of all germ cells which are known to not express dsx but show trace expression in FCA data (we note that these clusters are otherwise removed from this analysis and only used to decide the threshold for other clusters in this analysis).

Trajectory inference of testis subsets (Fig. 6E–G)

We used slingshot to infer a possible branching trajectory in subsets of the testis cells. Specifically, 1) for the spermatogonia-spermatocyte trajectory we used clusters annotated as spermatogonia or spermatocytes, 2) for spermatids we used clusters annotated as early/late spermatids, and 3) for early cyst cells, we used cyst stem cells, early cyst cells and the two spermatocyte cyst cell branches. As input for slingshot, we used Seurat's FindClusters function with resolution 0.4 to find clusters, and the first 20 PCA components

as dimensionality reduction. We also provided the start cell of the trajectory as "cyst stem cell", "spermatogonium", and "early elongation stage spermatid". To determine differentially expressed genes, we used the "calculate_overall_feature_importance" function from dyno (https://github.com/dynverse/dyno) and filtered genes based on a feature importance of at least 0.1 and a log2fold change along any point in the trajectory of at least 0.5. To map the trajectory onto the UMAP embedding, we used the project_trajectory function, also implemented in dyno.

Metabolic clustering using ASAP

For probing the FCA data for an enrichment in Fatty acid synthesis (FAS) and Fatty acid degradation (FAD) metabolic genes, we first downloaded the latest KEGG assignments of genes of *Drosophila* melanogaster (DM) which were mapped to KEGG pathways (FAS: map00061, and FAD: map00071) on December 2020. We next used the ASAP (Automated Single-cell Analysis Pipeline) platform to probe the FCA datasets for a transcriptional enrichment of the genes of these metabolic pathways (*18*). ASAP generates a score from the difference of expression of the genes in the FAS or FAD gene sets as compared to background genes. In short, for each gene in the gene set, the function will take n random genes from the same expression quantile and add them to a background gene set. The gene set score is calculated as the difference of average expression of the genes in the module score and the genes in the background, for each cell. Scores close to zero indicate a similar expression, positive scores indicate higher expression and negative scores indicate lower expression of the genes in the gene set than the background genes. This function was adapted from the AddModuleScore function from the Seurat package , and was entirely recorded in Java. We have used the non-normalized parsed Fat body v2 data as input matrix, 24 bins, 100 background genes and set the seed to 42. We plotted the results using the visualization feature implemented in ASAP and colored the cells according to the score values (*18*). Finally, we used the FCA 0.4 resolution clustering information available on Scope to delineate the frontiers between the different cell clusters.

Transcription factor pleiotropy analysis

Markers were calculated with the wilcoxon test, comparing every cell type against all other cells. Next only genes with pval adj<0.05 and average $\log_2$[foldchange] > 1 were selected as selective markers.

Ovary data integration (fig. S37)

Four scRNA-seq datasets of the adult ovary were used for the data integration: current FCA data and three other published datasets (*41*, *42*, *55*) merged and batch corrected using Seurat v4.0.1. Datasets were processed with Seurat v4.0.1 in RStudio Version 1.4.1103. Batch correction was performed as described (*51*) 4. Clustering of unannotated cell types was performed using the FindClusters command in Seurat v4.0.1 with a resolution factor of 0.6. Image processing was performed with FIJI. Fly lines were ordered from BDSC: *sick-Gal4* (#76195), *Wnt4-Gal4* (#67449), *UAS-RedStinger, UAS-Flp, Ubi-(FRT.STOP.FRT) -Stinger* (#28281).

FCA Consortium 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).

H. Li is a CPRIT Scholar in Cancer Research (RR200063) and supported by National Institutes of Health (R00AG062746). L. Luo is a Howard Hughes Medical Institute (HHMI) investigator and supported by the National Institutes of Health (R01DC005982). S. Aerts was supported by KU Leuven (grant no. C14/18/092) and FWO (grant no. G094121N). J. Janssens is supported by a PhD fellowship of The Research Foundation – Flanders (FWO, 1199518N). B. Deplancke, V. Gardeux, and F. David were supported by an EPFL Open Science Grant and a Precision Health and related Technologies Grant (no. PHRT-502). F. Schnorrer is supported by the Centre National de la Recherche Scientifique (CNRS), Aix-Marseille University (ANR-11-IDEX-0001-02), the European Research Council under the European's Union Horizon 2020 Programme (ERC-2019-SyG 856118), the French National Research Agency (ANR-18-CE45-0016-01) and the Turing Centre for Living Systems (ANR-16-CONV-0001). S. Quake is a Chan Zuckerberg Biohub Investigator. N. Perrimon is a Howard Hughes Medical Institute (HHMI) investigator. C. McLaughlin is a HHMI fellow of the Damon Runyon Cancer Research Foundation (DRG 2390- 20). M. T. Fuller and C. W. Berry were supported by NIH grants 1R01GM122951 and R35GM136433, the Reed-Hodgson Professorship in Human Biology to MTF and NIH Training grant T32AR007422 for CWB. S. K. Kim and L. Kockel were supported by NIH awards (R01 DK107507; R01 DK108817; U01 DK123743; P30 DK116074 to S.K.K.), the JDRF Northern California Center of Excellence (to S.K.K. and M. Hebrok), the H.L. Snyder Foundation and Elser Trust, and the Stanford Diabetes Research Center. K. Rust was supported by the Deutsche Forschungsgemeinschaft (DFG, project number 419293565) and the Promoting Scientific Independence (PSI) Program of the Philipps-University Marburg, Germany. C. Ribeiro is supported by the Champalimaud Foundation, "la Caixa" Banking Foundation (LCF/PR/HR17/52150002), Fundação para a Ciência e Tecnologia (02/SAICT/30081/2017), and the Global Consortium for Reproductive Longevity and Equality through the Buck Institute, made possible by the Bia-Echo Foundation (GCRLE-1420). H. Krause was supported by a grant from the Canadian Institutes of Health Research (MOP-133743). J. A. Brill was supported by the Natural Science and Engineering Research Council (RGPIN-2016-06775). B. Oliver was supported in part by the Intramural Research Program of the NIH, The National Institute of Diabetes and Digestive and Kidney Diseases (DK015600). T. Przytycka was supported in part by the Intramural Research Program of the NIH, National Library of Medicine, (LM200887). A. K. Groves is supported by the National Institutes of Health (R01DC014932). S. Yamamoto is supported by the National Institutes of Health (U54NS093793, R01DC014932, R24OD022005, U54OD030165, R01AG071557, R01HG011795). M. Wolfner was supported by NIH grant R37HD038921, and by an NIH grant of which A.G. Clark is joint-PI, R01HD059060. Mehmet Neset Özel, Nikolaos Konstantinides and Claude Desplan were supported by NIH grant R01EY017916. Claude Desplan was also supported by NYU Abu Dhabi Research Institute grant to the Center for Genomics and Systems Biology (ADHPG-CGSB). R. Lehmann was supported as

a Howard Hughes Medical Institute (HHMI) investigator and by the National Institutes of Health (R37 HD41900). E. Donà was supported by a Sir Henry Wellcome Postdoctoral Fellowship (110232/Z/15/Z). G. S. X. E. Jefferis was supported by the MRC (MC-U105188491) and an ERC Consolidator grant (649111). S.L. Holtz and R.I. Wilson are supported a grant from the NIH (R01NS101157). S.L.H. was also supported by a fellowship from the NIH (F31NS106982). R.I.W. is an investigator of the Howard Hughes Medical Institute (HHMI). S. DiNardo is supported by NIH R35GM136270. Scott Waddell and Christoph D. Treiber were funded by a Wellcome Principal Research Fellowship (200846/Z/16/Z), Wellcome Collaborative Award (209235/Z/17/Z) and European Research Council Advanced Grant (789274) to S.W. R. Bodmer is supported by NIH HL054732, AG045428, AG058075, HL149992, G. Vogler is supported my HL54732 and DOD W81XWH-21-1-0259, M. Tamayo is supported by DOD W81XWH-21-1-0259, K. Birker is supported by AHA 835002, J. Kezos is supported by NIH F32 Hl56607. T. Kornberg is supported by NIH R35GM122548. I.Papatheodorou, N. Georg, J. Manning are supported by EMBL States, BBSRC/NSF BB/T014563/1 and WT Single Cell Atlas: 221401/Z/20/Z. "The research was funded in part by the Wellcome Trust [221401/Z/20/Z]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission." "This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) [Grant number: BB/T014563/1]" Z. Carvalho-Santos is financed by national funds through the FCT in the framework of the financing of the Norma Transitória DL 57/2016. I. Tastekin is financed by Marie Skłodowska-Curie Actions postdoctoral fellowship MSCA-IF-EF-ST 867459. R. Cardoso-Figueiredo is financed by the FCT doctoral fellowship SFRH/BD/143862/2019. D. Dhakan is financed by the La Caixa Foundation, LCF/PR/HR17/52150002. Zepeng Yao was supported by a Jane Coffin Childs Fellowship. T. G. Nystul is supported by NIH R35GM136348. Erika L. Matunis is supported by NIH R35GM136665 and HD052937. Thomas R. Clandinin was supported by 5R01EY02263808, 5U19NS104655-04, 3R01NS11006003S1, and by 5P30EY02687704. Jiwon Shim was supported by the Samsung Science and Technology Foundation (SSTF-BA1701-15). Aaron M. Allen, Megan C. Neville and Stephen F. Goodwin were supported by a Wellcome Investigator Award (106189/Z/14/Z) and a Wellcome Collaborative Award (209235/Z/17/Z) to S.F.G. Joshua Shing Shun Li was is supported by a Croucher fellowship for Postdoctoral Research from the Croucher Foundation. Devika Agarwal was supported by the Wellcome Collaborative Award (209235/Z/17/Z) to S.W. , S.F.G and David Sims. Helen White-Cooper was supported by a the Biotechnology and Biological Sciences Research Council (BBSRC) (Grant number BB/P001564/1). O'Brien Lab information: Lucy O'Brien was supported by the American Cancer Society (ACS) RSG-17-167-01-DDC and the National Institutes of Health (NIH) (R01GM116000-01A1). Erin Nicole Sanders was supported by NSF GRFP DGE-1656518 and by a Stanford Developmental Biology and Genetics NIH Training Grant (2T32GM00779038). Paola Moreno Roman was supported by a Stanford Bio-X Bowes Graduate Fellowship and by a Stanford DARE (Diversifying Academia Recruiting Excellence) Fellowship and the National Institutes of Health (NIH) (R01GM116000-01A1). Anna A. Kim was supported by the Swedish Research Council under grant number 2017-06156 and a seed grant from the Stanford Bio-X. Yu-Han Su was supported by the National Institutes of Health (NIH) (R01GM116000-01A1). Anthony Galenza was supported by the National Institutes of Health (NIH) (R01GM116000-01A1). Sang Ngo was supported by the American Cancer

Society (ACS) RSG-17-167-01-DDC and the National Institutes of Health (NIH) (R01GM116000-01A1). Aparna Sherlekar was supported by the American Cancer Society (ACS) RSG-17-167-01-DDC and the National Institutes of Health (NIH) (R01GM116000-01A1). Andrew Thomas Labott was supported by the National Institutes of Health (NIH) (R01GM116000-01A1). Samantha Gumbin was supported by a Stanford Graduate Fellowship, Smith Fellow.

Supplementary Figures and Legends

FCA Figure S1

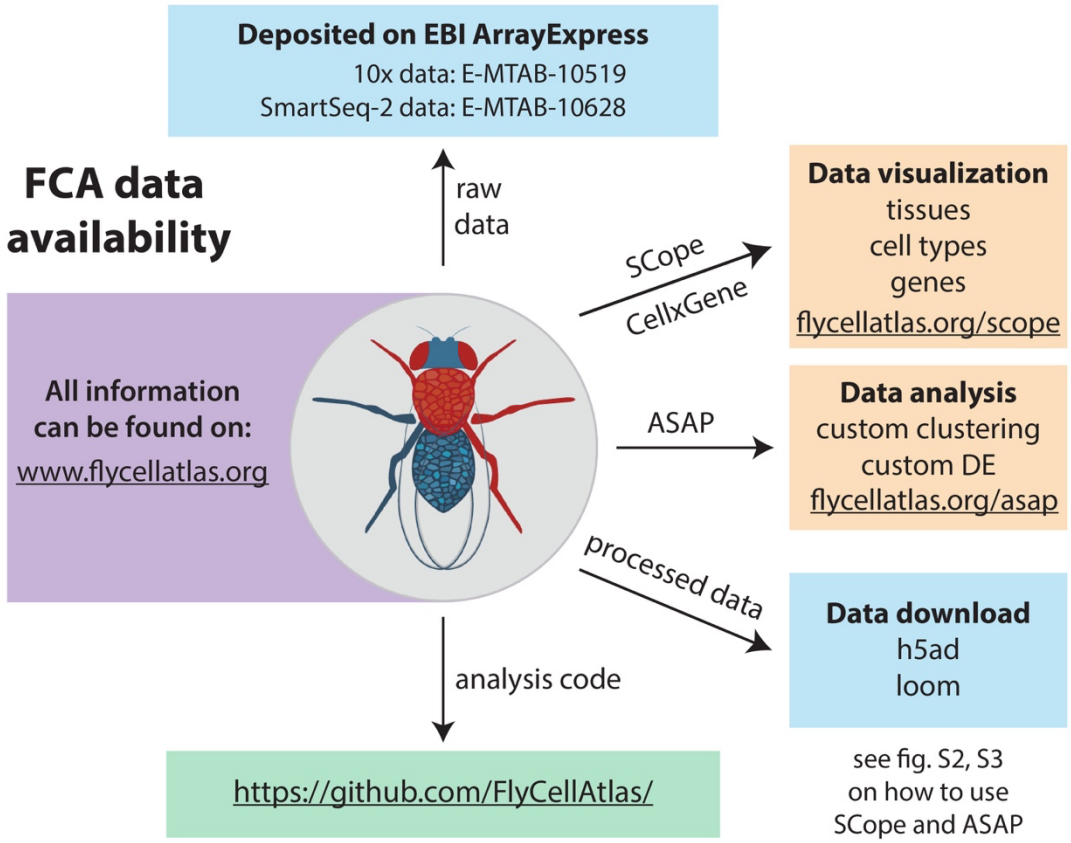


Figure S1. Summary of FCA data availability. See fig. S2 and S3 for more details.

FCA Figure S2

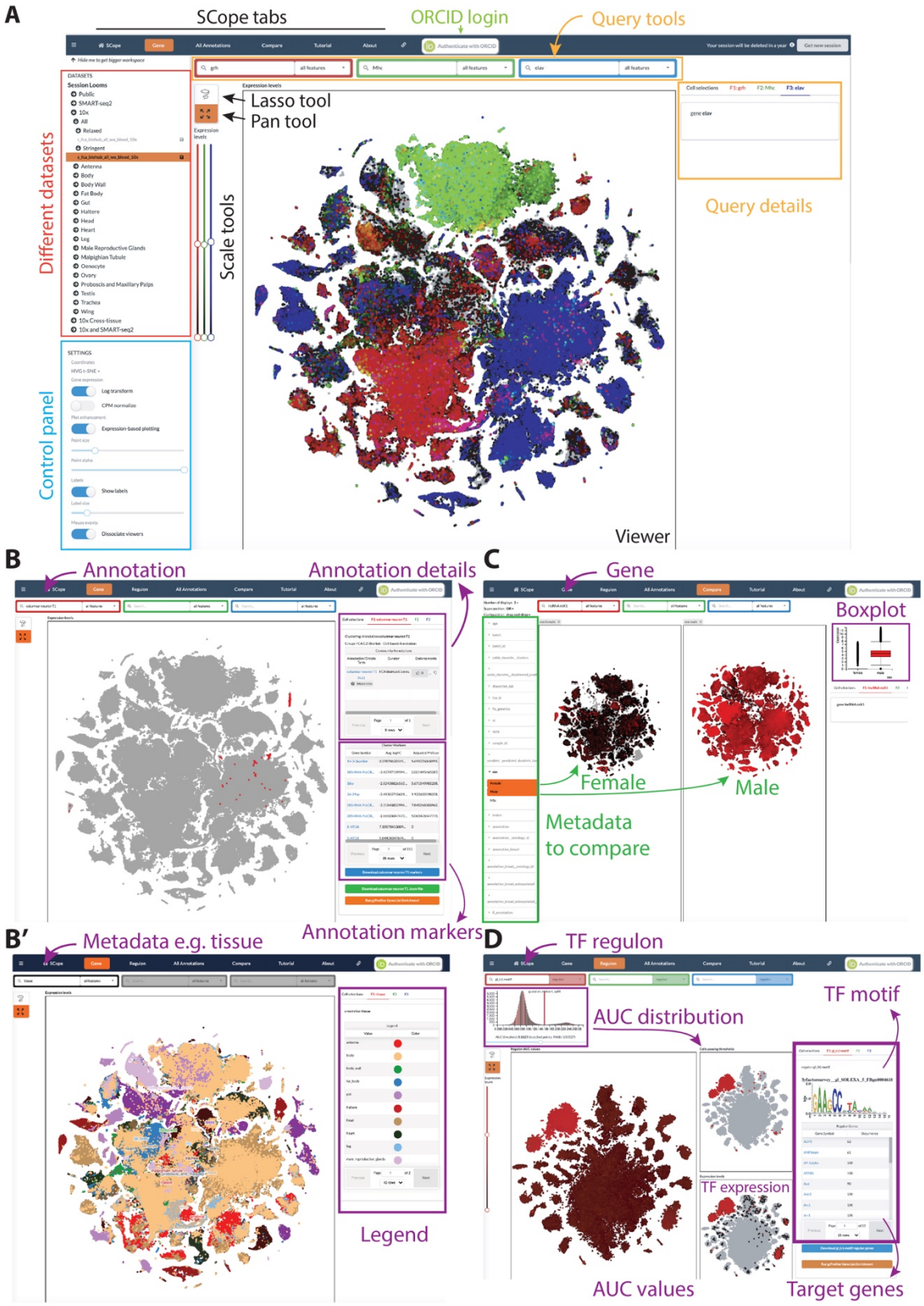
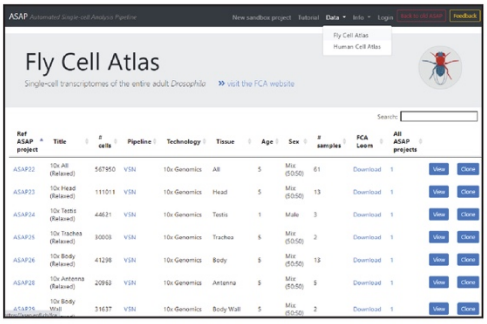


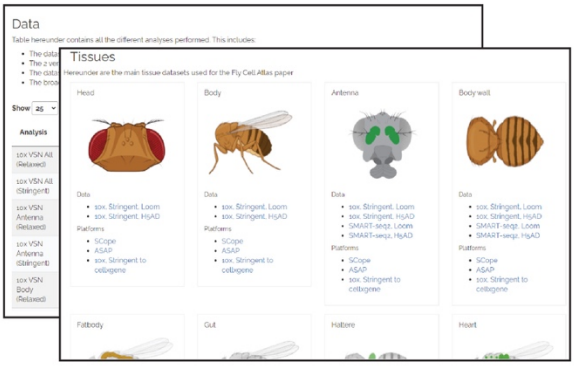
Figure S2. How to use SCope for visualizing the FCA datasets. This is a short overview of how the FCA datasets are accessible through the SCope platform (https://flycellatlas.org/scope).
(A) At the top, different modes can be selected, "Gene" being the default option. Further, a user can login with their ORCID. Navigation between different datasets is possible through the hierarchical tree on the left, showing both Relaxed and Stringent datasets for every tissue. The selected tissue is visualized in the main viewer, whose settings can be adapted in the Control panel: different embeddings can be selected by clicking on "Coordinates" (choices are tSNE, UMAP, PCA, SCENIC tSNE and SCENIC UMAP), and the appearance of the points can be modified (size and transparency [alpha]). Gene expression can be plotted using up to three query boxes, by default one for each primary color. Colors can be modified using the Scale tools or by selecting different normalization options in the Control panel. A user can select cell subpopulations for closer inspection with the Lasso tool, and move and zoom in the viewer using the Pan tool.
(B) The "Gene" mode also allows the user to look for annotations. Cells belonging to the selected annotation will be colored, and annotation details including marker genes will pop up on the right. The marker genes can be sorted by AvgLogFC or adjusted p-value. Download buttons are present to download marker genes or a subset of the loom file. Finally, the marker genes can be sent to gProfiler for gene ontology enrichment. Similarly, clusters and their marker genes can be queried in different resolutions.
(B') Metadata can also be queried in the "Gene" mode, coloring cells based on metadata (e.g. tissue, sex) with a legend on the right and labels over the median location.
(C) The "Compare" mode, allows to split the data based on metadata, allowing to compare gene expression. Additionally, a boxplot appears for more quantitative comparisons.
(D) The "Regulon" mode allows to visualize gene regulatory networks linked to TFs. A regulon (GRN) consists of a TF and its motif, with predicted target genes that are co-expressed and whose locus is enriched for the TF motif. Network activity in cells is scored as AUC with AUCell. The AUC distribution is shown as a histogram, and an optimal threshold can be set manually. The viewer shows both the raw AUC values (left), the cells passing the AUC threshold (top right), and the raw TF expression (bottom right). The "Regulon" tab is present for all tissues, but not for the combined dataset.
For more details, please refer to this video tutorial for using SCope (https://www.youtube.com/watch?v=yNETQVaSJYM&t=349s).

A

Access FCA datasets directly from ASAP (asap.epfl.ch)
Top bar > Data > FlyCellAtlas



Or from ASAP links on the FCA main portal (flycellatlas.org)
Top bar > Data



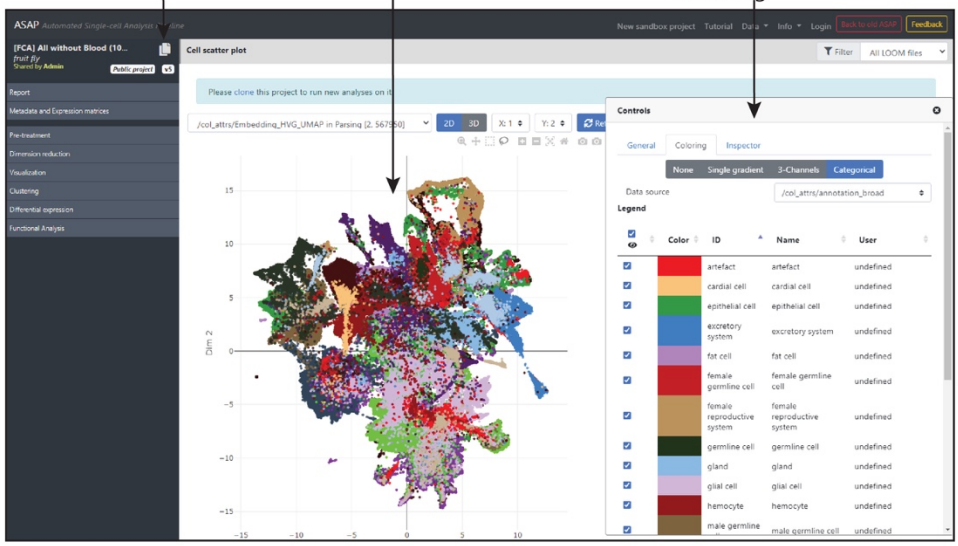
B

Selecting any tissue/project will then open the project in ASAP and jump directly to the precalculated UMAP (that can be changed to t-SNE/PCA). In this view, you can:

Clone / share the project (to modify it)

Browse the cells interactively or select cells using lasso

Color the plot according to
- gene expression
- metadata (sex, batch, etc..)
- cell type annotation
- clustering

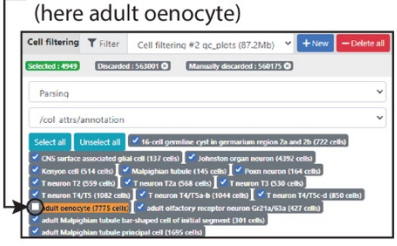


Run additional analyses online such as cell filtering (e.g. subclustering), clustering, differential expression, ...

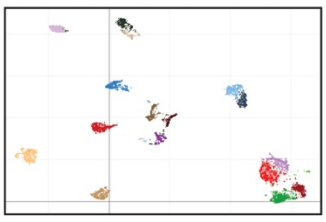
C

Case study: you can filter cells and select only one cluster for subclustering and marker gene discovery

1. Pre-treatment > Cell Filtering
Select cluster(s) of interest to keep
(here adult oenocyte)



2. Dimension Reduction
PCA > UMAP > Clustering
(here 19 subclusters are found)



3. Differential expression
Marker genes of each subcluster
(here top 10 up-/down-regulated)

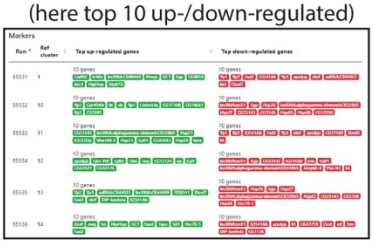


Figure S3. How to use ASAP for analyzing the FCA datasets. This is a short overview of how the FCA datasets are accessible through the ASAP platform.
(A) The datasets can be accessed directly from the main ASAP website (https://flycellatlas.org/asap) where there is a dedicated page (Top bar > Data > FlyCellAtlas) listing all projects and different information (number of cells, technology used, tissue, etc…). The projects are also listed at the main Fly Cell Atlas portal (https://flycellatlas.org), and linked out to the main ASAP website, for direct view of the data. For each dataset, the user can view the project, i.e. visualize and interact with the results of the analysis pipeline and cell type annotation. The user can also clone the project in its user space to be able to modify it or create new analyses.
(B) This is the main view when the user accesses a project (from any source). The main visualization is a UMAP but t-SNE and PCA can be visualized as well. In this view, the user can color the cells according to many criteria such as gene expression, metadata (sex, age, batch, etc.), annotated cell type, clustering results, and more. The left menu contains the single-cell analysis standard pipeline to run additional/new analyses on the dataset. For example, the user can perform a new differential expression to find marker genes of each cluster or each annotation. The user can also select cells of interest (by rectangular or lasso selection) and find marker genes that are specific to this selection.
(C) This use case is featuring re-analysis of the "All" dataset (https://asap.epfl.ch/projects/ASAP22) by 1). Performing a new "Pre-treatment > Cell filtering" and selecting a cluster/annotation of interest. Here we selected the "adult oenocyte" cluster from the "annotation" metadata. It generates a new subset of 7775 cells (from 567950 cells, initially). 2). Then, we performed a new PCA with 50PCs on this subset (Dimension Reduction > PCA) and, once computed, we ran a UMAP (Dimension Reduction > UMAP) and a clustering (Clustering > Seurat) on the 50PCs of the PCA. Then, the figure shows the visualization of the UMAP, colored using the 19 clusters found by the clustering method. 3). Finally, we ran a differential expression analysis (Differential expression > Seurat – Wilcoxon) for all clusters (vs complementary) to find the marker genes of each subcluster. This view shows the top 10 up- and down-regulated genes, for each cluster (full table is also available). The user can change the thresholds (p-value, FDR, fold-change), highlight transcription factors or surface markers, or annotate the clusters. Of note, clicking on a gene will display a description of the gene and provide a link out to the Ensembl database.

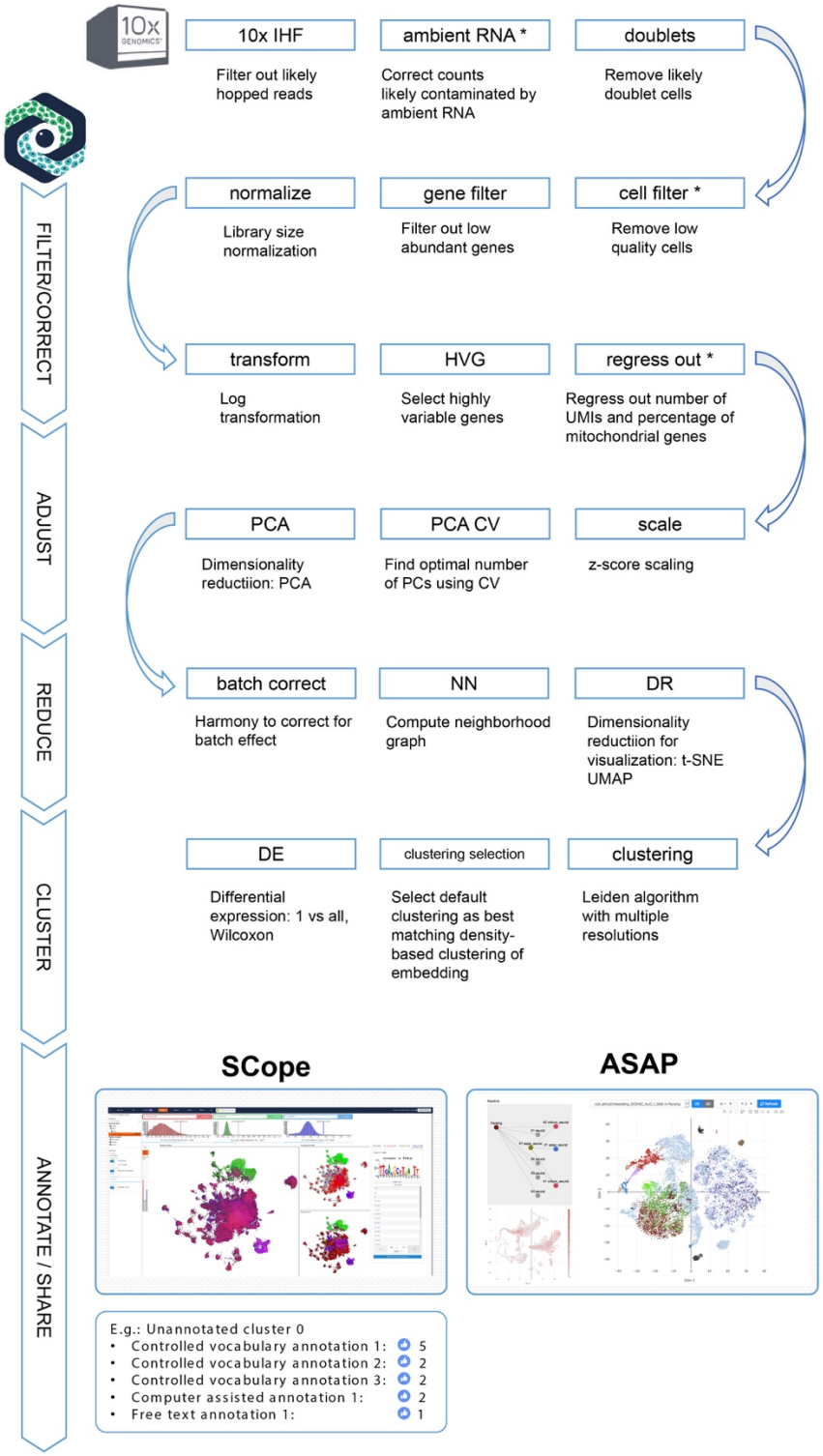


Figure S4. Steps for 10x data processing, from raw sequencing data to cluster analysis. Processed data are annotated and shared through SCope and ASAP. See Methods for detailed description.

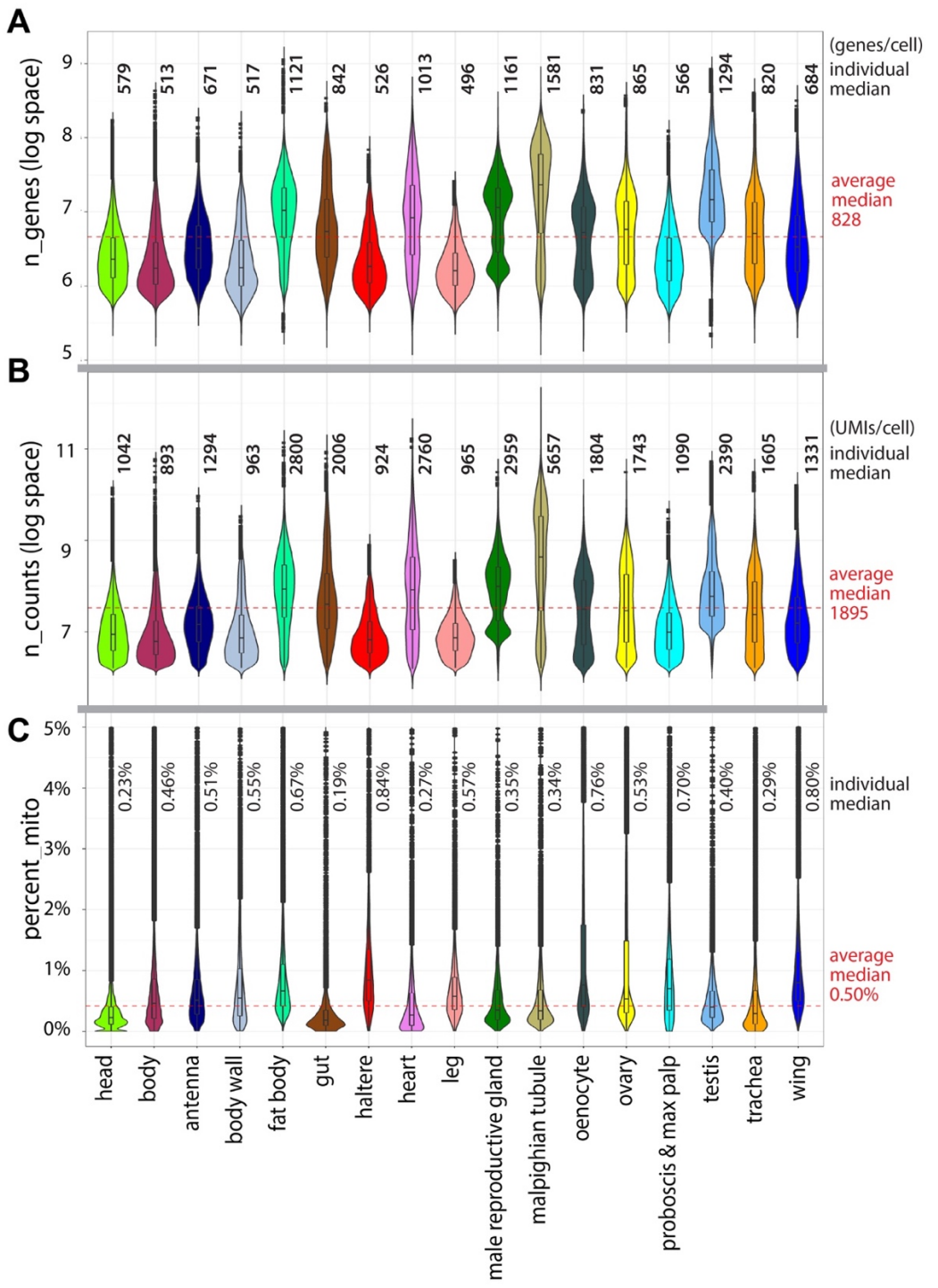


Figure S5. Quality control of 10x data.
(A) The average median UMI count is 1895 UMIs per cell.
(B) The average median gene detection is 828 genes per cell.
(C) The average median percentage of mitochondrial genes for all samples is 0.50%.
Numbers for individual tissues are indicated.

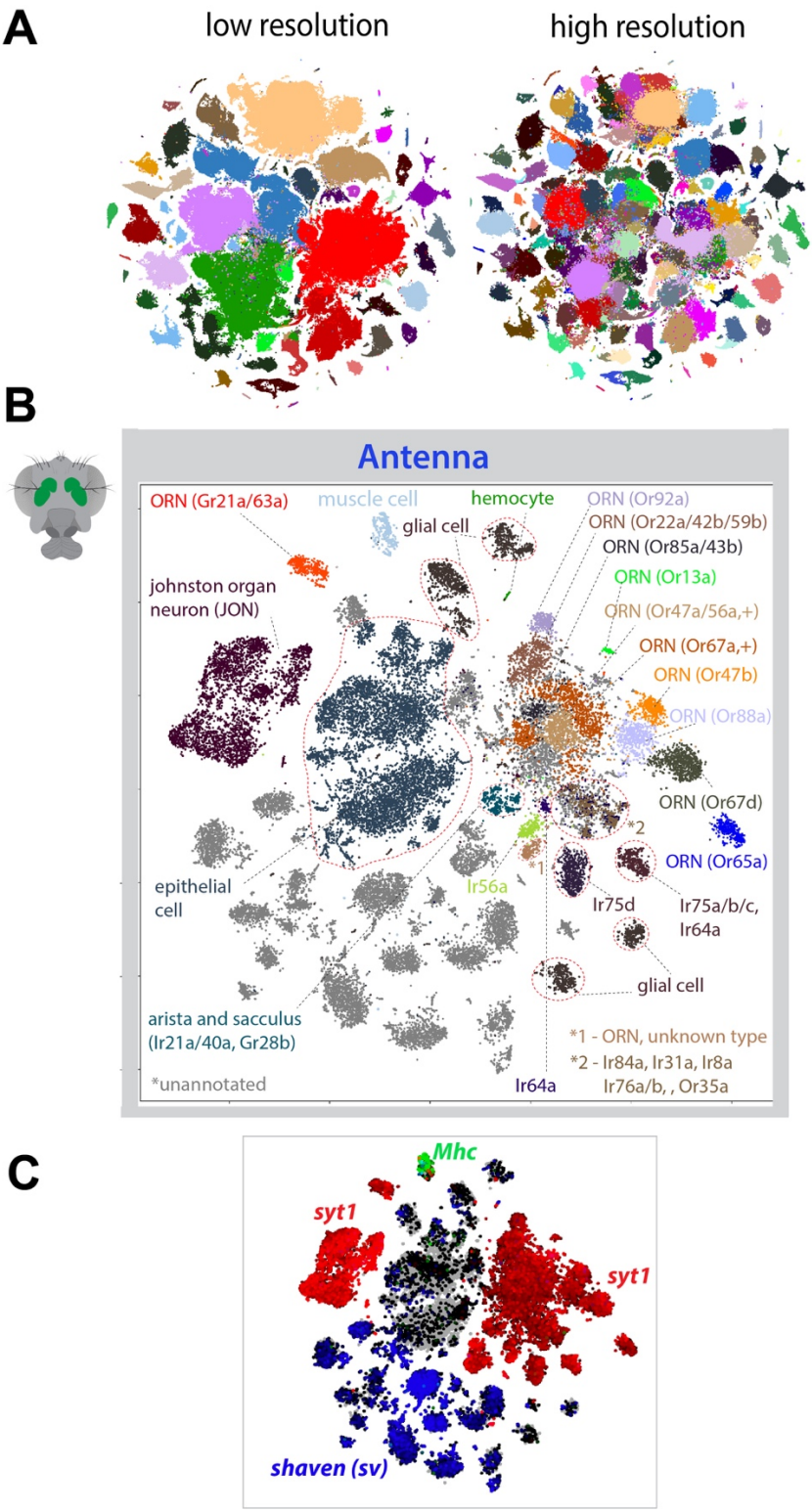


Figure S6. Different clustering resolutions and cell type annotation in the antenna.

(A) Different clustering resolutions (using Leiden) are used for annotating cell types, because some cell types are present at low clustering resolution, and others appear only at higher resolution.
(B) tSNE plot with annotations for the fly antenna from the *Stringent* 10x dataset. All three antennal segments were dissected for single-nucleus sequencing. ORN: olfactory receptor neuron.
(C) The unannotated clusters in the antenna are largely *shaven+,* likely to be different types of non-neuronal and non-glial supporting cells from different segments of the antenna.

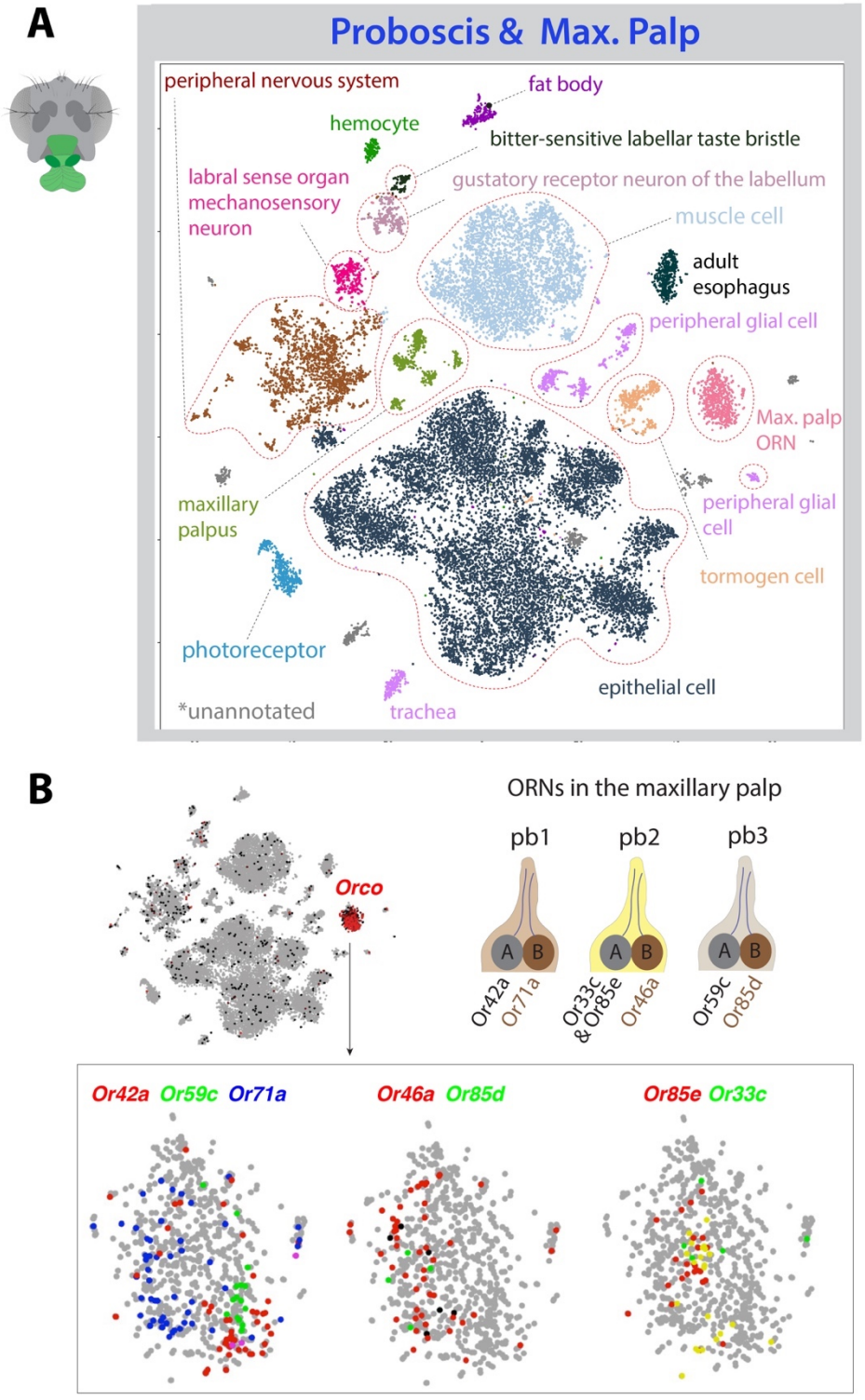


Figure S7. Cell type annotation in the proboscis and maxillary palp.

(A) tSNE plot with annotations for the fly proboscis and maxillary palp from the *Stringent* 10x dataset. ORN: olfactory receptor neuron.
(B) Expression of olfactory receptor co-receptor (*orco*) in one cluster annotated as maxillary palp ORNs. All 7 known olfactory receptor genes can be detected, including two receptors, Or33c and Or85e, that are co-expressed in the same ORN (*56*). Palpal basiconic 1 (pb1), pb2, and pb3 are three different types of sensilla in the maxillary palp.

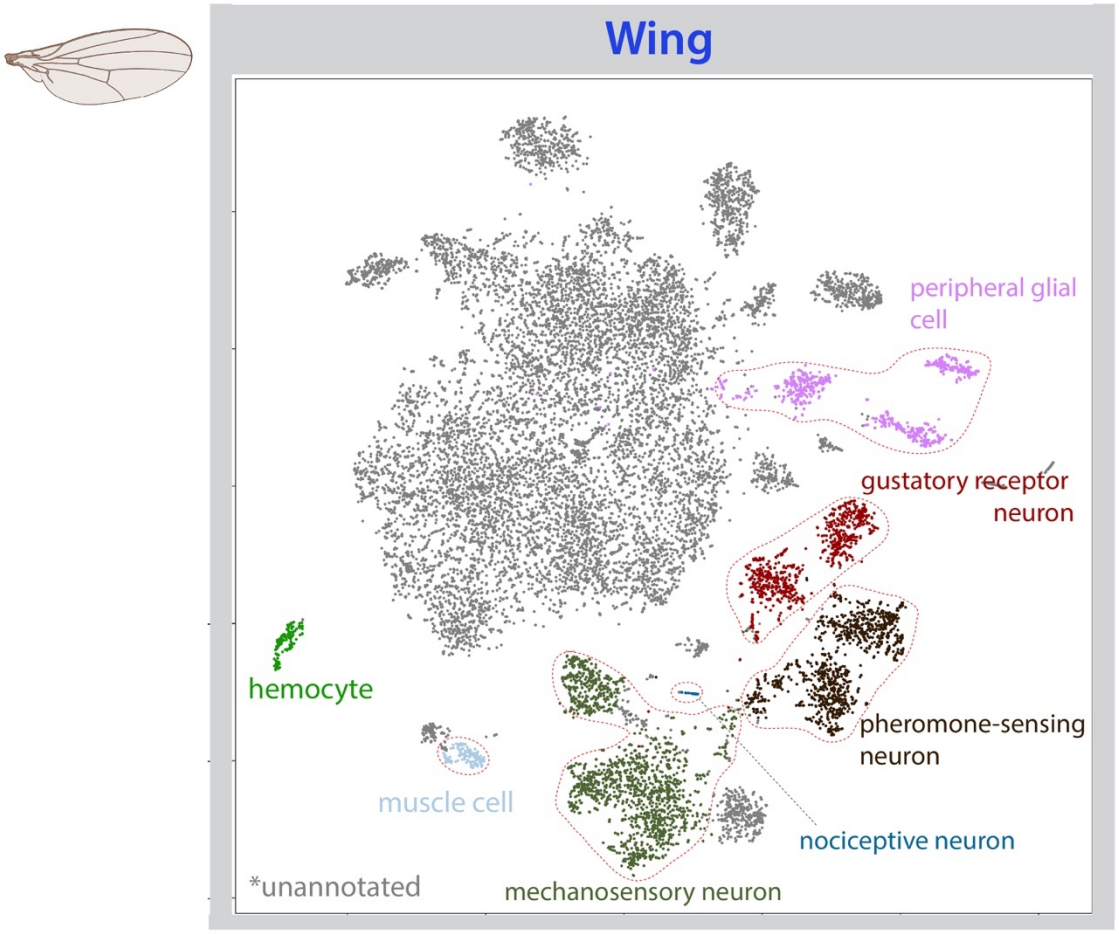


Figure S8. Cell type annotation in the wing. tSNE plot with annotations for the fly wing from the *Stringent* 10x dataset. Note a large group of cells are currently unannotated, which are likely to be epithelial cells.

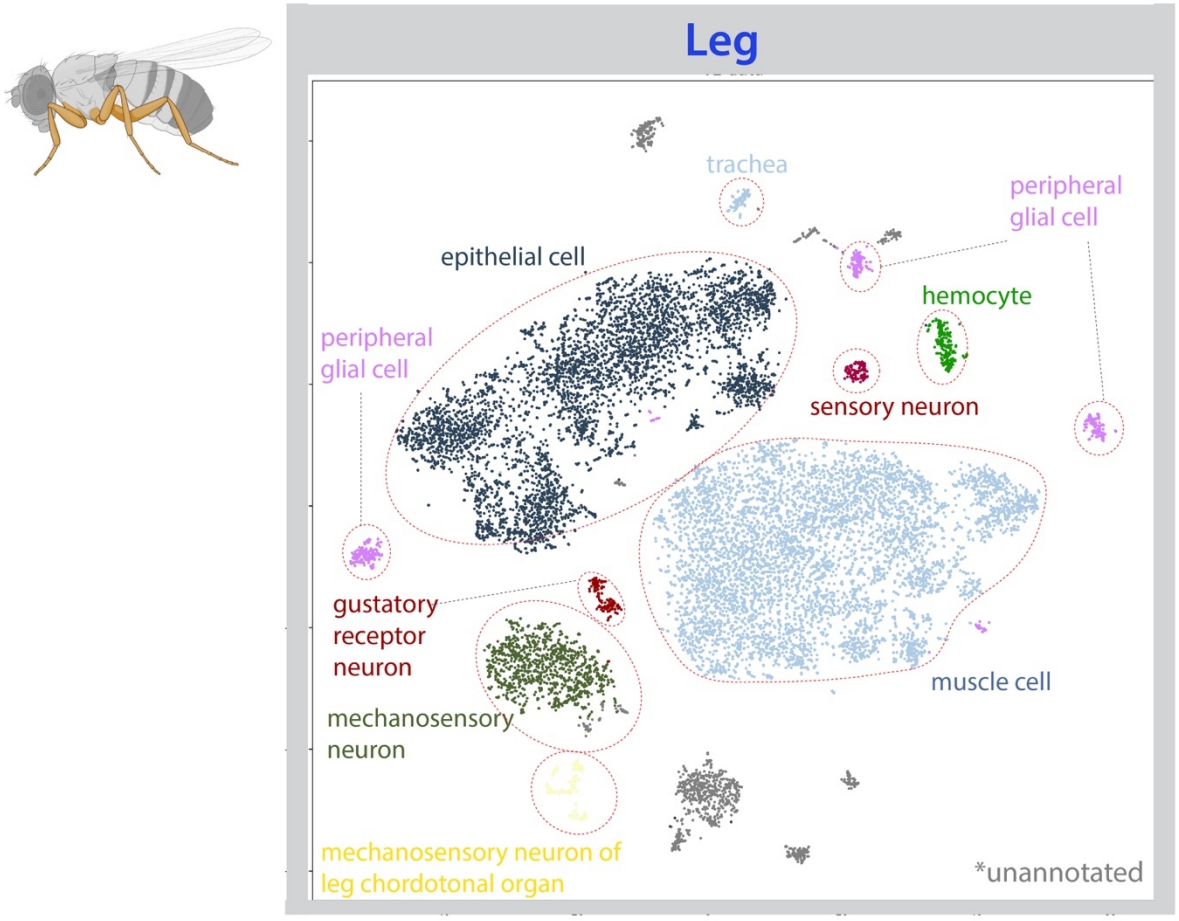


Figure S9. Cell type annotation in the leg. tSNE plot with annotations for the fly leg from the *Stringent* 10x dataset. All six fly legs were dissected and pooled for single-nucleus sequencing.

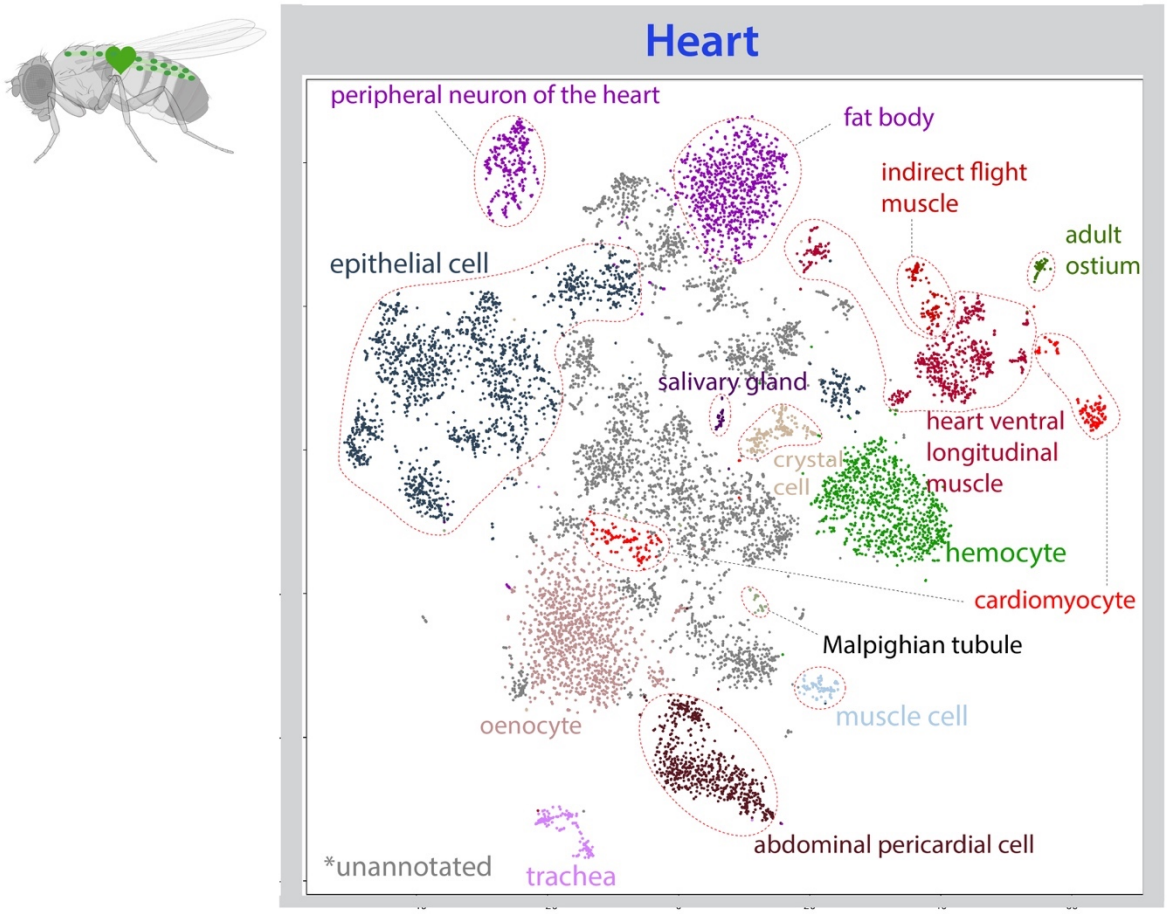


Figure S10. Cell type annotation in the heart. tSNE plot with annotations for the fly heart from the *Stringent* 10x dataset.

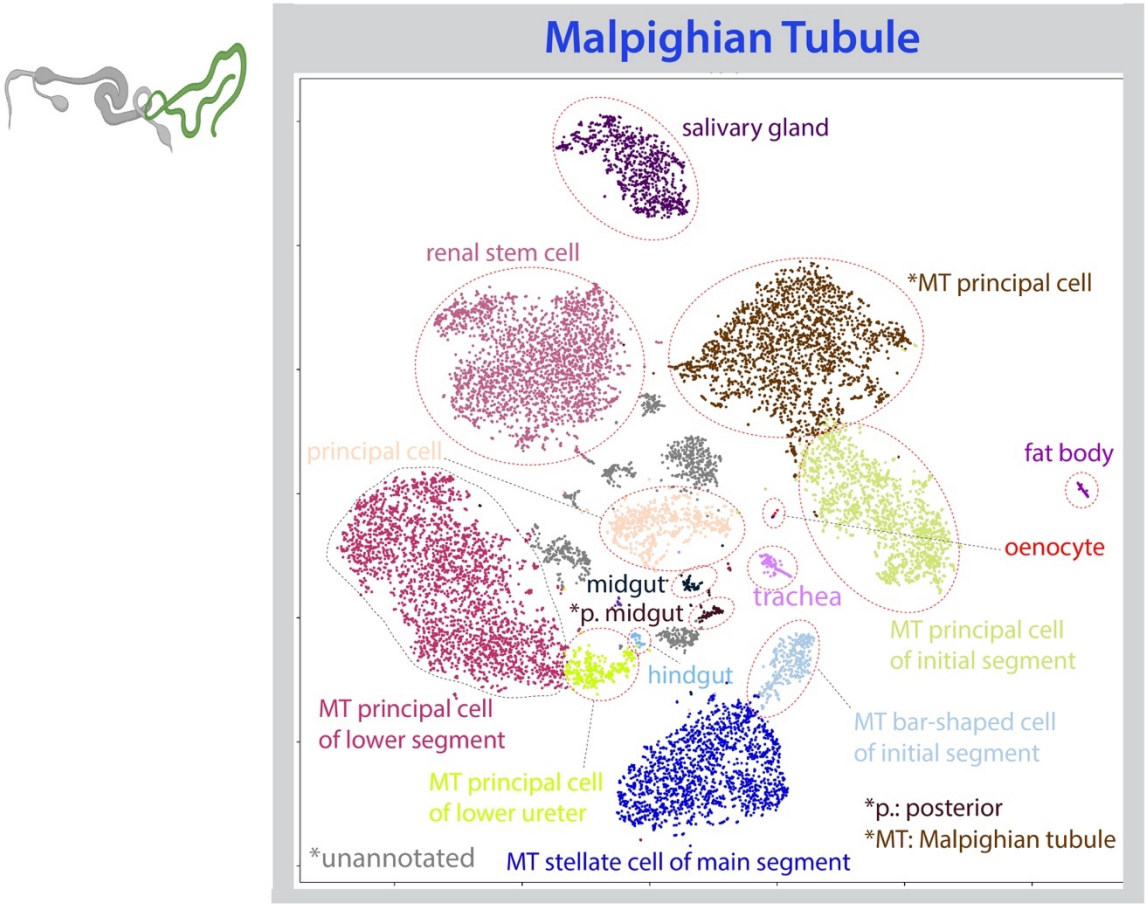


Figure S11. Cell type annotation in the Malpighian tubule. tSNE plot with annotations for the fly Malpighian tubule (MT) from the *Stringent* 10x dataset.

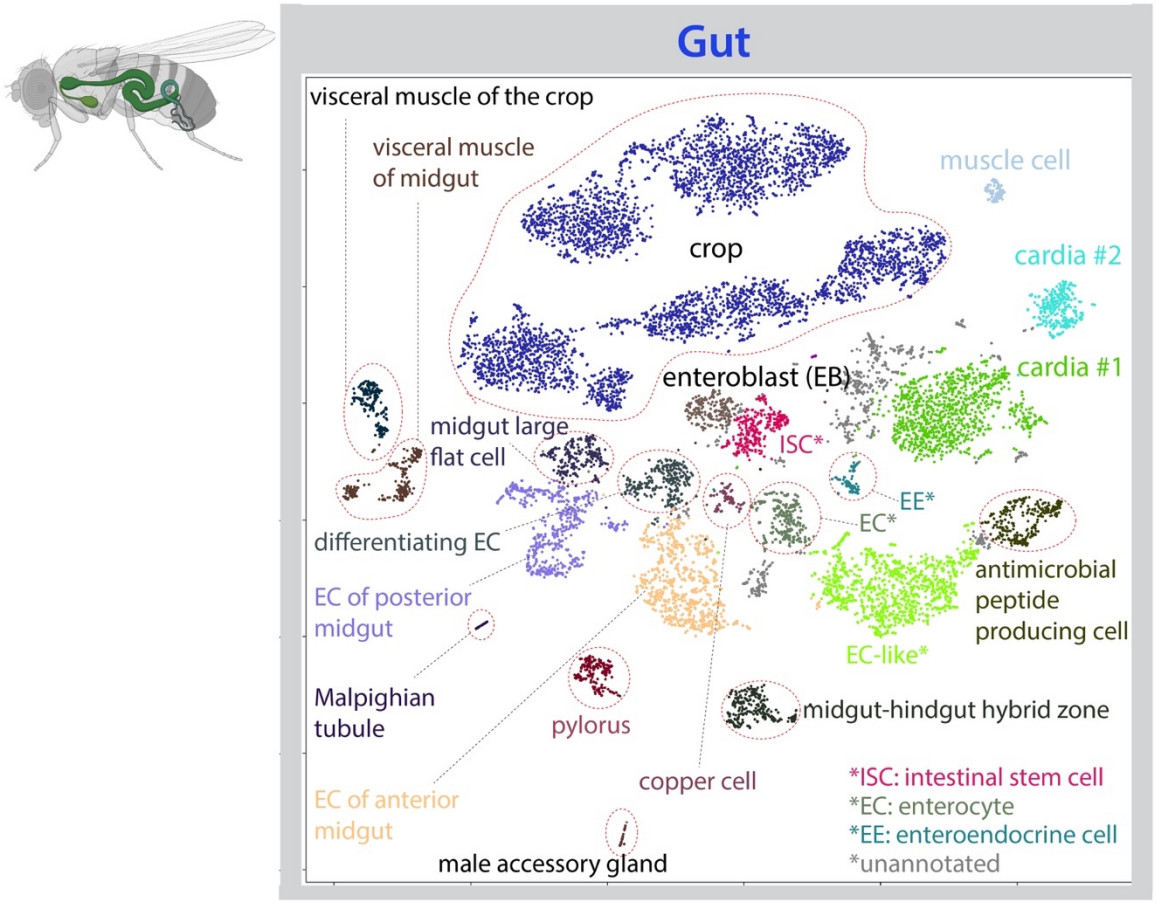


Figure S12. Cell type annotation in the gut. tSNE plot with annotations for the fly gut from the *Stringent* 10x dataset.

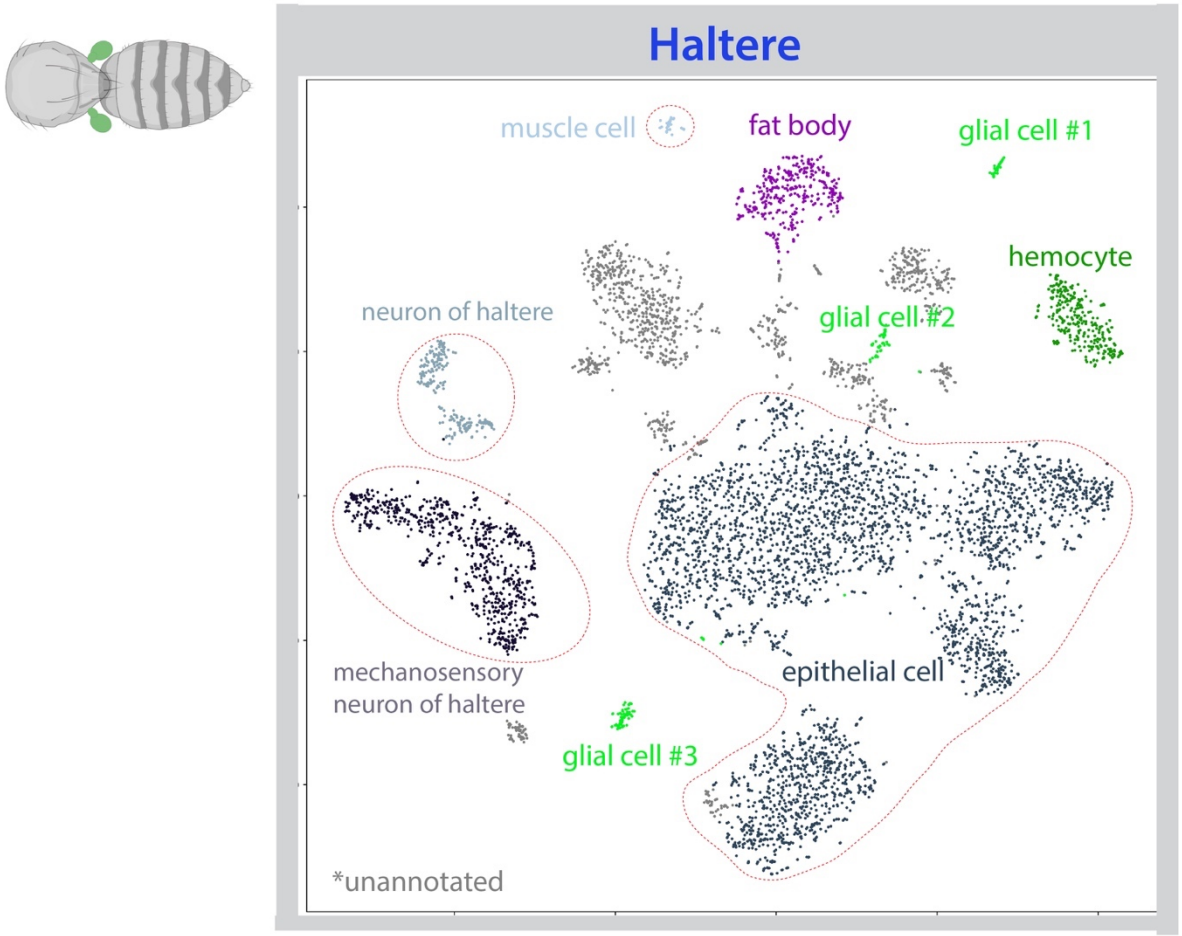


Figure S13. Cell type annotation in the haltere. tSNE plot with annotations for the fly haltere from the *Stringent* 10x dataset.

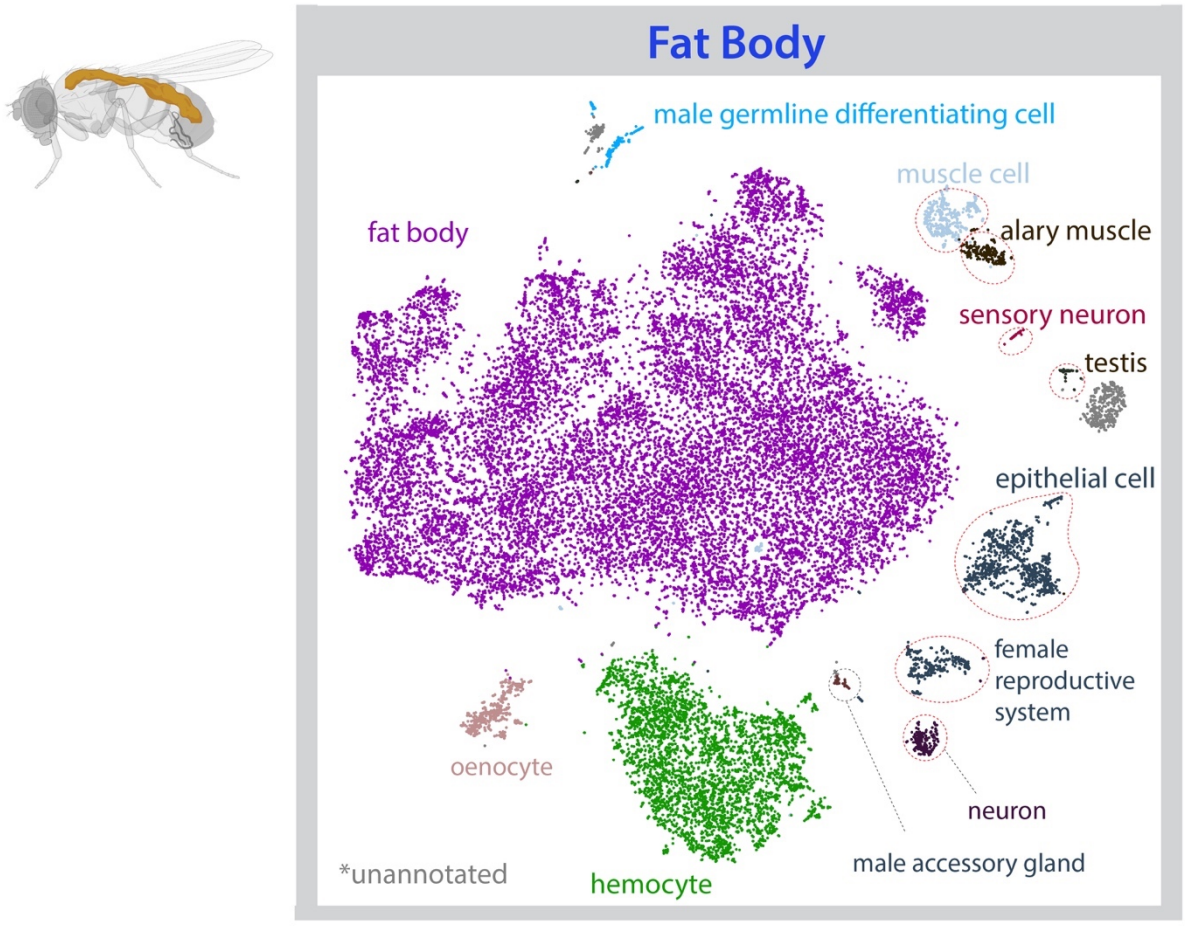


Figure S14. Cell type annotation in the fat body. tSNE plot with annotations for the fly fat body cells from the *Stringent* 10x dataset. Fat body cells are FAC-sorted based on the nuclear GFP signal; flies are *Cg-GAL4 > UAS-lamGFP*.

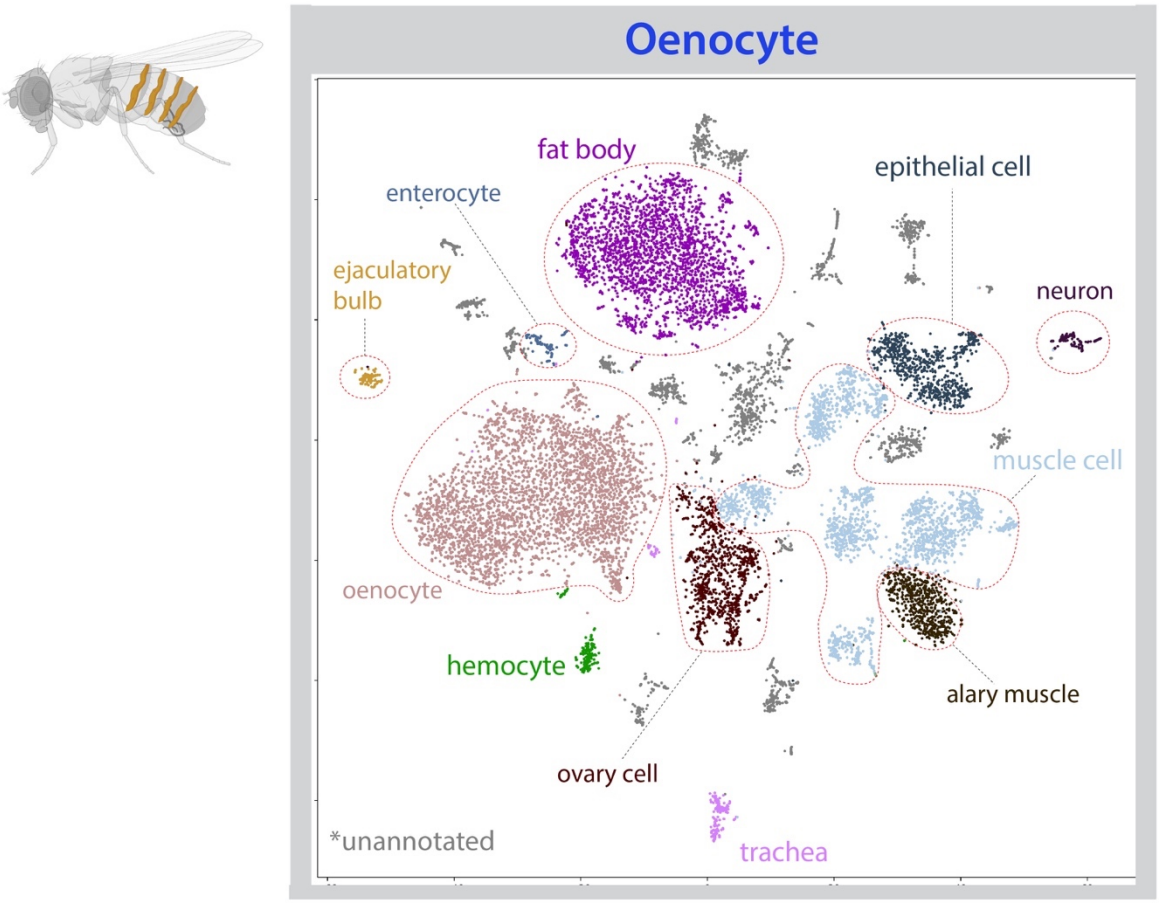


Figure S15. Cell type annotation in the oenocyte. tSNE plot with annotations for the fly oenocyte from the *Stringent* 10x dataset. Oenoctyes are FAC-sorted based on the nuclear GFP signal; flies are *PromE800-GAL4 > UAS-unc84GFP*.

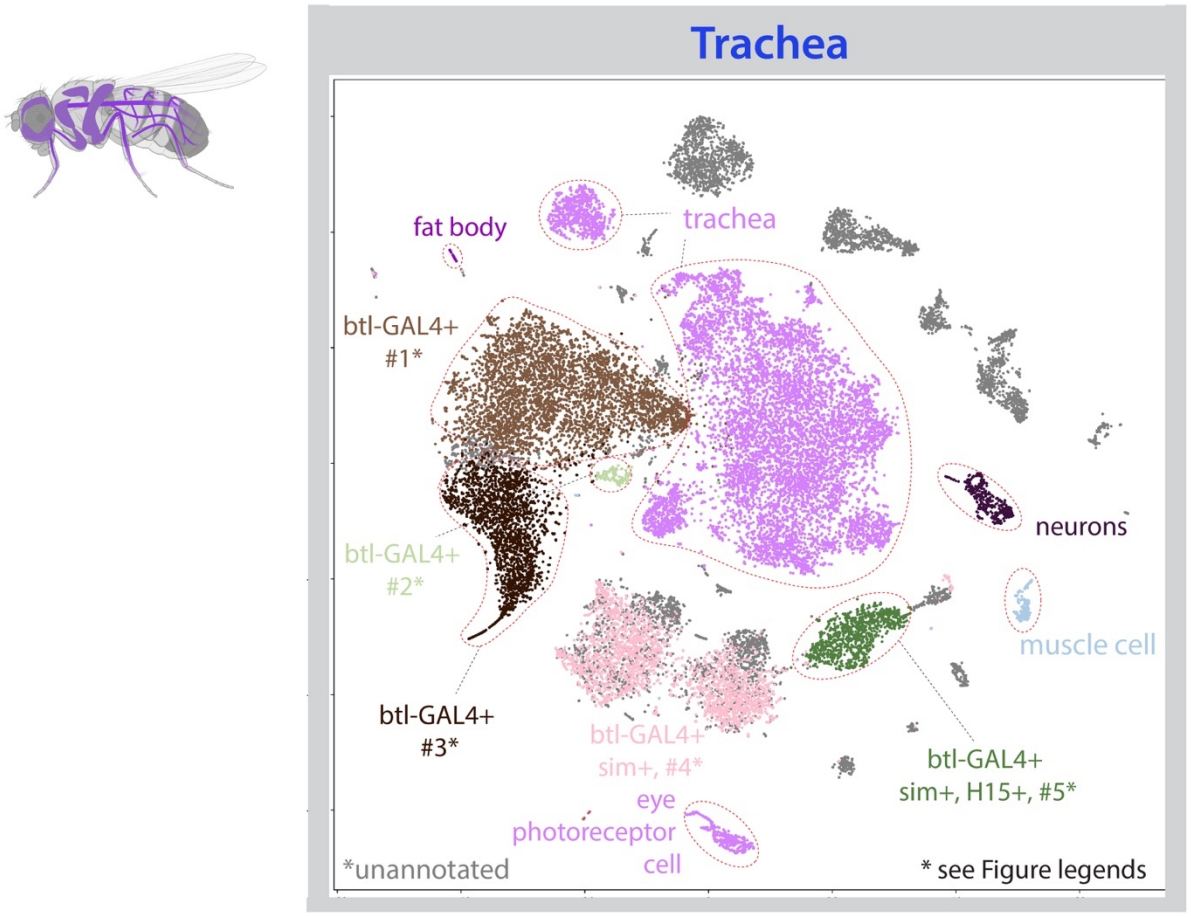


Figure S16.Cell type annotation in the trachea. tSNE plot with annotations for the fly trachea from the *Stringent* 10x dataset. Tracheal cells are FAC-sorted based on the nuclear GFP signal; flies are *btl-GAL4 > UAS-lamGFP*.

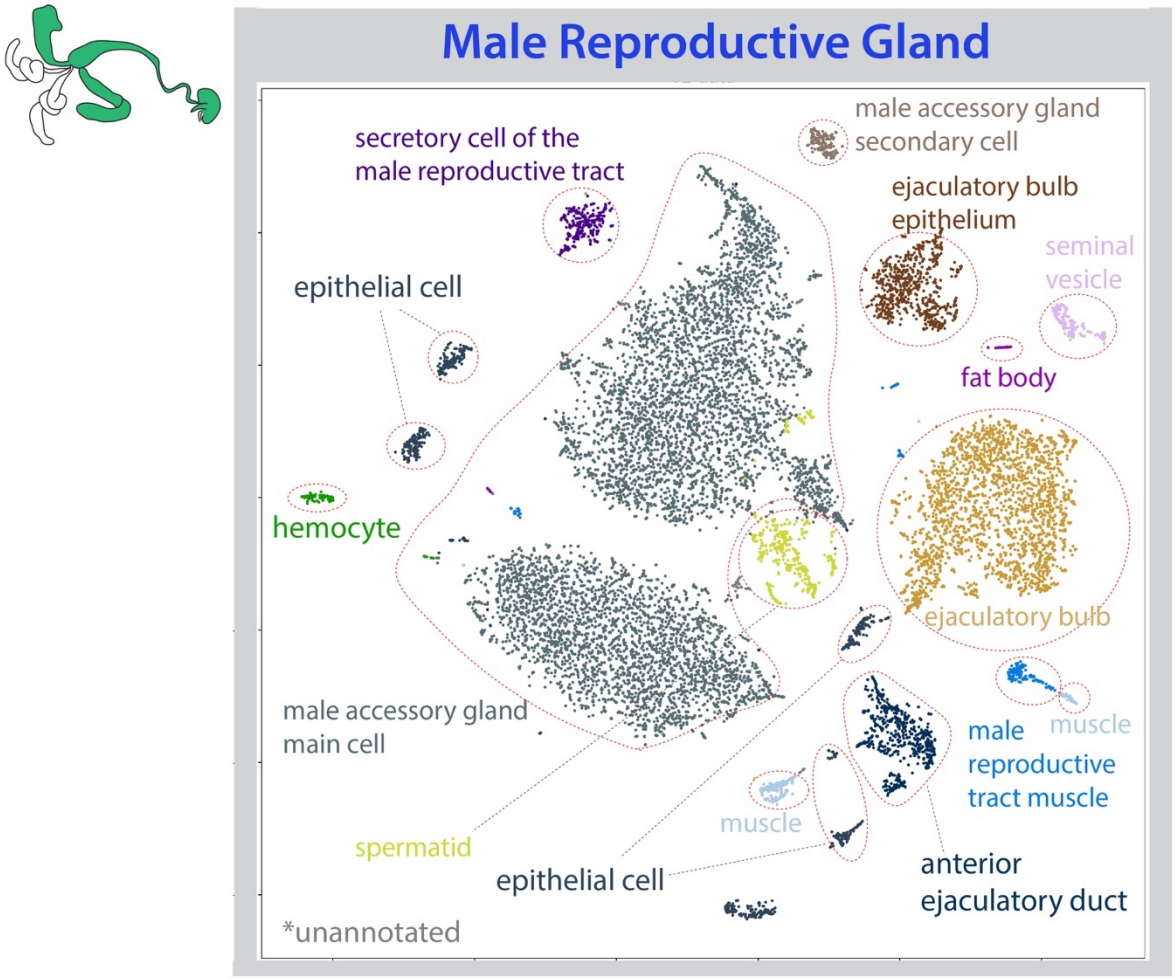


Figure S17. Cell type annotation in the male reproductive glands. tSNE plot with annotations for the fly male reproductive glands from the *Stringent* 10x dataset. The sequenced cells are from dissected male accessory glands, ejaculatory ducts and ejaculatory bulbs.

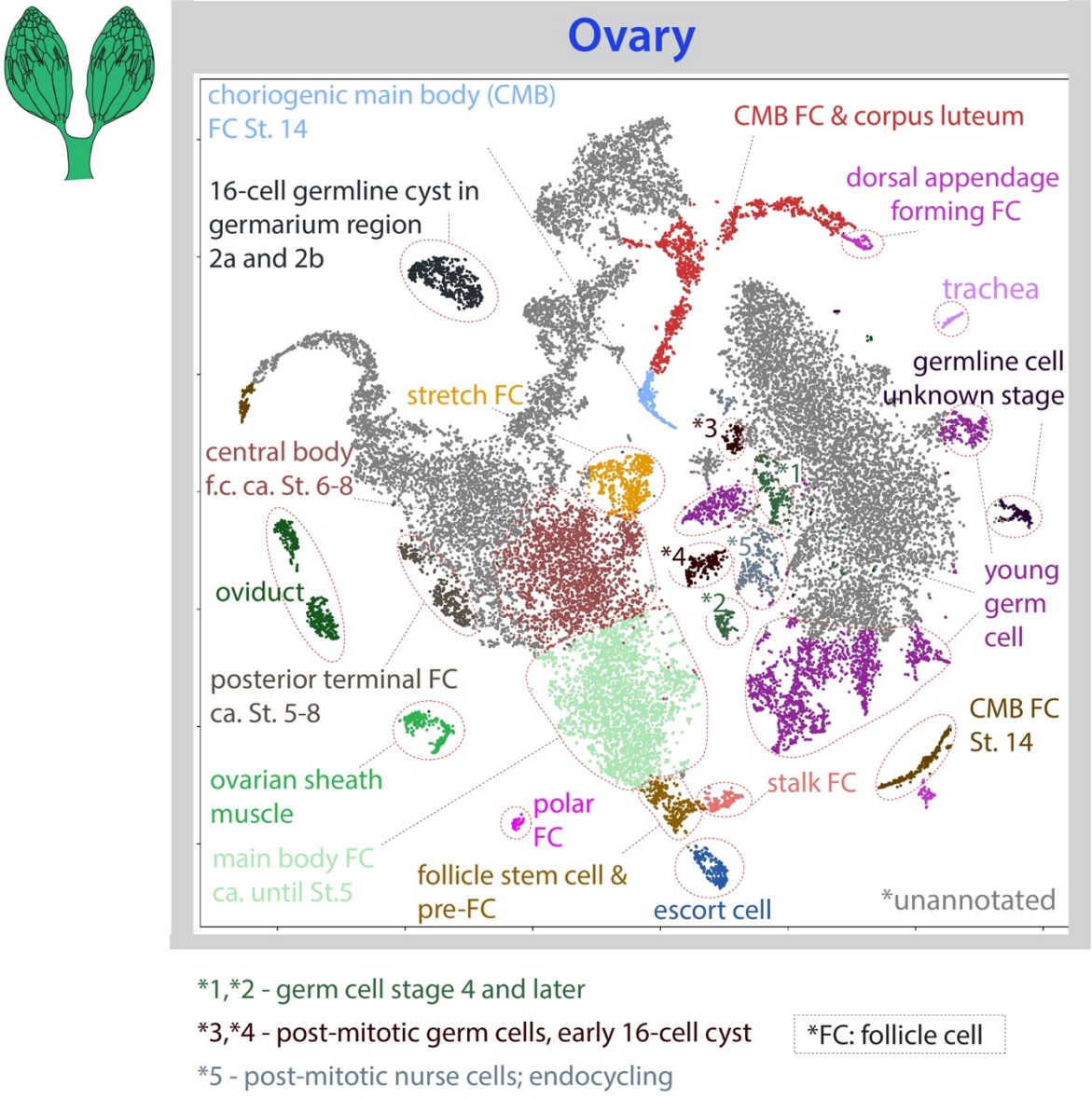


Figure S18. Cell type annotation in the ovary. tSNE plot with annotations for the fly ovary from the *Stringent* 10x dataset.

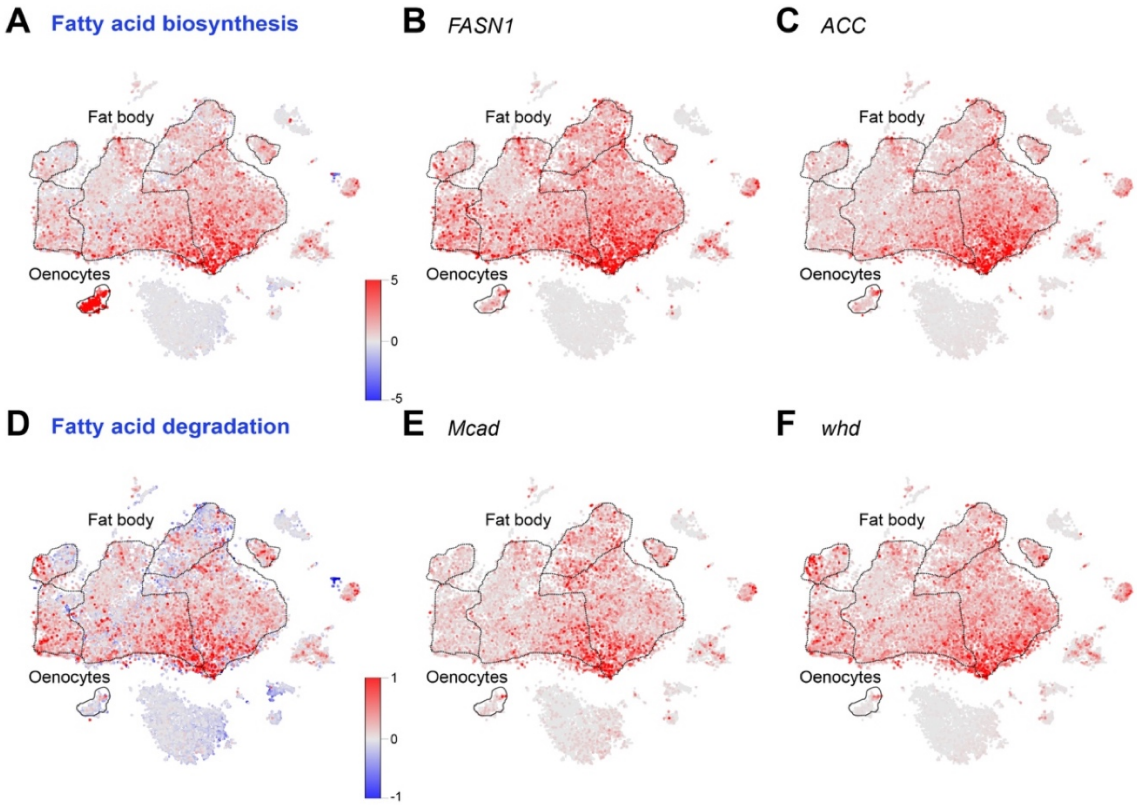


Figure S19. Metabolic pathway enrichment reveals the existence of cell subpopulations suggesting tissue specific functional specialization.
(A) Fatty acid biosynthesis pathway enrichment analysis as performed by ModuleScore (see Methods) in ASAP reveals strong homogeneous positive enrichment in oenocytes, while the fat body shows non-homogeneous enrichment across all cells. Red colors correspond to positive enrichment while blue colors correspond to negative enrichment.
(B, C) Similar profiles were obtained if using single genes in this pathway, *FASN1* and *ACC*.
(D) Fatty acid degradation pathway enrichment as revealed by ModuleScore analysis in ASAP shows low enrichment in oenocytes, while in the fat body there is a positive enrichment in specific subpopulations of cells. Red colors correspond to positive enrichment while blue colors correspond to negative enrichment.
(E, F) Similar profiles were obtained if using single genes in this pathway, *Mcad* and *whd.*

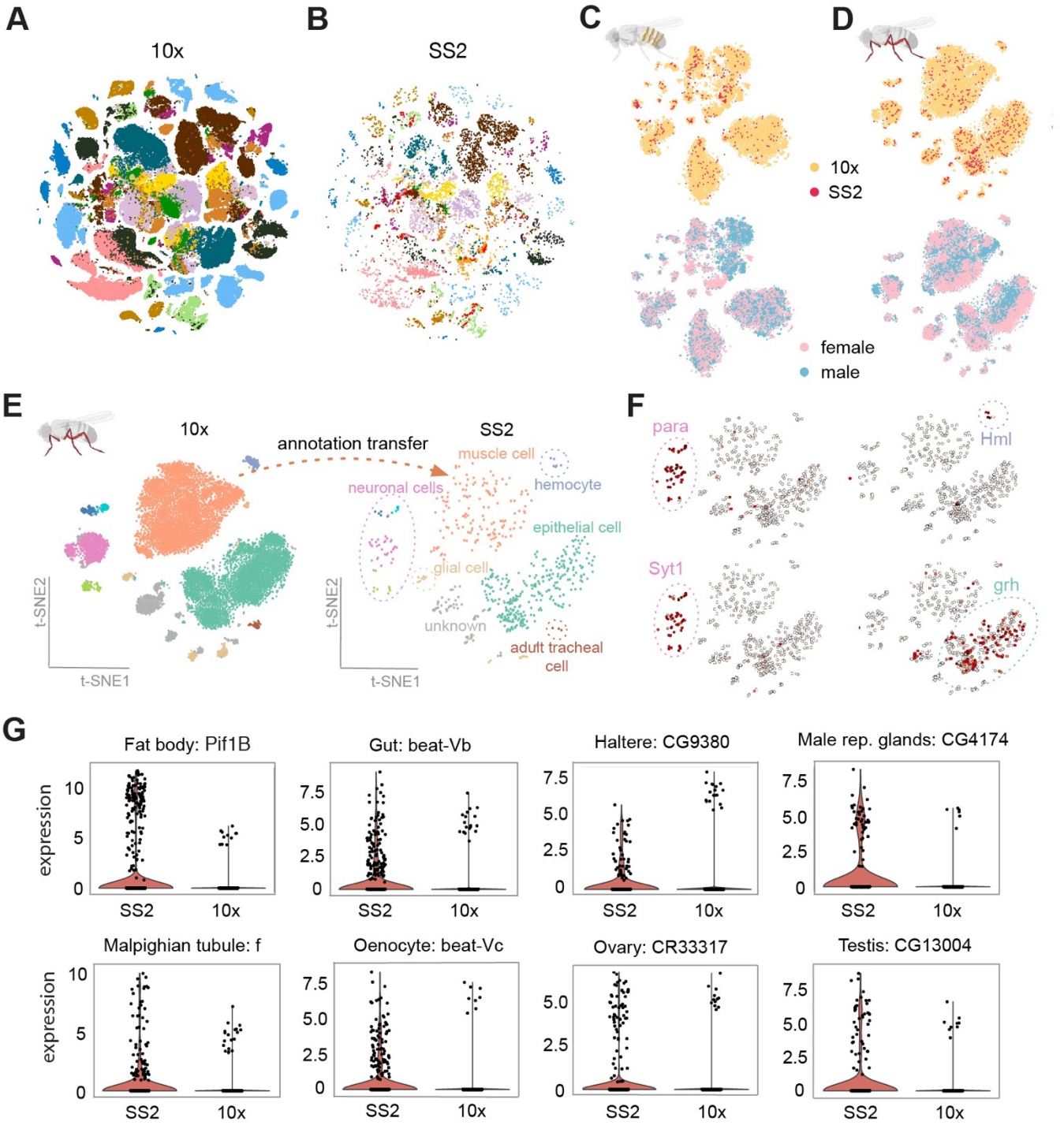


Figure S20. Integration of Smart-seq2 (SS2) and 10x Genomics data.
(A) tSNE visualization of individually dissected tissues using 10x Genomics and integrated with Smart-seq2 data. Colors denote different tissues.
(B) tSNE visualization of individually dissected tissues using Smart-seq2 and integrated with 10x Genomics data. Colors denote different tissues.

(C,D) Examples of integrated data for (C) oenocyte, and (D) leg. Cells are colored by technology (top) and by gender (bottom).
(E) Overview of computational pipeline for annotating Smart-seq2 data using leg as an example. After integrating 10x Genomics and Smart-seq2 data, we train a classifier on 10x Genomics data (left) and transfer annotations to Smart-seq2 data (right). Colors indicate different cell types.
(F) Validation of Smart-seq2 annotations by known marker genes. Cells annotated as neuronal cells correctly express *para* and *Syt1* neuronal markers, while cells annotated as hemocyte and epitelial correctly express *Hml* and *grh* markers, respectively.
(G) Examples of genes expressed in Smart-seq2 cells, but their expression is barely captured with 10x Genomics.

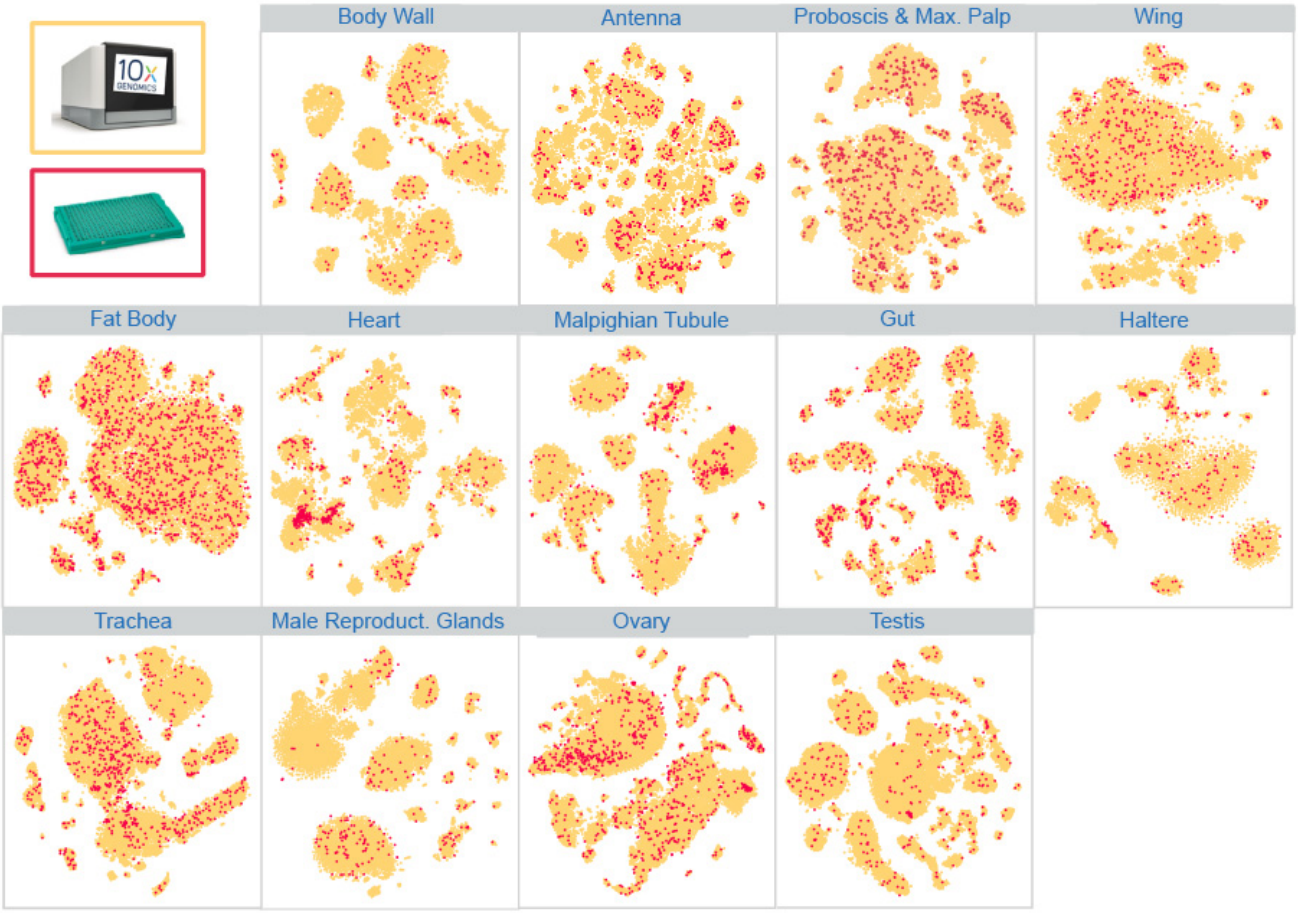
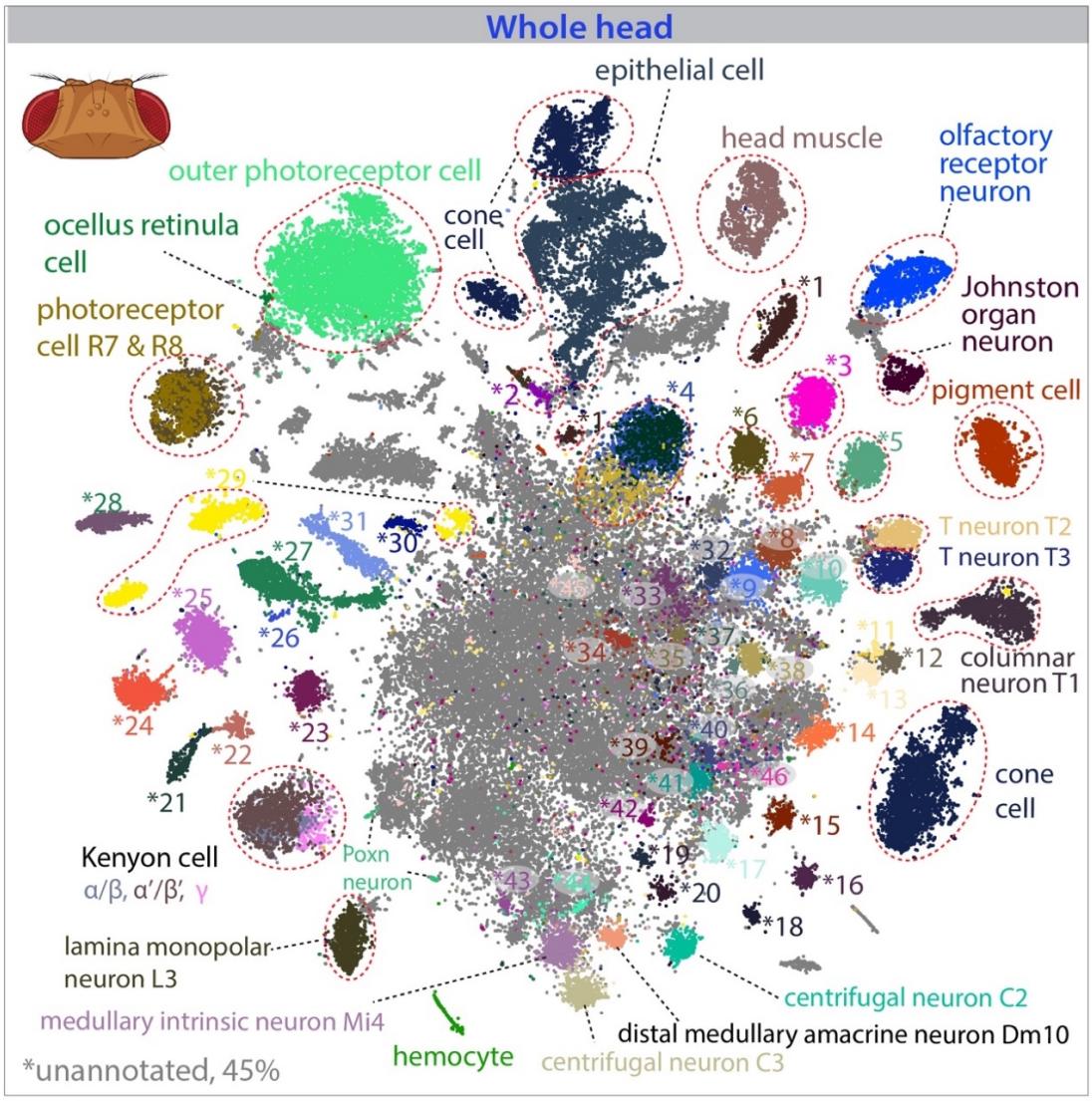


Figure S21. Tissue-level integration of 10x Genomics and Smart-seq2 datasets.
tSNE visualizations of 13 individually dissected tissues. Yellow color denotes cells from 10x Genomics and red color denotes cells from Smart-seq2. Remaining tissues (oenocyte and leg) are visualized in fig. **S17 C,D.**

FCA Figure S22

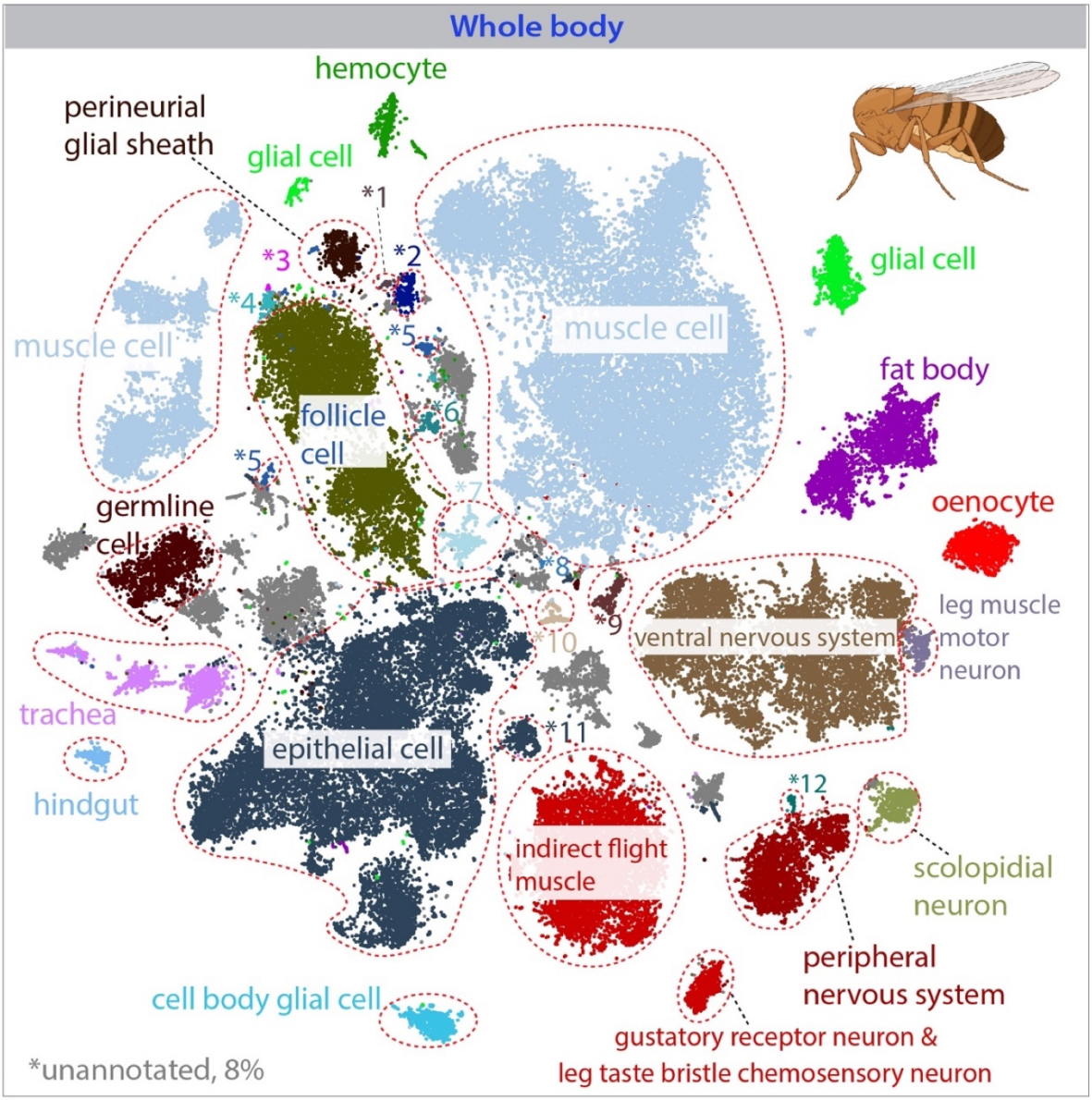


total 81 annotations

*1 adult lamina epithelial/marginal glial cell
*2 fat body
*3 medullary intrinsic neuron Mi1
*4 T neuron T4/T5 [see sub-types later]
*5 transmedullary neuron Tm1
*6 transmedullary neuron Tm2
*7 lamina monopolar neuron L5
*8 transmedullary neuron Tm4
*9 transmedullary neuron Tm3a
*10 T neuron T2a
*11 octopaminergic/tyraminergic neuron
*12 dopaminergic neuron
*13 medullary intrinsic neuron Mi15
*14 transmedullary Y neuron TmY8
*15 distal medullary amacrine neuron Dm8/Dm11
*16 lamina intrinsic amacrine neuron Lai
*17 transmedullary Y neuron TmY5a
*18 distal medullary amacrine neuron Dm9
*19 lamina wide-field 1 neuron
*20 lamina wide-field 2 neuron
*21 optic-lobe-associated cortex glial cell
*22 adult optic chiasma glial cell
*23 lamina monopolar neuron L4
*24 lamina monopolar neuron L1
*25 lamina monopolar neuron L2
*26 adult brain cell body glial cell
*27 ensheathing glial cell
*28 auditory sensory neuron
*29 adult brain perineurial glial cell
*30 subperineurial glial cell
*31 adult reticular neuropil associated glial cell
*32 transmedullary neuron Tm9
*33 transmedullary neuron Tm20
*34 lobula columnar neuron LC10
*35 transmedullary Y neuron TmY4
*36 transmedullary neuron Tm29
*37 lobula columnar neuron LC12
*38 lobula columnar neuron LC17
*39 distal medullary amacrine neuron Dm12
*40 distal medullary amacrine neuron Dm3
*41 medullary intrinsic neuron Mi9
*42 transmedullary Y neuron TmY14
*43 proximal medullary amacrine neuron Pm2
*44 proximal medullary amacrine neuron Pm4
*45 photoreceptor-like
*46 transmedullary neuron Tm5c

Figure S22. tSNE plot with annotations for the fly head from the *Stringent* 10x dataset. A large number of cells in the middle are unannotated cells, most of which are neurons from the central brain. The annotations not indicated in the plot are listed on the right.

FCA Figure S23

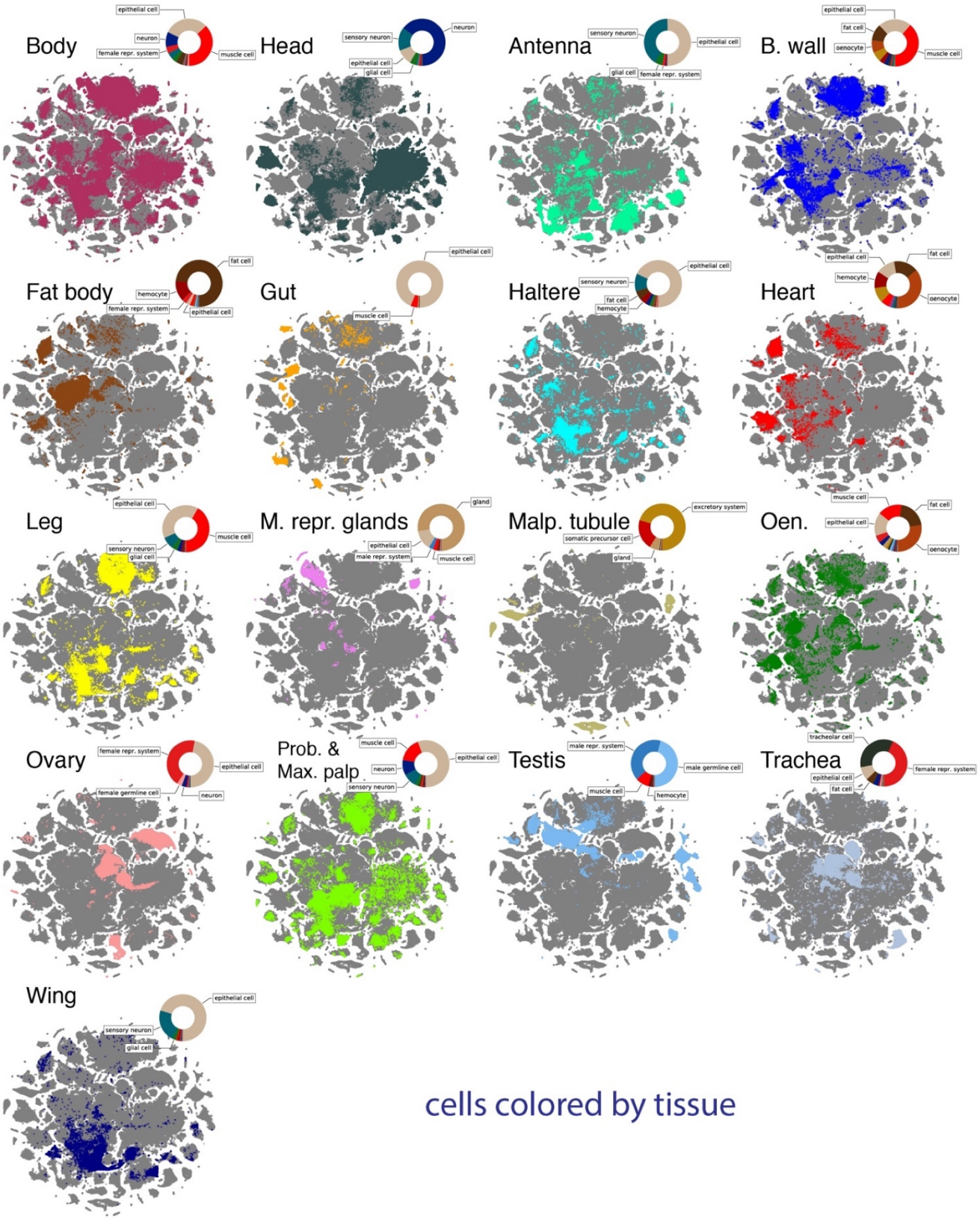


total 33 annotations

*1 CNS surface associated glial cell
*2 subperineurial glial cell
*3 polar follicle cell
*4 prefollicle cell and stalk follicle cell
*5 escort cell
*6 enteroendocrine cell
*7 eo support cell
*8 spermatocyte
*9 male accessory gland
*10 reticular neuropil associated glial cell
*11 female reproductive system
*12 multidendritic neuron

Figure S23. tSNE plot with annotations for the fly body from the *Stringent* 10x dataset. Note that only the top abundant cell types are annotated, and many of them can be further divided into different subtypes. The annotations not indicated in the plot are listed on the right.

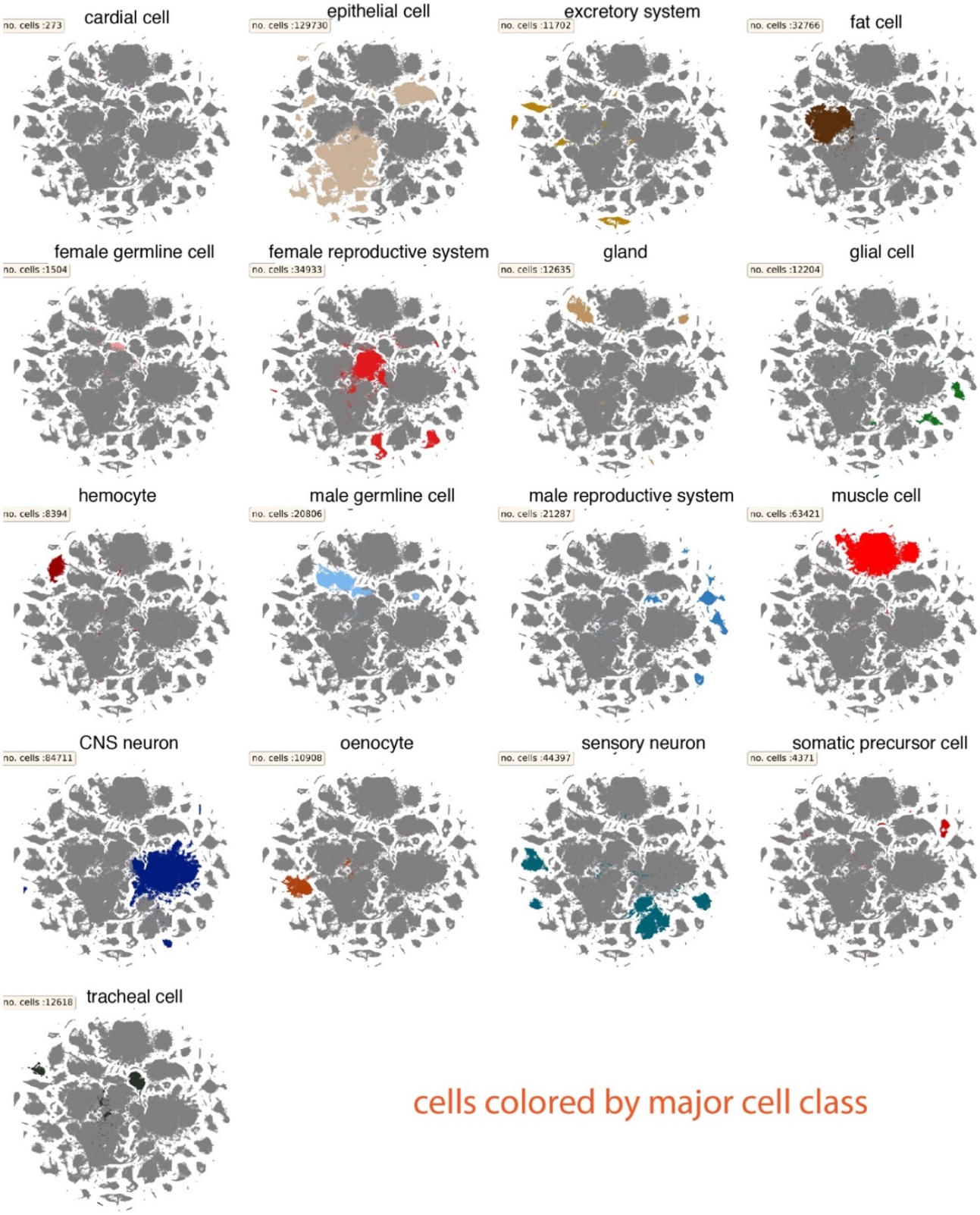
FCA Figure S24



cells colored by tissue

Figure S24. tSNE plots of all cells from the *Stringent* 10x dataset. Each tissue is highlighted in a different color. Pie charts show the top common cell types for each tissue, such as epithelial cells, neurons, and muscle cells, and so on. Note that some cells from two tissues overlap in one cluster, indicating these two tissues share one cell type. For example, the head and body share cell types, such as muscle, CNS neurons, and epithelial cells (see fig. S25). B. wall for body wall; M. repr. glands for male reproductive glands; Malp. tubule for Malpighian tubule; prob. max. palp for proboscis and maxillary palp.

FCA Figure S25



cells colored by major cell class

Figure S25. tSNE plots of all cells from the *Stringent* 10x dataset. Cell types from broad categories are highlighted and cell numbers are indicated.

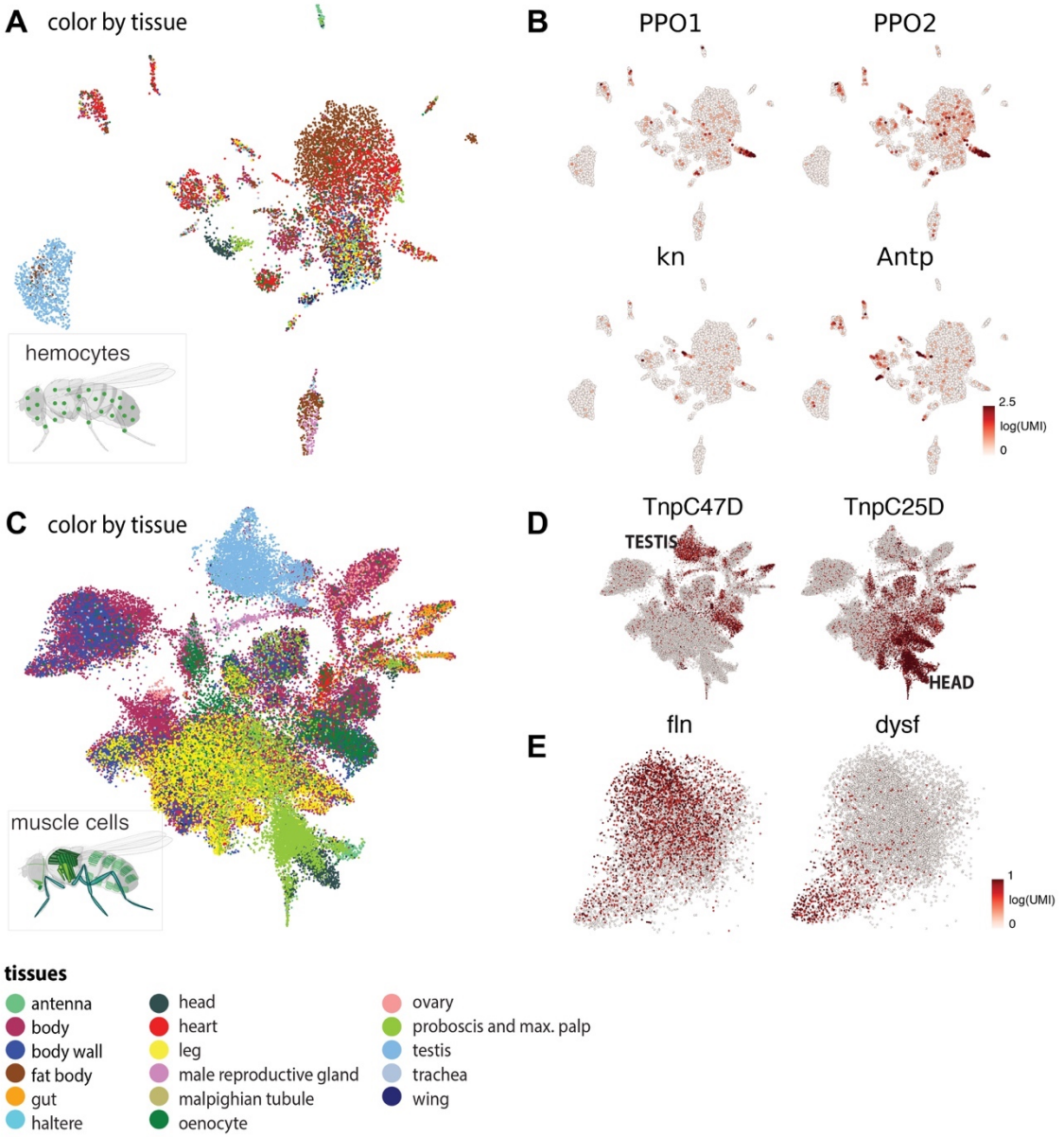
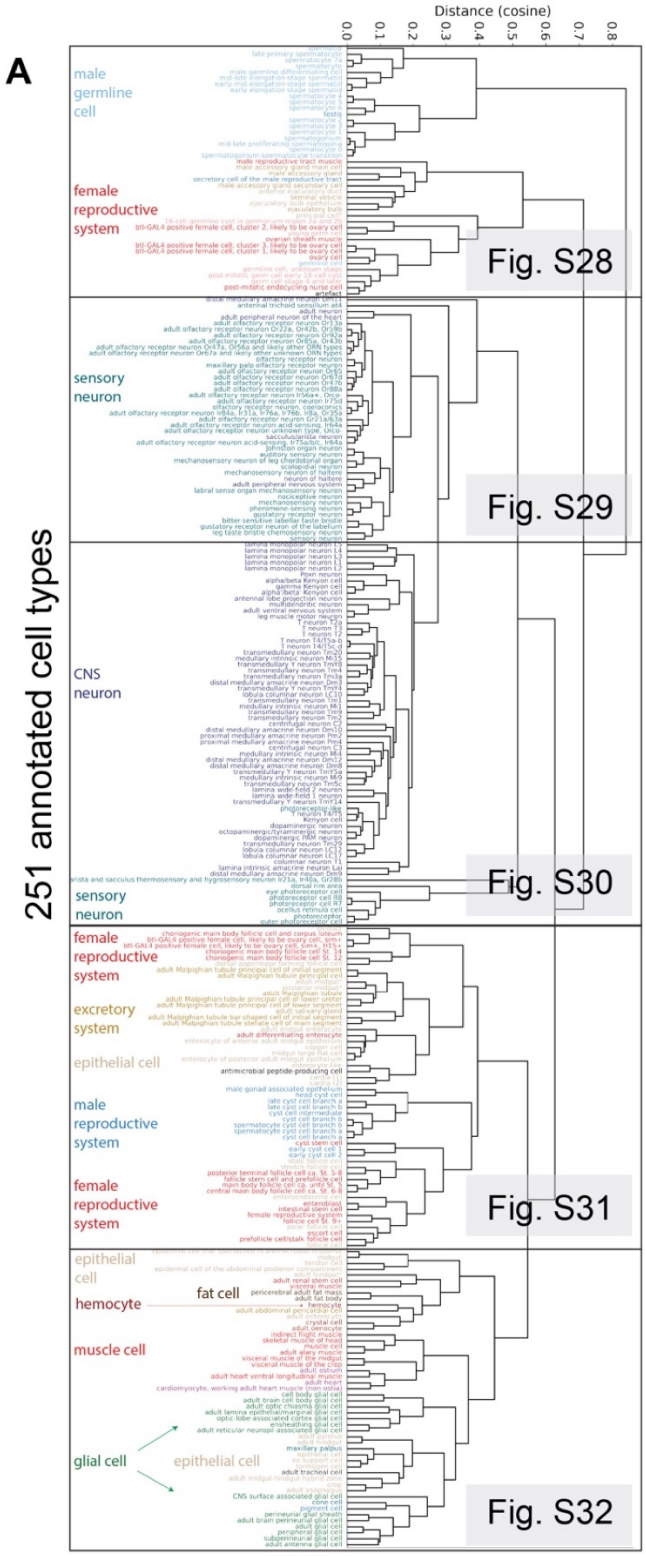
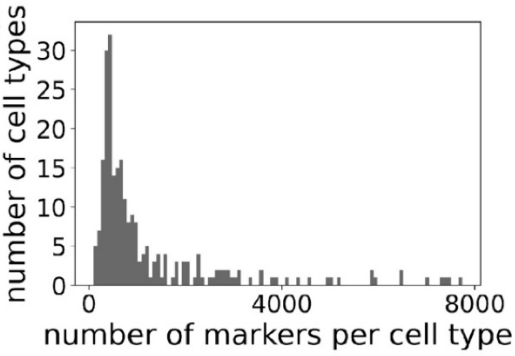
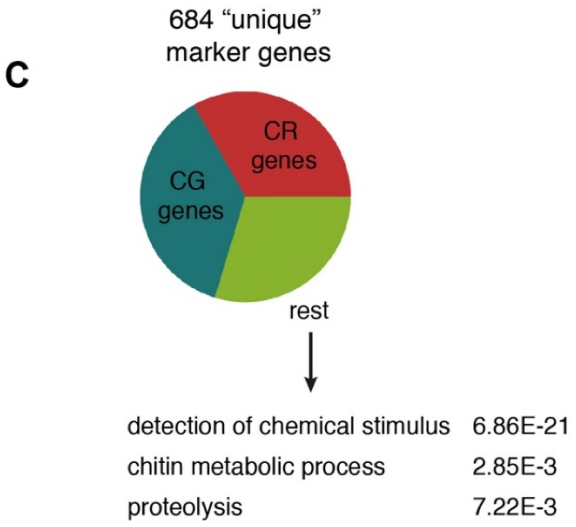


Figure S26. Cross-tissue analysis of hemocytes and muscle cells.
(A) tSNE plot of all annotated hemocytes colored by tissue types.
(B) Expression of *PPO1* and *PPO2* labeling crystal cells. Expression of *Antp* and *kn labeling* the presumptive lymph gland posterior signaling center.
(C) tSNE plot of all annotated muscle cells colored by tissue types.
(D) Expression of *TnpC47D* and and *TnpC25D* in all annotated muscle cells.
(E) Expression of *fln* and *dysf* showing gradients in the indirect flight muscle cluster.


A
Distance (cosine)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
251 annotated cell types
male germline cell
female reproductive system
Fig. S28
sensory neuron
Fig. S29
CNS neuron
sensory neuron
Fig. S30
female reproductive system
excretory system
epithelial cell
male reproductive system
female reproductive system
Fig. S31
epithelial cell
fat cell
hemocyte
muscle cell
glial cell
epithelial cell
Fig. S32


B
number of cell types
0
5
10
15
20
25
30
0
4000
8000
number of markers per cell type


684 “unique” marker genes
C
CR genes
CG genes
rest
detection of chemical stimulus 6.86E-21
chitin metabolic process 2.85E-3
proteolysis 7.22E-3
number of genes
0
2000
4000
6000
8000
10000
12000
14000
0
60
120
number of celltypes the gene is a marker in

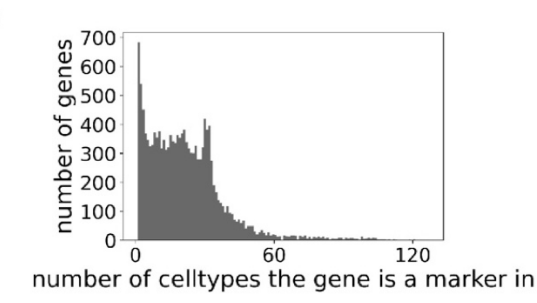

D
number of genes
0
100
200
300
400
500
600
700
0
60
120
number of celltypes the gene is a marker in

Figure S27. Organism-wide gene expression comparison
(A) Dendrogram showing the cosine similarity of the transcriptomes of the different annotated cell types. Cell types are colored based on broad classes. Enlarged details shown in fig. S28-S32.
(B) Histogram showing the number of markers calculated per cell type (avg. logfc>1, pval adj<0.05).
(C) Pieplot showing the 684 marker genes detected in only one cell type. Majority of unique marker genes are unknown CG and CR numbers, while the known marker genes are mostly linked to receptors. (pval adj shown as calculated by FlyMine). Insert shows a cumulative plot of the uniqueness of marker genes: 684 genes are markers in only one cell type, while almost all genes (~14k) can be found as markers in multiple cell types.
(D) Histogram showing the number of cell types a gene is expressed in on x-axis and number of genes on y-axis.

FCA Figure S28

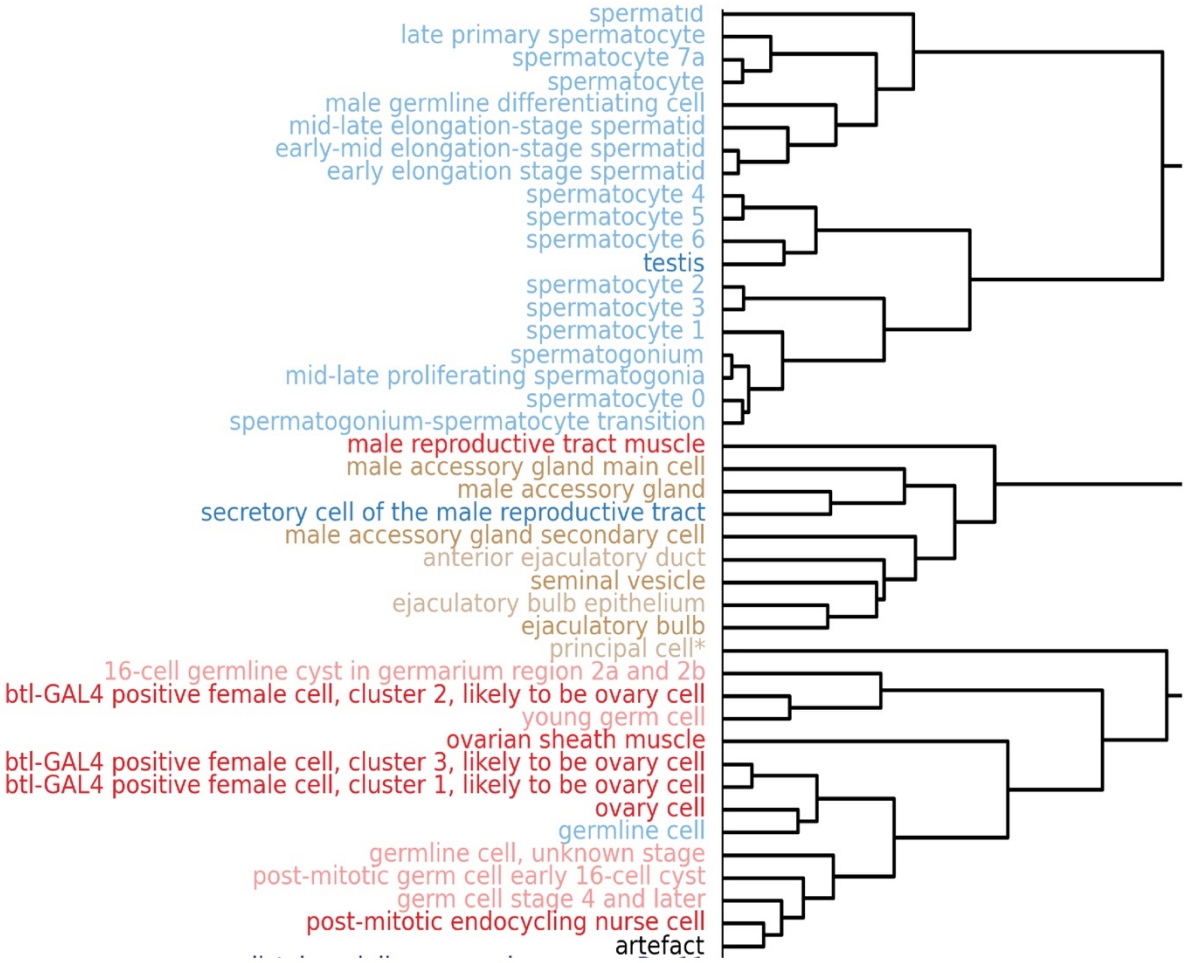


Figure S28. Dendrogram showing the cosine similarity of the transcriptomes of the different annotated cell types. Cell types are colored based on broad classes. Part 1 from fig. S27A.

FCA Figure S29

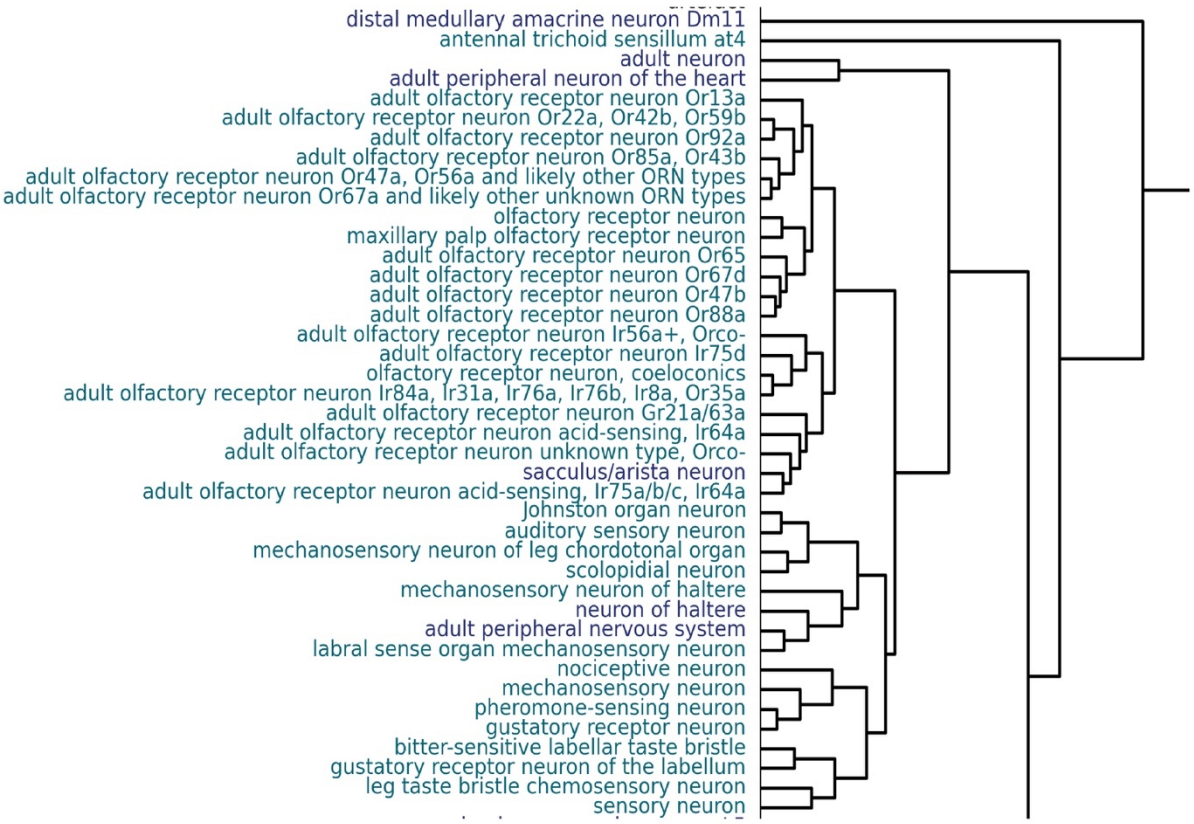


Figure S29. Dendrogram showing the cosine similarity of the transcriptomes of the different annotated cell types. Cell types are colored based on broad classes. Part 2 from fig. S27A.

FCA Figure S30

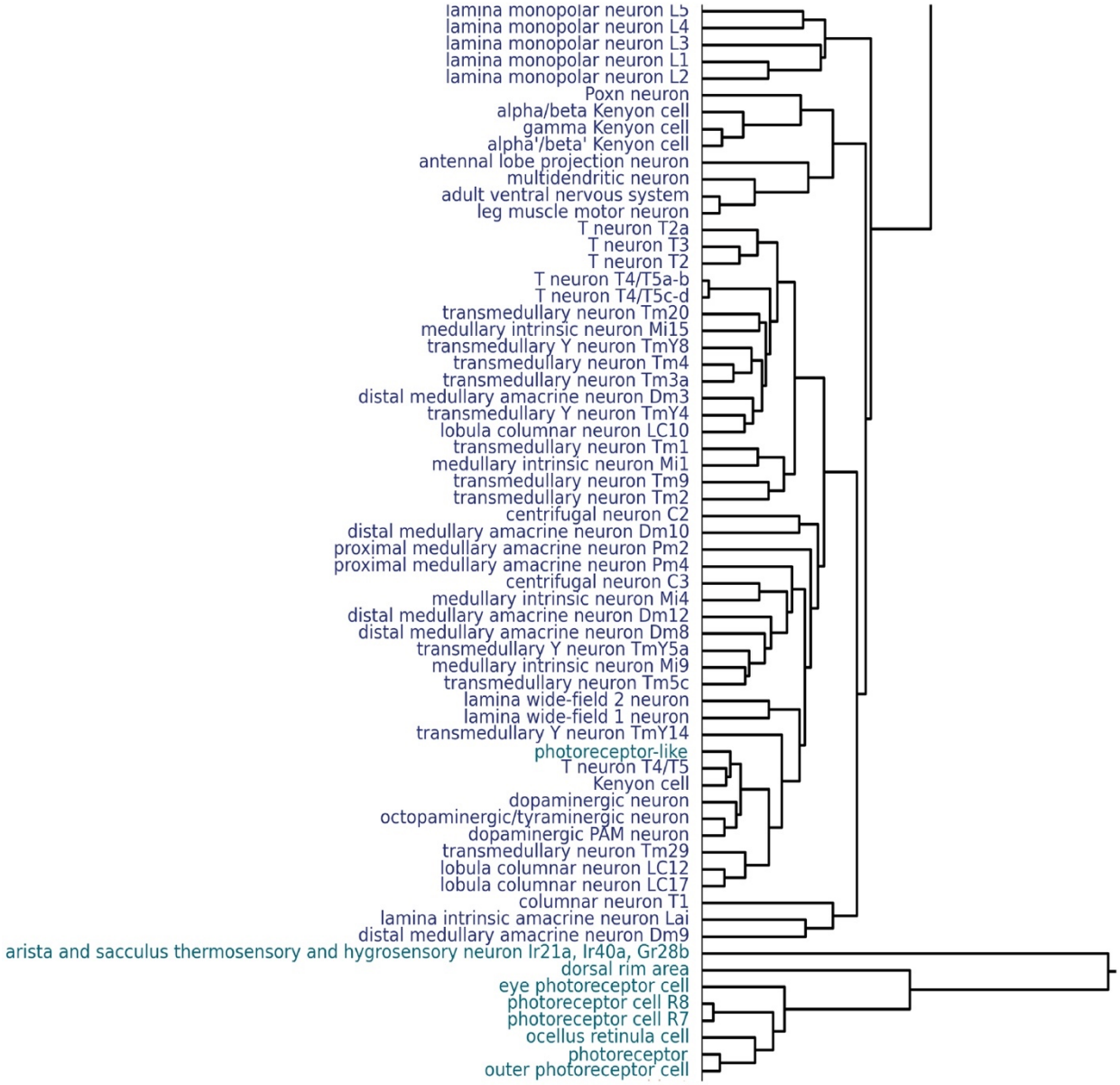


Figure S30. Dendrogram showing the cosine similarity of the transcriptomes of the different annotated cell types. Cell types are colored based on broad classes. Part 3 from fig. S27A.

FCA Figure S31

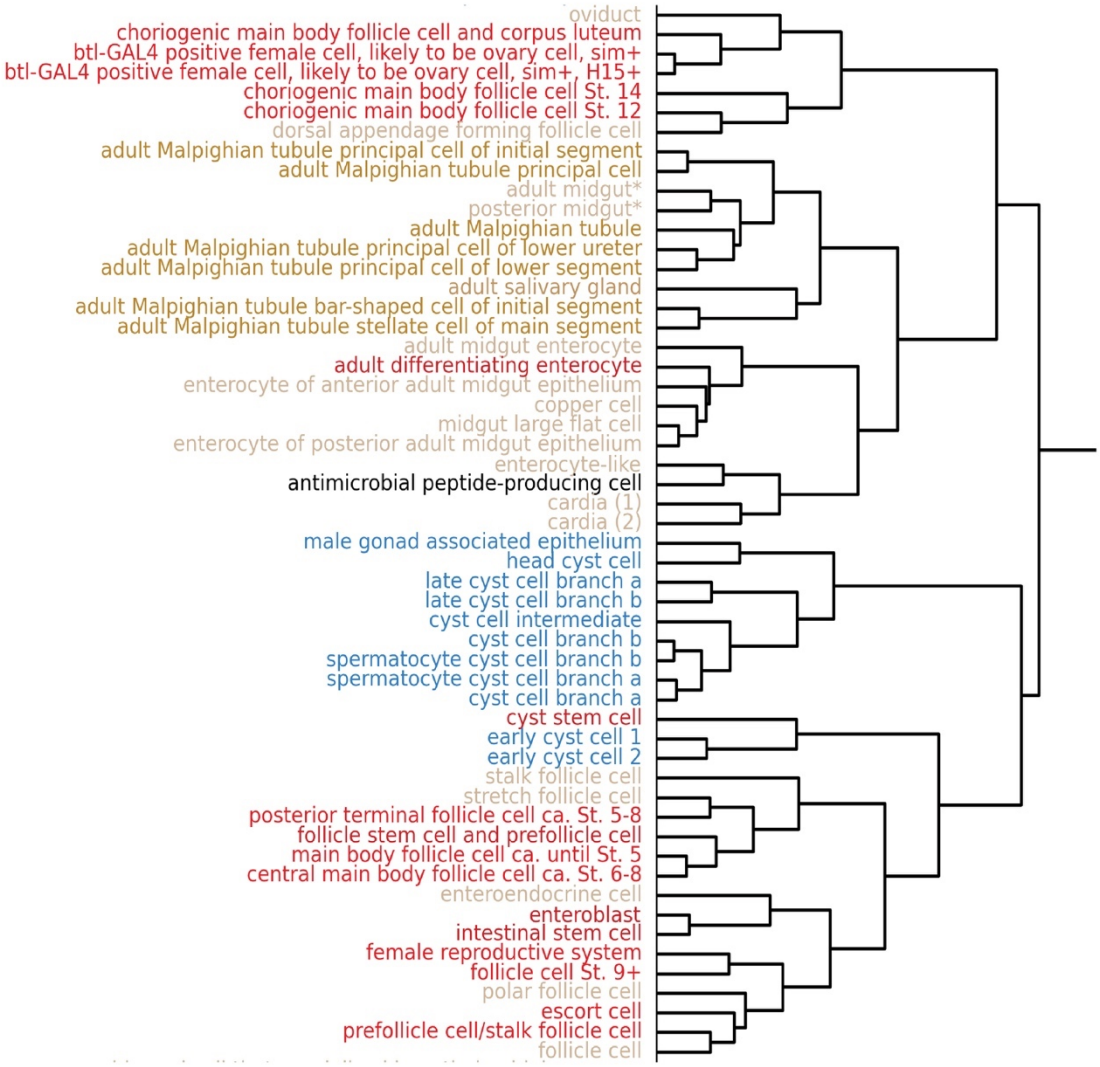


Figure S31. Dendrogram showing the cosine similarity of the transcriptomes of the different annotated cell types. Cell types are colored based on broad classes. Part 4 from fig. S27A.

FCA Figure S32

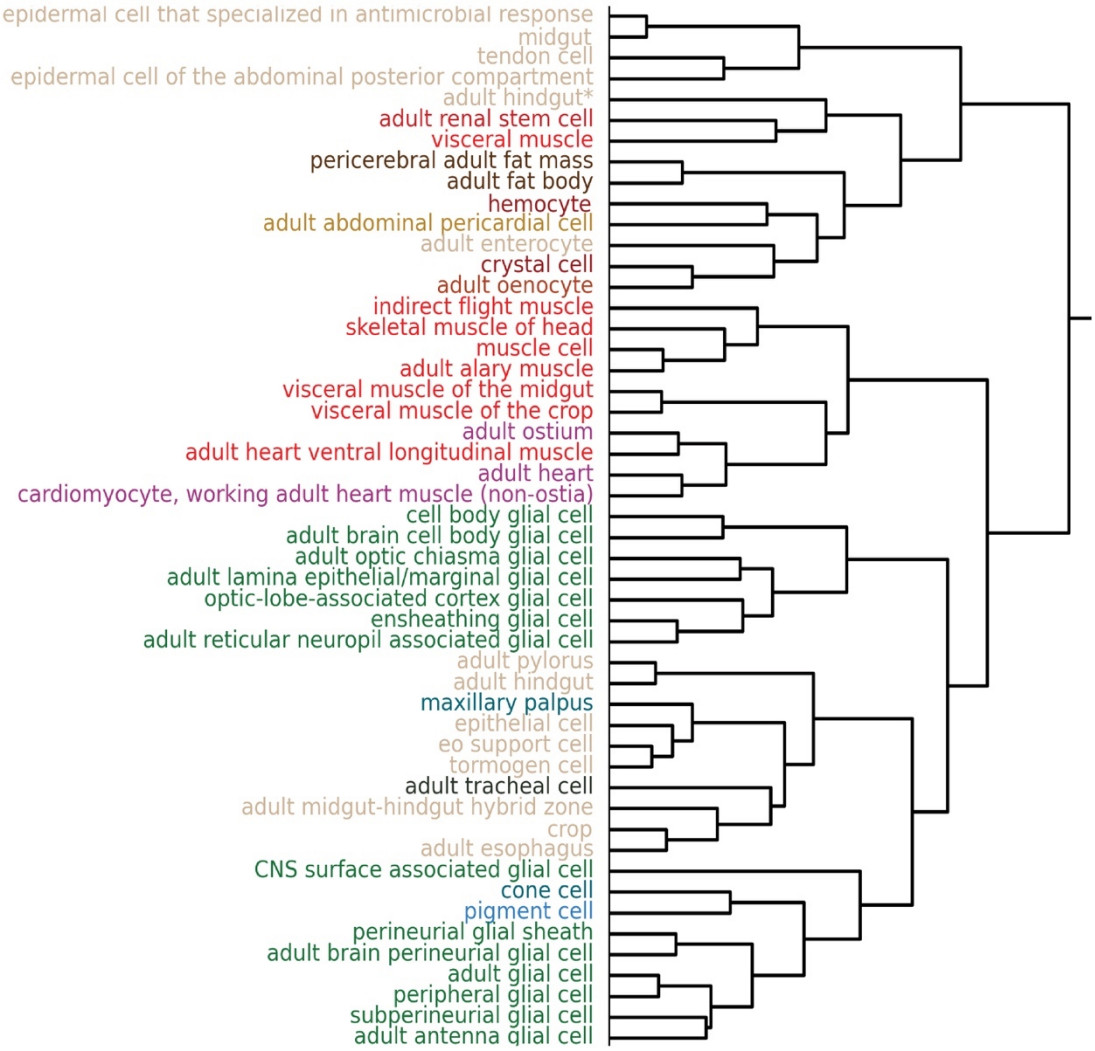


Figure S32. Dendrogram showing the cosine similarity of the transcriptomes of the different annotated cell types. Cell types are colored based on broad classes. Part 5 from fig. S27A.

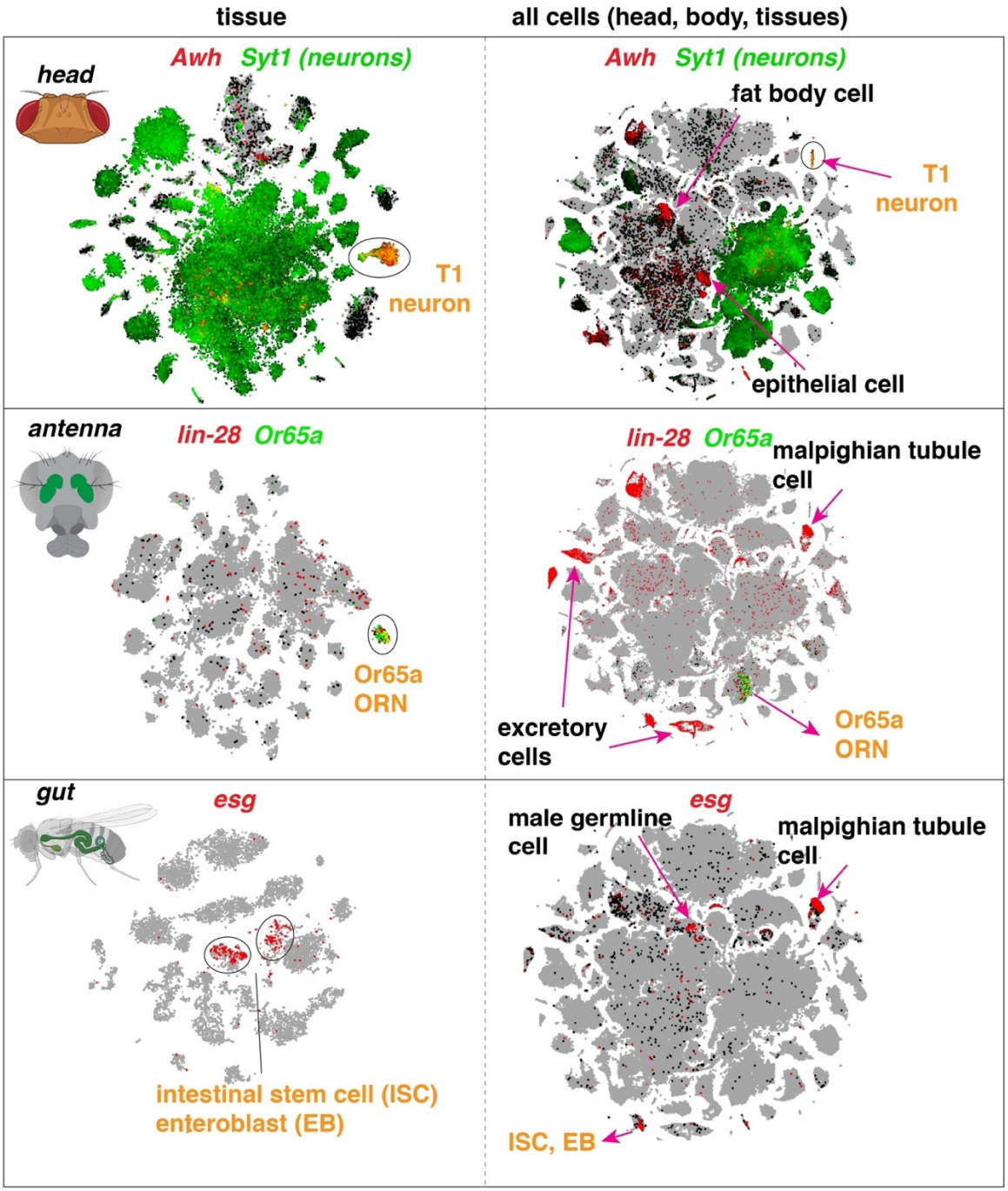


Figure S33. Cell-type specific markers in individual tissues and in all cells. Some cell type-specific markers identified in a specific tissue may have broader expression outside that tissue. *Awh* is specifically expressed in T1 neurons within the head, but also shows expression in fat body cells and epithelial cells. *lin-28* is specifically expressed in *Or65a* olfactory receptor neurons (ORNs) within the antenna, but also shows expression in malpighian tubule cells. *esg* is specifically expressed in intestinal stem cells (ISCs) and enteroblasts (EBs) within the gut, but also shows expression in Malpighian tubule stem cells and in the testis.

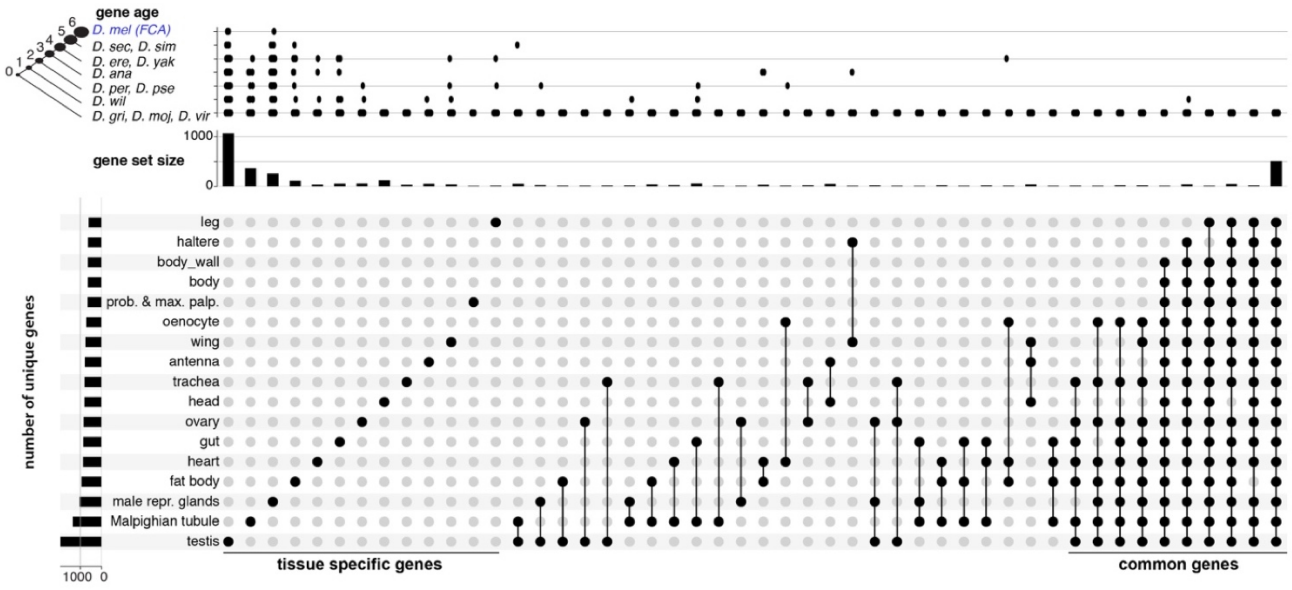


Figure S34. Common genes and tissue-specific genes shown by the UpSetPlot. Comparison of genes expressed per tissue (mean log2CPM>1) shows highly unique gene expression in the testis, Malpighian tubule, and male reproductive glands, while also highlighting a common module of conserved, ubiquitously expressed genes. 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).

FCA Figure S35

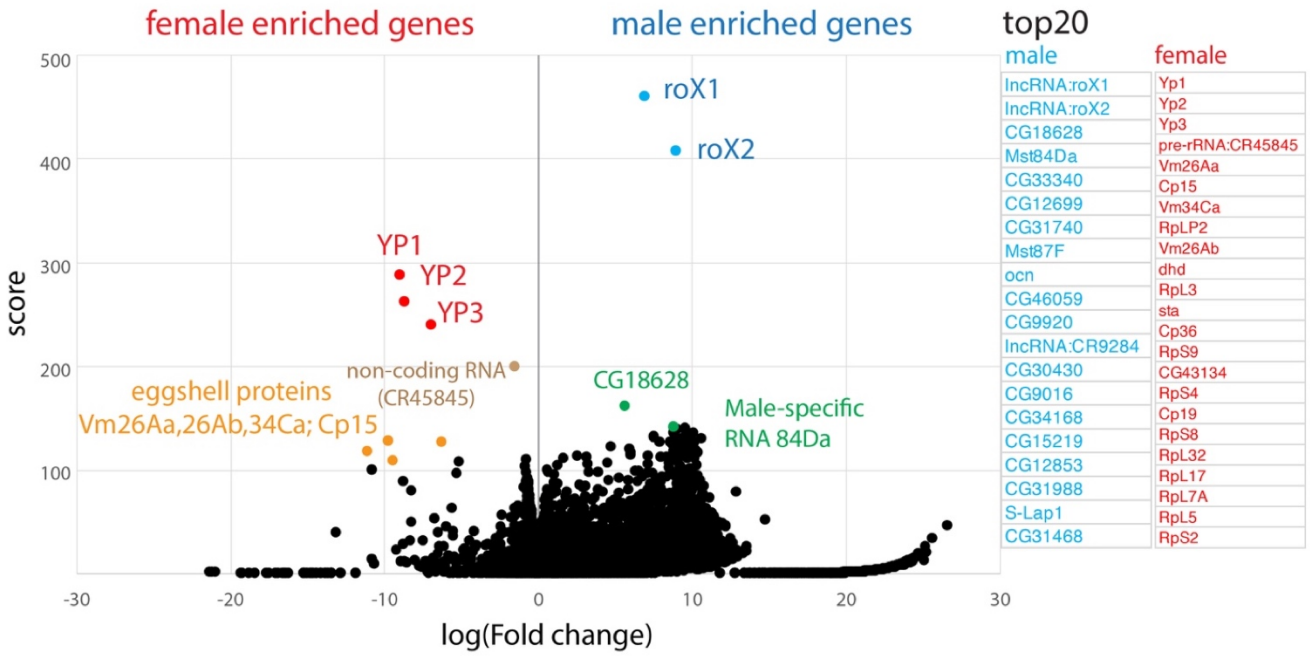


Figure S35. Volcano plot showing male and female enriched genes. Differential expression was performed between all male against all female cells in the dataset, using the Wilcoxon test in Scanpy. Score is the underlying z-value used to calculate the p-value. Fold change is in Log_2 scale. Male and female enriched genes with top scores (20) are shown on the right. Known male specific makers (*roX1, roX2*) and female specific genes (*Yp1, Yp2, Yp3*) have the highest scores as previously seen (*39*, *57*), validating the quality of the data. A large number of CG genes (poorly studied or uncharacterized genes) are on the male enriched list (*58*), suggesting their potential sex-related functions.

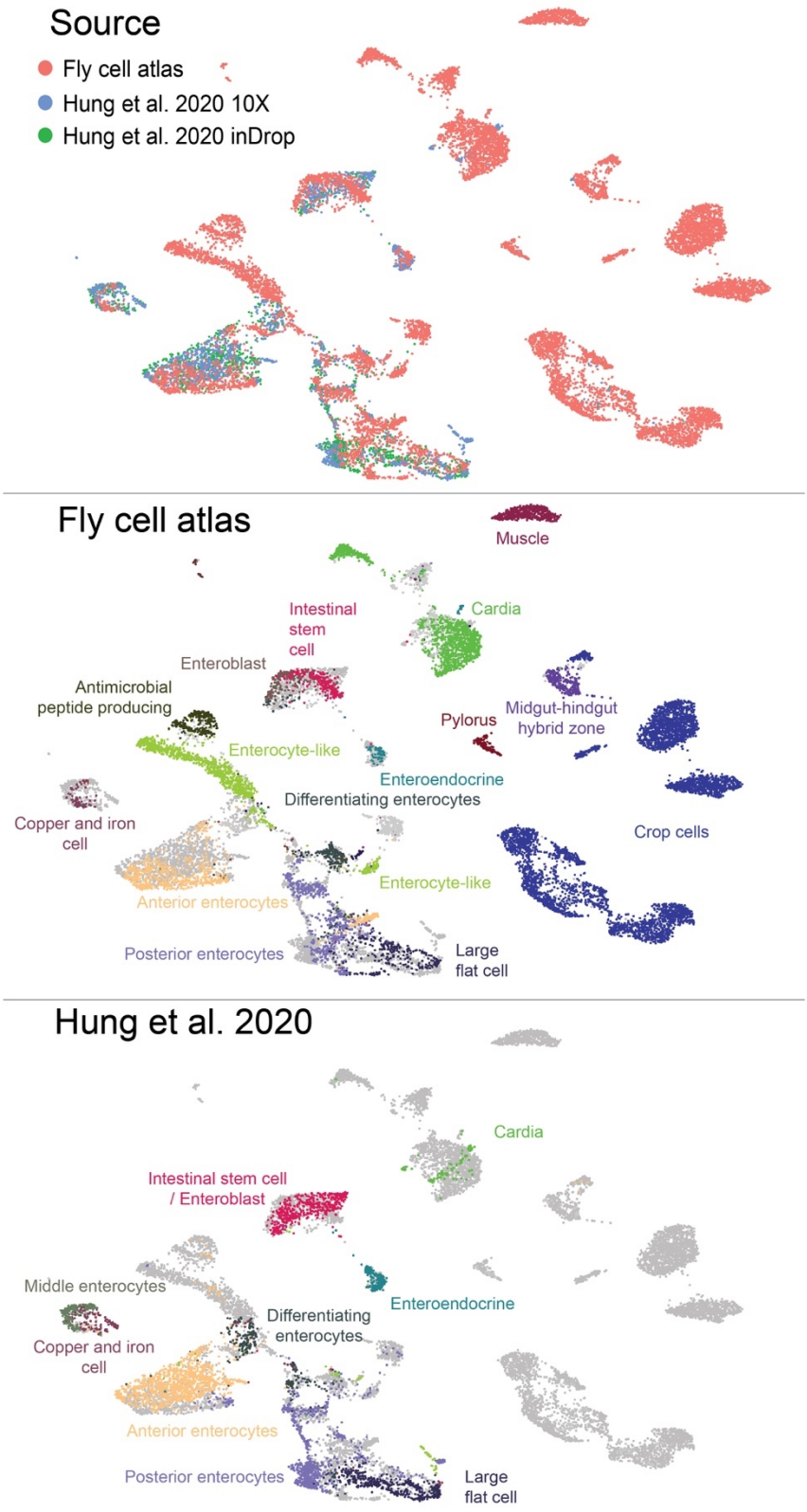

Source
Fly cell atlas
Hung et al. 2020 10X
Hung et al. 2020 inDrop
Fly cell atlas
Muscle
Intestinal stem cell
Cardia
Enteroblast
Antimicrobial peptide producing
Pylorus
Midgut-hindgut hybrid zone
Enterocyte-like
Enteroendocrine
Differentiating enterocytes
Copper and iron cell
Crop cells
Enterocyte-like
Anterior enterocytes
Posterior enterocytes
Large flat cell
Hung et al. 2020
Cardia
Intestinal stem cell / Enteroblast
Middle enterocytes
Enteroendocrine
Differentiating enterocytes
Copper and iron cell
Anterior enterocytes
Posterior enterocytes
Large flat cell

Figure S36. Integration of FCA snRNA-seq data and published scRNA-seq data of the gut. The published data are from two scRNA-seq platforms, 10x and inDrop (*40*). Data integration was performed using Harmony (*16*) using the first 30 PCA dimensions. From this analysis, we were able to identify all previously known cell types in the gut. In addition, we were able to characterize more cell types, including visceral muscle cells and 5 subtypes of crop cells.
Note that for Hung et al gut sample, the crop and midgut/hindgut junction (where the Malpighian tubules branch out of the gut) and Malpighian tubules were removed. For the FCA gut sample, we included the midgut/hindgut junction and the crop.

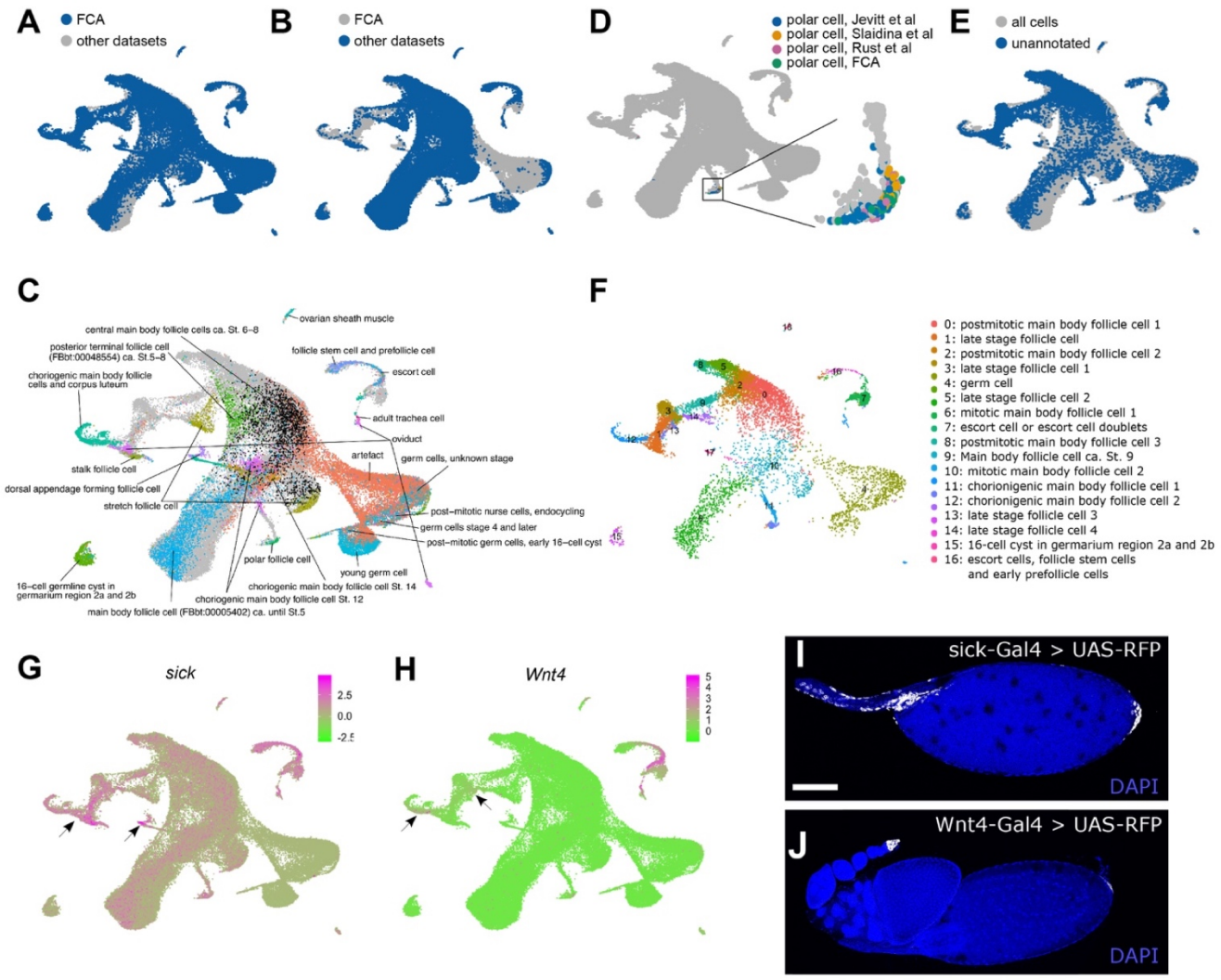


Figure S37. Integration of FCA snRNA-seq data and published scRNA-seq data of the ovary

(A) FCA cells are highlighted in blue, and other cells are colored in gray.

(B) Cells from the other three datasets are shown in blue, and FCA cells are displayed in gray.

(C) Annotated FCA clusters as noted. Unannotated cells and cells from other datasets are in gray.

(D) Polar cells identified in all datasets are highlighted and a magnified region of the UMAP plot containing polar cell clusters.

(E) Unannotated FCA cells are labeled blue, all other cells are shown in gray.

(F) Unannotated cells clustered independently. Presumptive cluster identities were determined by expression of marker genes as well as co-clustering with previously determined cell types.

(G, H) Expression of *sickie (sick)* and *Wnt4* labeling late stage terminal follicle cells indicated by arrows.

(I, J) Confocal images of *sick-GAL4* driving UAS-RFP showing expression in all late stage terminal follicle cells and of *Wnt4-GAL4* driving UAS-RFP showing expression in low levels in posterior terminal follicle cells and in high levels in escort cells. Confocal images are maximum intensity projections.

Confocal images are maximum intensity projections. Primary antibody, rat anti-RFP (ChromoTek 5F8, 1:1000); secondary antibody, goat anti-rat 555 (Thermo Fisher Scientific A-21434, 1:1000).
All plots are from UMAP. Three published adult ovarian scRNA-seq datasets are from (*41*, *42*, *55*). Datasets were integrated and batch corrected using Seurat v4.0.1. Scale bars in G and H depict average expression levels in log(((UMI + 1)/total UMI)×10^4). Scale bar in I and J, 100 μm.

References and Notes

1. T. H. Morgan, SEX LIMITED INHERITANCE IN DROSOPHILA. *Science*. **32**, 120–122 (1910).
2. 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).
3. 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).
4. 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).
5. 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).
6. 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).
7. 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).
8. 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).
9. 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).
10. H. Li, Single-cell RNA sequencing in Drosophila: Technologies and applications. *Wiley Interdiscip. Rev. Dev. Biol.* **10**, e396 (2021).
11. 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.
12. 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).
13. 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).
14. M. De Waegeneer, C. C. Flerin, K. Davie, G. Hulselmans, vib-singlecell-nf/vsn-pipelines: v0.26.0 (v0.26.0). Zenodo. https://doi.org/10.5281/zenodo.5055627. *Zenodo* (2021).
15. 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).
16. 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).
17. 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).
18. 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).
19. M. Costa, S. Reeve, G. Grumbling, D. Osumi-Sutherland, The Drosophila anatomy ontology. *J. Biomed. Semantics*. **4**, 32 (2013).
20. 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).
21. 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).
22. 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).
23. 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).
24. 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).
25. 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).
26. V. A. Pavlov, K. J. Tracey, The cholinergic anti-inflammatory pathway. *Brain Behav. Immun.* **19**, 493–499 (2005).
27. 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).
28. 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).
29. 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).
30. 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).
31. 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).
32. J. Mattila, V. Hietakangas, Regulation of Carbohydrate Energy Metabolism in Drosophila melanogaster. *Genetics*. **207**, 1231–1253 (2017).
33. 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).
34. H. Kaessmann, Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**, 1313–1326 (2010).
35. 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).
36. E. B. Lewis, A gene complex controlling segmentation in Drosophila. *Nature*. **276**, 565–570 (1978).
37. 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).
38. H. K. Salz, J. W. Erickson, Sex determination in Drosophila: The view from the top. *Fly (Austin)*. **4**, 60–70 (2010).
39. 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).
40. 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).
41. 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).
42. 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).
43. 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).

44. 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).
45. 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.
46. 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).
47. J. Janssens, S. Aibar, I. I. Taskiran, J. N. Ismail, A. E. Gomez, G. Aughey, K. I. Spanier, F. V. De Rop, C. B. González-Blas, M. Dionne, K. Grimes, X. J. Quan, D. Papasokrati, G. Hulselmans, S. Makhzami, M. De Waegeneer, V. Christiaens, T. Southall, S. Aerts, Decoding gene regulation in the fly brain. *Nature* (2022), doi:10.1038/s41586-021-04262-z.
48. 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).
49. 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).
50. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
51. 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).
52. 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).
53. 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).
54. 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).
55. 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).
56. 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).
57. J. C. Lucchesi, M. I. Kuroda, Dosage compensation in Drosophila. *Cold Spring Harb. Perspect. Biol.* **7** (2015), doi:10.1101/cshperspect.a019398.
58. 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).