

Supplementary figures

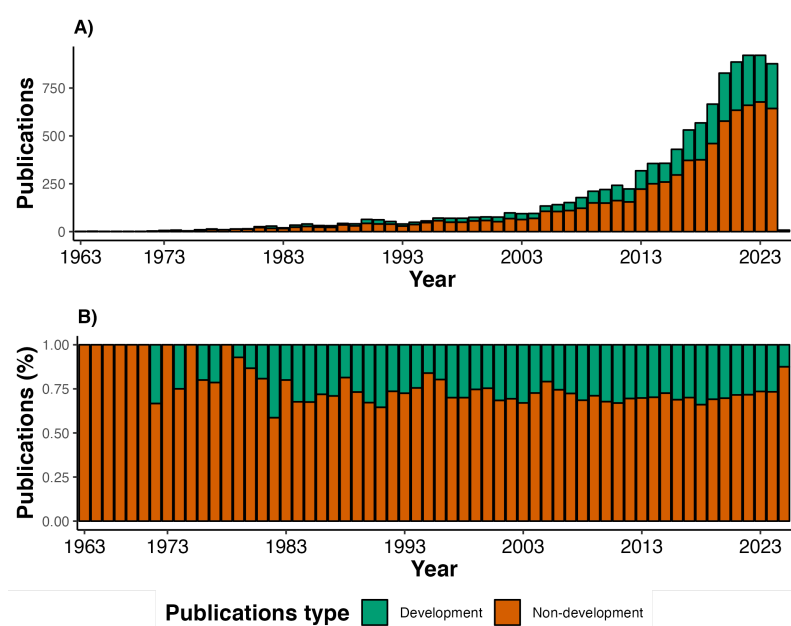


Fig. S1: PubMed results including keywords for sex differences and brain development. Results over the years for PubMed search: "sex differences" AND "brain development" as of late 2024. While the number of publications investigating brain development and sex differences has increased, it is still the minority of publications.

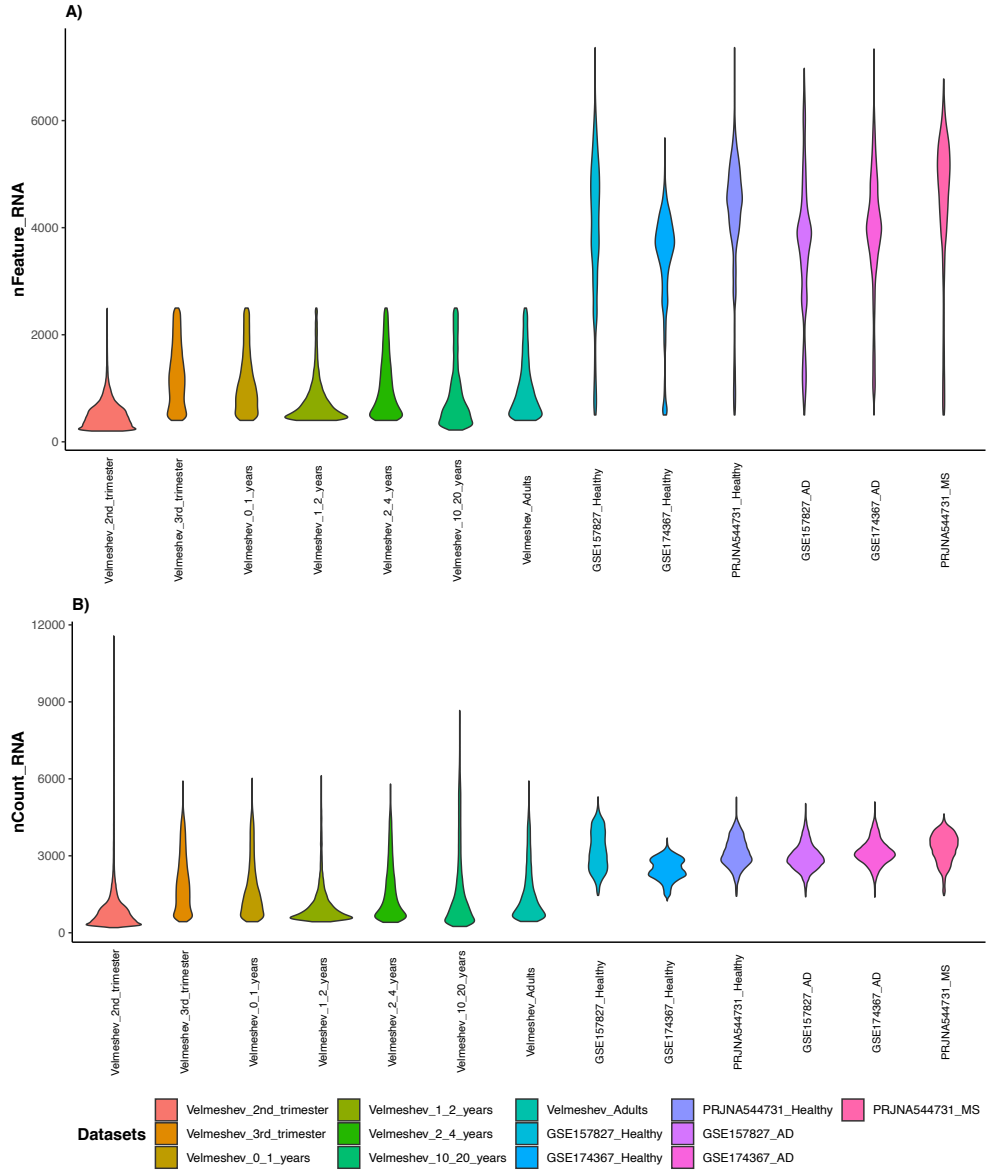


Fig. S2: Summary figures of gene coverage parameters in the data sources.
A) Genes detected per cells (nFeature.RNA) and B) nCount_RNA distributions in the data sources.

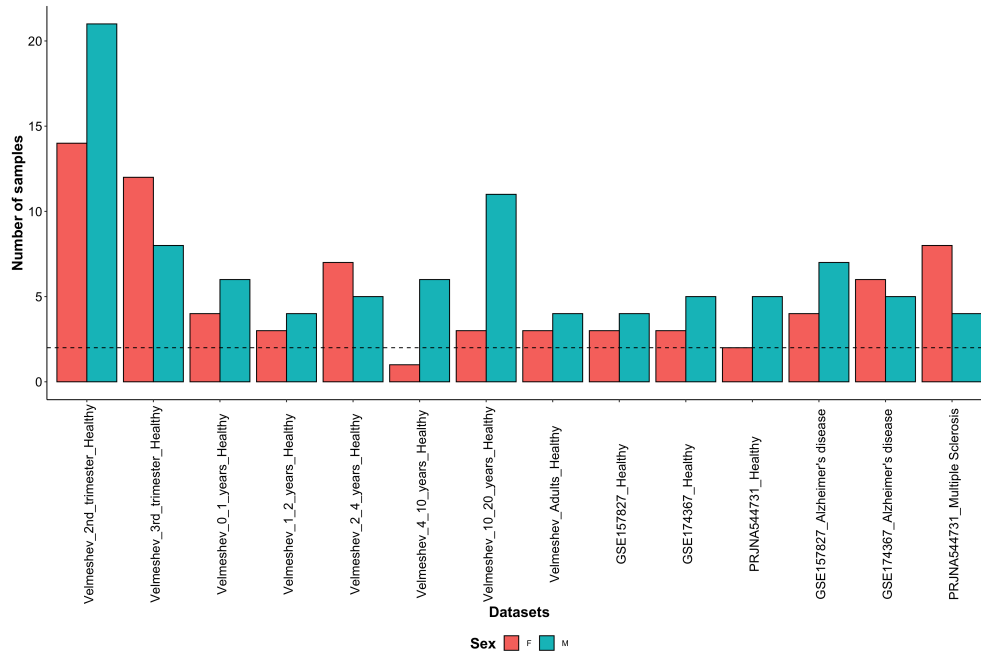


Fig. S3: Number of samples in each dataset analyzed, separated by sex. All datasets but one - Velmeshev 4-10 years - had at least $n=2$ (dashed line) of both female and male samples.

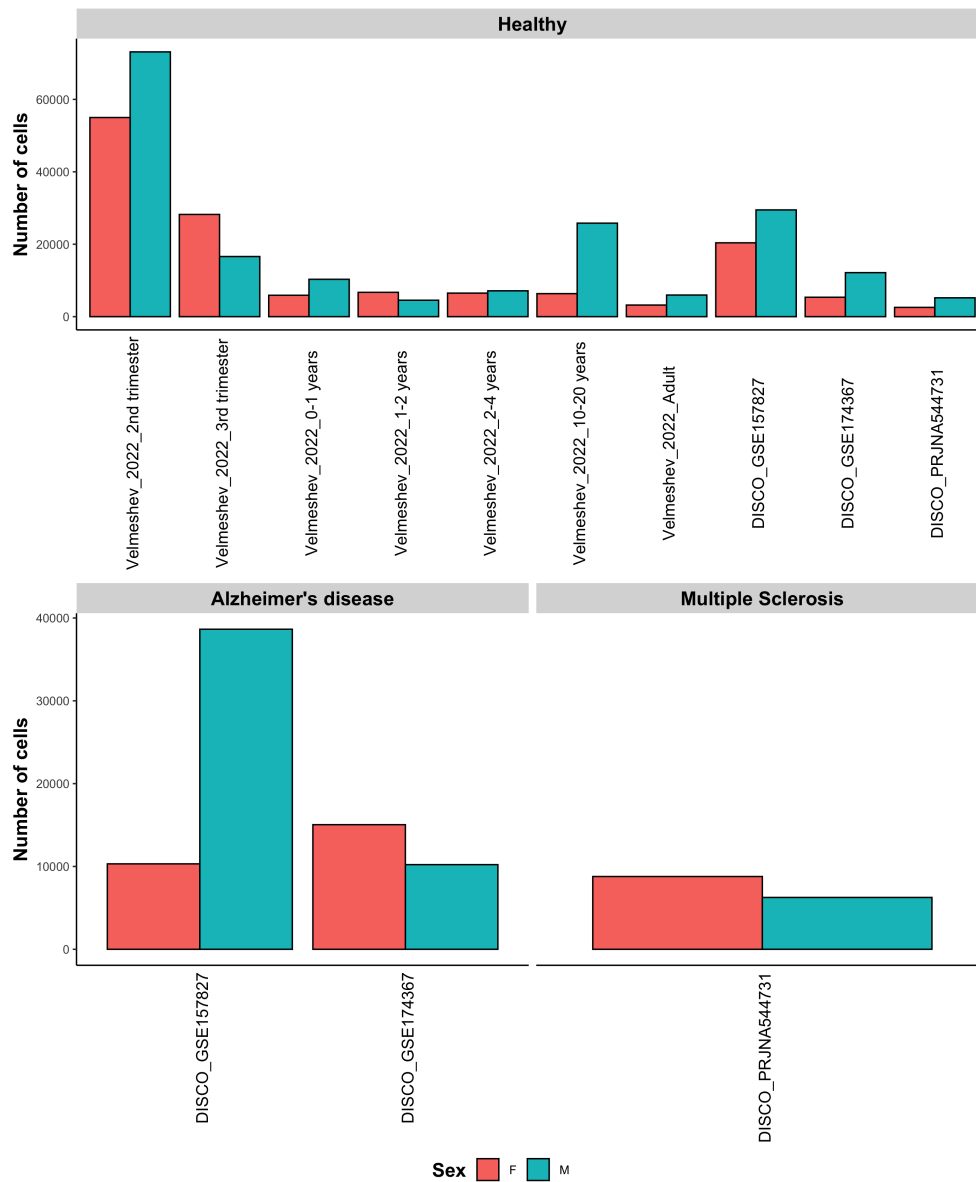


Fig. S4: Number of cells in each dataset analyzed, separated by sex.

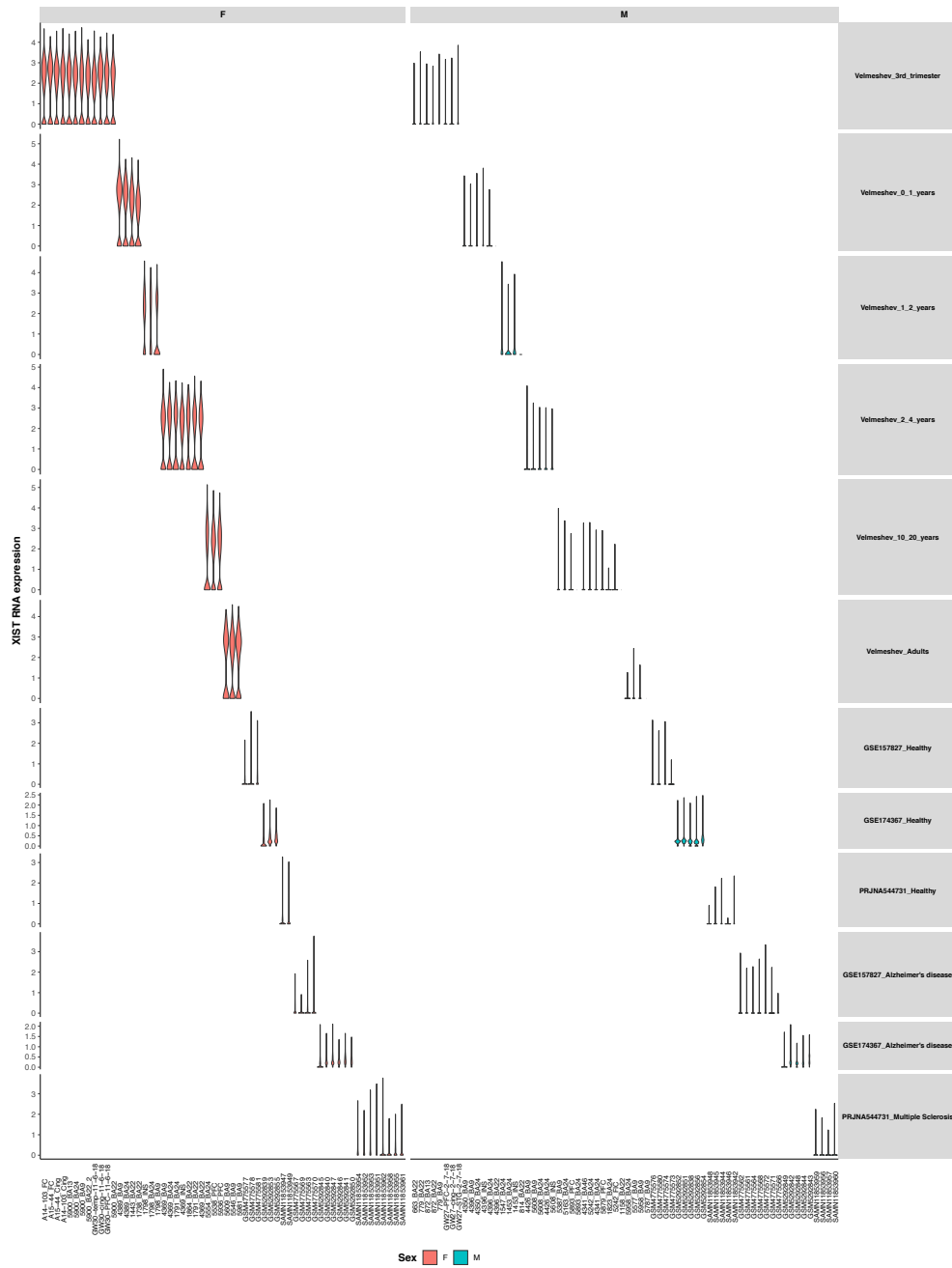


Fig. S5: XIST expression in the individual samples analyzed, organized by sex and dataset. Overall, all females samples showed a higher expression of XIST compared to males, as expected.

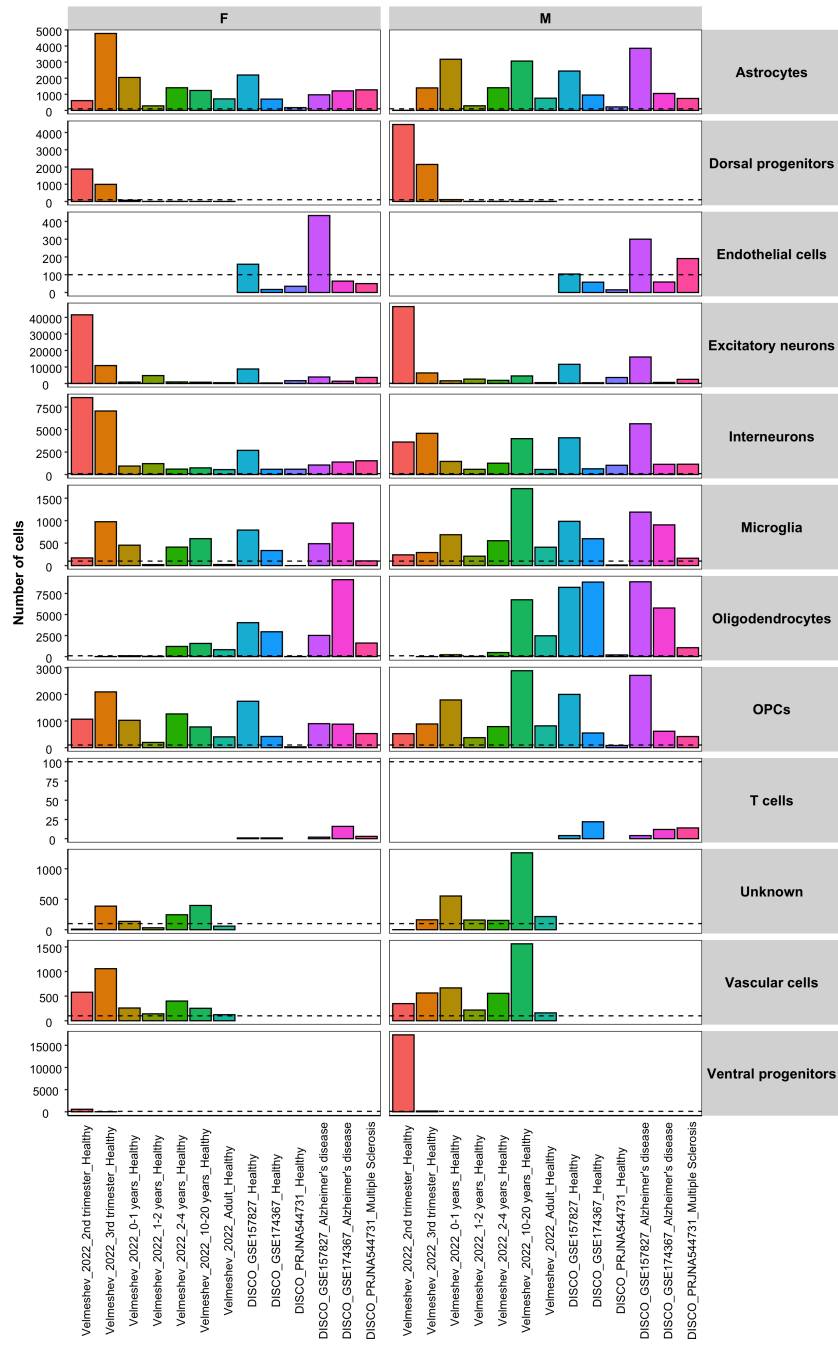


Fig. S6: Number of cells in each dataset and cell type analyzed, separated by sex. Most cell types had 100 or more cells (dashed line: 100 cells) in both female and male samples. Some cell types, e.g. T cells, were not high enough in numbers to be included in the downstream analysis.

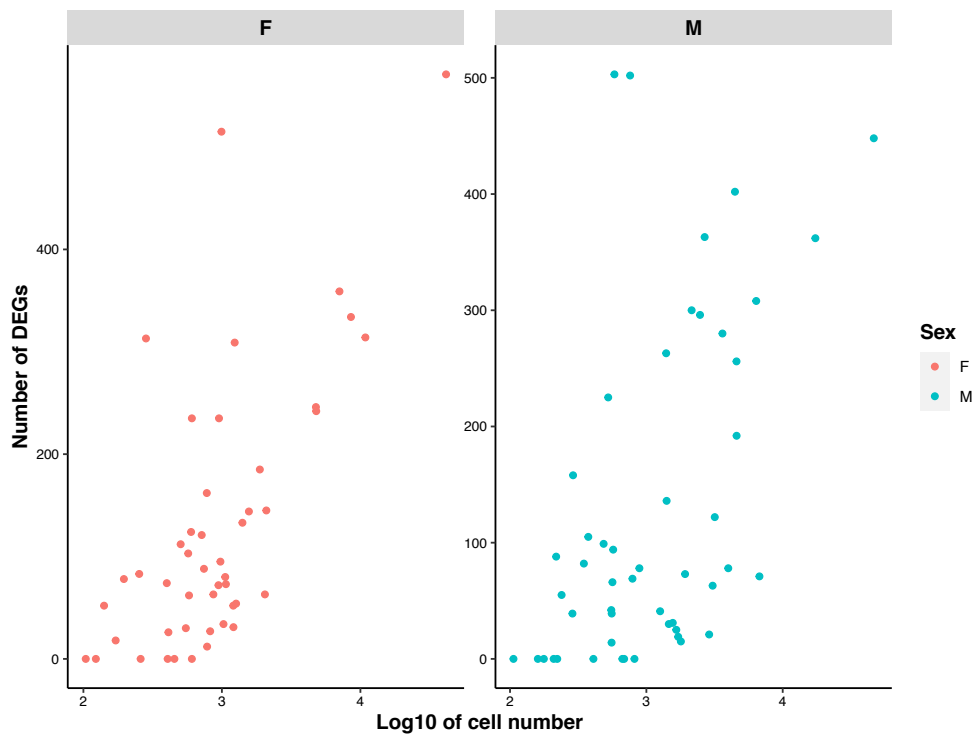


Fig. S7: Number of cells in log10 scale *versus* number of DEGs, split by sex. As seen from the plot, the higher the number of cells (x -axis), the higher the number of SG-biased genes (y -axis), in either sex.

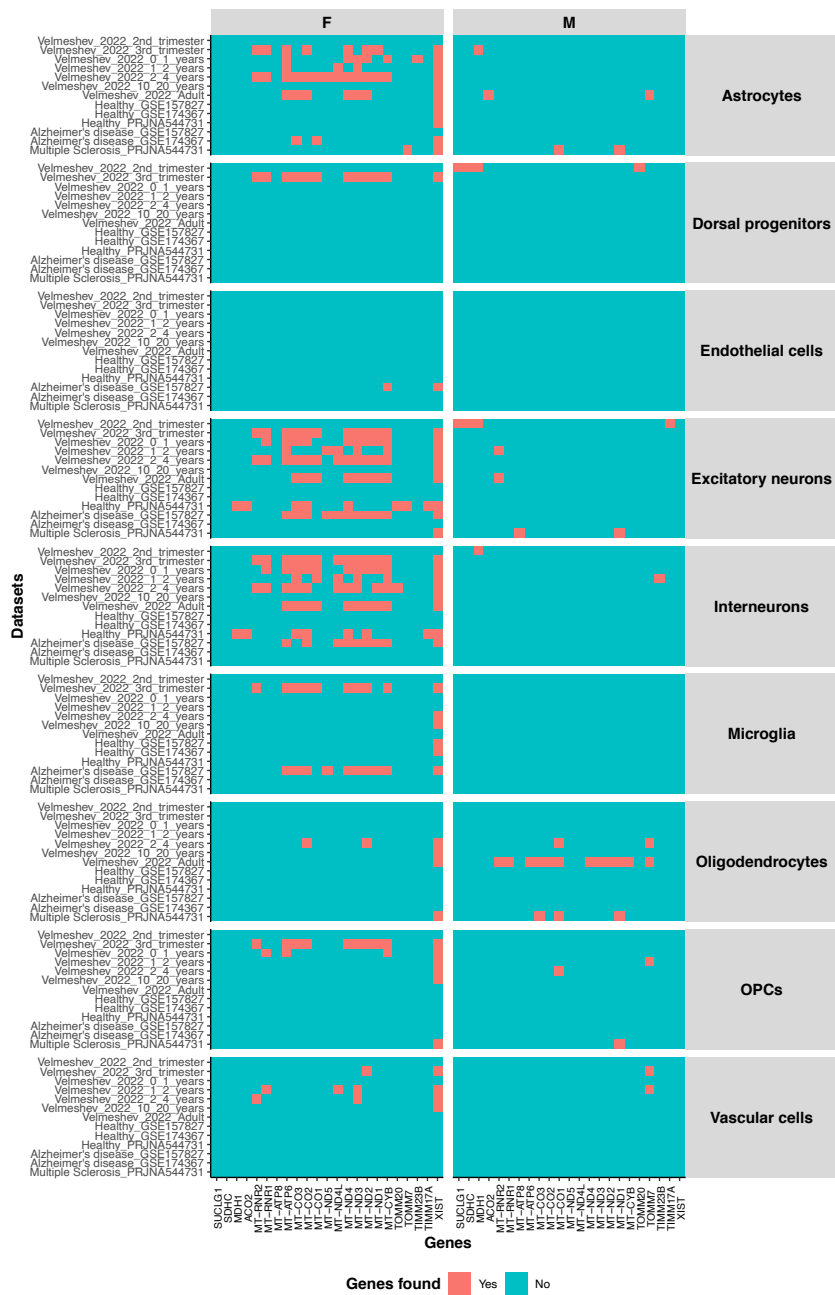


Fig. S8: The mitochondrial (MT) and tricarboxylic acid (TCA) cycle genes are expressed between sexes in a sex-biased manner. The heatmap displays the presence of MT and TCA genes in SG-biased genes lists from each cell type and dataset. MT, but not TCA or MT housekeeping genes (TIMM, TOMM) are more present in female-biased genes than in males.

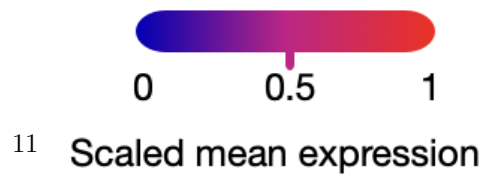
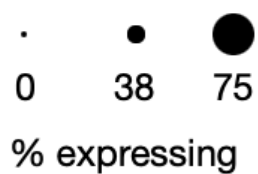
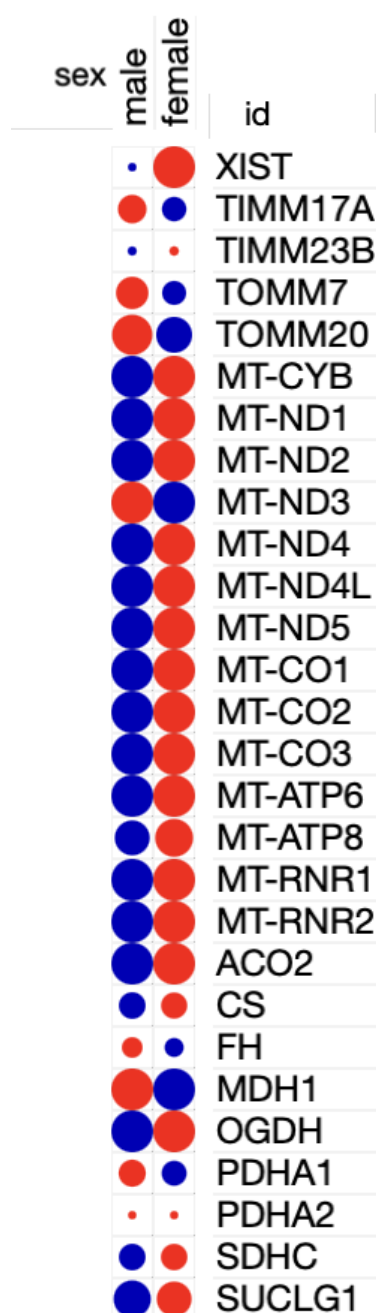


Fig. S9: Mitochondrial genes and the associated nuclear genes are more expressed in females than in males. This dot plot represents the expression of the same genes we explored regarding mitochondrial female bias in another single-cell study about dopaminergic neurons from Parkinson's disease patients and healthy controls ([Kamath et al., 2022](#)), using the [Broad Institute Single-Cell Portal](#).



Fig. S10: SG-biased genes are mainly expressed in few cellular compartments, and the distribution is quite similar between the sexes. This dot plot represents the hyper-geometric enrichment of cellular compartments in each dataset, cell type and sex. The size of the dots indicates the number of SG-biased in each cellular compartment, and the color indicates the p-values after hyper-geometric enrichment. **Legend:** NS: not significant; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$.

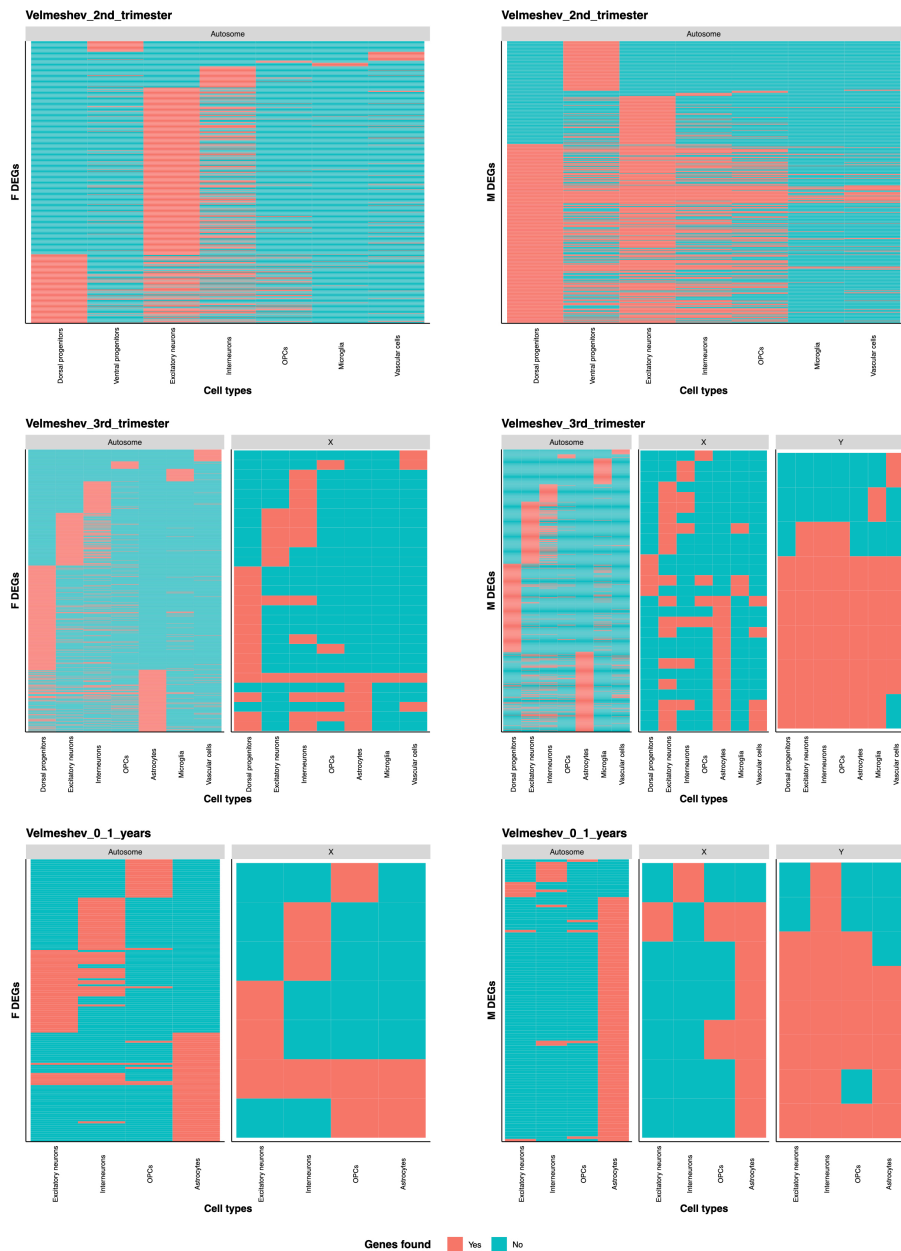


Fig. S11: The SG-biased genes are mostly autosomal and cell type-specific. Heatmaps showing the distribution of SG-biased genes across cell types and sex within each dataset, with the DEGs grouped by chromosome (autosome, X or Y). Presence indicates whether the gene is found in each specific dataset-sex-cell type combination.

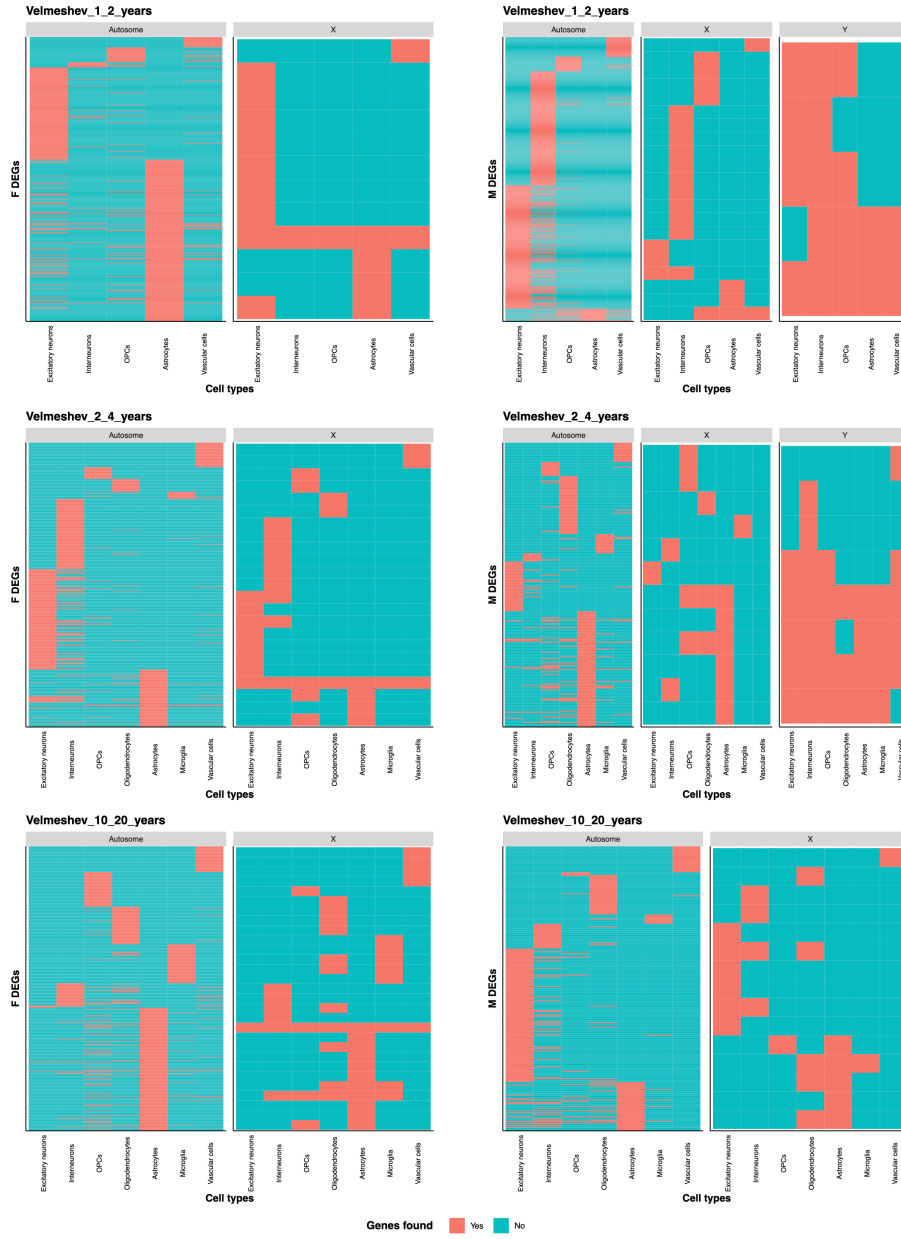


Fig. S12: The SG-biased genes are mostly autosomal and cell type-specific.

Heatmaps showing the distribution of SG-biased genes across cell types and sex within each dataset, with the DEGs grouped by chromosome (autosome, X or Y). Presence indicates whether the gene is found in each specific dataset-sex-cell type combination.

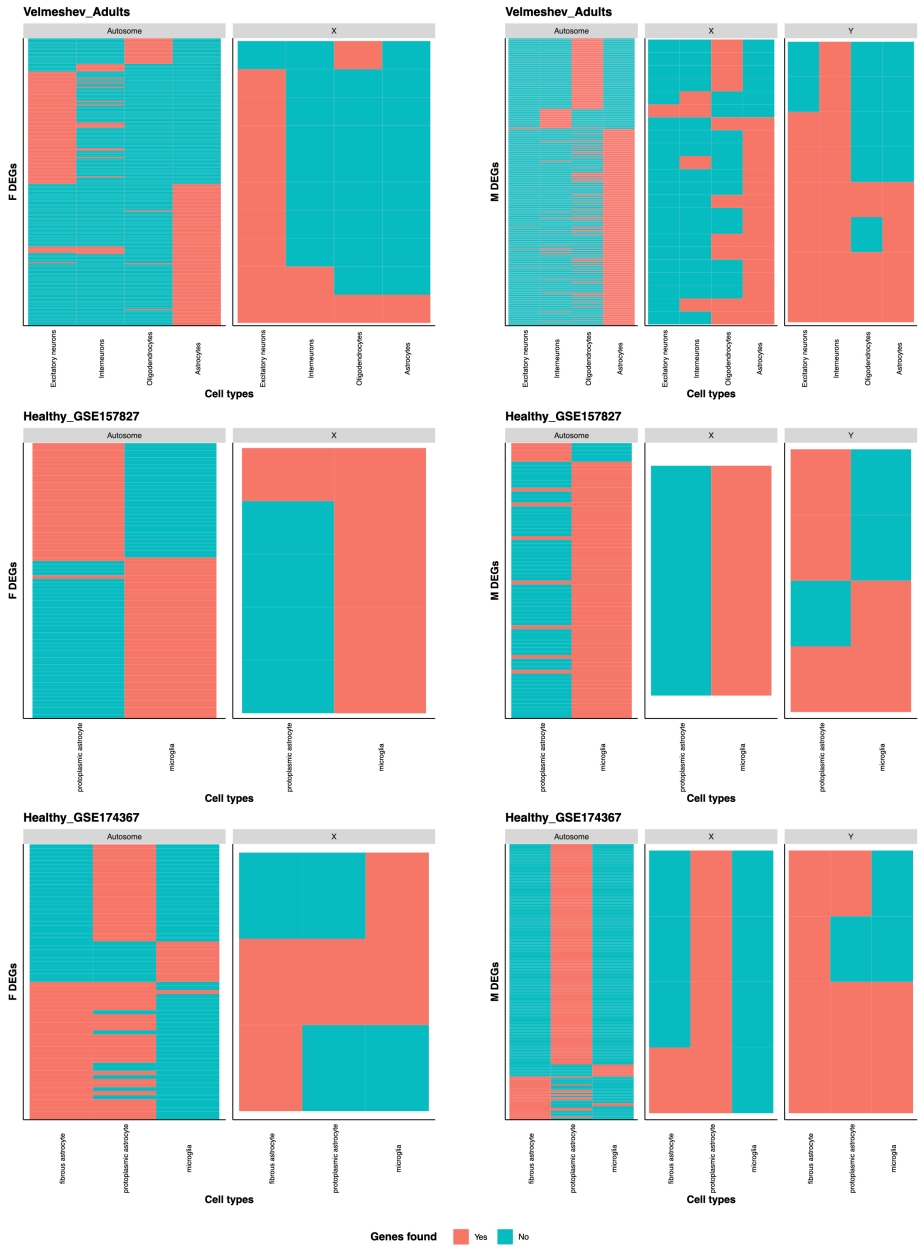


Fig. S13: The SG-biased genes are mostly autosomal and cell type-specific. Heatmaps showing the distribution of SG-biased genes across cell types and sex within each dataset, with the DEGs grouped by chromosome (autosome, X or Y). Presence indicates whether the gene is found in each specific dataset-sex-cell type combination.



Fig. S14: The SG-biased genes are mostly autosomal and cell type-specific. Heatmaps showing the distribution of SG-biased genes across cell types and sex within each dataset, with the DEGs grouped by chromosome (autosome, X or Y). Presence indicates whether the gene is found in each specific dataset-sex-cell type combination.

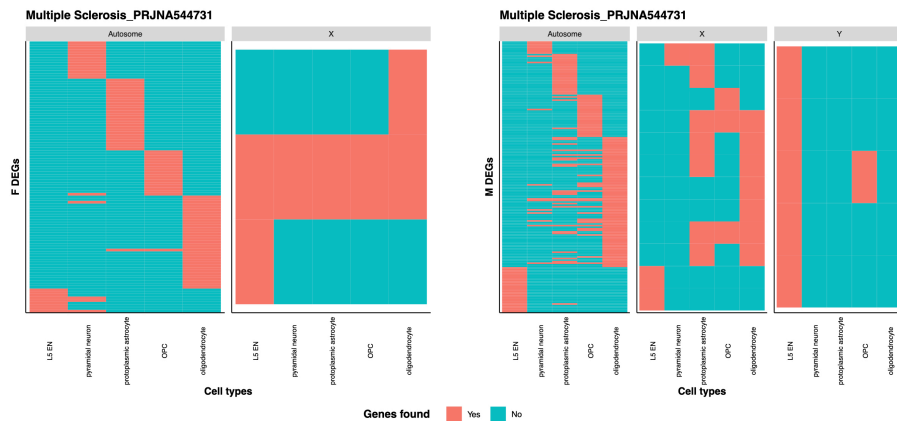


Fig. S15: The SG-biased genes are mostly autosomal and cell type-specific. Heatmaps showing the distribution of SG-biased genes across cell types and sex within each dataset, with the DEGs grouped by chromosome (autosome, X or Y). Presence indicates whether the gene is found in each specific dataset-sex-cell type combination.

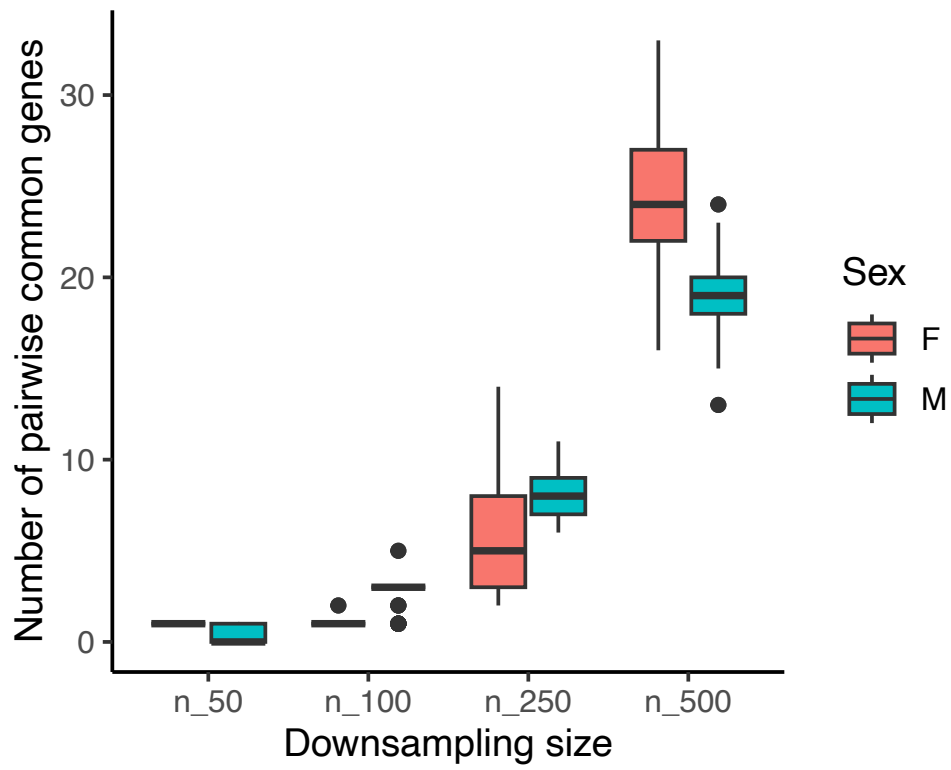


Fig. S16: Down-sampling of cell populations shows consistency in the found differential genes. Multiple rounds of down-sampling of excitatory neurons, using of different sizes of sampling, show that while the sampling size affects the number of pairwise common DEGs found, it does not affect the distribution of such comparisons.

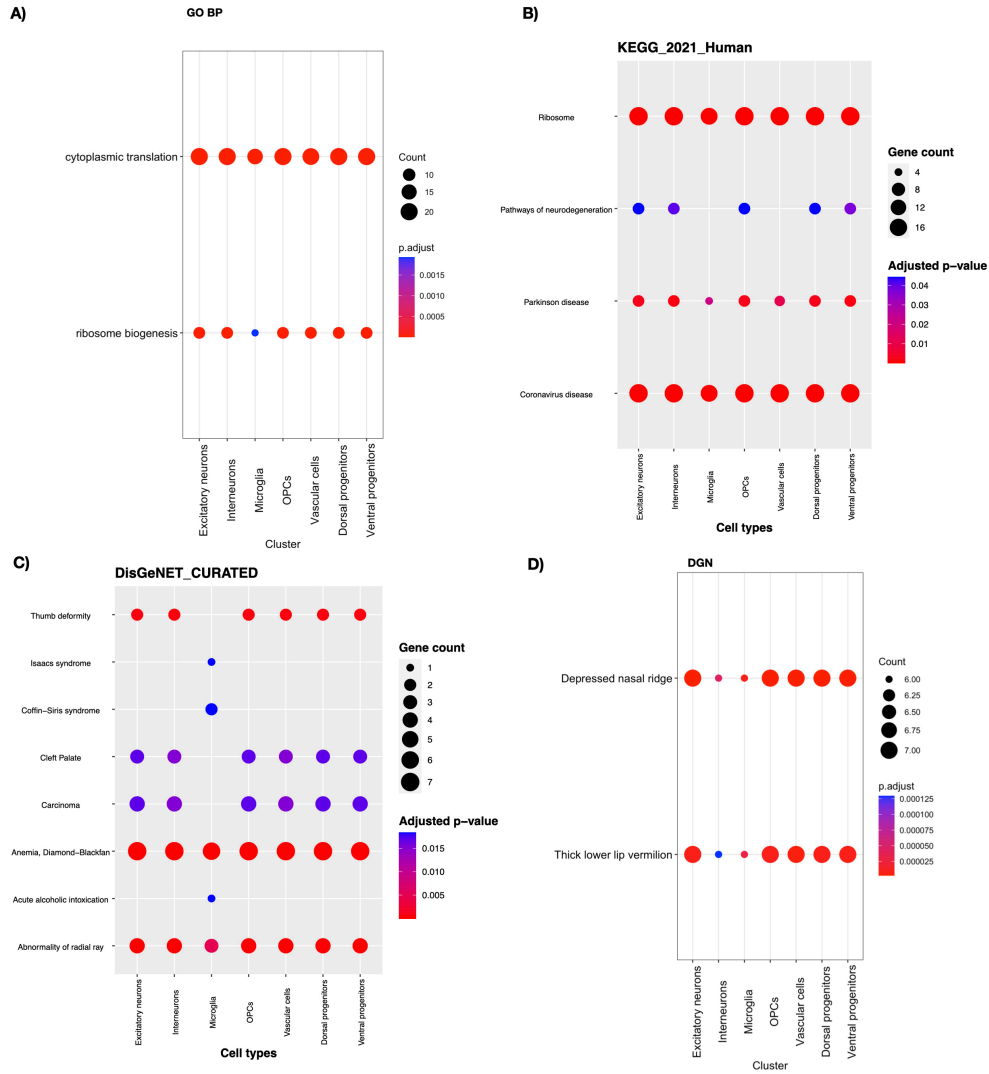


Fig. S17: Shared male-biased genes in the second trimester showed enrichment for ribosomal and developmental processes. A-D) Gene ontology biological processes (A), KEGG pathways (B), DisGeNET diseases CURATED (C) and clusterProfiler (D) enriched terms in the second trimester in the male-biased genes. The dot size indicate show many genes were found to belong to each GO BP term, and the color is the adjusted p-value (Benjamini-Hochberg correction).

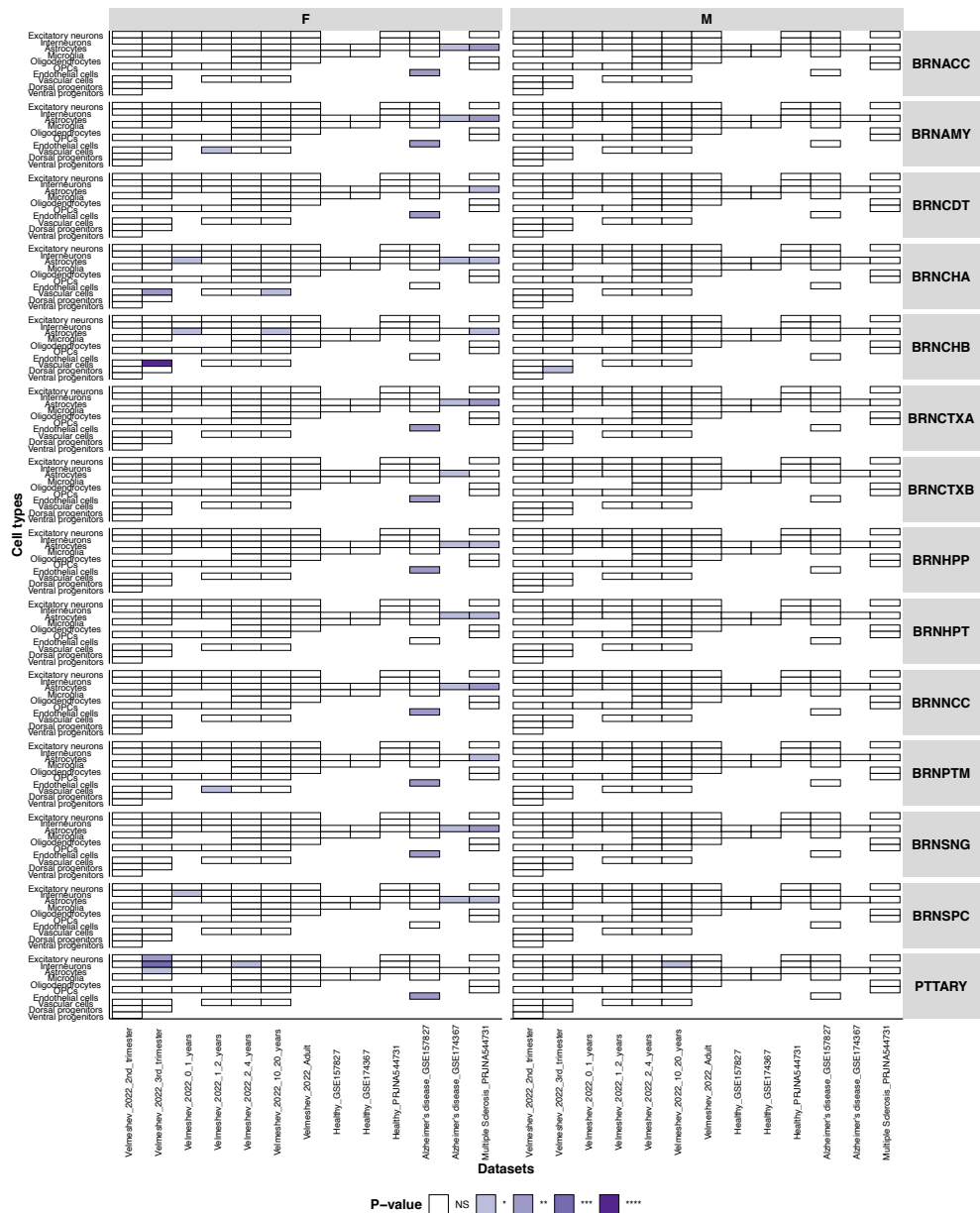


Fig. S20: SG-biased genes are developmental stage specific, with little overlap between datasets. This presence heatmap shows whether the SG-biased genes are found in the groups within the same cell type, divided by sex. The heatmaps for endothelial cells and ventral progenitors are missing due to them being present only in one group, AD GSE157827 and third trimester respectively.



Fig. S21: SG-biased genes are developmental stage specific, with little overlap between datasets. The bar plots indicate the number of common SG-biased genes between the reference dataset (facet labels) and all other datasets where the cell type is found.

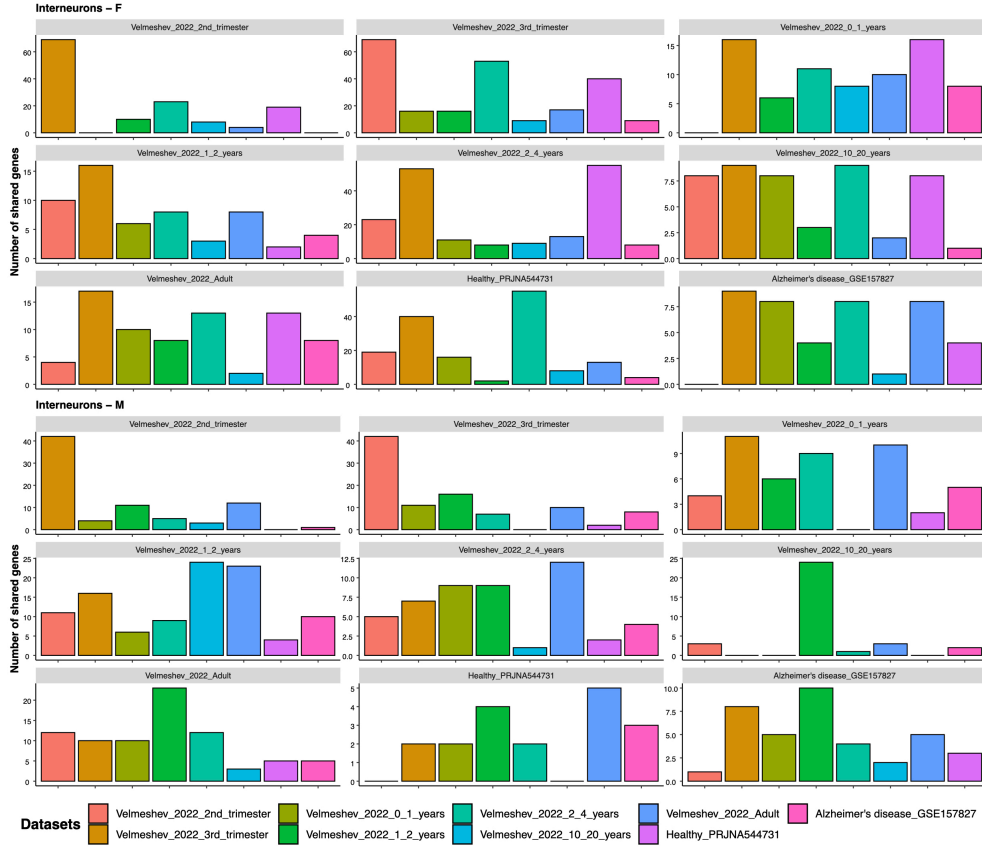


Fig. S22: SG-biased genes are developmental stage specific, with little overlap between datasets. The bar plots indicate the number of common SG-biased genes between the reference dataset (facet labels) and all other datasets where the cell type is found.

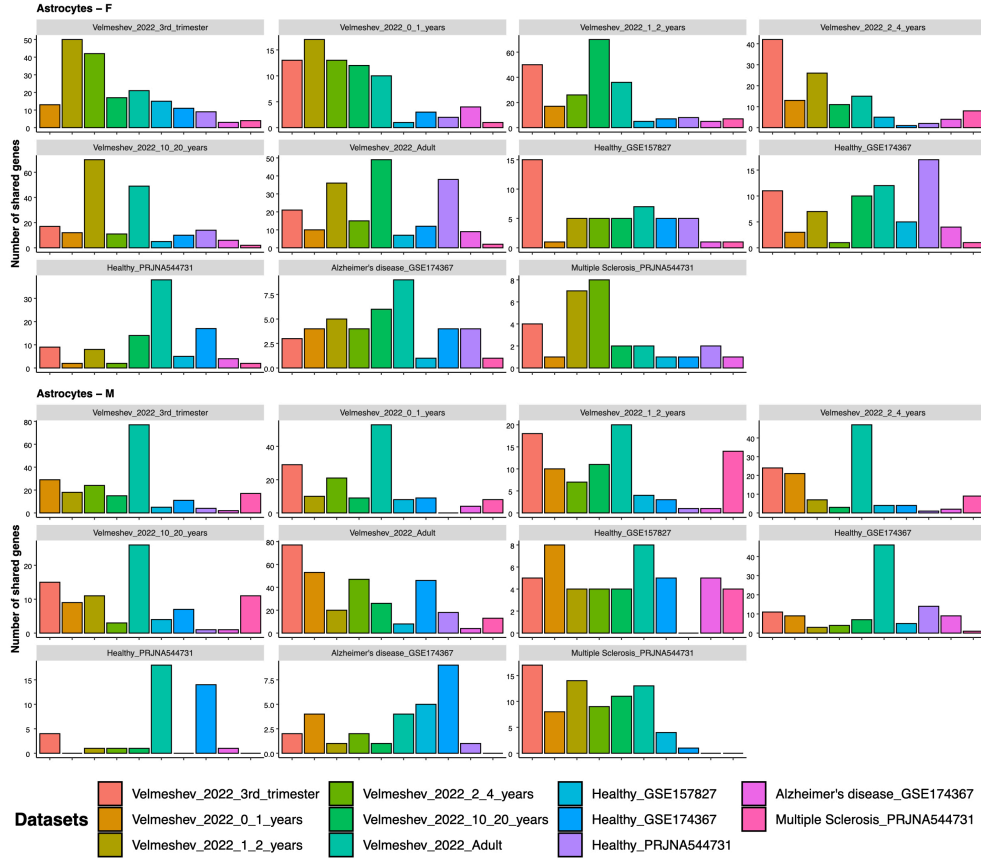


Fig. S23: SG-biased genes are developmental stage specific, with little overlap between datasets. The bar plots indicate the number of common SG-biased genes between the reference dataset (facet labels) and all other datasets where the cell type is found.

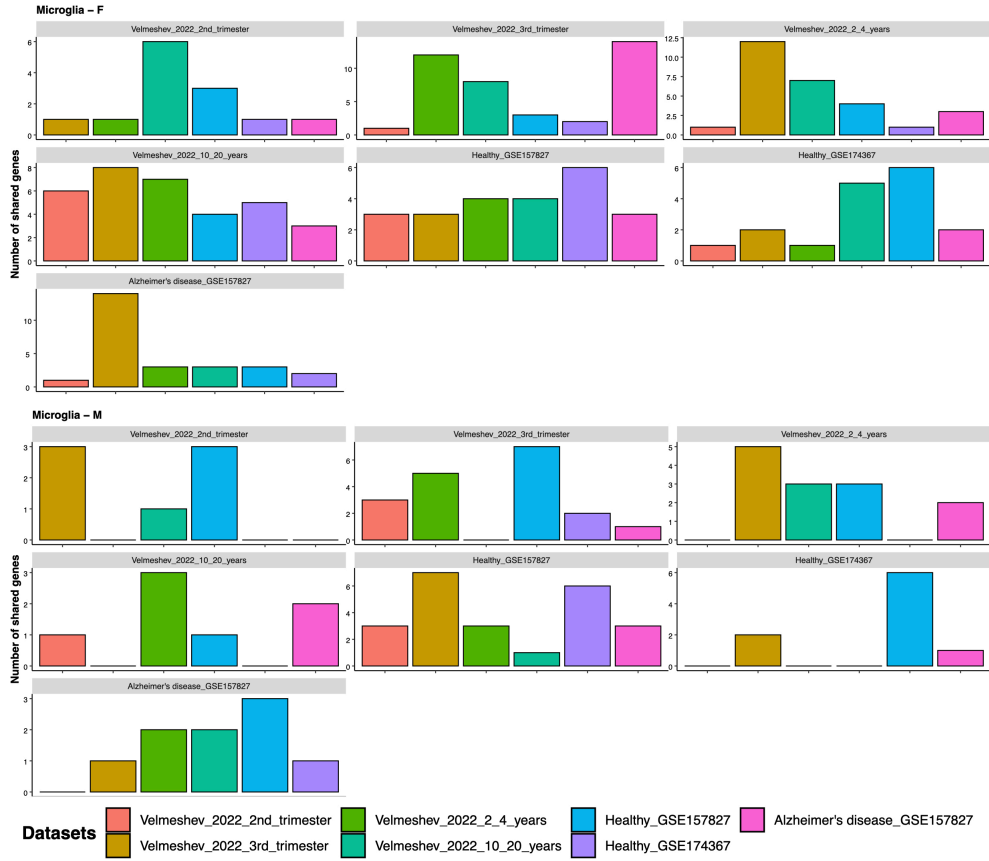


Fig. S24: SG-biased genes are developmental stage specific, with little overlap between datasets. The bar plots indicate the number of common SG-biased genes between the reference dataset (facet labels) and all other datasets where the cell type is found.

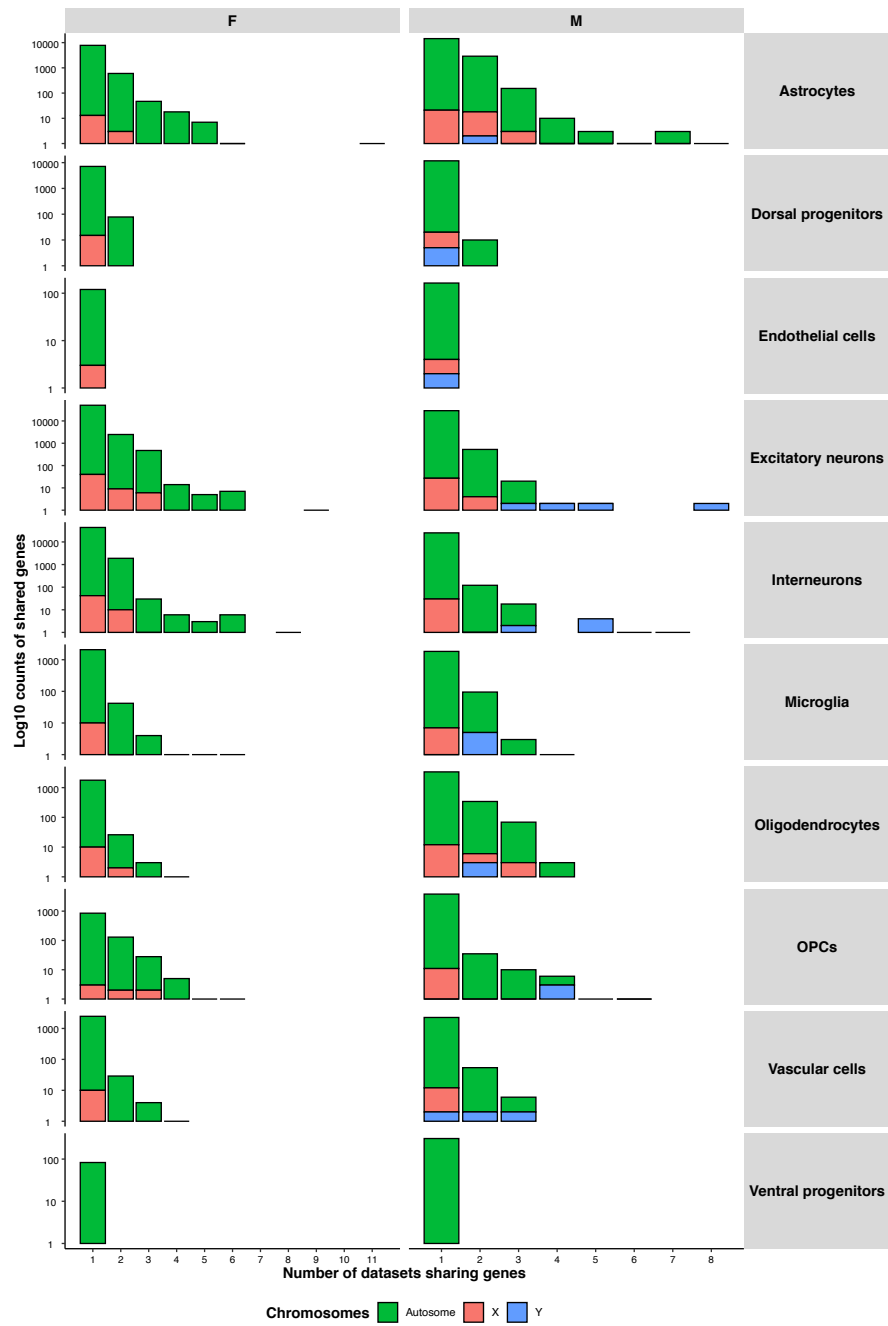


Fig. S25: Most of the SG-biased genes belong to the autosome, while the most shared SG-biased genes belong to the sex chromosome instead. This bar plot shows the distribution of shared SG-biased genes among groups for each cell type, expressed as log10 absolute counts and grouped by chromosome (autosome, X or Y). As the number of groups sharing the SG-biased genes increases, the number of SG-biased genes decreases, and the fraction of these genes which belong to the sex chromosome increases as well.

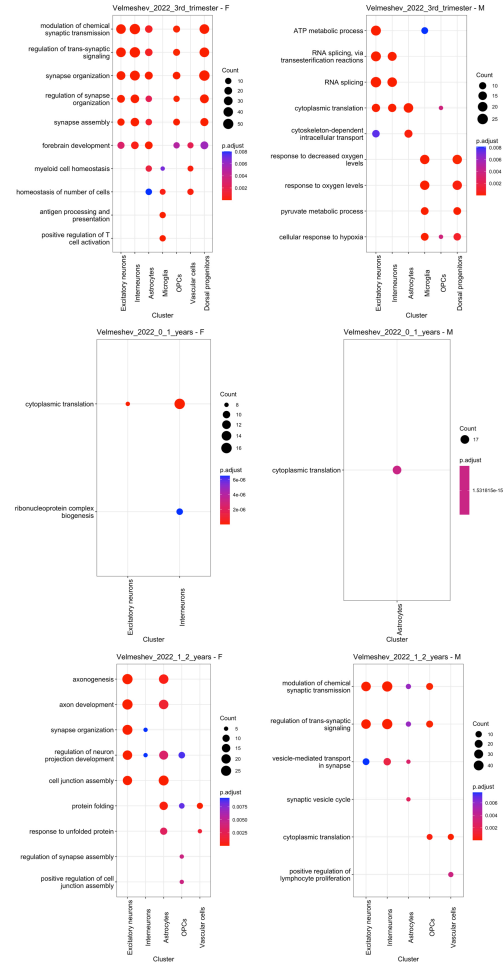


Fig. S26: Most gene ontology biological processes (GO BPs) were shared across cell types in each dataset-sex combination. These plots show the GO BPs enrichment in each dataset, in the different cell types and for each sex. The dot size indicate show many genes were found to belong to each GO BP term, and the color is the adjusted p-value (Benjamini-Hochberg correction).

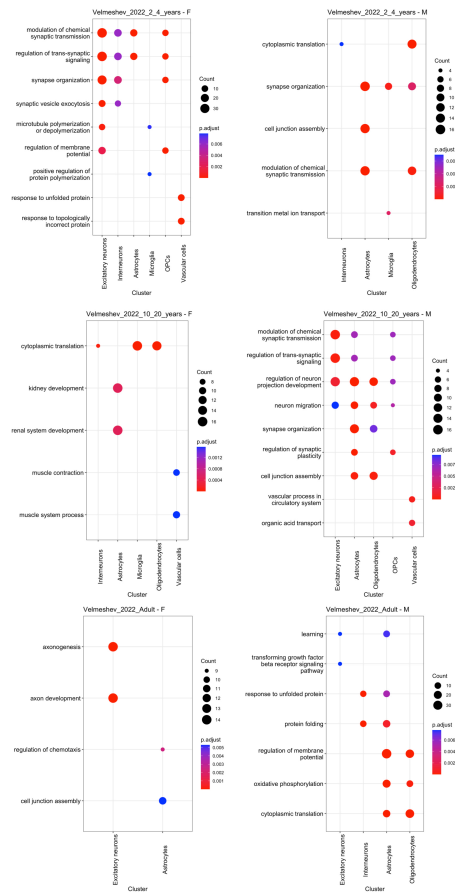


Fig. S27: Most gene ontology biological processes (GO BPs) were shared across cell types in each dataset-sex combination. These plots show the GO BPs enrichment in each dataset, in the different cell types and for each sex. The dot size indicate show many genes were found to belong to each GO BP term, and the color is the adjusted p-value (Benjamini-Hochberg correction).

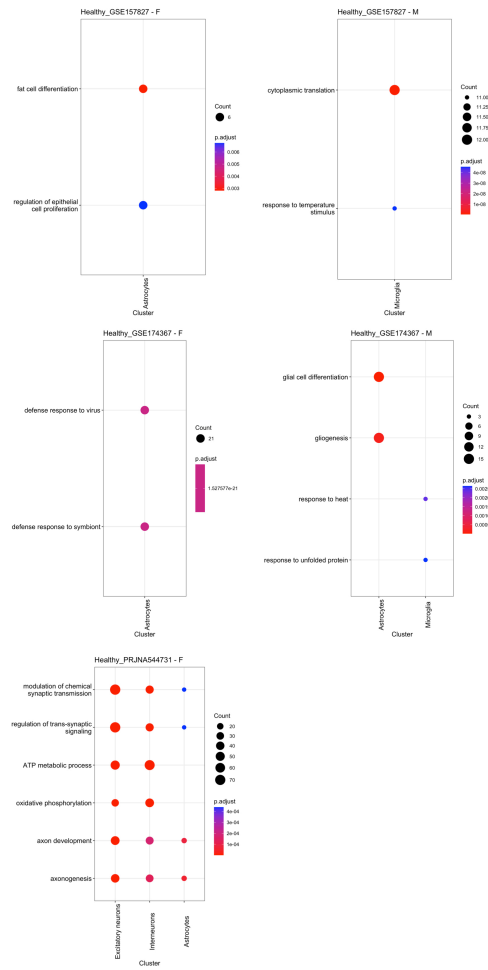


Fig. S28: Most gene ontology biological processes (GO BPs) were shared across cell types in each dataset-sex combination. These plots show the GO BPs enrichment in each dataset, in the different cell types and for each sex. The dot size indicate show many genes were found to belong to each GO BP term, and the color is the adjusted p-value (Benjamini-Hochberg correction).

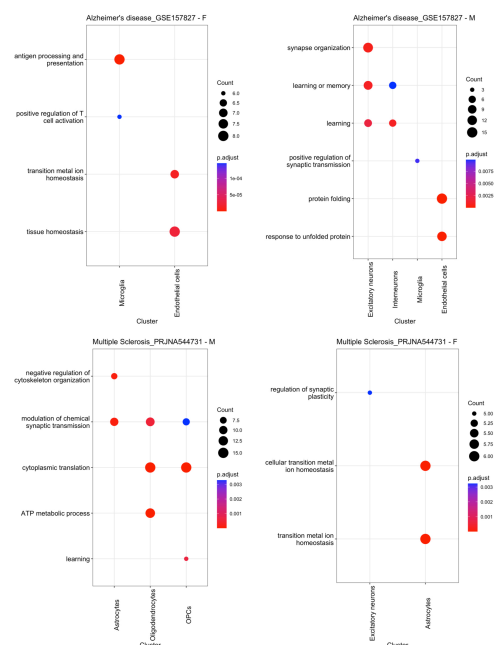


Fig. S29: Most gene ontology biological processes (GO BPs) were shared across cell types in each dataset-sex combination. These plots show the GO BPs enrichment in each dataset, in the different cell types and for each sex. The dot size indicate show many genes were found to belong to each GO BP term, and the color is the adjusted p-value (Benjamini-Hochberg correction).

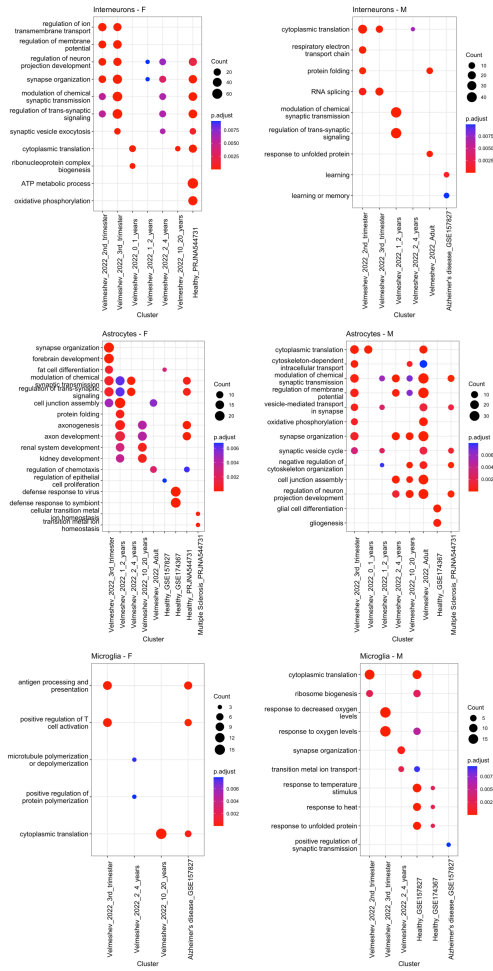


Fig. S30: Most gene ontology biological processes (GO BPs) were shared across datasets in each cell type-sex combination. These plots show the GO BPs enrichment in each cell type, across datasets and for each sex. The dot size indicate show many genes were found to belong to each GO BP term, and the color is the adjusted p-value (Benjamini-Hochberg correction).

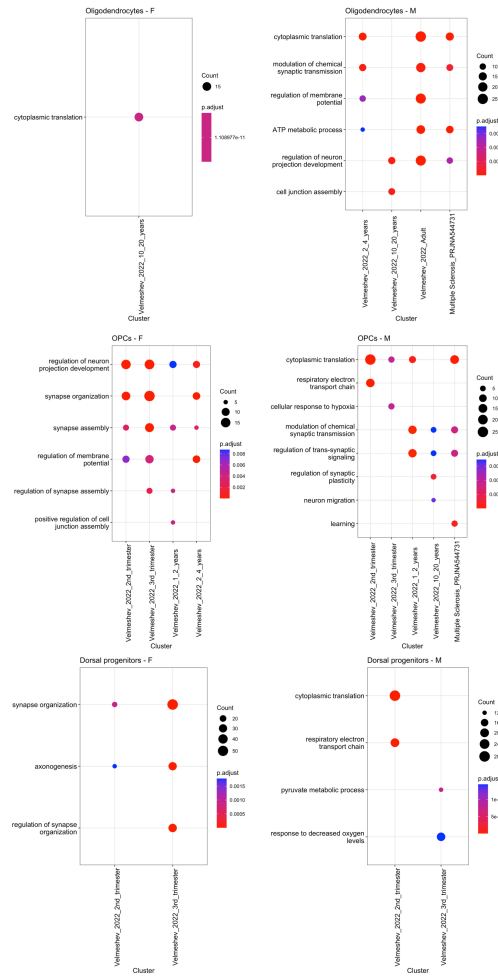


Fig. S31: Most gene ontology biological processes (GO BPs) were shared across datasets in each cell type-sex combination. These plots show the GO BPs enrichment in each cell type, across datasets and for each sex. The dot size indicate show many genes were found to belong to each GO BP term, and the color is the adjusted p-value (Benjamini-Hochberg correction).

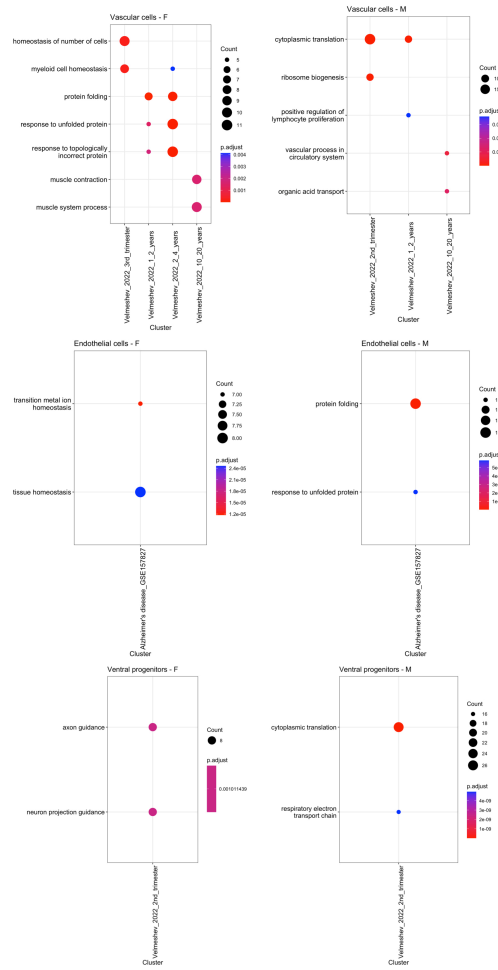


Fig. S32: Most gene ontology biological processes (GO BPs) were shared across datasets in each cell type-sex combination. These plots show the GO BPs enrichment in each cell type, across datasets and for each sex. The dot size indicate show many genes were found to belong to each GO BP term, and the color is the adjusted p-value (Benjamini-Hochberg correction).

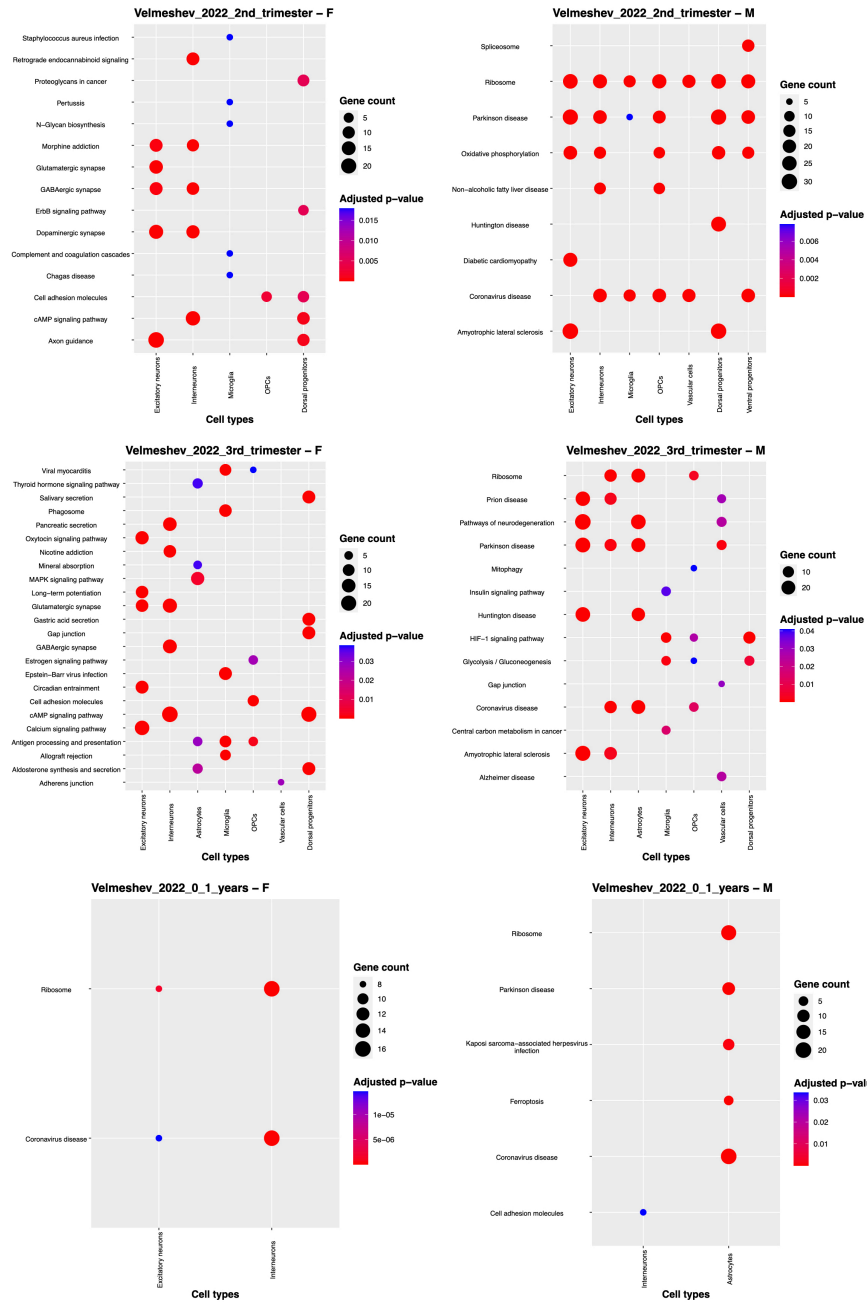


Fig. S33: Most Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were shared across cell types in each dataset-sex combination. These plots show the KEGG pathways enrichment in each dataset, across cell types and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction).

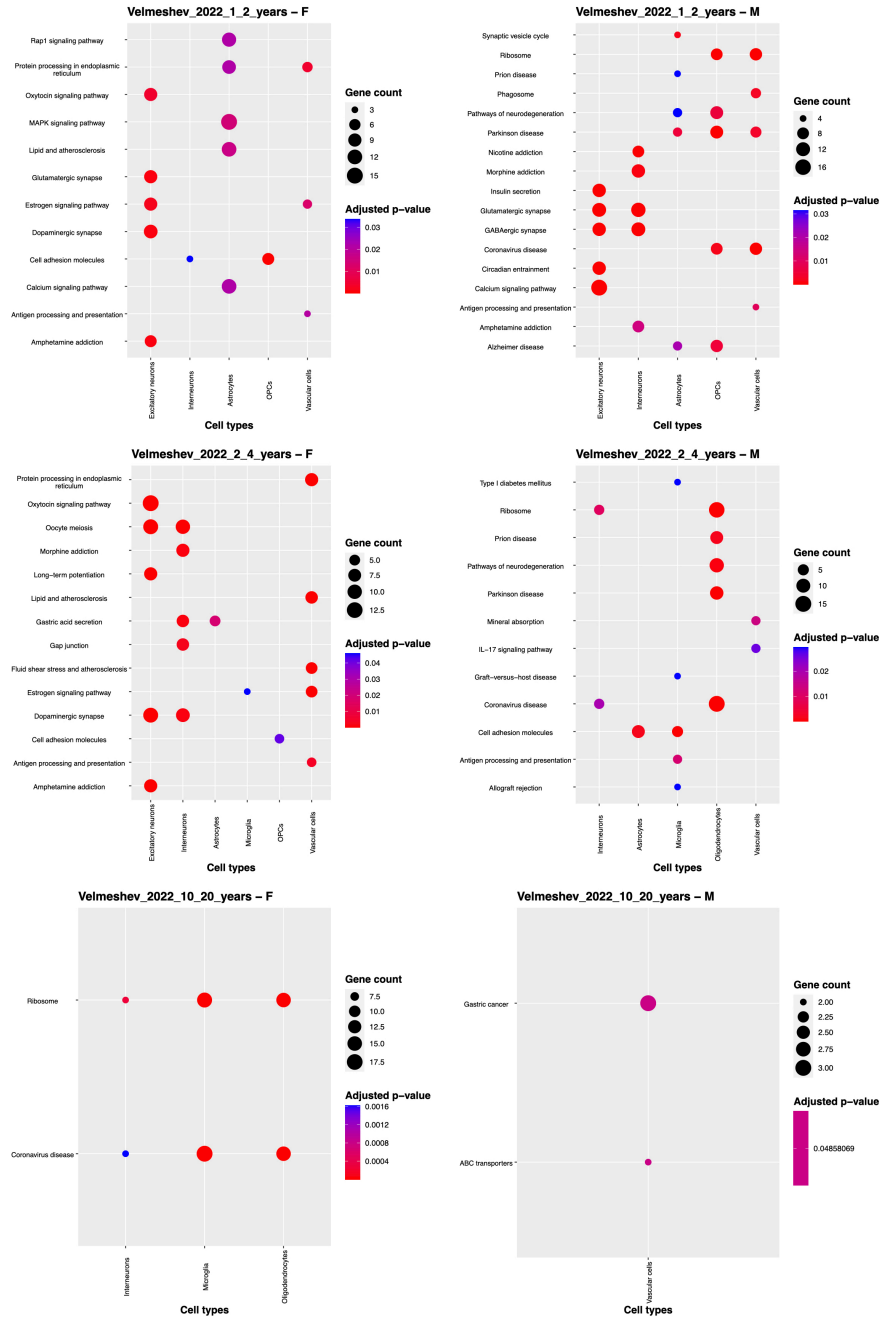


Fig. S34: Most Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were shared across cell types in each dataset-sex combination. These plots show the KEGG pathways enrichment in each dataset, across cell types and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction).

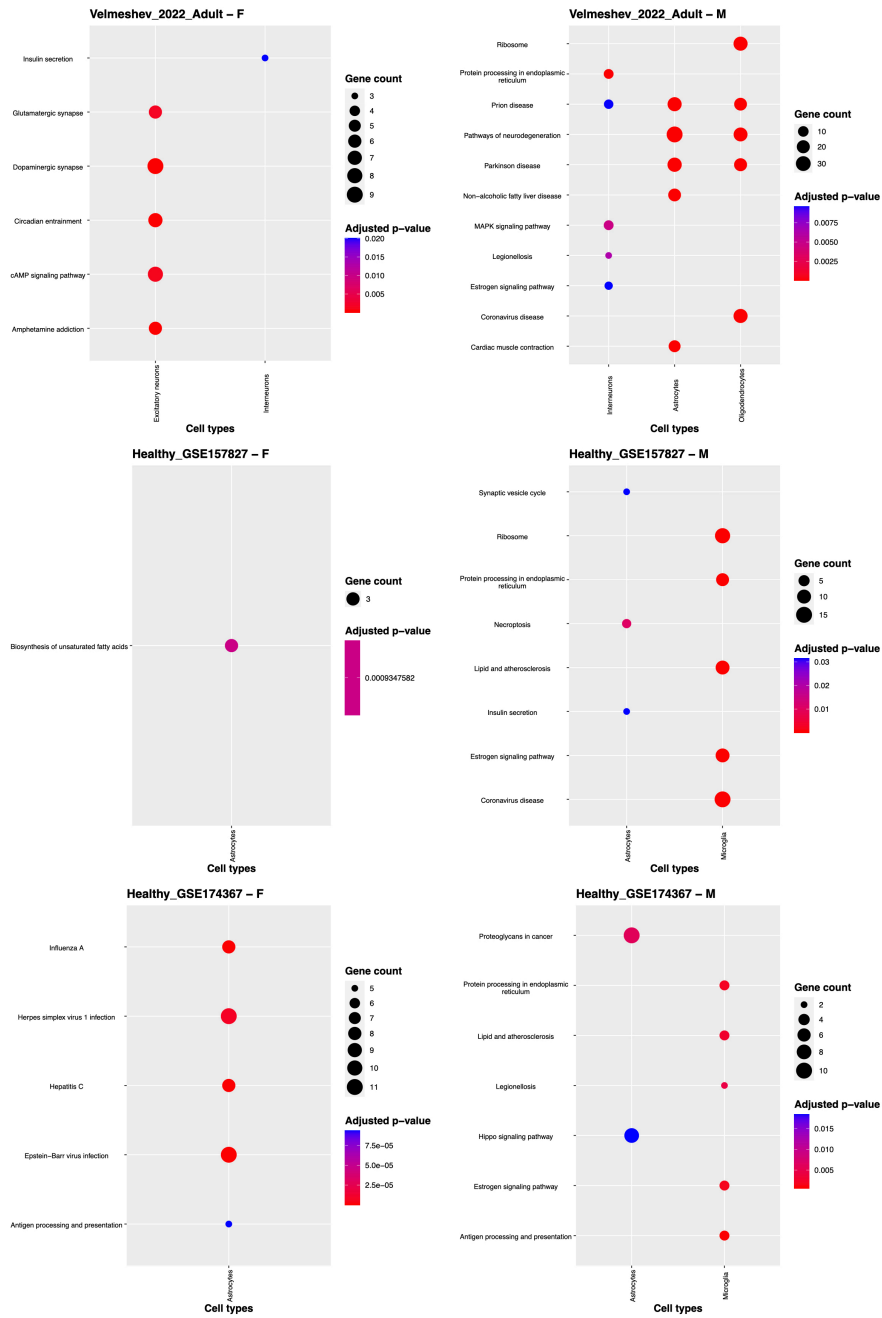


Fig. S35: Most Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were shared across cell types in each dataset-sex combination. These plots show the KEGG pathways enrichment in each dataset, across cell types and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction).

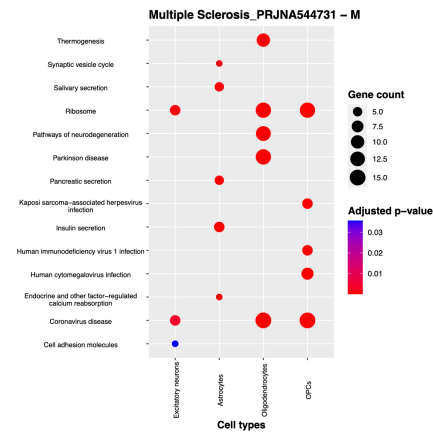
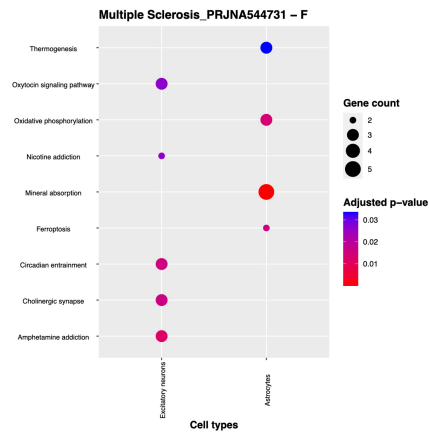
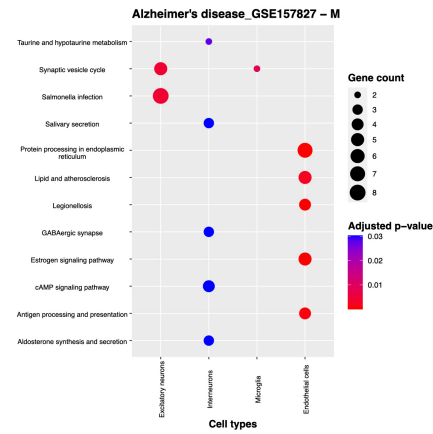
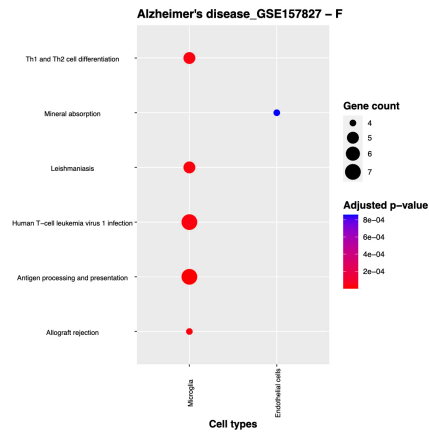
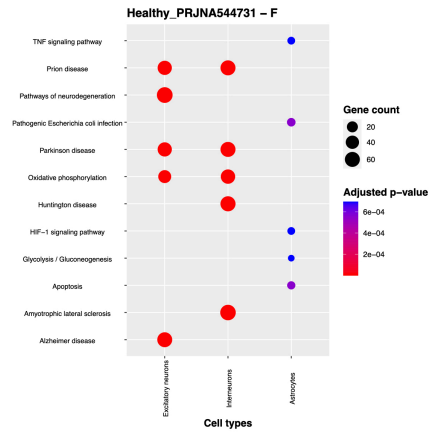


Fig. S36: Most Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were shared across cell types in each dataset-sex combination. These plots show the KEGG pathways enrichment in each dataset, across cell types and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction).

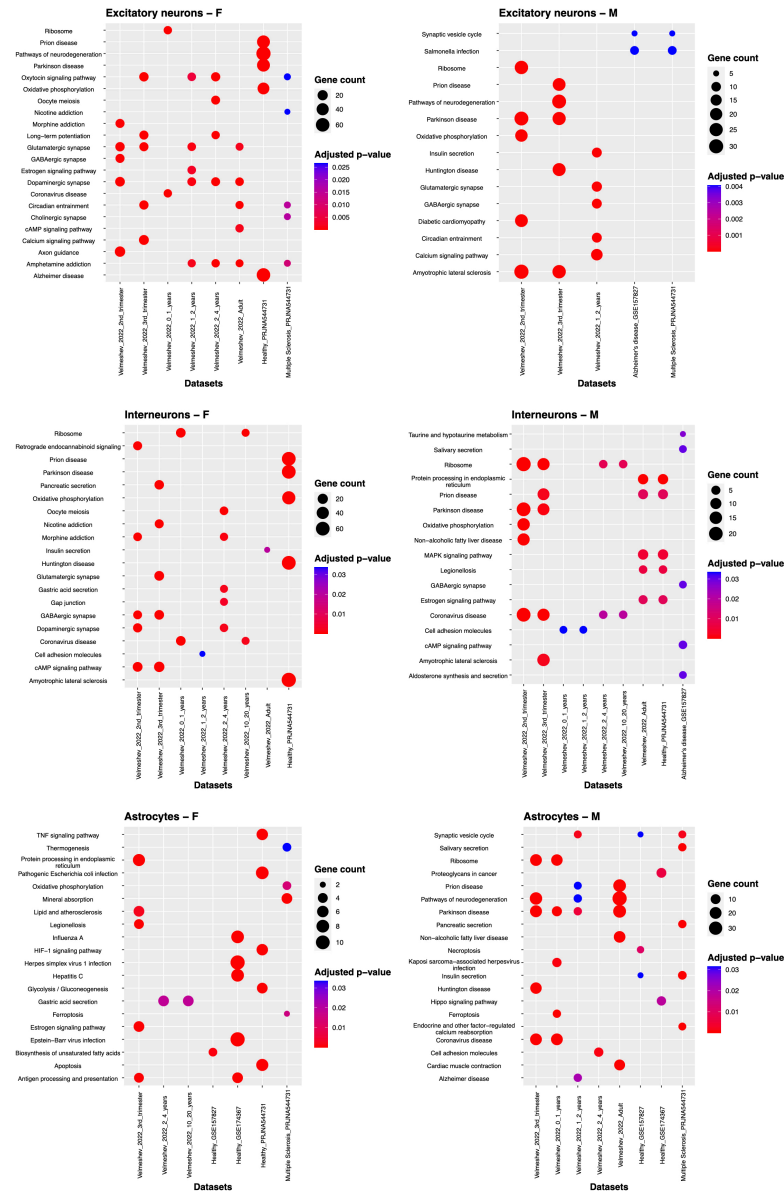


Fig. S37: Most Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were shared across datasets in each cell type-sex combination. These plots show the KEGG pathways enrichment in each cell type, across datasets and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction).

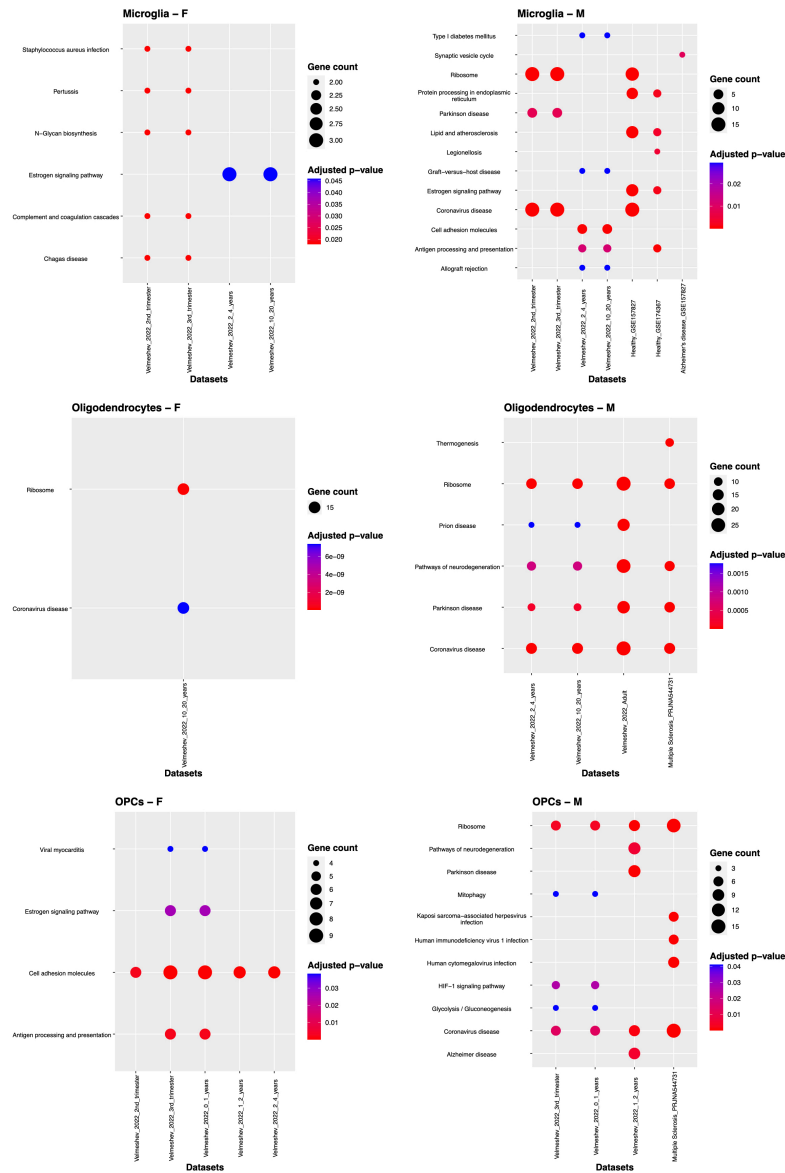


Fig. S38: Most Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were shared across datasets in each cell type-sex combination. These plots show the KEGG pathways enrichment in each cell type, across datasets and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction).

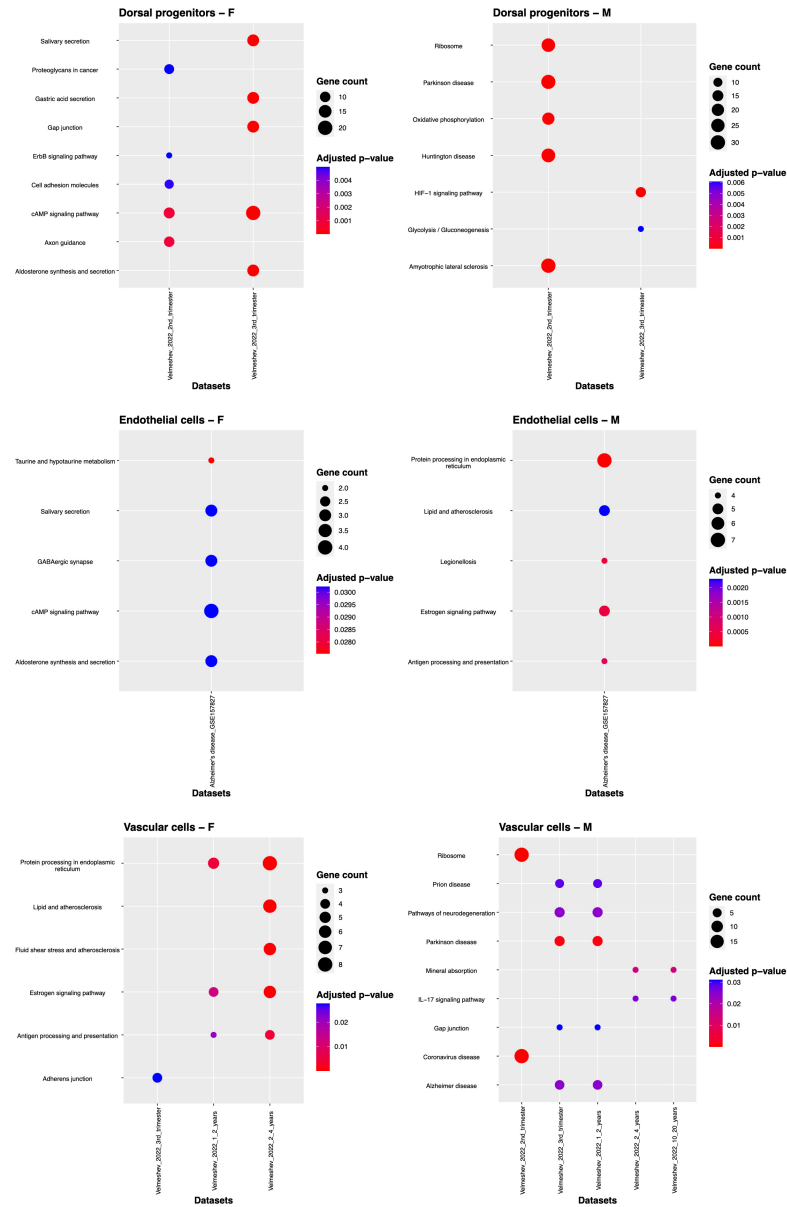


Fig. S39: Most Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were shared across datasets in each cell type-sex combination. These plots show the KEGG pathways enrichment in each cell type, across datasets and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction).

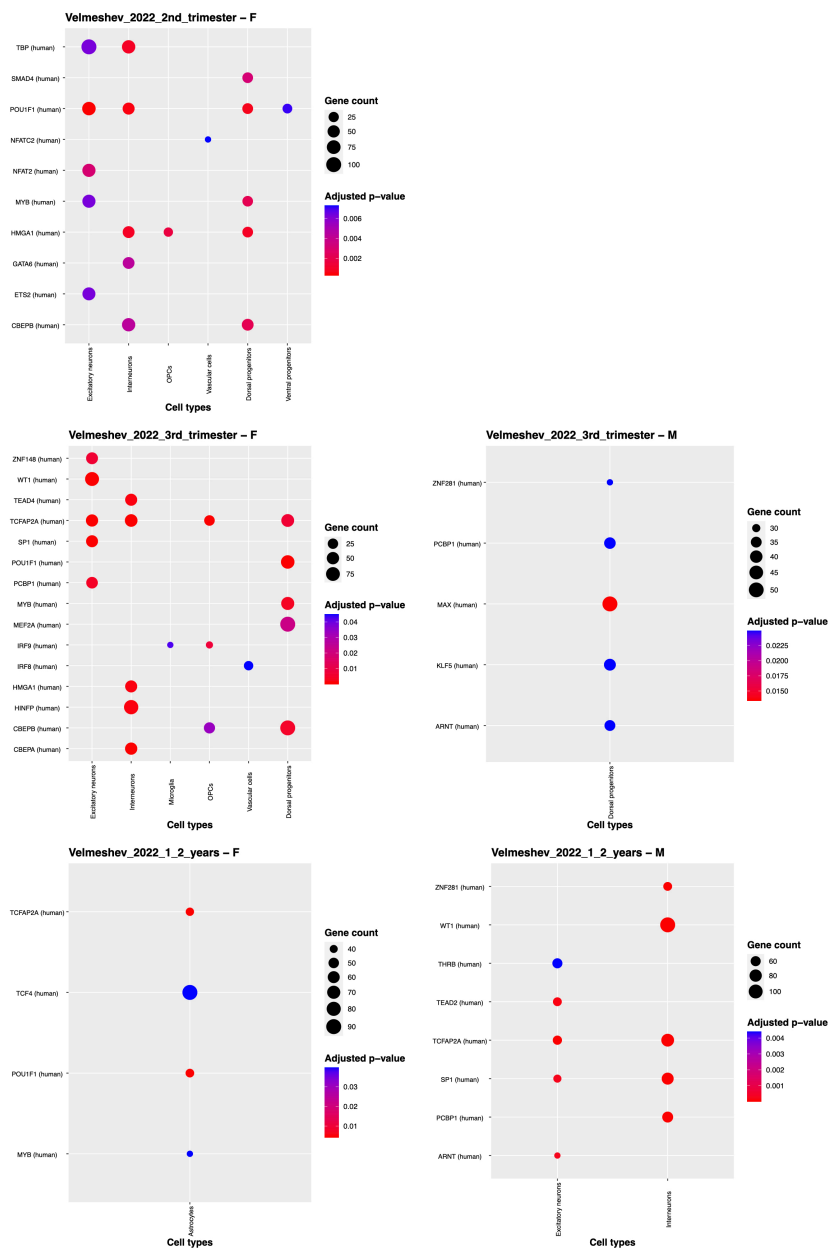


Fig. S40: TRANSFAC and JASPAR PWMs binding sites enrichment showed some overlap across cell types. These plots show the TRANSFAC and JASPAR PWMs binding sites enrichment in each dataset, across cell types and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction). TCFAP2A was the only TF binding site to be enriched in at least half of the datasets in females.

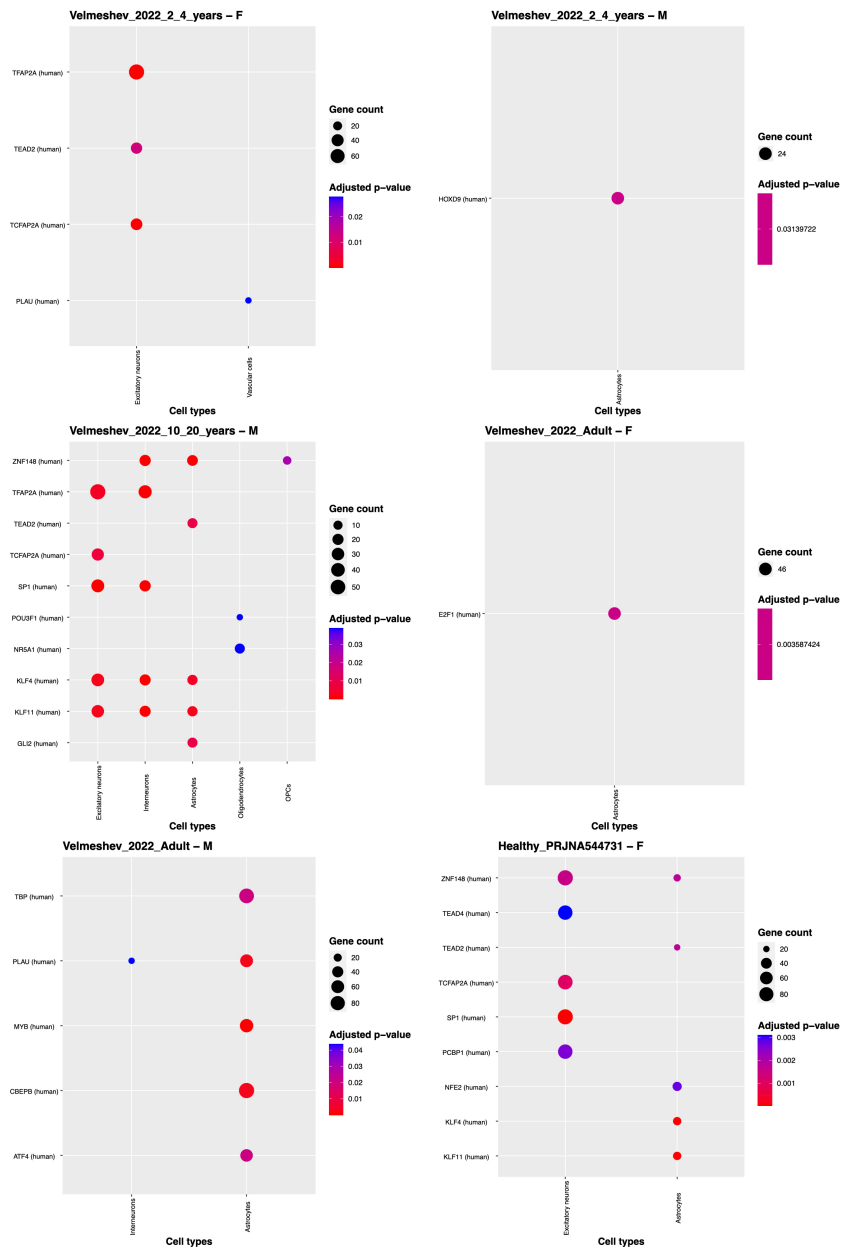


Fig. S41: TRANSFAC and JASPAR PWMs binding sites enrichment showed some overlap across cell types. These plots show the TRANSFAC and JASPAR PWMs binding sites enrichment in each dataset, across cell types and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction). TCFAP2A was the only TF binding site to be enriched in at least half of the datasets in females.

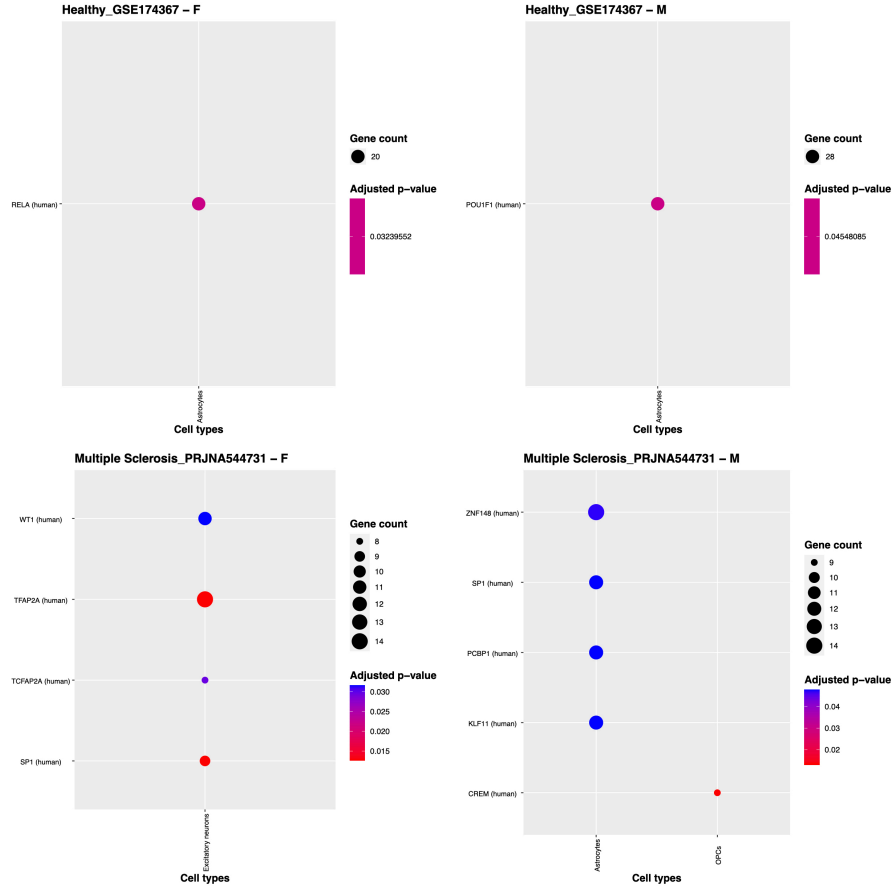


Fig. S42: TRANSFAC and JASPAR PWMs binding sites enrichment showed some overlap across cell types. These plots show the TRANSFAC and JASPAR PWMs binding sites enrichment in each dataset, across cell types and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction). TCFAP2A was the only TF binding site to be enriched in at least half of the datasets in females.

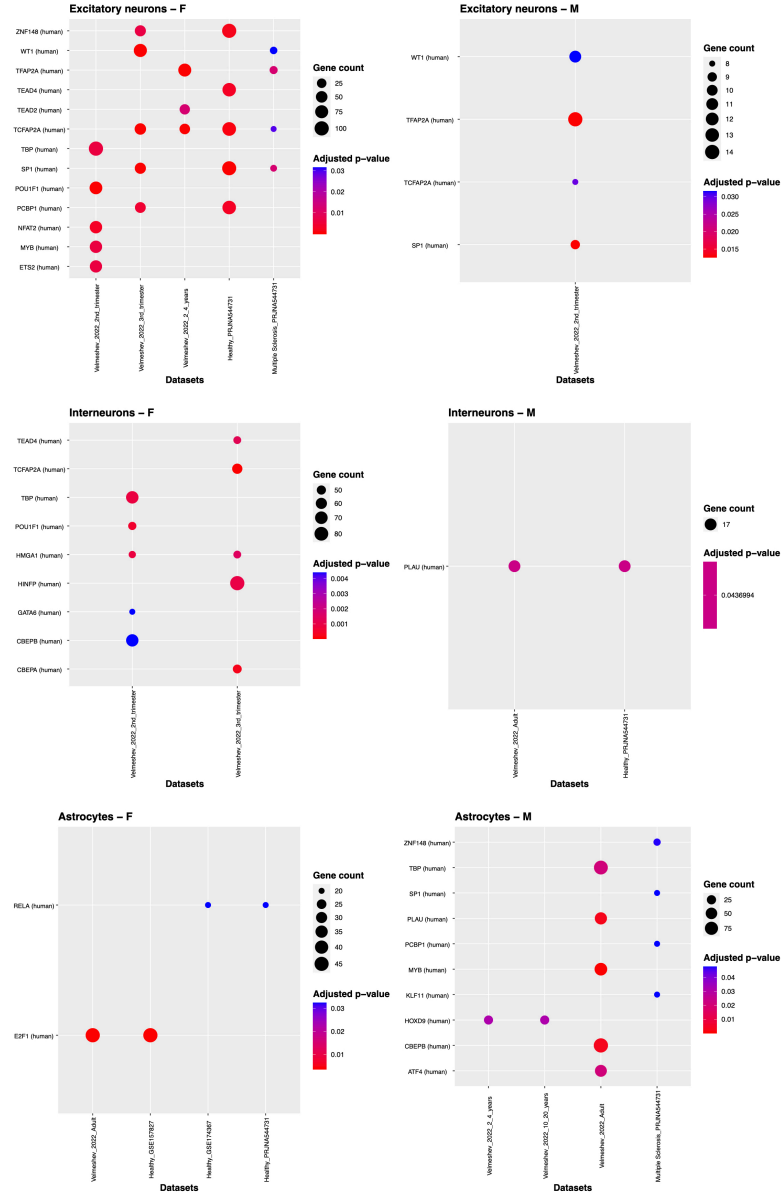


Fig. S43: TRANSFAC and JASPAR PWMs binding sites enrichment showed some overlap across datasets. These plots show the TRANSFAC and JASPAR PWMs binding sites enrichment in each cell type, across datasets and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction). TCFAP2A and POU1F1 were the only TF binding sites to be enriched in at least half of the datasets in females.

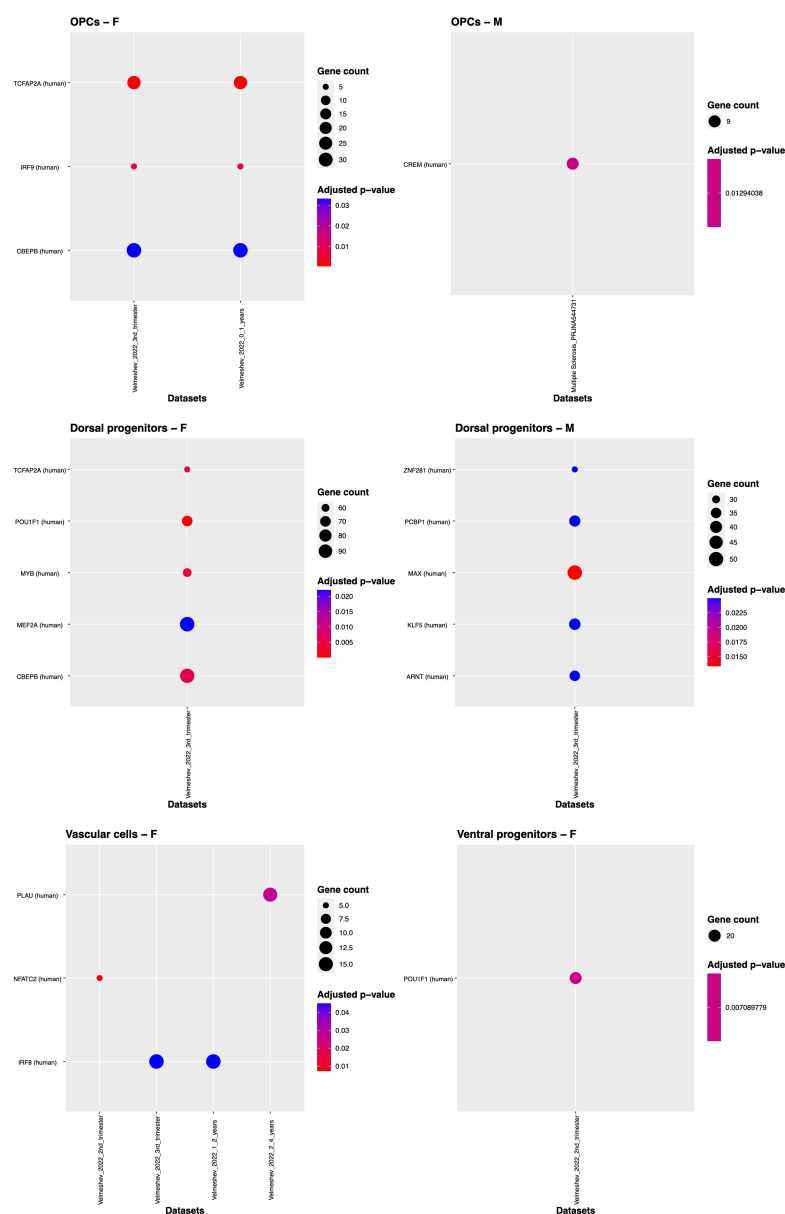


Fig. S44: TRANSFAC and JASPAR PWMs binding sites enrichment showed some overlap across datasets. These plots show the TRANSFAC and JASPAR PWMs binding sites enrichment in each cell type, across datasets and for each sex. The dot size indicate show many genes were found to belong to each pathway, and the color is the adjusted p-value (Benjamini-Hochberg correction). TCFAP2A and POU1F1 were the only TF binding sites to be enriched in at least half of the datasets in females.

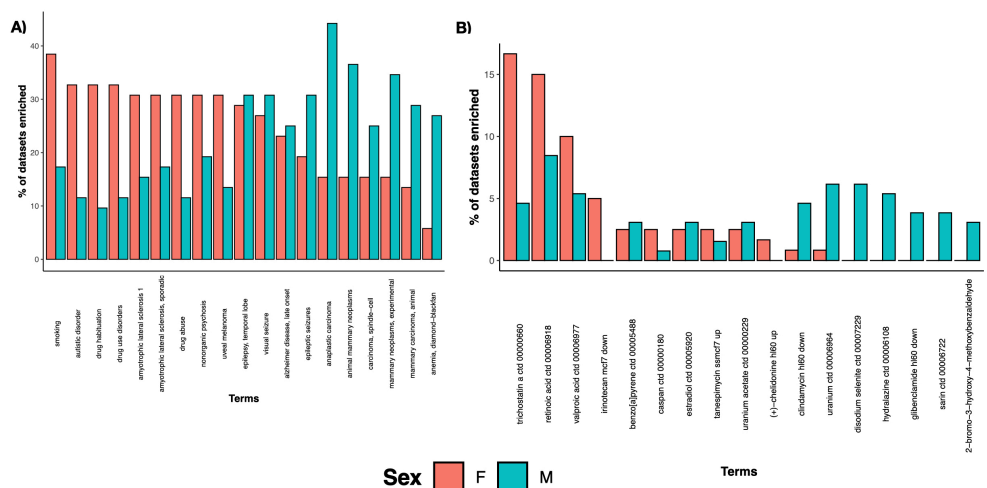


Fig. S45: SG-biased genes show biased enrichment for diseases (including autism) and drugs. A) The 10 most frequent disease-related terms in each sex, regardless of dataset, cell type and database used, is represented in this bar plot, shown as percentage of how many datasets were enriched for each term. B) This bar plot shows the 10 most frequent drug-related terms in each sex, regardless of dataset and cell type. Females show higher percentages of datasets sharing the same drug terms, although most of the same terms could be found in males. On the other hand, male-biased drug terms seemed to show a specificity, with some of them found solely in males.

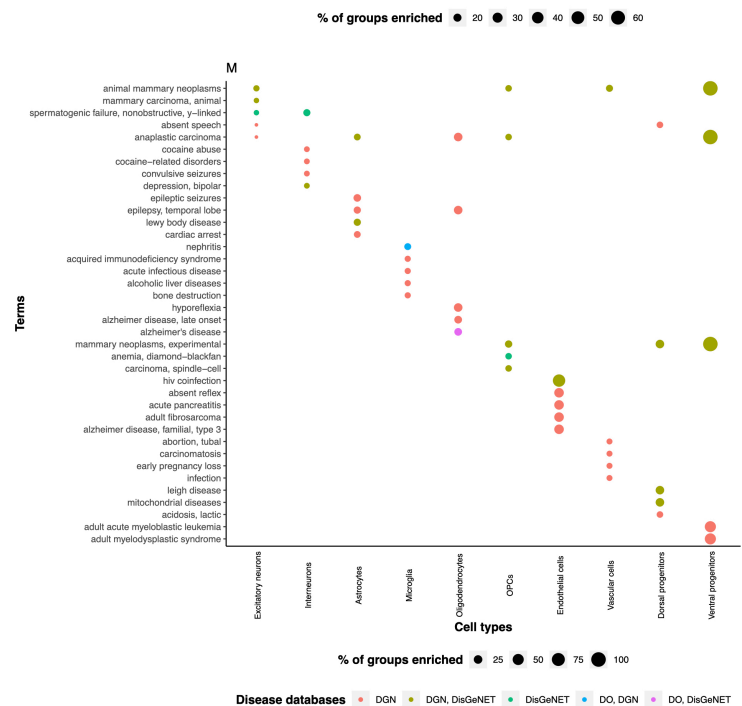
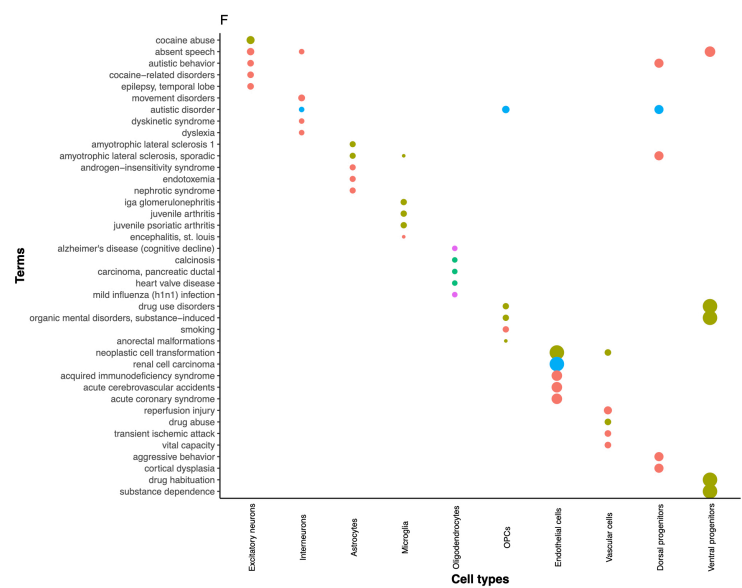


Fig. S46: The disease-related term were mostly sex-biased, and found in more than one disease database. These dot plots show the top 5 disease-related enriched terms, divided by sex, cell type and database which showed the enrichment in the terms. The dot size indicates how many datasets presented each term, and the color indicates in which combination of databases was the term found.

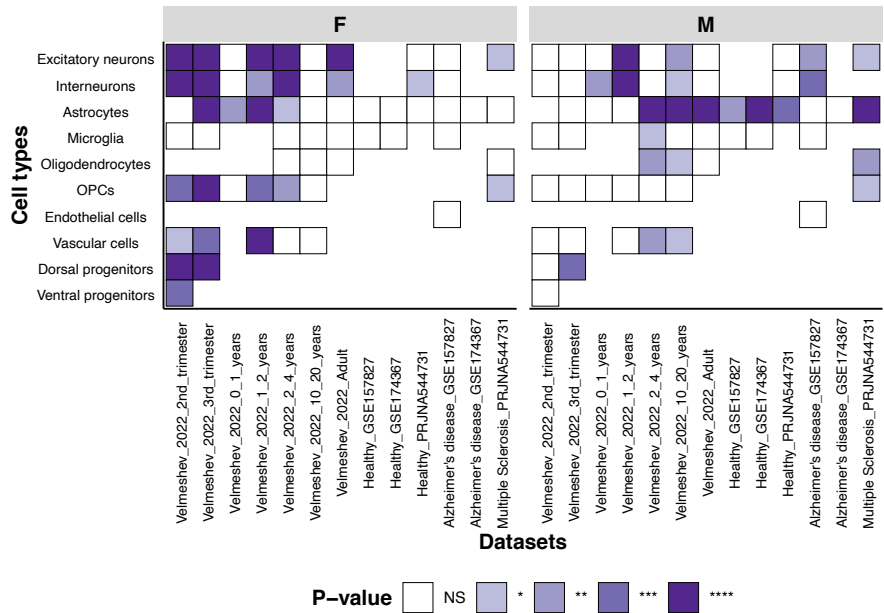


Fig. S47: Female-biased genes showed enrichment in the early age groups for autism-related genes. This heatmap shows the hyper-geometric enrichment of SFARI genes in the SG-biased DEGs. Most of the SFARI genes enrichment was found in females, an in earlier age datasets. **Legend:** NS: not significant; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$.

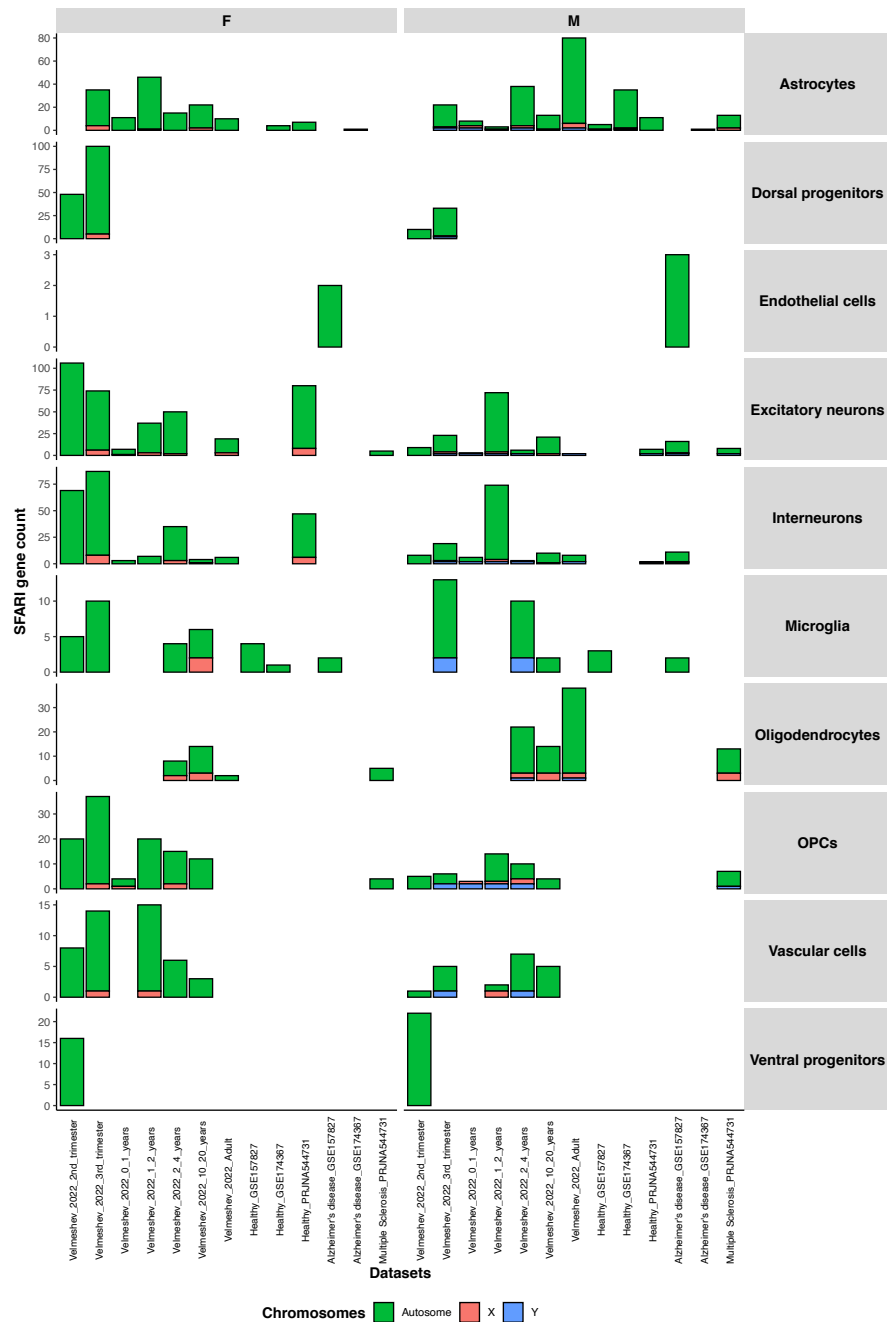


Fig. S48: Most SG-biased DEG lists contained genes related to autism, based on the SFARI reference database, and such genes were mostly autosomal. These bar plots show the SFARI absolute gene count found in each SG-biased DEG lists, divided by sex, cell type and dataset. Overall, the gene counts were the highest in female-biased genes in the fetal datasets, while the male-biased genes showed higher counts in later datasets. Additionally, most of the SFARI genes were autosomal, and only few SFARI genes belonged to the X or Y chromosomes.

Neuropsychiatric diseases

	Neuropsychiatric diseases		
	F	M	
Datasets	Velmeshev_2022_3rd_trimester Velmeshev_2022_0_1_years Velmeshev_2022_1_2_years Velmeshev_2022_2_4_years Velmeshev_2022_10_20_years Velmeshev_2022_Adult Healthy_GSE157827 Healthy_PRLNA544731 Alzheimer's disease_GSE174367 Multiple Sclerosis_PRLNA544731		Astrocytes
	Velmeshev_2022_2nd_trimester Velmeshev_2022_3rd_trimester		Dorsal progenitors
	Alzheimer's disease_GSE157827		Endothelial cells
	Velmeshev_2022_2nd_trimester Velmeshev_2022_3rd_trimester Velmeshev_2022_0_1_years Velmeshev_2022_1_2_years Velmeshev_2022_2_4_years Velmeshev_2022_10_20_years Velmeshev_2022_Adult Healthy_PRLNA544731 Alzheimer's disease_GSE157827 Multiple Sclerosis_PRLNA544731		Excitatory neurons
	Velmeshev_2022_2nd_trimester Velmeshev_2022_3rd_trimester Velmeshev_2022_0_1_years Velmeshev_2022_1_2_years Velmeshev_2022_2_4_years Velmeshev_2022_10_20_years Velmeshev_2022_Adult Healthy_PRLNA544731 Alzheimer's disease_GSE157827		Interneurons
	Velmeshev_2022_2nd_trimester Velmeshev_2022_3rd_trimester Velmeshev_2022_2_4_years Velmeshev_2022_10_20_years Healthy_GSE157827 Healthy_GSE174367 Alzheimer's disease_GSE157827		Microglia
	Velmeshev_2022_2_4_years Velmeshev_2022_10_20_years Velmeshev_2022_Adult Multiple Sclerosis_PRLNA544731		Oligodendrocytes
	Velmeshev_2022_2nd_trimester Velmeshev_2022_3rd_trimester Velmeshev_2022_0_1_years Velmeshev_2022_1_2_years Velmeshev_2022_2_4_years Velmeshev_2022_10_20_years Multiple Sclerosis_PRLNA544731		OPCs
	Velmeshev_2022_2nd_trimester Velmeshev_2022_3rd_trimester Velmeshev_2022_1_2_years Velmeshev_2022_2_4_years Velmeshev_2022_10_20_years		Vascular cells
	Velmeshev_2022_2nd_trimester		Ventral progenitors
Disease-associated genes			
Genes found		Yes No	

Fig. S49: Neuropsychiatric disease-related genes were mainly found in female-biased genes. This heatmap shows the presence of neuropsychiatric disease-associated genes (DAGs) among the SG-biased genes. Only two genes were found among the DAGs, KDM6A and PCDH11X, both related to AD. Interestingly, both were found mainly from fetal to young adulthood datasets, but not in the disease datasets. **Abbreviations:** AD: Alzheimer's disease; ASD: autism spectrum disorder.

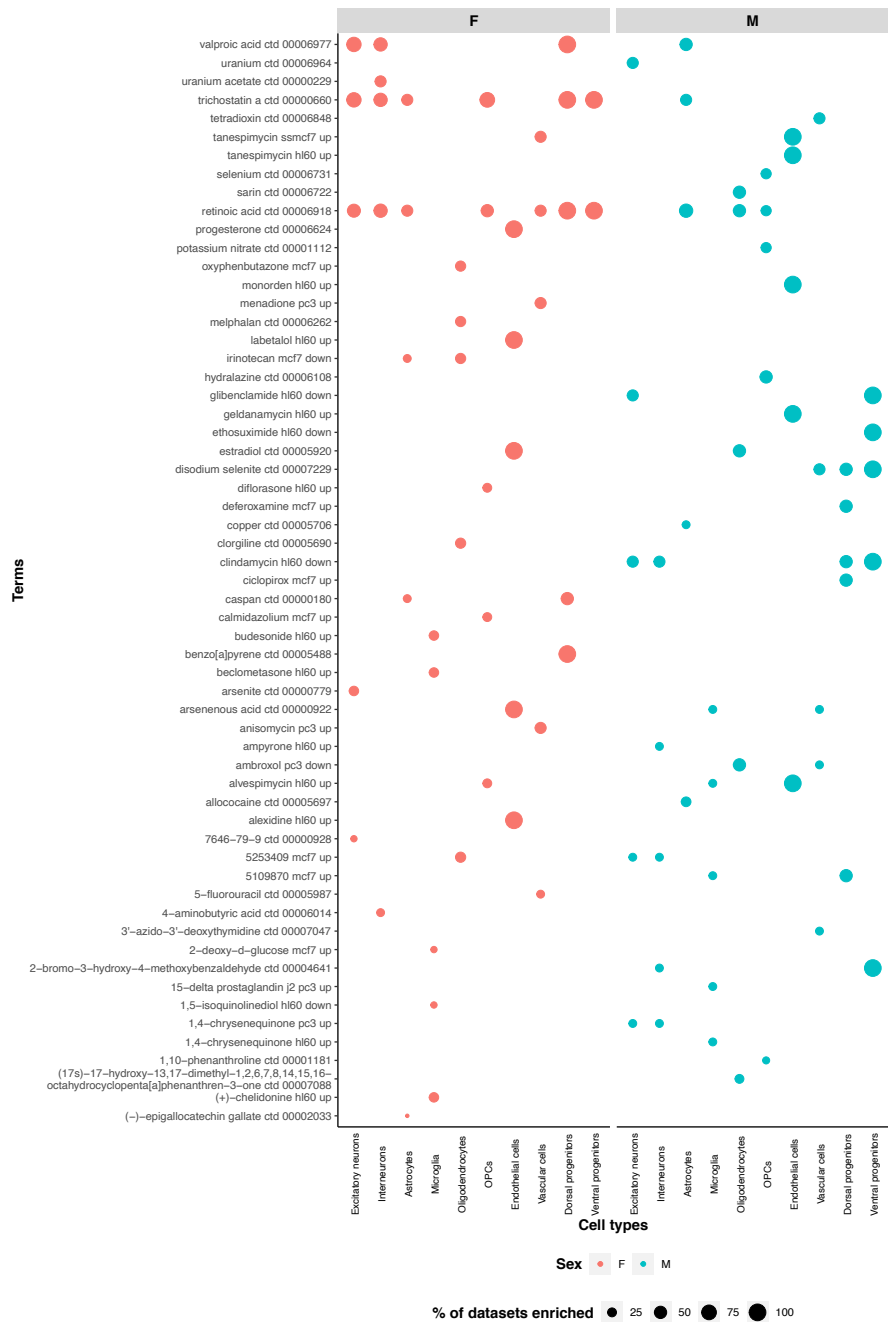


Fig. S50: Most enriched drug terms were sex-specific, with some exceptions. These dot plots show the top 5 enriched drugs, divided by sex and cell type. The dot size indicates how many datasets were enriched for each term, and the color the sex where the term was found to be enriched in.

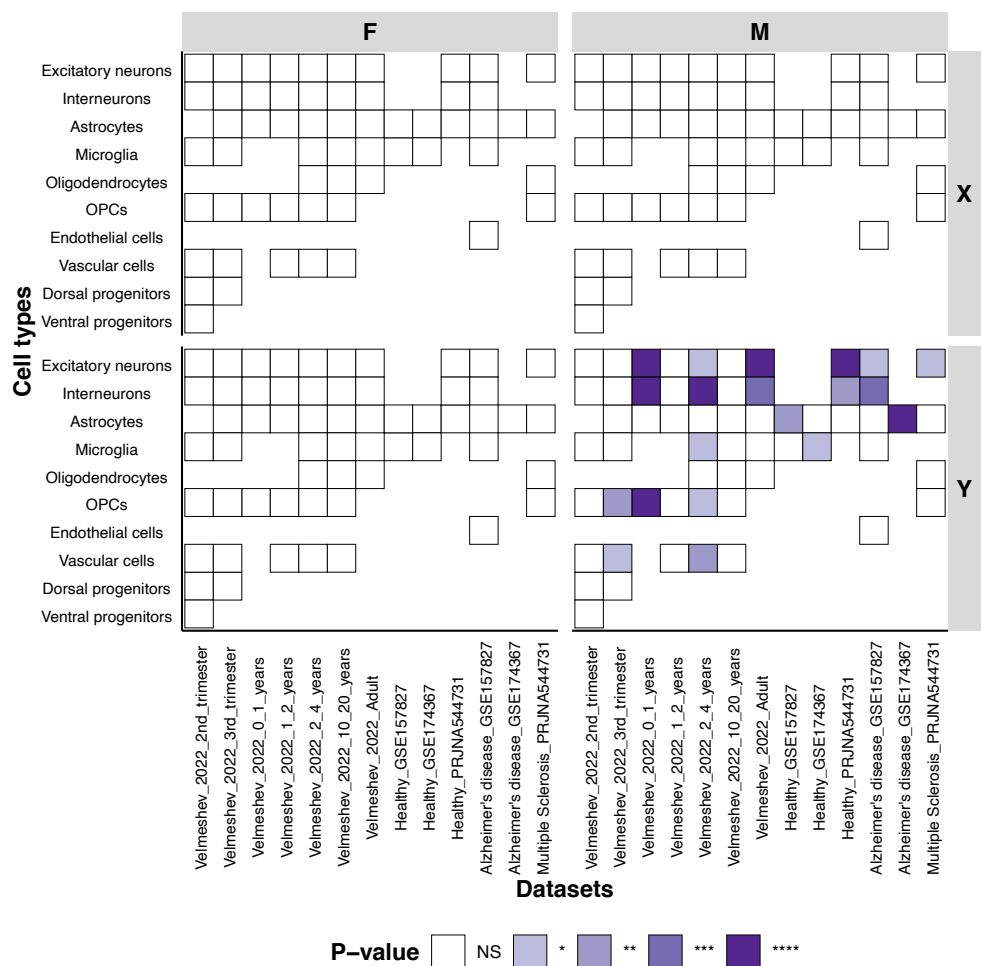


Fig. S51: There was little Y chromosome enrichment in the male-biased genes, and no enrichment for X in female-biased genes. This heatmap represents the significant hyper-geometric enrichment of sex chromosomes in the SG-biased genes. Grey cells represent cell types not analyzed in the datasets. The only enrichment we could find was for Y in males, as expected, although we found enrichment specifically in neuronal and glial populations. **Legend:** NS: not significant; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$.

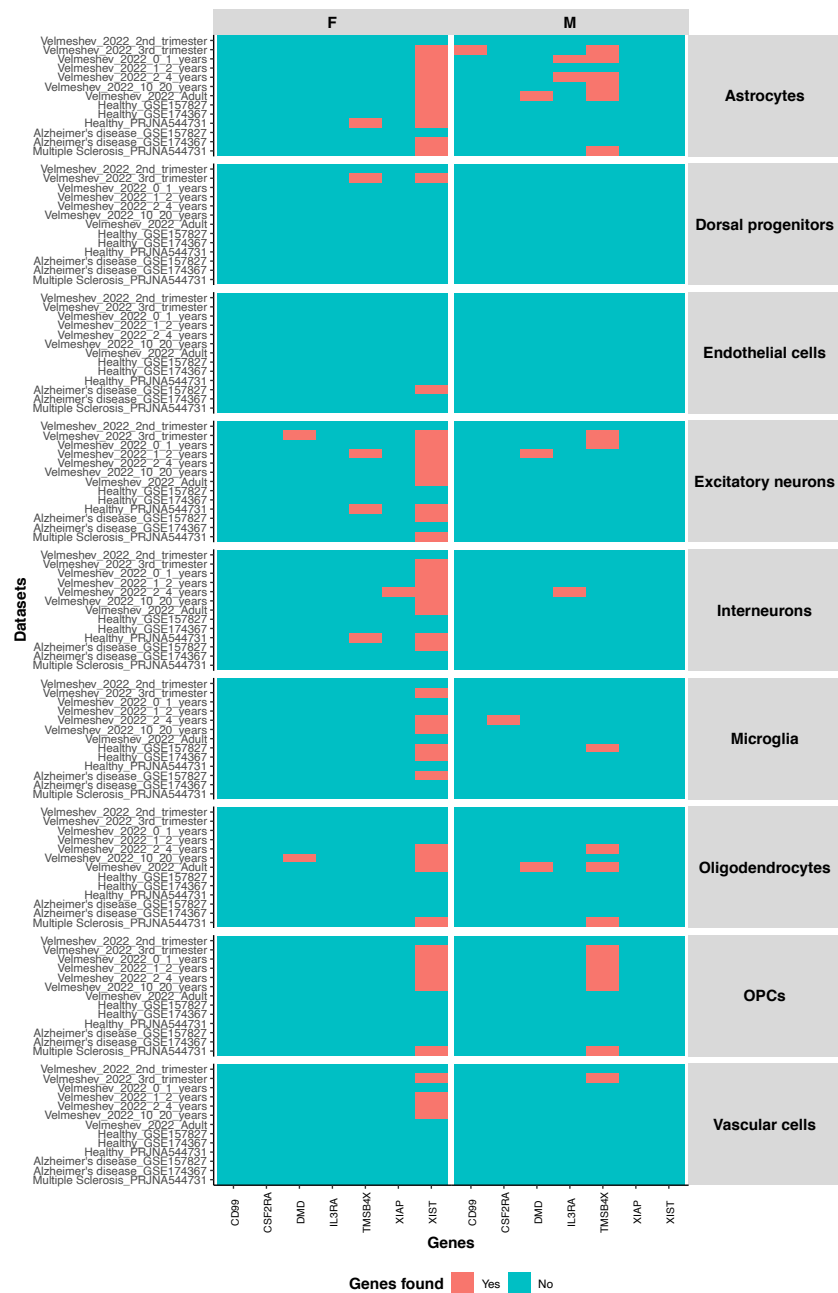
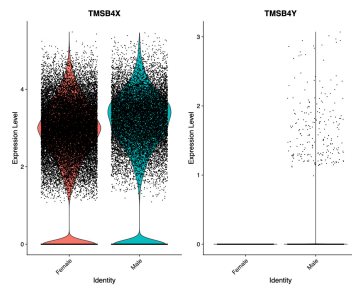
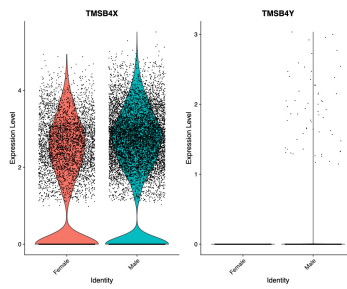


Fig. S52: XCI genes are mostly not found in either females or males, with the exception of XIST and TMSB4X. The heatmap show the presence of genes X-inactivation in SG-biased genes lists from each cell type and dataset. As expected, XIST is solely found in the female-biased genes. However, TMSB4X, which we expected to behave as XIST, was instead mostly found in male- and not female-biased genes.

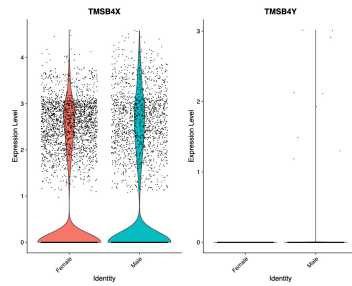
Velmeshev 3rd trimester



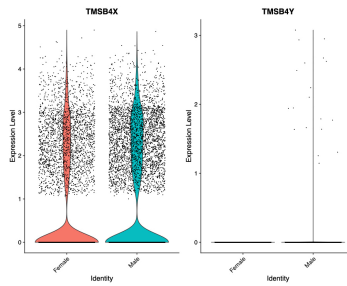
Velmeshev 0-1 years



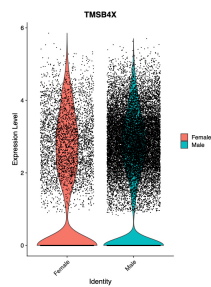
Velmeshev 1-2 years



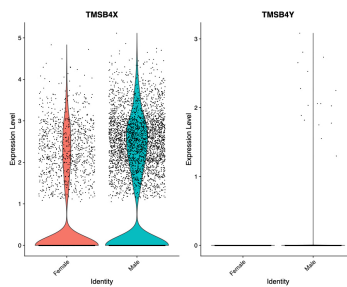
Velmeshev 2-4 years



Velmeshev 10-20 years



Velmeshev Adults



DISCO

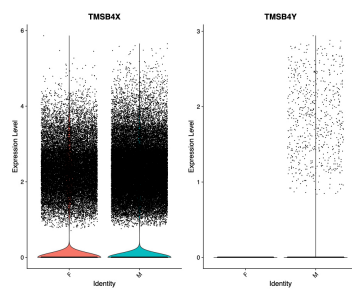


Fig. S53: Expression of TMSB4X and TMSB4Y in the sexes in each dataset.
The X- and Y-linked genes were expressed in almost every dataset, indicating that the biased expression we observed for TMSB4X is due to actual differential expression and not other causes, e.g. misalignment. As seen previously, the second trimester did not have X and Y chromosome genes, therefore is missing.

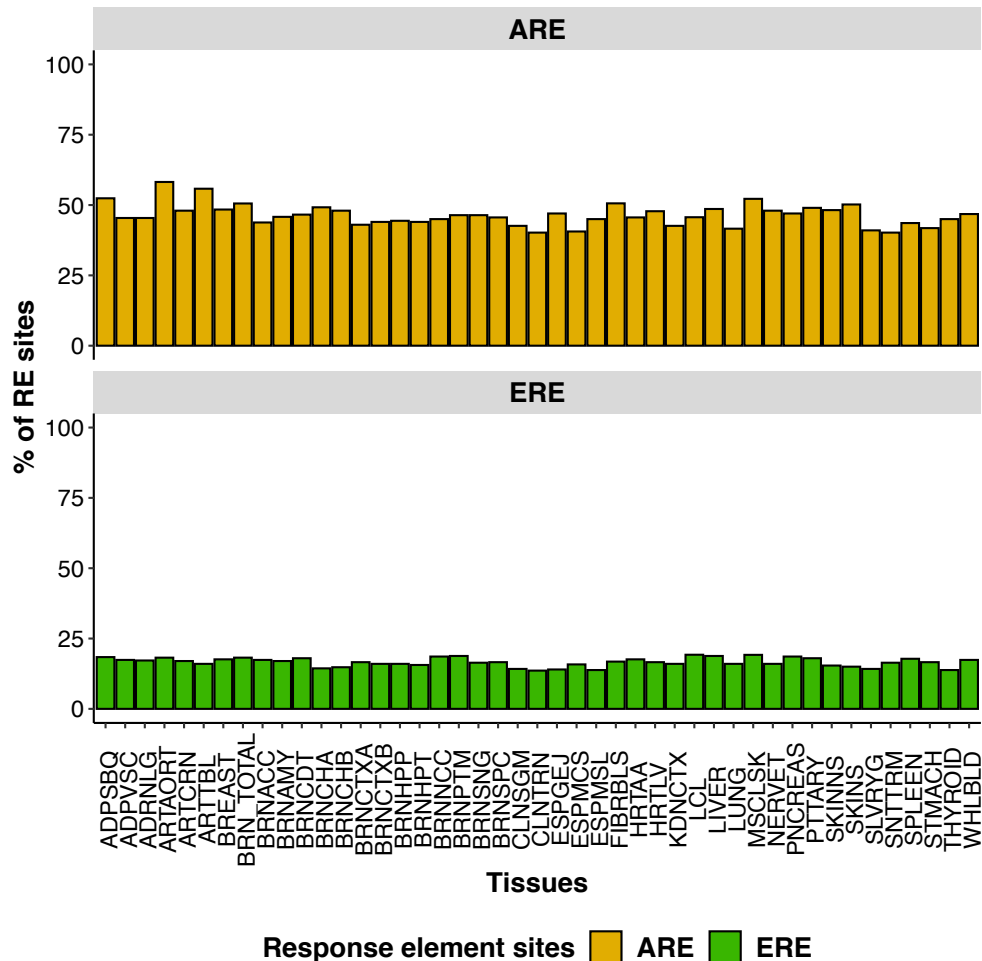


Fig. S54: Percentage of the top 500 human organ-specific genes which express sex hormones response elements. The organ-specific genes were calculated and ranked according to local false sign rate previously (Oliva et al., 2020). For each of the organs/tissues, we calculated how many genes were presenting AREs (top) and EREs (bottom). No modifications were performed on the groups or the labeling from the original study, except for the introduction of 'BRN_TOTAL', in which we grouped together the distinct genes from all brain regions combined, to investigate whether we could a overall brain-specificity. **Abbreviations:** ARE: androgen response element; ERE: estrogen response element; ADPSBQ: Subcutaneous adipose; ADPVSC: Visceral omentum; ADRNLG: Adrenal gland; ARTAORT: Aorta; ARTCRN: Coronary artery; ARTTBL: Tibial artery; BREAST: Breast mammary tissue; BRN_TOTAL: All brain regions; BRNACC: Anterior cingulate cortex; BRNAMY: Amygdala; BRNCDT: Caudate [basal ganglia]; BRNCHA/BRNCHB: Cerebellum/Cerebellar hemisphere; BRNCTXA/BRNCTXB: Cortex/Frontal cortex; BRNHPP: Hippocampus; BRNHPT: Hypothalamus; BRNNCC: Nucleus accumbens [basal ganglia]; BRNPTM: Putamen [basal ganglia]; BRNSNG: Substantia nigra; BRNSPC: Spinal cord [cervical c-1]; CLNSGM: Sigmoid colon; CLNTRN: Transverse colon; ESPGEJ: Gastroesophageal junction; ESPMCS: Esophagus mucosa; ESPMSL: Esophagus muscularis; FIBRBLS: Cultured fibroblasts; HRTAA: Atrial appendage; HRTLIV: Left ventricle; KDNCTX: Kidney cortex; LCL: lymphoblastoid cell lines; LIVER: Liver; LUNG: Lung; MSCLSK: Skeletal muscle; NERVET: Tibial nerve; PNCREAS: Pancreas; PTTARY: Pituitary; SKINNS: Not sun-exposed skin (suprapubic); SKNS: Sun-exposed skin (lower leg); SLVRYG: Minor salivary gland; SNTTRM: Small intestine terminal ileum; SPLEEN: Spleen; STMACH: Stomach; THYROID: Thyroid; WHLBLD: Whole blood.



Fig. S55: Androgen response element (ARE) sites are strongly present across cell types, datasets and sexes. These bar plots show the ARE sites expression in the SG-biased genes, divided in full, half and full-half sites. ARE sites percentages were overall around 75%, much higher than expected, with few exceptions (e.g. female-biased genes in the neuronal populations in the AD GSE157827 dataset). The majority of SG-biased genes had both half and full sites for AREs. **Abbreviations:** ARE: androgen response element.

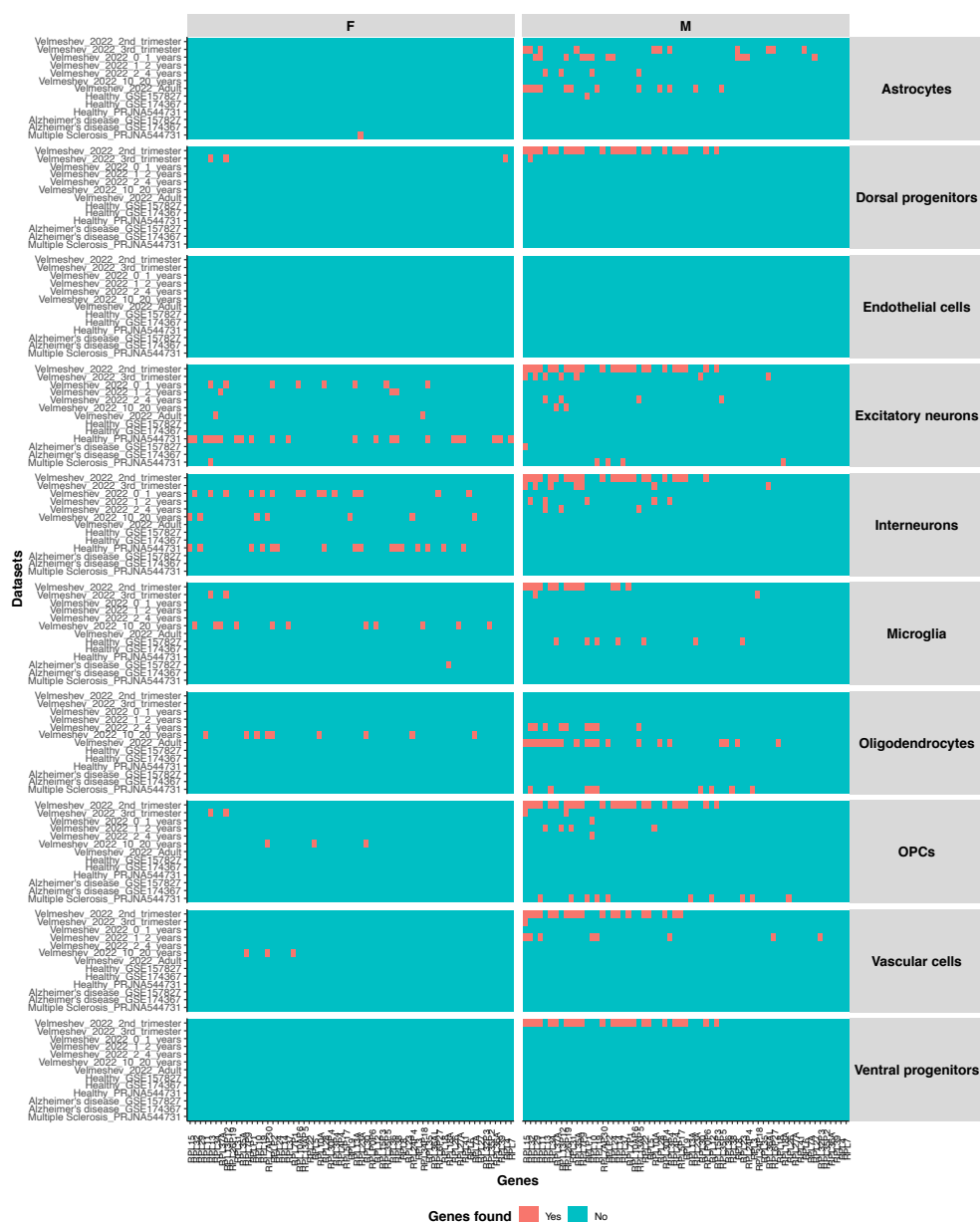


Fig. S56: Presence of L ribosomal proteins (RPL) genes in the SG-biased genes. Heatmap showing the presence of RPL genes across cell types and sex within each dataset. Presence indicates whether the gene is found in each specific dataset-sex-cell type combination.

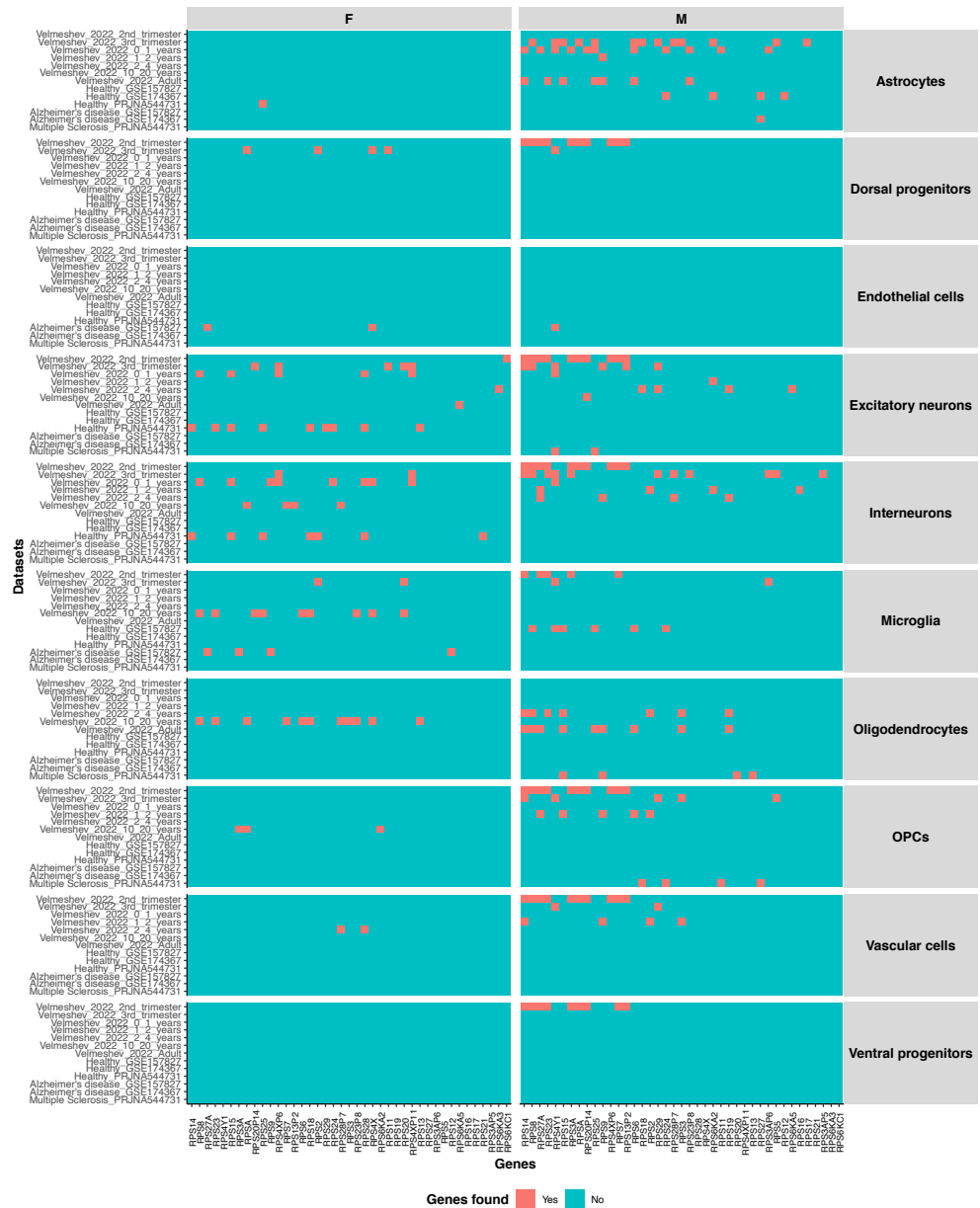


Fig. S57: Presence of S ribosomal proteins (RPS) genes in the SG-biased genes. Heatmap showing the presence of RPS genes across cell types and sex within each dataset. Presence indicates whether the gene is found in each specific dataset-sex-cell type combination.

