Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Single-cell analysis of sex and gender differences in the human brain during development and disease

Aura Zelco (aura.zelco@gmail.com)
Department of Clinical Science, Computational Biology Unit, University of Bergen, Bergen, Norway
https://orcid.org/0000-0002-5851-6355
Anagha Joshi (Anagha.Joshi@uib.no)
Department of Clinical Science, Computational Biology Unit, University of Bergen, Bergen, Norway

Research Article

Keywords: sex and gender differences, human brain, transcriptome, single-nucleus RNA-seq, brain development, data integration

DOI: https://doi.org/10.21203/rs.3.rs-3023302/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Single-cell analysis of sex and gender differences in the human brain during development and disease

Aura Zelco[1] and Anagha Joshi[1*]

[1*]Department of Clinical Science, Computational Biology Unit, University of Bergen, Bergen, Norway.

*Corresponding author(s). E-mail(s): anagha.joshi@uib.no; Contributing authors: aura.zelco@gmail.com;

Abstract

Sex and gender differences in human brain are of interest to society and science as numerous processes are impacted by them, including brain development, behavior, and the occurrence, prevalence, and therapeutic response of diseases. By assembling publicly accessible data from the *in utero* to elderly age in healthy, Alzheimer's disease and multiple sclerosis samples, we thoroughly analyzed SG changes through brain development and brain disorders at a single-cell level. We identified and characterised SG-biased genes in ten key types of brain cells across nine age and illness groups, by analysing a total of 419,885 single-nucleus RNA-sequencing samples from 161 human brains (72 females, 89 males). SG-biased genes were located mostly on the autosomes and showed some enrichment for Y chromosome but not X. Most SG-biased genes were unique to cell types and were enriched for cell type specific markers. Furthermore, SG-biased genes showed little overlap across age and disease groups. Despite a very low gene overlap, there was a high functional overlap for SG-biased genes across cell types as well as across age and disease. Female-biased genes were enriched for brain-related processes, and male-biased genes were enriched for metabolic pathways. Mitochondrial genes were consistently female-biased. Both male- and female-biased genes were highly enriched for androgen (not estrogen) response elements and interestingly thymosin targets were enriched consistently only in male-biased genes. Finally, we have created a web resource with full characterisation of SG-biased genes at different thresholds for the scientific community, to validate findings and generate further new hypotheses, available at https://joshiapps.cbu.uib.no/SRB_app/.

Keywords: sex and gender differences, human brain, transcriptome, single-nucleus RNA-seq, brain development, data integration

1 Background

The existence, scope, and importance of sex differences in the human brain are contentious issues in both the scientific community and the public. In the scientific literature, terminology for sex and gender are frequently employed confusingly. Although sex is increasingly being considered a biological variable in scientific investigation, phenotypes often depend on both sex and gender. We use a widely accepted definition of gender as the influence of socio-cultural distinctions and identity, and sex as a dimorphic state based on the existence of X and Y chromosomes in the DNA. Importantly, sex and gender differences in the brain might originate from either the sex chromosomes, the individual's gender identity or both. Therefore, hereafter we will call them as sex and gender (SG) differences.

Differences between sexes and genders (SG) evolve dynamically through developmental processes that impact numerous phenotypes. SG differences in the human brain have implications in a variety of processes, including brain development, behavior, and the presentation, prevalence, and therapeutic response of disease. The brain is not sexually dimorphic, but there are many documented anatomical [1–6] and behavioural [7–9] differences between the sexes and genders, including some controversies (e.g. [10]). Male and female SG differences in behaviour emerge at early age, further widening during adolescence [11]. These differences have an impact on a variety of physiological processes, including the regulation of the immune system, the response to injury or disease, and the ability to produce and respond to hormones [7, 12–17]. These differences are thought to be driven by hormones, genetics, and environmental factors. Hormones play a crucial role in regulating growth, maturation, and behaviour, and levels of hormones vary between males and females, affecting the growth and differentiation of neurons and the formation of neural circuits. For example, higher levels of estrogen in females have been linked to increased brain regions associated with memory and emotional processing, while higher levels of testosterone in males have been associated with spatial ability and aggressive behaviour [18, 19]. Despite the presence of consistent large behavioural sex and gender differences, the molecular or structural correlates of these differences has so far led to a limited success [20].

Big consortia's efforts and the low cost of sequencing technologies allowed for the accessibility of genome-wide expression data across tissues from hundreds of different people. This data enabled identification of SG differences at the whole genome level in a variety of organs and tissues, including the brain [21–23]. Accordingly the studies investigating the role of SG differences in brain have dramatically increased in recent years (Fig. S1), they tend to focus on a specific disease, cell type or brain region to explore SG differences [24–26]. Previous findings indicate that there are small but consistent SG differences in several brain regions, including the cortex [21]. In addition, SG biases in transcriptome and gene regulatory networks have been observed in the second trimester of pregnancy, very early in the development process [27, 28]. During brain development, maturation, and neurological diseases, differences in the number, distribution, and activity of particular cells, such as neurons and immune cells, have been observed between males and females [7, 12–17]. There were however inconsistencies noted between cell type biases across sex and gender [23]. This might be partly because these studies rely on bulk RNA-seq technology, limiting the characterization

of cell type-specific SG differences in a complex organ such as the brain. Therefore, studies are needed to fully understand the nature and extent of sex differences at a single-cell level. Despite many recent studies have generated single-cell/nucleus data in the human brain at across ages and healthy/disease statuses (e.g. [29–34]), they do not specifically explore SG differences.

Importantly, brain region and age contribute more to the global differences in gene expression than sex and gender [35]. It is therefore vitally important to obtain age-stratified samples for identification of sex and gender differences. Accordingly, when the spatio-temporal dynamics of the human brain transcriptome was explored in 16 brain regions, from embryonic development to late adulthood, it demonstrated that developmentally and spatially regulated differences in gene- and exon-level expression exist between male and female brains, albeit in bulk RNA-seq samples [35]. Thus, characterization of the SG differences across brain cell types throughout the life span is still missing. Furthermore, SG affect many regions of brain, therefore brain region choice is of high relevance. The cortex is one of the brain regions where most of anatomical SG differences are found [5, 6, 12, 36], and with the most available datasets for single-cell analysis. In this study, we therefore set out to investigate and characterize in depth the SG differences during the brain maturation, from fetal to adult life, using available human single-nucleus RNA-sequencing datasets from human cortical region. We also included samples from patients diagnosed with Alzheimer's disease and multiple sclerosis, to investigate SG differences in cognitive and neurological diseases.

2 Results

2.1 An atlas of cell type specific SG-biased genes in human brain through life stages and disease

To identify and characterise SG-biased gene expression in human brain at a cellular level, we collected large publicly available single-nucleus RNA sequencing data from 419,885 nuclei (after filtering) from 161 human brain samples (72 females, 89 males) collected during entire human lifespan and disease from studies including both males and females (see section 5, Fig. S2, S3). We generated consistent annotation of the cell types across studies (see section 5.2, Supplementary file 1 for details) and defined 10 major cell types, namely dorsal and ventral progenitors, interneurons, excitatory neurons, astrocytes, T cells, microglia, endothelial cells, oligodendrocytes, oligodendrocyte precursors (OPCs), vascular cells. The data was grouped into 10 human life stages or age groups, namely second trimester, third trimester, 0-1 years, 1-2 years, 2-4 years, 10-20 years and adults (4 datasets). The diseases included were Alzheimer's disease (AD, 2 datasets) and multiple sclerosis (MS, 1 dataset). We validated the sex of the samples from the metadata using X chromosome gene expression (Fig. S4). We thus stratified the samples by biological sex (female/male), however we will henceforth use the sex and gender (SG) terminology because the exact source of the differential expression is uncertain: sex, gender or both. The number of cells varied not only between sexes within the same dataset, but also across different datasets (Fig. S3, S5). To guarantee robust results, we used a threshold of minimum 100 cells in each cell type/sex in each dataset. Thus, some cell types were not further analyzed in some

or all datasets e.g. T cells (Fig. S5). We therefore continued with 10 major cell types; not all were present in all datasets, including the physiological absence of certain developmental cell types in later ages (e.g. dorsal and ventral progenitors).

We obtained SG-biased differentially expressed genes (DEGs) between males and females using default parameters from `Seurat` (see section 5 for details), and generated a total of 260 cell type-specific SG-biased gene lists in human brain through life stages and disease. We further combined them in 166 gene lists by merging sub-cell types to follow a more generalized cell annotation. The number of DEGs varied greatly across the datasets (Fig. 1) with the number of cells in each dataset influencing greatly the number of DEGs (Fig. S6). No major differences in SG-biased genes were noted between males and females within a dataset with some exceptions e.g. excitatory neurons in the healthy adults and brain pathologies.

2.2 Mitochondrial genes show a female-biased gene expression

We firstly explored the genes highly and consistently SG-biased across datasets, both for females and males (Fig. 2). The gene with the highest SG difference in expression was XIST, as expected. The other female-biased SG genes were X chromosome gene (JPX) and mitochondrial genes and an autosomal gene CADM2 (chromosome 3), which has been recently associated with a range of behavioural traits [37]. Similarly, genes located on the Y chromosome formed the majority of uniquely expressed in males. Other three male-biased SG genes (TMSB10, HINT1, STMN1) reside on the autosomal chromosomes. STMN1 has been known to be a marker for immature neurons and it was recently found to have an increased expression of its protein in AD [38]. HINT1 has been found to be up-regulated in the dorsolateral prefrontal cortex of schizophrenic patients [39], and TMSB10 was identified as possible biomarker in glioma [40]. Interestingly, TMSB4X is a X-linked gene, and a target of thymosin (a non-sex hormone), as TMSB10.

As several mitochondrial (MT) genes were consistently found in the female-biased genes across datasets, we explored further the presence of such genes in the SG-biased genes. MT genes were mostly expressed in females across cell types and age groups, with a few exceptions (Fig. S7). MT ribosomial genes such as MT-RNR1 and MT-RNR2 were differentially expressed across many cell types in females. In astrocytes, excitatory neurons and interneurons MT genes consistently expressed more frequently in females. Neurons are known to be enriched for mitochondria, while astrocytes are spatially located preferentially close to synapses. Furthermore, it has been hypothesized that mitochondria are involved in the glucose metabolism of neurons [41] and that they regulate the timing of neuronal development [42]. MT-ATP8, which encodes for ATP Synthase F0 Subunit 8, was the only mitochondrial gene expressed in male-biased genes in the excitatory neurons in the MS dataset, and not found in the female-biased genes. Interestingly, nuclear housekeeping genes belonging to the Translocase Of Inner Mitochondrial Membrane/Translocase Of Outer Mitochondrial Membrane (TIMM/TOMM) complex did not show a strong SG-biased pattern similar to MT genes (Fig. S7).

Studies have shown before a female bias regarding mitochondrial activity, and it has been suggested that also enzymes involved in the tricarboxylic acid (TCA) cycle

gene expression is sex biased as the mitochondrial complex II is involved in both the respiratory chain and the TCA, using succinate as a substrate [43]. We therefore investigated whether TCA enzymes (https://maayanlab.cloud/Harmonizome/gene_set/TCA+cycle/PANTHER+Pathways) showed SG-biased expression and found no such bias (Fig. S7). This suggests that the SG difference in the MT genes were likely not due to difference in number of mitochondria or TCA-related expression, but rather reflecting differences in energy demand.

We further systematically explored whether SG-biased genes were preferentially located in specific cellular locations. We obtained cellular localization data from the Cell Atlas [44]. Interestingly, male-biased genes were enriched in endoplasmic reticulum in males, especially in non-neuronal populations (Fig. S8).Most of the SG-biased genes were present in the cytosol, and we noted very little difference in cellular location distribution between females and males, and across the datasets, except for the mitochondria, more frequent in females than in males (Fig. S8).

2.3 SG-biased genes are largely cell type specific

We checked whether the SG-biased genes were shared across cell types in each dataset by calculating the overlap of gene lists across cell types. We noted that most SG-biased genes were highly specific to each cell type (Fig. S9, S10, S11, S12, S13). However, across datasets, SG-biased genes showed higher overlap with cell types closely associated in developmental lineage, as expected. Specifically, developmentally closely related cell types showed a high SG-biased genes overlap than distant related cell types. e.g. excitatory neurons and interneurons displayed a higher overlap of SG-biased genes compared to excitatory neurons and other non-neuronal cell types (Fig. S9, S10, S11, S12, S13).

We further explored this by using the original cell type annotation from individual datasets, which annotated further sub cell types. For example, in DISCO datasets, the neuronal population is further classified into many sub types for both excitatory neurons and interneurons. Similar to previous observation, the SG-biased genes were overall cell type-specific, even within the same major cell type. For example, the various sub cell types of excitatory neurons showed distinct DEGs, with very few genes shared across the sub populations (Fig. S9, S10, S11, S12, S13).

Cell type-specific SG-biased genes belonged to the autosomal genes, and genes located on the X- and Y chromosomes tended to be shared across cell types (Fig. S9, S10, S11, S12, S13). One of the most shared SG-biased genes across all cell types and datasets was XIST, which is not surprising giving its specific expression in females (Fig. 2, S4). The same was observed for Y-expressed SG-biased genes such as DDX3Y and USP9Y.

We observed a cluster of shared male-biased genes across cell types in the second trimester dataset (Fig. 3A). The functional enrichment analysis of these shared genes showed that they were mainly involved in ribosomal and neurological diseases, including developmental pathologies (Fig. S14). To validate this interesting observation, we collected SG-biased genes from bulk and integrated single-cell RNA-sequencing data from second trimester samples [27, 33, 34], and compared with the second trimester

datasets present in our analysis (Fig. S15). To firstly estimate the overlap of SG-biased genes across datasets, we pooled the cell types in the single-cell studies, similar to the experimental settings of the bulk RNA-seq data. We then compared both the female- and male-biased genes from our second trimester data with the bulk RNA-seq [27]. Most male- and female-biased genes in the second trimester did not overlap with the O'Brien data. The same was observed when the Eze-Nowakowski integrated samples were compared (Fig. S15). CKS1B and FSTL1 were the common genes between the male-biased genes and the O'Brien data. CKS1B and FSTL1 are associated with metabolic functions such as cyclin-dependent kinases and calcium ion binding. No common female-biased genes were found in the three datasets (Supplementary file 2).

As SG-biased genes showed little overlap across cell types, we checked for the cell type-specific enrichment for known cell type markers, comparing the SG-biased genes with known genes from previously published data [45]. Overall, for each reference cell type gene signature (astrocytes, endothelial cells, microglia, neurons and oligodendrocytes) a stronger overlap was observed in the corresponding cell types in our datasets. One exception was astrocytes from healthy females and MS male patients showed higher percentage of endothelial cells and neuron markers respectively (Fig. 3B).

Since only samples from the cortex were included in this study, we investigated whether the SG-biased genes were specific to the cortex. A recent bulk RNA-seq study investigated sex differences in multiple tissues, including several brain regions [21]. We calculated the enrichment of the DEGs from this study in SG-biased genes, separately for each brain region. We found that the DEGs from the study were overall enriched in the female- but not in the male-biased genes (Fig. S16). Interestingly, the enrichment was focused in the early datasets, but not in the age-comparable datasets to the original study (healthy adults, Fig. S16). Moreover, SG-biased genes were found to be enriched in all brain region-specific DEGs from the study i.e. SG-biased genes were not cortex-specific (Fig. S17).

2.4 SG-biased genes are developmental stage specific

We further noted that most SG-biased genes were developmental stage specific (Fig. 3C, S18, and a very few DEGs for each cell types were shared among at least 75% of the datasets (Supplementary file 3). Most shared genes across datasets in a given cell type were located on the X and Y chromosomes (Supplementary file 3), while the majority of the unique genes were located on the autosomes, similar to shared genes across cell types in a specific age group described in previous section. However, some autosomal genes were shared SG-biased genes across datasets, such as RGS1 in females (microglia), CADM2 and MT genes (interneurons) in females, and several genes in vascular cells (e.g. STMN1 in males) and oligodendrocytes, both in females and males. An increased RGS1 expression 6 hours after birth was found to be correlated with a worse outcome at 18 months after neonatal encephalopathy, where 65% of neonates was male [46].

In each cell type for each SG, gene overlap analysis showed that the groups closer in age had a higher overlap of genes compared to the other groups (Fig. S19, S20, S21, S22). Additionally, the shared genes were mainly autosomal, although the genes shared by most datasets belonged to X or Y chromosomes (Fig. S23). For some datasets,

such a temporal relationship was not noted. For example, SG-biased genes in MS and AD datasets had a higher overlap with SG-biased genes at younger ages than adult in astrocytes and microglia (Fig. S21, S22). Overall, significant overlap of SG-biased genes across related developmental stages across multiple cell types provided another proof of confidence in the SG-biased gene lists.

Thus, the SG-biased genes not only were specifically expressed in each cell type across datasets, but also showed low overlap across age and disease groups within a cell type. A few highly shared genes across datasets mostly belonged to the X and Y chromosome.

2.5 Functional implications of SG-biased genes

To understand the likely functional implications of SG-biased genes, we calculated gene ontology (GO) enrichment for the biological processes (BPs) terms for all SG-biased gene lists. We firstly checked overlap of enriched BPs across cell types in each group. Though there was not much gene overlap of SG-biased genes across cell types and datasets (Fig. S19, S20, S21, S22, S23), many BPs were shared among different cell types, indicating similar pathways enriched at specific dataset (Fig. 4A-B, S24, S25, S26, S27, Supplementary file 4). Across datasets, female-biased enrichment was mainly focused on brain-related BPs, such as neuron development, axonogenesis and synapse-related terms. Instead, BP enrichment for males were mainly related to metabolism and cellular respiration. This difference was particularly pronounced in the earlier age datasets (second trimester to 2-4 years), while in later age datasets we observed brain-related functions enriched also in males, and especially in the MS and AD disease datasets (Fig. 4A-B, S24, S25, S26, S27). We checked the overlap of GO BP enrichments across the age groups for each cell type as well (Fig. 4C-D, S28, S29, S30, Supplementary file 5). We again noted a high overlap of BP terms across datasets (Fig. 4C-D, S28, S29, S30). Excitatory neurons from females consistently enriched in the synapse- and axon-related BPs, which were not enriched in the excitatory neurons in males. Some BPs were enriched only in a specific developmental stage as well. For example, respiratory electron transport chain and other metabolic processes were enriched in excitatory neurons in males during the fetal stage, but not later. Moreover, these processes were not enriched in the excitatory neurons in females in any dataset (Fig. 4C-D, S28, S29, S30). We noted a consistent pattern across cell types, where male-biased genes mainly showed enrichment for metabolic processes, while in females the majority of enriched BPs were related to brain processes. For example, astrocytes in females showed an enrichment in axon development and axonogenesis, while in males the enriched BPs were mostly metabolism-focused processes. In previous studies, female-biased genes were also found to be mostly related to brain functions, while male-biased genes were more enriched in metabolism and nucleus-related terms [23].

We further investigated enriched pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. We noted a high overlap of enriched pathways across cell types in the fetal datasets, and cell types from older age groups did not display the same level of overlap (Fig. S31, S32, S33, S34, Supplementary file 6). Interestingly, most of the terms across all datasets were diseases and pathologies rather than

physiological pathways, both neurological diseases and infections from a variety of pathogens, among which the most recurrent were prion, Parkinson's and Coronavirus diseases (Fig. S35, S36, S37, Supplementary file 7).

We also investigated the enrichment of transcription factor (TF) binding sites using the TRANSFAC and JASPAR PWMs database in `enrichR` [47, 48]. When TF enrichments for the cell types in each dataset were compared, TCFAP2A binding sites were enriched in female DEGs in 50% of the datasets. No highly shared TF binding sites were found in males (Fig. S38, S39, S40, Supplementary files 8 and 9). Similarly, the overlap of TF enrichments in each cell types across datasets showed that only two TFs were enriched at least in 50% of the cell types: TCFAP2A and POU1F1, both in females and no TFs were found in 50% of the cell types in males (Fig. S41, S42, Supplementary files 10 and 11). TCFAP2A target enrichment was found in excitatory neurons, interneurons, astrocytes, OPCs and dorsal progenitors, and mostly during the early development, from the second trimester of gestation to 2-4 years of age, and also in female MS patients (Fig. S41, S42). TCFAP2A enrichment was also found in males, namely in excitatory neurons and interneurons at 10-20 years of age. While TCFAP2A is expressed ubiquitously, the knowledge of its function in the brain is limited to early facial development [49]. POU1F1 targets instead were almost uniquely enriched in females in early development (from the second trimester until 1-2 years of age). POU1F1 was found enriched in females across many cell types: excitatory neurons, interneurons, astrocytes, dorsal and ventral progenitors. POU1F1 was also found in male astrocytes and oligodendrocytes, respectively in healthy adults (GSE174367) and Velmeshev 10-20 years of age. POU1F1, also known as Growth Hormone Factor 1, is very important in early mammalian development, and its mutations can lead to pituitary hormone deficiencies and dwarfism [50, 51]. Therefore, despite being located on the autosome (chromosome 3), POU1F1 may be part of the gene regulatory network which promotes faster maturation in females compared to males during the early life.

2.6 Pathological implications of SG-biased genes

As SG-biased genes showed pathology-related term enrichment in pathway analysis, we explored further using the following disease databases: disease ontology (DO), DisGeNet and DisGeNET CURATED, and GWAS catalog 2019 (Supplementary files 12, 13, 14 and 15). After calculating the enrichment in each disease database, we selected the most common enriched terms in females and males in all datasets (Fig. 5A, Supplementary file 16). The most common enriched terms for females were drug use disorders and abuse, smoking, autistic behaviour and amyotrophic lateral sclerosis, among other brain-related pathologies. In males, the most common enriched disease terms were neoplasms, epilepsy, Alzheimer's disease onset and other brain-related pathologies as well as for cancer-related terms. Additionally, we investigated the most common enriched terms for each cell type (Supplementary file 17). In females, astrocytes had enrichment for amyotrophic lateral sclerosis and androgen-insensitivity syndrome, while in males epilepsy-related terms were mostly enriched (Fig. 5A, S43). Excitatory neurons and interneurons were enriched in females for autistic behaviour and drug addiction-related disorders, while males displayed enriched terms for cancer-related terms, spermatogenic failure and epilepsy.

Autism is typically considered a male-biased disorder [52]. In contrast, we noted female-biased genes enriched for autism. We therefore used the SFARI autism genes database (https://gene.sfari.org/database/human-gene/) [53] and calculated the number of SG-biased genes involved in autism (Fig. S44, S45). During early development (the second trimester and 1-2 years of age), nearly all cell types showed a higher autism-related gene count in females compared to males (Fig. S44, S45). In older age groups, either both SGs had a similar number of autism-related genes, (e.g. in microglia), or males had more autism-related genes (e.g. in endothelial cells, and astrocytes). Additionally, most of these genes were located on the autosomes, indicating that the sex chromosomes were not the source of this bias (Fig. S45).

We also investigated the enrichment of disease-associated genes from the literature [54] in our SG-biased genes. Notably, we found higher overlap of disease-associated genes with the female-biased genes than with the male-biased (Fig. S46). Two genes, KDM6A and PCDH11X had higher expression in females, and have been associated with a higher protection from the brain-related diseases [54]. Furthermore, KDM6A was expressed more ubiquitously in glial cells (astrocytes, oligodendrocytes), microglia and vascular cells (Fig. S46), and was expressed mainly in the younger datasets (third trimester until 20 years of age). PCDH11X showed a cell type-specificity, mainly neurons (Fig. S46) and was predominantly in the early childhood.

Lastly, we investigated if the SG-biased genes were enriched for drug targets, using the DSigDB database from `enrichR`. We noted that most of the enriched drug targets were present in both sexes (Fig. 5C, Supplementary files 18 and 19). The three most frequent enriched drugs in females were trichostatin, retinoic acid and valproic acid targets. Moreover, they were enriched in nearly all cell types in females. Males showed enrichment for these drugs as well, although to a lesser extent (Fig. S47, Supplementary file 20). Trichostatin is a histone deacetylase (HDAC) inhibitor, involved in epigenetic control of brain gene expression and masculinization [55, 56]. Valproic acid is another HDAC inhibitor, and its administration during pregnancy has been linked to an increased incidence of autism [57, 58]. Retinoic acid is a metabolite of vitamin A, and deficiency of this vitamin is associated with both physiological cognitive decline as well as well as AD [59] and autism spectrum disorder [60, 61].

2.7 Origins of the sex bias

We finally explored likely source of SG differential gene expression. We have noted before that the majority of SG-biased genes were located on the autosomes (Fig. S9, S10, S11, S12, S13, S23). Sex-biased genes in some cell types (vascular cells, OPCs, microglia, interneurons and excitatory neurons) were enriched for the Y chromosome for the male-biased genes, while no enrichment was found for the X chromosome (Fig. S48). In previous microarray results, 91% of SG-biased genes in the dorsolateral prefrontal cortex were autosomal, higher than in other regions [62], suggesting that indeed the sex bias in the cortex may be less dependent on the sex chromosomes than in other brain regions. We further explored if the genes escaping X chromosome inactivation (XCI) [63] had a female-bias (Fig. S49). Firstly, only a few XCI genes were expressed SG differentially in brain. Furthermore, TMSB4X was more expressed in males than in females (Fig. 2). TMSB4X has been previously reported to be more

expressed in patients with bipolar and major depressive disorders [64], and showed promises of being a biomarker in childhood brain tumour [65]. Interestingly, its Y-linked counterpart, TMSB4Y, was ubiquitously expressed at much lower levels in all datasets (Fig. S50).

Sex-divergent brain development is believed to be caused primarily by the testosterone exposure *in utero*, where early testosterone exposure impacts social behaviours e.g. sexual orientation and gender identity, and brain structure and functions [16, 66]. We calculated the enrichment of the targets of 53 hormones, including testosterone, in the SG-biased genes, using a reference database [67]. The hormones enriched in more than one dataset were likely to be true positives (Fig. 6, Supplementary file 21). Overall, many hormone targets were found in several cell types but in females only, with one main exception (thymosin). Interestingly, testosterone targets were enriched in one of AD groups in microglia in females and in vascular cells during the second trimester, also in females. No enrichment was observed for testosterone targets in males. Estradiol-oestradiol and progesterone both showed enrichment in microglia in one of the AD datasets. Corticotropin releasing hormone also showed enrichment mostly in females (Fig. 6).

A non-sex hormone thymosin was the most enriched hormone, across multiple cell types but mainly in males (Fig. 6). Previous studies have investigated the role of specifically thymosin β10 and β4 in the developing brain and in neuroembryogenesis [68–70]. These two products of thymosin are encoded respectively by TMSB10 and TMSB4X, which were among the top 10 most unique genes in males (Fig. 2) and the X-escaping genes also expressed in males (TMSB4X, Fig. 2, S49).

We noted no enrichment for testosterone targets, which could have been due to a small number of testosterone targets in the reference database. Sex hormones, such as testosterone and estrogen, influence gene expression through androgen and estrogen nuclear receptors, respectively. These receptors affect the regulation of genes via the presence on promoters of androgen and estrogen response elements (AREs and EREs) [71, 72]. We investigated the enrichment for ERE and ARE binding sites using previously published data [73, 74]. We did not observe an enrichment for ERE sites in either males or females across all datasets (Fig. 7). However, we observed a consistent enrichment of ARE sites in almost all cell types and all datasets (Fig. 7, S51), independent of the total number of SG-biased genes. More specifically, overall 75% and more of SG-biased genes in both females and males showed full, half or full-half ARE sites (Fig. 7, S51). We did not observe striking differences between females and males in any dataset or cell type, indicating that the SG-biased genes are influenced by testosterone, independent of sex, cell type, age or disease. This is even more striking considering the difference in testosterone levels between males and females [75], and the general decline of testosterone levels in males after puberty [76, 77].

In summary, sex difference in the human brain cortex might be driven by the influence of hormones, specifically thymosin and testosterone, than sex chromosomes.

2.8 Rshiny resource to explore for the community

We developed a web resource to provide the entire analysis for the scientific community and demonstrate the robustness of the analysis by allowing different thresholds for

analysis. SGHumanBrainApp (https://joshiapps.cbu.uib.no/SRB_app/) is a web application built with R shiny [78], an interactive tool for the user to explore the datasets included in this study, together with SG analysis from bulk RNA-seq data in brain from a previous study [23]. There are six tabs in the application: a general explanatory tab about the application and the studies included. The second tab provides information about the datasets included in this study (such as number of samples, ages, number of cells, etc). The third tab allows the users to explore the main results from our analysis on the SG-biased genes. In the third tab, users can select among a range of thresholds for adjusted p-value and fold change(FC). Once the thresholds are selected, users can explore an overview of the SG-biased genes (number of SG-biased genes, chromosome fractions, cellular location), heatmaps of genes of interest (most frequent sex-specific DEGs, mitochondrial and X-escaping genes), expression of cell and disease markers, hormone targets enrichment, percentage of genes expressing ARE and ERE sites, and the functional enrichment (GO - biological process, cellular components, molecular function - and KEGG pathways). All the plots can be individually downloaded. Additionally, users can download all the plots, the unfiltered or filtered SG-biased genes, together in one ZIP folder. The fourth and fifth tabs contain the SG analysis from the bulk RNA-seq study [23], where the user can either explore SG DEGs in individual brain regions (fourth tab) or across all brain regions included in the bulk RNA-seq study (fifth tab), also using different thresholds. The sixth tab contains information about the source code to generate the application. The source code of the SGHumanBrainApp application is available at https://github.com/aurazelco/SGHumanBrainApp. The scripts used to generate the output files for the single-cell DEGs tab can also be found in the repository.

3 Discussion

SG affects many regions of brain. As cortex is one of the brain regions with most of anatomical SG differences [5, 6, 12, 36], we focused here on the cortical region to characterize SG differences throughout healthy lifespan from *in utero* to aging, including two diseases, AD and MS, the latter known to be highly SG-biased in incidence and progression [79, 80] using single-nucleus transcriptomic data. Gene expression signature is highly specific to a cell or tissue type and within a cell or tissue type, age affects gene expression more than sex [35]. It is therefore essential to identify sex differences by stratifying according to age and cell type complexity, especially in a complex and ever-changing organ such as the brain. Indeed, SG-biased genes were present in most of the cell populations: neurons (excitatory, interneurons), glial cells (astrocytes, oligodendrocytes, OPCs), astrocytes and vascular cells in most datasets and dorsal and ventral progenitors only in the fetal datasets for physiological reasons. Endothelial and T cells were either too few cells (at least 100 cells for each sex) or had few DEGs to be included for further analysis. The SG-biased genes were cell type-specific, developmental stage-specific, and had a low overlap across datasets. A small overlap in SG-biased genes across cell types and developmental stages are expected as previous studies have shown that indeed the cell types populations separated according to age [29] or

disease [31, 32]. Despite the SG-biased genes were cell type- and developmental stage-specific, the enriched terms for gene ontology, KEGG and diseases highly overlapped. Overall, most female-biased genes showed enrichment for brain-related biological processes and pathways. male-biased genes were more enriched for metabolic functions and pathways. This is in agreement with previous findings where male-biased genes were enriched for membrane and nuclear structures, while female-biased genes were more enriched for neuronal processes [23, 81].

One of the interesting findings was a female bias for mitochondrial genes, across cell types and developmental stages. A recent study showed significant hypomethylation on mitochondrial DNA on multiple *loci* in females compared to males, likely leading to up-regulation of mitochondrial genes [82]. Additionally, estrogens have shown to regulate mitochondrial functions, providing support for a female-biased mitochondrial expression [83]. Mitochondrial gene up-regulation was especially present in neuronal population, as it has shown that mitochondrial metabolism is strictly connected to neuronal development [42], and neurons are among the most energy-demanding cell types in the brain [41].

Disease enrichment analysis showed female-biased genes enriched for autism, amyotrophic lateral sclerosis 1 and drug abuse, while male-biased genes enriched for neoplasms. Interestingly, both sexes showed enrichment for epilepsy, albeit slightly more in male-biased genes. Previous results indicate that male DEGs showed enrichment for autism-related genes [81], however more recent results showed female-biased enrichment for autism-related genes [23]. However, females might be under-diagnosed when it comes to autism, because of the "female camouflage effect", describing when females mask autism symptoms in social situations [84, 85]. Also, "female protective effect" has been shown in autism, where females require more mutations to develop autism [86].

Sex is thought to affect the brain via divergent gene expression from sex chromosomes and the role of by sex hormones [12, 66]. The SG-biased genes were mainly present on the autosomes, with surprisingly little Y and no X chromosome enrichment. However, researchers reported that 91% of the SG-biased genes in the dorsolateral prefrontal cortex were indeed autosomal, more than in other brain regions included in the study [62]. These results suggest that SG differences, at least in the cortex, might originate mainly from the autosomal genes. Importantly we noted that the most common SG-biased genes across cell types and ages were present on sex chromosomes (Fig. 2). This is in agreement with the previous findings that showed X-linked genes most frequently sex deferentially expressed across different brain regions [21, 23]. The current paradigm is that brains are masculinised by the expression of sex-biased genes from conception and pubertal surges in testosterone in males [76, 77]. Similarly female-bias is manifested by the sex-biased genes and estrogen surge at puberty driving the SG differences [8, 52, 87–89]. For instance, it has been suggested that physiological testosterone decline in adult males and its absence in females throughout brain development enhance the risk for AD [90]. We thus focused on sex and non-sex hormones to find potential sources of the SG bias. The enrichment analysis of SG-biased genes for hormone targets, identified a few hormones. We observed microglia from female AD patients enriched for testosterone, progesterone, and estradiol-oestradiol targets.

Corticotropin releasing hormone was enriched across multiple cell types in females. Surprisingly, sex hormone targets were not highly enriched in the SG-biased genes. A possible reason for some hormones not being enriched might be that the hormone-target database was rather conservative and furthermore the expression of hormones through brain regions varies greatly, e.g. oxytocin [91]. Both estrogen and androgens (testosterone), can influence expression via nuclear receptors and specific response elements, called androgen and estrogen response elements (ARE, ERE) [71, 72]. AREs and EREs have documented sex biased expression in the brain, with females expressing more EREs and males more AREs [92], and it has been hypothesized before that estrogen through EREs can influence SG gene expression differences in mice [93]. We found that almost ubiquitously, the SG-biased genes presented more than 75% of ARE binding sites, while the percentage of ERE sites was less than 25%, with no difference between female- and male-biased genes for both AREs and EREs. Our results agree with previous findings, where similar percentages were found across multiple brain regions [23]. Our findings suggest a similar mechanism, however testosterone- and not estrogen-dependent, in which testosterone may differentially influence the expression in both females and males.

However, the most striking sex bias was observed for thymosin. Thymosin targets were enriched in SG-biased genes in 9 of 10 cell types, across multiple ages and both in healthy individuals and MS patients. Thymosin is a hormone produced by the thymus, for the maturation process of T cells. However, its products has been found in other tissues, and have been studied in relation to brain pathologies [40, 64, 65, 94]. Indeed, TMSB10 and TMSB4X, thymosin products, were among the most frequent male-biased genes in our analysis. Interestingly, TMSB4X is an X-escaping genes [95]. However, contrary to the other X-escaping genes, were mostly absent from the SG-biased genes (XIST excluded), it was found more among the male-biased genes. TMSB4X has been studied as a candidate biomarker for childhood brain tumours [65], and it has been shown to be up-regulated in bipolar disorder and even more in major depressive disorder [64]. On the other hand, researchers have shown that thymosin $\beta 4$, encoded by TMSB4X, has neuroprotective properties, and it has been hypothesized it could a therapeutical target for neurological diseases [94, 96]. However, the role of TMSB4X in both physiological and pathological conditions in the sexes is still not fully characterised.

The lack of sex-stratified studies, partly because of long tradition of male-biased research sampling in pre-clinical studies and clinical trials [97], made difficult to compare some of the novel results, such as ARE enrichment in our sex-biased DEGs. We validated the consistency of our finding by applying various thresholds (explorable in the web application). Another caveat in the analysis is variability is the sampling of brain region. While all the datasets and samples/cells included were from the cortex, the cortex itself can be divided in sub-regions. The lack of detailed sample annotation in the metadata from the original studies was a bottleneck. Furthermore, other brain regions such as hippocampus which plays a central role in behaviour and known SG differences in anatomy [6, 98], or other specific brain regions have shown to regulate signs of aggression and mating dependently from sex or estrous state [99–112]. Such regions would be of interest to study in future. Thus, adding additional studies with

data from specific brain sub-regions may make the analysis more robust, as well as expand to other disease states other than AD and MS.

4 Conclusions

In conclusion, we have characterized in depth SG differences in the human brain transcriptome spanning the whole lifespan, and included also AD and MS patient data. We showed that the SG-biased gene expression is present not only among brain cell types, but also within the same cell type across all ages. We found female-biased upregulation of mitochondrial genes, while male-biased genes were enriched for thymosin hormone targets. We laso noted enrichment of androgen response elements in SG-biased genes in both males and females. The main results included in this study are publicly available and can be interrogated further at https://joshiapps.cbu.uib.no/SRB_app/.

5 Methods

5.1 Data and Code Availability

The datasets included in this analysis were obtained from the following publicly available databases: University of California Santa Cruz (UCSC) Cell Browser [113], and DISCO v1.0 [114]. We found only four projects [29–32] providing single-cell data in human cortex through lifespan. Other studies were explored but then excluded for reasons such as incomplete metadata (sex, age, cell type annotation, etc) and/or insufficient number of samples, such as the 4-10 years Velmeshev dataset [29]. The included projects were chosen since they covered the whole life span, including the fetal and neonatal period; adult datasets were also included to investigate temporal changes between the sexes at later stage in life, and because of the easier availability of disease-related data. Furthermore, we focused on cortical samples, for biological and sample availability reasons. Hereafter, dataset refers to samples from a given age and disease group, thus the included studies were split in multiple datasets, either by age or disease status. The scripts and code used to run the whole analysis, including figures present in this article, can be found at https://github.com/aurazelco/SGDifferencesHumanBrain.

5.2 Single-cell data processing

The main data analysis was performed using `R` (4.1.1, 4.2.1) [115, 116], with the `Seurat` package (4.2.0, 4.3.0) [117]. Multiple versions are due to running the analysis on different machines; Apple MacBook Air M1 2020 macOS BigSur v11.6.8 (4.2.0) and OracleLinux v8.7 Fedora (4.3.0).

A standard single-cell data analysis protocol was used for the Velmeshev data, removing cells with a percentage of mitochondrial genes higher than 5% and with too few (less than 200) or too many (2500) genes. This was followed by normalization, scaling and clustering according to `Seurat` recommendations. We selected only cortical samples in all datasets. The sex of the samples was verified by checking the expression

of XIST, a X chromosome gene mainly expressed in females (Fig. S4). There was no expression of XIST or other X and Y genes in the second trimester dataset, the sex metadata therefore could not be confirmed in this case.

The Velmeshev single-nucleus study [29] provided data for the most datasets, divided by age range; namely, second trimester, third trimester, 0-1 years, 1-2 years, 2-4 years, 4-10 years, 10-20 years and Adults. The 4-10 years dataset was excluded as it had only one female sample (Fig. S2). Each dataset was analyzed individually. The one instance where all datasets were visualized together was to confirm that our cell type distribution in the Uniform Manifold Approximation and Projection (UMAP) plot was similar to the original paper (data not shown).

From the DISCO database [114], we removed from the `brain v1.0` database the projects which did not have metadata about the age, and kept only the samples from the cortex. This resulted in three single-nucleus RNA-seq projects with both female and male samples in the final selection: GSE157827 [30], GSE174367 [31] and PRJNA544731 [32]. The DISCO projects also included samples from patients suffering from Alzheimer's disease (AD) [30, 31] and multiple sclerosis (MS) [32]. No modification was done to the original clustering, which were visually confirmed using the UMAP as in the original version from DISCO (data not shown).

Consistent cell annotation across datasets was one of the challenging steps of the analysis. When analyzing the individual datasets, the original cell annotation was kept as present in the metadata. When the results from the differential expression gene analysis were compared and integrated, the different annotations were organized manually (Supplementary file 1 for details).

5.3 Differential Expression Gene Analysis

The `Seurat` package was used to calculate the differentially expressed genes (DEGs), using the default `FindMarkers` parameters (logfc.threshold 0.25, min.pct 0.1). Briefly, in each dataset, each cell type that had at least 100 cells for each of the female and male sex was analyzed, thus obtaining two lists of DEGs (one per sex) in each cell type. We selected only the up-regulated genes in each sex. We therefore stratified the samples by biological sex (female/male), however we used the SG-biased genes terminology because we cannot know for sure the source of the differential expression; sex, gender or both. These SG-biased DEG lists were then filtered further, so that only the genes with fold change with absolute values of at least 1.2 and adjusted p-value (Bonferroni correction) of 0.05 or lower were kept. Furthermore, only cell types with more than 10 DEGs after filtering were included for further analysis. For each dataset aforementioned in section 5.2, the different ages and disease conditions were analyzed separately, including separating the three projects in the DISCO dataset.

5.4 Mitochondrial genes

We investigated the presence of mitochondrial genes in the SG-biased genes by selecting genes which names started with "`MT-`". We also included "`TIMM`" and "`TOMM`" in the search. Since a recent study implicated a female bias in the expression of Krebs or tricarboxylic acid (TCA) cycle, in connection with mitochondrial activity, we added also

these enzymes from the following source (https://maayanlab.cloud/Harmonizome/gene_set/TCA+cycle/PANTHER+Pathways). We included "XIST" as a control for female.

5.5 Cellular compartment enrichment

Using the Cell Atlas [44], we mapped the SG-biased genes according to their cellular compartment. For each dataset, the number of SG-biased genes were for each sex and cell type found in each cellular location were used to calculate enrichment in each cellular compartment using hyper-geometric test in R (`stats` package, 4.1.1). The enrichment was considered significant when p-value < 0.05.

5.6 Cell type markers

We evaluated the cell type specificity of the SG-biased genes in each sex and dataset, using a reference [45]. Briefly, we calculated the percentage of known markers from five brain populations (astrocytes, neurons, endothelial cells, microglia and oligodendrocytes) in the each cell type, for each dataset in each sex.

5.7 Second trimester integration analysis

For validation of the second trimester dataset, we compared our SG-biased genes with previous results from bulk RNA-seq [27] and single-cell RNA-seq datasets from UCSC Cell Browser [33, 34, 113]. The Eze [33] and Nowakowski [34] single-cell datasets were filtered and integrated according to Seurat recommended integration pipeline, in order to obtain a more complete second trimester reference dataset. The samples from the first and third trimester were excluded for low number of cells/samples. For the cell type annotation used while merging the Eze and Nowakowski's datasets, the more general of the two annotations from Eze et al. was kept. The SG-biased genes were calculated as previously described (see section 5.3). The package `ggVennDiagram` (v1.2.2, [118]) was used to create the Venn diagrams.

5.8 Functional, pathological and transcription enrichment

We used the SG-biased genes to study functional enrichment, in each cell type across all datasets and for each dataset across all cell types, both divided by sex. For this analysis we used the following packages: `clusterProfiler` v4.2.2 [119, 120], `disgenet2r` v0.99.2 [121], `enrichR` v3.1 [47, 48]. We used `clusterProfiler` to investigate the enrichment for Gene Ontology (GO) of biological processes (BP), and diseases from the Disease Ontology (DO) and DisGeNet. `disgenet2r` was also used to obtain the enrichment using the CURATED DisGeNet database, while `enrichR` was used to compare pathways from Kyoto Encyclopedia of Genes and Genomes (KEGG), GWAS Catalog 2019 for SNP enrichment, and the drug enrichment using DSigDB for drug-related enrichment. Additionally, we also investigated the enrichment of transcription factor bind sites using the TRANSFAC and JASPAR PWMs database in `enrichR`. Only the top 5 significant terms were displayed in the plots for the comparison across the datasets or the cell types. All adjusted p-values from the enrichment analyses

were obtained by running Fisher's exact tests with the Benjamini-Hochberg (BH), also knows as false discovery rate (FDR), correction. The enrichment was considered significant when p-value < 0.05.

Additionally, we also investigated the presence of known autism-related genes using the SFARI database (https://gene.sfari.org/database/human-gene/) [53]. We calculated the hyper-geometric enrichment for each dataset according to SFARI nomenclature using `stats` (4.1.1) [115, 116]. The enrichment was considered significant when p-value < 0.05.

5.9 Neuropsychiatric diseases enrichment

We investigated if genes known to be associated with neuropsychiatric diseases [54] were present in the SG DEG lists, and generated a heatmap to display the results.

5.10 Brain region specificity

We checked whether our SG-biased genes were expressed specifically in the cortex, or also in other brain regions. We extracted the DEGs from all the brain regions included in a bulk RNA-seq study [21], namely the cortex, basal ganglia, hypothalamus, amygdala, hippocampus, cerebellum, substantia nigra, spinal cord and pituitary gland. We then calculated the enrichment through hyper-geometric distribution from `stats` (4.1.1) [115, 116] and the significance was reached when p-value < 0.05.

5.11 Sex Chromosomes Enrichment analysis

The SG-biased genes were mapped to the genome with the use of the `biomaRt` package (2.50.3, [122, 123]), and enrichment of the X or Y chromosome compared to the autosomal chromosomes was calculated using the hyper-geometric test from `stats` (4.1.1) [115, 116]. An enrichment was considered significant when the p-value < 0.05. The p-values were averaged to obtain one value for each cell type when comparing the datasets.

5.12 Hormone targets enrichment

We next investigated the enrichment for hormone targets in the SG-biased genes by using data for 53 hormones and their respective targets from a public dataset, Hormone-Gene version 1 [67], available at https://github.com/BIRDSgroup/BioEmbedS. Some hormones had less than 10 targets and were therefore excluded from the enrichment. We calculated the enrichment through hyper-geometric distribution from `stats` (4.1.1) [115, 116]. The enrichment was considered significant when p-value < 0.05 and if the hormone targets were found in more than one SG-biased list.

5.13 Sex hormones response element analysis

We obtained list of genes with presence of androgen or estrogen response elements (AREs and EREs respectively) from previous published data [73, 74]. We calculated the number of genes among the SG-biased genes with ERE and ARE sites, and then

represented the results as percentages of the SG-biased genes, while the genes with no RE sites labelled as "`None`". The percentages for ARE sites were further divided according if the genes had a full, half or both half and full sites for AREs.

Figure Legends

Figure 1. SG-biased genes are present in multiple brain cell types, and across multiple ages and pathologies. We confirmed the presence of SG-biased genes across cell types and datasets in both sexes but the number of SG-DEGs varied greatly. Cell type icons were obtained from BioRender.com, and brain images, indicating the ages included in this study, were modified from [124].

Figure 2. The most different SG-biased genes between sexes are mainly X- and Y-linked genes, with some exceptions. The heatmap displays the presence (not the actual expression) of the 20 most different genes between sexes (10 per sex) in SG-biased genes lists from each cell type and dataset. Most of the genes belong to either the X or Y chromosome. However, among the genes most uniquely expressed by females, there are mitochondrial genes.

Figure 3. The SG-biased genes are cell type- and developmental stage-specific, and enriched for cell type specific markers. A) Heatmaps for SG-biased genes across cell types in the second trimester datasets. B) This bar plot shows the percentages of cell type markers found in the SG-biased genes in each cell type, by sex. For most of the cell types, the highest percentages corresponded to the expected cell type. For example, the excitatory neurons and interneurons have the highest percentages of the neuronal markers. C) Heatmaps for SG-biased genes across datasets for excitatory neurons and astrocytes.

Figure 4. The SG-biased genes, despite being cell type- and developmental stage-specific, show overlap of enriched biological processes across both the cell types and the datasets. A,B) Both female- (A) and male-biased genes (B) showed overlap of enriched terms across different cell types, here as shown in the second trimester dataset. C, D) Similar findings were observed when instead of cell types, we compared multiple datasets in each cell type. Again, we observed an overlap in enriched biological processes both in female- (C) and male-biased genes (D) in excitatory neurons across datasets. 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).

Figure 5. 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. **Legend:** NS: not significant; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$; ****: $p<0.0001$.

Figure 6. The SG-biased genes enrichment for hormone targets. This heatmap shows the p-values of enrichment for hormone targets from the literature in the SG-biased genes. Albeit most hormones are enriched in both sexes, there are some SG specific enrichments. For example, cortisol and progesterone are solely enriched in females. On the other hand, thymosin was strongly enriched in males, much more than in females, across multiple cell types and datasets. **Legend:** NS: not significant; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$; ****: $p<0.0001$.

Figure 7. Androgen response element binding sites, but not estrogen, are consistently enriched in SG-biased genes across sexes, cell types and datasets. Androgen response element (ARE) and estrogen response element (ERE) sites presence in the SG-biased genes, by sex, cell type and dataset. The majority of the SG-biased genes had ARE binding sites, in both sexes. ERE sites instead were present in less than one fourth of SG-biased genes overall. "None" indicates SG-biased DEGs which did not have neither ARE nor ERE binding sites.

6 Supplementary information

This paper is accompanied by the following supplementary materials:

- **Supplementary_figures.pdf:** PDF file containing all supplementary figures
- **Supplementary_file_1.csv:** unified annotation among different snRNA-seq studies
- **Supplementary_file_2.csv:** intersected genes among second trimester studies
- **Supplementary_file_3.csv:** shared genes in each cell type and sex, across different datasets
- **Supplementary_file_4.xlsx:** gene ontology BP enrichment across groups
- **Supplementary_file_5.xlsx:** gene ontology BP enrichment across cell types
- **Supplementary_file_6.xlsx:** KEGG enrichment across groups
- **Supplementary_file_7.xlsx:** KEGG enrichment across cell types
- **Supplementary_file_8.xlsx:** transcription factor binding sites enrichment across groups
- **Supplementary_file_9.csv:** transcription factor binding sites enrichment shared across groups
- **Supplementary_file_10.xlsx:** transcription factor binding sites enrichment across cell types

- **Supplementary_file_11.csv:** transcription factor binding sites enrichment shared across cell types
- **Supplementary_file_12.xlsx:** disease ontology enrichment across cell types
- **Supplementary_file_13.xlsx:** DisGeNET enrichment across cell types
- **Supplementary_file_14.xlsx:** DisGeNET2r enrichment across cell types
- **Supplementary_file_15.xlsx:** GWAS enrichment across cell types
- **Supplementary_file_16.csv:** top 10 most common enriched diseases across datasets, regardless of cell type
- **Supplementary_file_17.csv:** top 5 most common enriched diseases across datasets for each cell type
- **Supplementary_file_18.xlsx:** DSigDB enrichment across cell types
- **Supplementary_file_19.csv:** top 10 most common enriched drugs across datasets, regardless of cell type
- **Supplementary_file_20.csv:** top 5 most common enriched drugs across datasets for each cell type
- **Supplementary_file_21.csv:** hormone targets enrichment

Declarations

- Funding -
- Competing interests - The authors declare that they have no competing interests
- Ethics approval - the original studies all obtained ethical approval, with details in the original scientific publications
- Consent to participate - not applicable
- Consent for publication - not applicable
- Availability of data and materials - the data from the original studies are all publicly available, the DISCO brain v1.0 at https://www.immunesinglecell.org [114] and the individual studies from UCSC Cell Browser at https://cells.ucsc.edu [113]
- Code availability - the scripts used to perform the DEG analysis can be found at https://github.com/aurazelco/SGDifferencesHumanBrain, while the Rshiny source code is available at https://github.com/aurazelco/SGHumanBrainApp
- Authors' contributions - AZ and AJ designed the study together. AZ was responsible for finding the data, performing the analysis and building the web application. AZ and AJ interpreted the data. AZ drafted the manuscript, while AJ substantively revised it. All authors read and approved the final manuscript.
- Acknowledgments - We acknowledge the donors and their families for consenting the sampling in the original studies. We thank the DISCO and UCSC Cell Browser databases for providing publicly available resources, without which this study would not have been possible. We acknowledge Pattama Wapeesittipan, whose scripts were used as a foundation for the DEGs analysis and the web application.

References

[1] Buffery, A. W. H. Sex differences in the development of hemispheric asymmetry of function in the human brain. *Brain Research* (1971).

[2] Ingalhalikar, M. *et al.* Sex differences in the structural connectome of the human brain. *Proc Natl Acad Sci U S A* **111**, 823–8 (2014). URL https://www.ncbi.nlm.nih.gov/pubmed/24297904. Eng Comparative Study 2013/12/04 Proc Natl Acad Sci U S A. 2014 Jan 14;111(2):823-8. doi: 10.1073/pnas.1316909110. Epub 2013 Dec 2.

[3] Kurth, F., Gaser, C. & Luders, E. Development of sex differences in the human brain. *Cogn Neurosci* **12**, 155–162 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/32902364. Eng England 2020/09/10 Cogn Neurosci. 2021 Jul-Oct;12(3-4):155-162. doi: 10.1080/17588928.2020.1800617. Epub 2020 Sep 9.

[4] Lewis, J. D. *et al.* Allometry in the corpus callosum in neonates: Sexual dimorphism. *Hum Brain Mapp* **43**, 4609–4619 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/35722945. Eng 2022/06/21 Hum Brain Mapp. 2022 Oct 15;43(15):4609-4619. doi: 10.1002/hbm.25977. Epub 2022 Jun 20.

[5] Fuhrmann, D., Madsen, K. S., Johansen, L. B., Baare, W. F. C. & Kievit, R. A. The midpoint of cortical thinning between late childhood and early adulthood differs between individuals and brain regions: Evidence from longitudinal modelling in a 12-wave neuroimaging sample. *Neuroimage* **261**, 119507 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/35882270. Eng 2022/07/27 Neuroimage. 2022 Nov 1;261:119507. doi: 10.1016/j.neuroimage.2022.119507. Epub 2022 Jul 23.

[6] Neufang, S. *et al.* Sex differences and the impact of steroid hormones on the developing human brain. *Cereb Cortex* **19**, 464–73 (2009). URL https://www.ncbi.nlm.nih.gov/pubmed/18550597. Eng Comparative Study 2008/06/14 Cereb Cortex. 2009 Feb;19(2):464-73. doi: 10.1093/cercor/bhn100. Epub 2008 Jun 11.

[7] Altemus, M., Sarvaiya, N. & Neill Epperson, C. Sex differences in anxiety and depression clinical perspectives. *Front Neuroendocrinol* **35**, 320–30 (2014). URL https://www.ncbi.nlm.nih.gov/pubmed/24887405. Eng Review 2014/06/03 Front Neuroendocrinol. 2014 Aug;35(3):320-30. doi: 10.1016/j.yfrne.2014.05.004. Epub 2014 Jun 2.

[8] Beking, T. *et al.* Prenatal and pubertal testosterone affect brain lateralization. *Psychoneuroendocrinology* **88**, 78–91 (2018). URL https://www.ncbi.nlm.nih.gov/pubmed/29195161. Eng England 2017/12/02 Psychoneuroendocrinology. 2018 Feb;88:78-91. doi: 10.1016/j.psyneuen.2017.10.027. Epub 2017 Nov 6.

[9] Peper, J. S., Burke, S. M. & Wierenga, L. M. Sex differences and brain development during puberty and adolescence. *Handb Clin Neurol* **175**, 25–54 (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/33008529. Eng Review Netherlands 2020/10/04 Handb Clin Neurol. 2020;175:25-54. doi: 10.1016/B978-0-444-64123-6.00003-5.

[10] DeCasien, A. R., Guma, E., Liu, S. & Raznahan, A. Sex differences in the human brain: a roadmap for more careful analysis and interpretation of a biological reality. *Biol Sex Differ* **13**, 43 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/35883159. Eng Comment Letter Meta-Analysis Review England 2022/07/27 Biol Sex Differ. 2022 Jul 26;13(1):43. doi: 10.1186/s13293-022-00448-w.

[11] Kaczkurkin, A. N., Raznahan, A. & Satterthwaite, T. D. Sex differences in the developing brain: insights from multimodal neuroimaging. *Neuropsychopharmacology* **44**, 71–85 (2019). URL https://www.ncbi.nlm.nih.gov/pubmed/29930385. Eng Review England 2018/06/23 Neuropsychopharmacology. 2019 Jan;44(1):71-85. doi: 10.1038/s41386-018-0111-z. Epub 2018 Jun 6.

[12] Arain, M. *et al.* Maturation of the adolescent brain. *Neuropsychiatr Dis Treat* **9**, 449–61 (2013). URL https://www.ncbi.nlm.nih.gov/pubmed/23579318. Eng New Zealand 2013/04/13 Neuropsychiatr Dis Treat. 2013;9:449-61. doi: 10.2147/NDT.S39776. Epub 2013 Apr 3.

[13] Bao, A. M. & Swaab, D. F. Sexual differentiation of the human brain: relation to gender identity, sexual orientation and neuropsychiatric disorders. *Front Neuroendocrinol* **32**, 214–26 (2011). URL https://www.ncbi.nlm.nih.gov/pubmed/21334362. Eng Review 2011/02/22 Front Neuroendocrinol. 2011 Apr;32(2):214-26. doi: 10.1016/j.yfrne.2011.02.007. Epub 2011 Feb 18.

[14] Clayton, J. A. Sex influences in neurological disorders: case studies and perspectives. *Dialogues in Clinical Neuroscience* (2016).

[15] Hanamsagar, R. & Bilbo, S. D. Sex differences in neurodevelopmental and neurodegenerative disorders: Focus on microglial function and neuroinflammation during development. *J Steroid Biochem Mol Biol* **160**, 127–33 (2016). URL https://www.ncbi.nlm.nih.gov/pubmed/26435451. Eng Review England 2015/10/06 J Steroid Biochem Mol Biol. 2016 Jun;160:127-33. doi: 10.1016/j.jsbmb.2015.09.039. Epub 2015 Oct 23.

[16] Hines, M. Neuroscience and sex/gender: Looking back and forward. *J Neurosci* **40**, 37–43 (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/31488609. Eng Review 2019/09/07 J Neurosci. 2020 Jan 2;40(1):37-43. doi: 10.1523/JNEUROSCI.0750-19.2019. Epub 2019 Sep 5.

[17] Turner, T. N. *et al.* Sex-based analysis of de novo variants in neurodevelopmental disorders. *Am J Hum Genet* **105**, 1274–1285 (2019). URL https://www.ncbi.nlm.nih.gov/pubmed/31785789. Eng 2019/12/02 Am J Hum Genet. 2019 Dec 5;105(6):1274-1285. doi: 10.1016/j.ajhg.2019.11.003. Epub 2019 Nov 27.

[18] Zaidi, Z. F. Gender differences in human brain: A review. *The Open Anatomy Journal* (2010).

[19] Sherwin, B. B. Estrogen and cognitive functioning in women. *Endocr Rev* **24**, 133–51 (2003). URL https://www.ncbi.nlm.nih.gov/pubmed/12700177. Eng Review 2003/04/18 Endocr Rev. 2003 Apr;24(2):133-51. doi: 10.1210/er.2001-0016.

[20] van Eijk, L. *et al.* Are sex differences in human brain structure associated with sex differences in behavior? *Psychol Sci* (2021).

[21] Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* **369** (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/32913072. Eng 2020/09/12 Science. 2020 Sep 11;369(6509):eaba3066. doi: 10.1126/science.aba3066.

[22] Gershoni, M. & Pietrokovski, S. The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biol* **15**, 7 (2017). URL https://www.ncbi.nlm.nih.gov/pubmed/28173793. Eng England 2017/02/09 BMC Biol. 2017 Feb 7;15(1):7. doi: 10.1186/s12915-017-0352-z.

[23] Wapeesittipan, P. & Joshi, A. Integrated analysis of robust sex-biased gene signatures in human brain. *Biology of Sex Differences* (2023 (preprint)).

[24] Rosenkrantz, T. S., Hussain, Z. & Fitch, R. H. Sex differences in brain injury and repair in newborn infants: Clinical evidence and biological mechanisms. *Front Pediatr* **7**, 211 (2019). URL https://www.ncbi.nlm.nih.gov/pubmed/31294000. Eng Review Switzerland 2019/07/12 Front Pediatr. 2019 Jun 26;7:211. doi: 10.3389/fped.2019.00211. eCollection 2019.

[25] Rebuli, M. E., Gibson, P., Rhodes, C. L., Cushing, B. S. & Patisaul, H. B. Sex differences in microglial colonization and vulnerabilities to endocrine disruption in the social brain. *Gen Comp Endocrinol* **238**, 39–46 (2016). URL https://www.ncbi.nlm.nih.gov/pubmed/27102938. Eng 2016/04/23 Gen Comp Endocrinol. 2016 Nov 1;238:39-46. doi: 10.1016/j.ygcen.2016.04.018. Epub 2016 Apr 19.

[26] Belonwu, S. A. *et al.* Sex-stratified single-cell rna-seq analysis identifies sex-specific and cell type-specific transcriptional responses in alzheimer's disease across two brain regions. *Mol Neurobiol* **59**, 276–293 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/34669146. Eng 2021/10/21 Mol Neurobiol. 2022 Jan;59(1):276-293. doi: 10.1007/s12035-021-02591-8. Epub 2021 Oct 20.

[27] O'Brien, H. E. *et al.* Sex differences in gene expression in the human fetal brain. *bioRxiv* (2019).

[28] de Toledo, V. H. C., Feltrin, A. S., Barbosa, A. R., Tahira, A. C. & Brentani, H. Sex differences in gene regulatory networks during mid-gestational brain development. *Front Hum Neurosci* **16**, 955607 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/36061507. Eng Switzerland 2022/09/06 Front Hum Neurosci. 2022 Aug 17;16:955607. doi: 10.3389/fnhum.2022.955607. eCollection 2022.

[29] Velmeshev, D. *et al.* Single-cell analysis of prenatal and postnatal human cortical development. *bioRxiv* (2022).

[30] Lau, S. F., Cao, H., Fu, A. K. Y. & Ip, N. Y. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in alzheimer's disease. *Proc Natl Acad Sci U S A* **117**, 25800–25809 (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/32989152. Eng Research Support, Non-U.S. Gov't 2020/09/30 Proc Natl Acad Sci U S A. 2020 Oct 13;117(41):25800-25809. doi: 10.1073/pnas.2008762117. Epub 2020 Sep 28.

[31] Morabito, S. *et al.* Single-nucleus chromatin accessibility and transcriptomic characterization of alzheimer's disease. *Nat Genet* **53**, 1143–1155 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/34239132. Eng 2021/07/10 Nat Genet. 2021 Aug;53(8):1143-1155. doi: 10.1038/s41588-021-00894-z. Epub 2021 Jul 8.

[32] Schirmer, L. *et al.* Neuronal vulnerability and multilineage diversity in multiple sclerosis. *Nature* **573**, 75–82 (2019). URL https://www.ncbi.nlm.nih.gov/pubmed/31316211. Eng England 2019/07/19 Nature. 2019 Sep;573(7772):75-82. doi: 10.1038/s41586-019-1404-z. Epub 2019 Jul 17.

[33] Eze, U. C., Bhaduri, A., Haeussler, M., Nowakowski, T. J. & Kriegstein, A. R. Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat Neurosci* **24**, 584–594 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/33723434. Eng 2021/03/17 Nat Neurosci. 2021 Apr;24(4):584-594. doi: 10.1038/s41593-020-00794-1. Epub 2021 Mar 15.

[34] Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* (2017).

[35] Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).

[36] Lombardo, M. V. *et al.* Fetal testosterone influences sexually dimorphic gray matter in the human brain. *J Neurosci* **32**, 674–80 (2012). URL https://www.ncbi.nlm.nih.gov/pubmed/22238103. Eng 2012/01/13 J Neurosci. 2012 Jan 11;32(2):674-80. doi: 10.1523/JNEUROSCI.4389-11.2012.

[37] Pasman, J. A. *et al.* The cadm2 gene and behavior: A phenome-wide scan in uk-biobank. *Behav Genet* **52**, 306–314 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/35867259. Eng 2022/07/23 Behav Genet. 2022 Sep;52(4-5):306-314. doi: 10.1007/s10519-022-10109-8. Epub 2022 Jul 22.

[38] Zhou, M. *et al.* Targeted mass spectrometry to quantify brain-derived cerebrospinal fluid biomarkers in alzheimer's disease. *Clin Proteomics* **17**, 19 (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/32514259. Eng England

2020/06/10 Clin Proteomics. 2020 May 29;17:19. doi: 10.1186/s12014-020-09285-8. eCollection 2020.

[39] Varadarajulu, J. *et al.* Differential expression of hint1 in schizophrenia brain tissue. *Eur Arch Psychiatry Clin Neurosci* **262**, 167–72 (2012). URL https://www.ncbi.nlm.nih.gov/pubmed/21553311. Eng Germany 2011/05/10 Eur Arch Psychiatry Clin Neurosci. 2012 Mar;262(2):167-72. doi: 10.1007/s00406-011-0216-4. Epub 2011 May 8.

[40] Xiong, Y. *et al.* Pancancer landscape analysis of the thymosin family identified tmsb10 as a potential prognostic biomarker and immunotherapy target in glioma. *Cancer Cell Int* **22**, 294 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/36163046. Eng England 2022/09/27 Cancer Cell Int. 2022 Sep 26;22(1):294. doi: 10.1186/s12935-022-02698-5.

[41] Frackowiak, R. S. J., Magistretti, P. J., Shulman, R. G., Altman, J. S. & Adams, M. Neuroenergenetics: Relevance in functional brain imaging. *Human Frontier Science Program Organization* (2002).

[42] Iwata, R. *et al.* Mitochondria metabolism sets the species-specific tempo of neuronal development. *Science* **379**, eabn4705 (2023). URL https://www.ncbi.nlm.nih.gov/pubmed/36705539. Eng 2023/01/28 Science. 2023 Feb 10;379(6632):eabn4705. doi: 10.1126/science.abn4705. Epub 2023 Feb 10.

[43] Lee, J. W., Profant, M. & Wang, C. Metabolic sex dimorphism of the brain at the gene, cell, and tissue level. *J Immunol* **208**, 212–220 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/35017210. Eng Review 2022/01/13 J Immunol. 2022 Jan 15;208(2):212-220. doi: 10.4049/jimmunol.2100853.

[44] Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356** (2017). URL https://www.ncbi.nlm.nih.gov/pubmed/28495876. Eng 2017/05/13 Science. 2017 May 26;356(6340):eaal3321. doi: 10.1126/science.aal3321. Epub 2017 May 11.

[45] McKenzie, A. T. *et al.* Brain cell type specific gene expression and co-expression network architectures. *Sci Rep* **8**, 8868 (2018). URL https://www.ncbi.nlm.nih.gov/pubmed/29892006. Eng Comparative Study England 2018/06/13 Sci Rep. 2018 Jun 11;8(1):8868. doi: 10.1038/s41598-018-27293-5.

[46] Montaldo, P. *et al.* Transcriptomic profile of adverse neurodevelopmental outcomes after neonatal encephalopathy. *Sci Rep* **10**, 13100 (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/32753750. Eng England 2020/08/06 Sci Rep. 2020 Aug 4;10(1):13100. doi: 10.1038/s41598-020-70131-w.

[47] Chen, E. Y. *et al.* Enrichr: interactive and collaborative html5 gene list enrichment analysis tool. *Bioinformatics* (2013).

[48] Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90–7 (2016). URL https://www.ncbi.nlm.nih.gov/pubmed/27141961. Eng England 2016/05/05 Nucleic Acids Res. 2016 Jul 8;44(W1):W90-7. doi: 10.1093/nar/gkw377. Epub 2016 May 3.

[49] Wang, Y., Song, L. & Zhou, C. J. The canonical wnt/ß-catenin signaling pathway regulates fgf signaling for early facial development. *Developmental Biology* **349**, 250–260 (2011).

[50] Bosch i Ara, L., Katugampola, H. & Dattani, M. T. Congenital hypopituitarism during the neonatal period: Epidemiology, pathogenesis, therapeutic options, and outcome. *Frontiers in Pediatrics* **8** (2021).

[51] Akiba, K. *et al.* Pou1f1/pou1f1 c.143-83a ¿g variant disrupts the branch site in pre-mrna and leads to dwarfism. *Endocrinology* **164** (2023).

[52] Lai, M. C. *et al.* Imaging sex/gender and autism in the brain: Etiological implications. *J Neurosci Res* **95**, 380–397 (2017). URL https://www.ncbi.nlm.nih.gov/pubmed/27870420. Eng J Neurosci Res. 2017 Jan 2;95(1-2):380-397. doi: 10.1002/jnr.23948.

[53] Banerjee-Basu, S. & Packer, A. Sfari gene: an evolving database for the autism research community. *Dis Model Mech* (2010).

[54] Chlamydas, S., Markouli, M., Strepkos, D. & Piperi, C. Epigenetic mechanisms regulate sex-specific bias in disease manifestations. *J Mol Med (Berl)* **100**, 1111–1123 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/35764820. Eng Review Germany 2022/06/29 J Mol Med (Berl). 2022 Aug;100(8):1111-1123. doi: 10.1007/s00109-022-02227-x. Epub 2022 Jun 29.

[55] Olsson, M. *et al.* Epigenetic regulation of tissue-type plasminogen activator in human brain tissue and brain-derived cells. *Gene Regul Syst Bio* **10**, 9–13 (2016). URL https://www.ncbi.nlm.nih.gov/pubmed/26823649. Eng 2016/01/30 Gene Regul Syst Bio. 2016 Jan 24;10:9-13. doi: 10.4137/GRSB.S30241. eCollection 2016.

[56] Matsuda, K. I. *et al.* Histone deacetylation during brain development is essential for permanent masculinization of sexual behavior. *Endocrinology* **152**, 2760–7 (2011). URL https://www.ncbi.nlm.nih.gov/pubmed/21586557. Eng 2011/05/19 Endocrinology. 2011 Jul;152(7):2760-7. doi: 10.1210/en.2011-0193. Epub 2011 May 17.

[57] Sharma, A. R. *et al.* Valproic acid and propionic acid modulated mechanical pathways associated with autism spectrum disorder at prenatal and neonatal exposure. *CNS Neurol Disord Drug Targets* (2022).

[58] Konopko, M. A., Densmore, A. L. & Krueger, B. K. Sexually dimorphic epigenetic regulation of brain-derived neurotrophic factor in fetal brain in the valproic acid model of autism spectrum disorder. *Dev Neurosci* **39**, 507–518 (2017). URL https://www.ncbi.nlm.nih.gov/pubmed/29073621. Eng Switzerland 2017/10/27 Dev Neurosci. 2017;39(6):507-518. doi: 10.1159/000481134. Epub 2017 Oct 27.

[59] Wołoszynowska-Fraser, M. U., Kouchmeshky, A. & McCaffery, P. Vitamin a and retinoic acid in cognition and cognitive disease. *Annu Rev Nutr* (2020).

[60] Hu, V. W., Sarachana, T., Sherrard, R. M. & Kocher, K. M. Investigation of sex differences in the expression of rora and its transcriptional targets in the brain as a potential contributor to the sex bias in autism. *Molecular Autism* (2015).

[61] Sayad, A., Noroozi, R., Omrani, M. D., Taheri, M. & Ghafouri-Fard, S. Retinoic acid-related orphan receptor alpha (rora) variants are associated with autism spectrum disorder. *Metab Brain Dis* **32**, 1595–1601 (2017). URL https://www.ncbi.nlm.nih.gov/pubmed/28608249. Ghafouri-Fard, Soudeh eng 2017/06/14 Metab Brain Dis. 2017 Oct;32(5):1595-1601. doi: 10.1007/s11011-017-0049-6. Epub 2017 Jun 12.

[62] Mayne, B. T. *et al.* Large scale gene expression meta-analysis reveals tissue-specific, sex-biased gene expression in humans. *Front Genet* **7**, 183 (2016). URL https://www.ncbi.nlm.nih.gov/pubmed/27790248. Eng Switzerland 2016/10/30 Front Genet. 2016 Oct 13;7:183. doi: 10.3389/fgene.2016.00183. eCollection 2016.

[63] Mousavi, M. J., Mahmoudi, M. & Ghotloo, S. Escape from x chromosome inactivation and female bias of autoimmune diseases. *Mol Med* **26**, 127 (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/33297945. Eng Review England 2020/12/11 Mol Med. 2020 Dec 9;26(1):127. doi: 10.1186/s10020-020-00256-1.

[64] Kim, H. *et al.* Identification of altered protein expression in major depressive disorder and bipolar disorder patients using liquid chromatography-tandem mass spectrometry. *Psychiatry Res* **299**, 113850 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/33711561. Eng Ireland 2021/03/13 Psychiatry Res. 2021 May;299:113850. doi: 10.1016/j.psychres.2021.113850. Epub 2021 Mar 2.

[65] Bruschi, M. *et al.* Potential biomarkers of childhood brain tumor identified by proteomics of cerebrospinal fluid from extraventricular drainage (evd). *Sci Rep* **11**, 1818 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/33469081. Eng England 2021/01/21 Sci Rep. 2021 Jan 19;11(1):1818. doi: 10.1038/s41598-020-80647-w.

[66] Bakker, J. The role of steroid hormones in the sexual differentiation of the human brain. *J Neuroendocrinol* **34**, e13050 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/34708466. Eng Review 2021/10/29 J Neuroendocrinol. 2022 Feb;34(2):e13050. doi: 10.1111/jne.13050. Epub 2021 Oct 27.

[67] Jadhav, A. *et al.* Predicting cross-tissue hormone–gene relations using balanced word embeddings. *Bioinformatics* **38**, 4771–4781 (2022).

[68] Dalakas, M. C. & Trapp, B. D. Thymosin beta 4 is a shared antigen between lymphoid cells and oligodendrocytes of normal human brain. *Ann Neurol* **19**, 349–55 (1986). URL https://www.ncbi.nlm.nih.gov/pubmed/3085577. Eng 1986/04/01 Ann Neurol. 1986 Apr;19(4):349-55. doi: 10.1002/ana.410190407.

[69] Hall, A. K. Developmental regulation of thymosin beta 10 mrna in the human brain. *Molecular Brain Research* (1991).

[70] Condon, M. R. & Hall, A. K. Expression of thymosin beta-4 and related genes in developing human brain. *J Mol Neurosci.* (1992).

[71] Claessens, F. *et al.* Diverse roles of androgen receptor (ar) domains in ar-mediated signaling. *Nucl Recept Signal* **6**, e008 (2008). URL https://www.ncbi.nlm.nih.gov/pubmed/18612376. Eng Review 2008/07/10 Nucl Recept Signal. 2008 Jun 27;6:e008. doi: 10.1621/nrs.06008.

[72] Rettberg, J. R., Yao, J. & Brinton, R. D. Estrogen: a master regulator of bioenergetic systems in the brain and body. *Front Neuroendocrinol* **35**, 8–30 (2014). URL https://www.ncbi.nlm.nih.gov/pubmed/23994581. Eng Review 2013/09/03 Front Neuroendocrinol. 2014 Jan;35(1):8-30. doi: 10.1016/j.yfrne.2013.08.001. Epub 2013 Aug 29.

[73] Bourdeau, V. *et al.* Genome-wide identification of high-affinity estrogen response elements in human and mouse. *Mol Endocrinol* **18**, 1411–27 (2004). URL https://www.ncbi.nlm.nih.gov/pubmed/15001666. Eng 2004/03/06 Mol Endocrinol. 2004 Jun;18(6):1411-27. doi: 10.1210/me.2003-0441. Epub 2004 Mar 4.

[74] Wilson, S., Qi, J. & Filipp, F. V. Refinement of the androgen response element based on chip-seq in androgen-insensitive and androgen-responsive prostate cancer cell lines. *Sci Rep* **6**, 32611 (2016). URL https://www.ncbi.nlm.nih.gov/pubmed/27623747. Eng England 2016/09/15 Sci Rep. 2016 Sep 14;6:32611. doi: 10.1038/srep32611.

[75] van Anders, S. M. Beyond masculinity: testosterone, gender/sex, and human social behavior in a comparative context. *Front Neuroendocrinol* **34**, 198–210 (2013). URL https://www.ncbi.nlm.nih.gov/pubmed/23867694. Eng Review 2013/07/23 Front Neuroendocrinol. 2013 Aug;34(3):198-210. doi: 10.1016/j.yfrne.2013.07.001. Epub 2013 Jul 16.

[76] Kuiri-Hanninen, T. *et al.* Increased activity of the hypothalamic-pituitary-testicular axis in infancy results in increased androgen action in premature boys. *J Clin Endocrinol Metab* **96**, 98–105 (2011). URL https://www.ncbi.nlm.nih.gov/pubmed/20881260. Eng 2010/10/01 J Clin Endocrinol Metab. 2011 Jan;96(1):98-105. doi: 10.1210/jc.2010-1359. Epub 2010 Sep 29.

[77] Forest, M. G., Cathiard, A. M. & Bertrand, J. A. Evidence of testicular activity in early infancy. *J Clin Endocrinol Metab* **37**, 148–151 (1973). URL https://doi.org/10.1210/jcem-37-1-148.

[78] Chang, W. *et al.* *shiny: Web Application Framework for R* (2023). URL https://shiny.rstudio.com/. R package version 1.7.4.9002.

[79] Leffler, J., Trend, S., Gorman, S. & Hart, P. H. Sex-specific environmental impacts on initiation and progression of multiple sclerosis. *Front Neurol* **13**, 835162 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/35185777. Eng Review Switzerland 2022/02/22 Front Neurol. 2022 Feb 3;13:835162. doi: 10.3389/fneur.2022.835162. eCollection 2022.

[80] Ysrraelit, M. C. & Correale, J. Impact of sex hormones on immune function and multiple sclerosis development. *Immunology* **156**, 9–22 (2019). URL https://www.ncbi.nlm.nih.gov/pubmed/30222193. Eng Review England 2018/09/18 Immunology. 2019 Jan;156(1):9-22. doi: 10.1111/imm.13004. Epub 2018 Oct 11.

[81] Shi, L., Zhang, Z. & Su, B. Sex biased gene expression profiling of human brains at major developmental stages. *Sci Rep* **6**, 21181 (2016). URL https://www.ncbi.nlm.nih.gov/pubmed/26880485. Eng England 2016/02/18 Sci Rep. 2016 Feb 16;6:21181. doi: 10.1038/srep21181.

[82] Devall, M. *et al.* Genome-wide characterization of mitochondrial dna methylation in human brain. *Front Endocrinol (Lausanne)* **13**, 1059120 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/36726473. Eng Switzerland 2023/02/03 Front Endocrinol (Lausanne). 2023 Jan 16;13:1059120. doi: 10.3389/fendo.2022.1059120. eCollection 2022.

[83] Klinge, C. M. Estrogenic control of mitochondrial function. *Redox Biol* **31**, 101435 (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/32001259. Eng Review Netherlands 2020/02/01 Redox Biol. 2020 Apr;31:101435. doi: 10.1016/j.redox.2020.101435. Epub 2020 Jan 23.

[84] Ruigrok, A. N. V. & Lai, M. C. Sex/gender differences in neurology and psychiatry: Autism. *Handb Clin Neurol* **175**, 283–297 (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/33008532. Eng Review Netherlands 2020/10/04 Handb Clin Neurol. 2020;175:283-297. doi: 10.1016/B978-0-444-64123-6.00020-5.

[85] Rynkiewicz, A. *et al.* An investigation of the 'female camouflage effect' in autism using a computerized ados-2 and a test of sex/gender differences. *Mol Autism* **7**, 10 (2016). URL https://www.ncbi.nlm.nih.gov/pubmed/26798446. Eng Comparative Study England 2016/01/23 Mol Autism. 2016 Jan 21;7:10. doi: 10.1186/s13229-016-0073-0. eCollection 2016.

[86] Wigdor, E. M. *et al.* The female protective effect against autism spectrum disorder. *Cell Genom* **2**, 100134 (2022). URL https://www.ncbi.nlm.nih.gov/

pubmed/36778135. Eng 2023/02/14 Cell Genom. 2022 Jun 8;2(6):100134. doi: 10.1016/j.xgen.2022.100134. eCollection 2022 Jun 8.

[87] Bramble, M. S., Lipson, A., Vashist, N. & Vilain, E. Effects of chromosomal sex and hormonal influences on shaping sex differences in brain and behavior: Lessons from cases of disorders of sex development. *J Neurosci Res* **95**, 65–74 (2017). URL https://www.ncbi.nlm.nih.gov/pubmed/27841933. Eng Review 2016/11/15 J Neurosci Res. 2017 Jan 2;95(1-2):65-74. doi: 10.1002/jnr.23832.

[88] Lombardo, M. V. *et al.* Fetal programming effects of testosterone on the reward system and behavioral approach tendencies in humans. *Biol Psychiatry* **72**, 839–47 (2012). URL https://www.ncbi.nlm.nih.gov/pubmed/22763187. Eng 2012/07/06 Biol Psychiatry. 2012 Nov 15;72(10):839-47. doi: 10.1016/j.biopsych.2012.05.027. Epub 2012 Jul 3.

[89] Christine Knickmeyer, R. & Baron-Cohen, S. Fetal testosterone and sex differences. *Early Hum Dev* **82**, 755–60 (2006). URL https://www.ncbi.nlm.nih.gov/pubmed/17084045. Eng Review Ireland 2006/11/07 Early Hum Dev. 2006 Dec;82(12):755-60. doi: 10.1016/j.earlhumdev.2006.09.014. Epub 2006 Nov 1.

[90] Pike, C. J. Sex and the development of alzheimer's disease. *J Neurosci Res* **95**, 671–680 (2017). URL https://www.ncbi.nlm.nih.gov/pubmed/27870425. Eng Review 2016/11/22 J Neurosci Res. 2017 Jan 2;95(1-2):671-680. doi: 10.1002/jnr.23827.

[91] Quintana, D. S. *et al.* Oxytocin pathway gene networks in the human brain. *Nat Commun* **10**, 668 (2019). URL https://www.ncbi.nlm.nih.gov/pubmed/30737392. Eng England 2019/02/10 Nat Commun. 2019 Feb 8;10(1):668. doi: 10.1038/s41467-019-08503-8.

[92] Gegenhuber, B. & Tollkuhn, J. Signatures of sex: Sex differences in gene expression in the vertebrate brain. *Wiley Interdiscip Rev Dev Biol* **9**, e348 (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/31106965. Eng Review 2019/05/21 Wiley Interdiscip Rev Dev Biol. 2020 Jan;9(1):e348. doi: 10.1002/wdev.348. Epub 2019 May 20.

[93] Gegenhuber, B., Wu, M. V., Bronstein, R. & Tollkuhn, J. Gene regulation by gonadal hormone receptors underlies brain sex differences. *Nature* **606**, 153–159 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/35508660. Eng England 2022/05/05 Nature. 2022 Jun;606(7912):153-159. doi: 10.1038/s41586-022-04686-1. Epub 2022 May 4.

[94] Zhang, G.-h., Murthy, K. D., Pare, R. B. & Qian, Y.-h. Protective effect of tbeta4 on central nervous system tissues and its developmental prospects. *Inflammation and Mental Health* (2020).

[95] Wainer Katsir, K. & Linial, M. Human genes escaping x-inactivation revealed by single cell expression data. *BMC Genomics* **20**, 201 (2019). URL https://www.ncbi.nlm.nih.gov/pubmed/30871455. Eng England 2019/03/16 BMC Genomics. 2019 Mar 12;20(1):201. doi: 10.1186/s12864-019-5507-6.

[96] Morris, D. C., Chopp, M., Zhang, L. & Zhang, Z. G. Thymosin beta4: a candidate for treatment of stroke? *Ann N Y Acad Sci* **1194**, 112–7 (2010). URL https://www.ncbi.nlm.nih.gov/pubmed/20536457. Eng Ann N Y Acad Sci. 2010 Apr;1194:112-7. doi: 10.1111/j.1749-6632.2010.05469.x.

[97] Hägg, S. & Jylhava, J. Sex differences in biological aging with a focus on human studies. *Elife* **10** (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/33982659. Eng Review England 2021/05/14 Elife. 2021 May 13;10:e63425. doi: 10.7554/eLife.63425.

[98] Kight, K. E. & McCarthy, M. M. Androgens and the developing hippocampus. *Biol Sex Differ* **11**, 30 (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/32487149. Eng Review England 2020/06/04 Biol Sex Differ. 2020 Jun 1;11(1):30. doi: 10.1186/s13293-020-00307-6.

[99] Knoedler, J. R. *et al.* A functional cellular framework for sex and estrous cycle-dependent gene expression and behavior. *Cell* **185**, 654–671 e22 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/35065713. Eng 2022/01/24 Cell. 2022 Feb 17;185(4):654-671.e22. doi: 10.1016/j.cell.2021.12.031. Epub 2022 Jan 21.

[100] Karigo, T. *et al.* Distinct hypothalamic control of same- and opposite-sex mounting behaviour in mice. *Nature* **589**, 258–263 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/33268894. Eng England 2020/12/04 Nature. 2021 Jan;589(7841):258-263. doi: 10.1038/s41586-020-2995-0. Epub 2020 Dec 2.

[101] Bayless, D. W. *et al.* Limbic neurons shape sex recognition and social behavior in sexually naive males. *Cell* **176**, 1190–1205 e20 (2019). URL https://www.ncbi.nlm.nih.gov/pubmed/30712868. Eng 2019/02/05 Cell. 2019 Feb 21;176(5):1190-1205.e20. doi: 10.1016/j.cell.2018.12.041. Epub 2019 Jan 31.

[102] Osakada, T. *et al.* Sexual rejection via a vomeronasal receptor-triggered limbic circuit. *Nat Commun* **9**, 4463 (2018). URL https://www.ncbi.nlm.nih.gov/pubmed/30367054. Eng England 2018/10/28 Nat Commun. 2018 Oct 26;9(1):4463. doi: 10.1038/s41467-018-07003-5.

[103] Wei, Y. C. *et al.* Medial preoptic area in mice is capable of mediating sexually dimorphic behaviors regardless of gender. *Nat Commun* **9**, 279 (2018). URL https://www.ncbi.nlm.nih.gov/pubmed/29348568. Eng England 2018/01/20 Nat Commun. 2018 Jan 18;9(1):279. doi: 10.1038/s41467-017-02648-0.

[104] Hashikawa, K. *et al.* Esr1(+) cells in the ventromedial hypothalamus control female aggression. *Nat Neurosci* **20**, 1580–1590 (2017). URL https://

www.ncbi.nlm.nih.gov/pubmed/28920934. Eng 2017/09/19 Nat Neurosci. 2017 Nov;20(11):1580-1590. doi: 10.1038/nn.4644. Epub 2017 Sep 18.

[105] Yao, S., Bergan, J., Lanjuin, A. & Dulac, C. Oxytocin signaling in the medial amygdala is required for sex discrimination of social cues. *Elife* **6** (2017). URL https://www.ncbi.nlm.nih.gov/pubmed/29231812. Eng England 2017/12/13 Elife. 2017 Dec 12;6:e31373. doi: 10.7554/eLife.31373.

[106] Ishii, K. K. *et al.* A labeled-line neural circuit for pheromone-mediated sexual behaviors in mice. *Neuron* **95**, 123–137 e8 (2017). URL https://www.ncbi.nlm.nih.gov/pubmed/28648498. Eng 2017/06/27 Neuron. 2017 Jul 5;95(1):123-137.e8. doi: 10.1016/j.neuron.2017.05.038. Epub 2017 Jun 22.

[107] Unger, E. K. *et al.* Medial amygdalar aromatase neurons regulate aggression in both sexes. *Cell Rep* **10**, 453–62 (2015). URL https://www.ncbi.nlm.nih.gov/pubmed/25620703. Eng 2015/01/27 Cell Rep. 2015 Feb 3;10(4):453-62. doi: 10.1016/j.celrep.2014.12.040. Epub 2015 Jan 22.

[108] Hong, W., Kim, D. W. & Anderson, D. J. Antagonistic control of social versus repetitive self-grooming behaviors by separable amygdala neuronal subsets. *Cell* **158**, 1348–1361 (2014). URL https://www.ncbi.nlm.nih.gov/pubmed/25215491. Eng 2014/09/13 Cell. 2014 Sep 11;158(6):1348-1361. doi: 10.1016/j.cell.2014.07.049.

[109] Lee, H. *et al.* Scalable control of mounting and attack by esr1+ neurons in the ventromedial hypothalamus. *Nature* **509**, 627–32 (2014). URL https://www.ncbi.nlm.nih.gov/pubmed/24739975. Eng England 2014/04/18 Nature. 2014 May 29;509(7502):627-32. doi: 10.1038/nature13169. Epub 2014 Apr 16.

[110] Yang, C. F. *et al.* Sexually dimorphic neurons in the ventromedial hypothalamus govern mating in both sexes and aggression in males. *Cell* **153**, 896–909 (2013). URL https://www.ncbi.nlm.nih.gov/pubmed/23663785. Eng 2013/05/15 Cell. 2013 May 9;153(4):896-909. doi: 10.1016/j.cell.2013.04.017.

[111] Lin, D. *et al.* Functional identification of an aggression locus in the mouse hypothalamus. *Nature* **470**, 221–6 (2011). URL https://www.ncbi.nlm.nih.gov/pubmed/21307935. Eng England 2011/02/11 Nature. 2011 Feb 10;470(7333):221-6. doi: 10.1038/nature09736.

[112] Musatov, S., Chen, W., Pfaff, D. W., Kaplitt, M. G. & Ogawa, S. Rnai-mediated silencing of estrogen receptor alpha in the ventromedial nucleus of hypothalamus abolishes female sexual behaviors. *Proc Natl Acad Sci U S A* **103**, 10456–10460 (2006). URL https://www.ncbi.nlm.nih.gov/pubmed/16803960. Eng 2006/06/29 Proc Natl Acad Sci U S A. 2006 Jul 5;103(27):10456-10460. doi: 10.1073/pnas.0603045103. Epub 2006 Jun 27.

[113] Speir, M. L. *et al.* Ucsc cell browser: Visualize your single-cell data. *Bioinformatics* **37**, 4578–80 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/34244710. Eng England 2021/07/11 Bioinformatics. 2021 Jul 9;37(23):4578-80. doi: 10.1093/bioinformatics/btab503.

[114] Li, M. *et al.* Disco: a database of deeply integrated human single-cell omics data. *Nucleic Acids Res* **50**, D596–D602 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/34791375. Eng England 2021/11/19 Nucleic Acids Res. 2022 Jan 7;50(D1):D596-D602. doi: 10.1093/nar/gkab1020.

[115] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2020). URL https://www.R-project.org/.

[116] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA (2019). URL http://www.rstudio.com/.

[117] Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 e29 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/34062119. Eng 2021/06/02 Cell. 2021 Jun 24;184(13):3573-3587.e29. doi: 10.1016/j.cell.2021.04.048. Epub 2021 May 31.

[118] Gao, C. H., Yu, G. & Cai, P. ggvenndiagram: An intuitive, easy-to-use, and highly customizable r package to generate venn diagram. *Front Genet* **12**, 706907 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/34557218. Eng Switzerland 2021/09/25 Front Genet. 2021 Sep 7;12:706907. doi: 10.3389/fgene.2021.706907. eCollection 2021.

[119] Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterprofiler: an r package for comparing biological themes among gene clusters. *OMICS* **16**, 284–7 (2012). URL https://www.ncbi.nlm.nih.gov/pubmed/22455463. Eng 2012/03/30 OMICS. 2012 May;16(5):284-7. doi: 10.1089/omi.2011.0118. Epub 2012 Mar 28.

[120] Wu, T. *et al.* clusterprofiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/34557778. Eng 2021/09/25 Innovation (Camb). 2021 Jul 1;2(3):100141. doi: 10.1016/j.xinn.2021.100141. eCollection 2021 Aug 28.

[121] Pinero, J. *et al.* The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* **48**, D845–D855 (2020). URL https://www.ncbi.nlm.nih.gov/pubmed/31680165. Eng England 2019/11/05 Nucleic Acids Res. 2020 Jan 8;48(D1):D845-D855. doi: 10.1093/nar/gkz1021.

[122] Durinck, S. *et al.* Biomart and bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–40 (2005). URL https://www.ncbi.nlm.nih.gov/pubmed/16082012. Eng England

2005/08/06 Bioinformatics. 2005 Aug 15;21(16):3439-40. doi: 10.1093/bioinformatics/bti525.

[123] Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nat Protoc* **4**, 1184–91 (2009). URL https://www.ncbi.nlm.nih.gov/pubmed/19617889. Eng England 2009/07/21 Nat Protoc. 2009;4(8):1184-91. doi: 10.1038/nprot.2009.97. Epub 2009 Jul 23.

[124] Konkel, L. The brain before birth: Using fmri to explore the secrets of fetal neurodevelopment. *Environ Health Perspect* **126**, 112001 (2018). URL https://www.ncbi.nlm.nih.gov/pubmed/30457876. Eng News 2018/11/21 Environ Health Perspect. 2018 Nov;126(11):112001. doi: 10.1289/EHP2268.

Figures

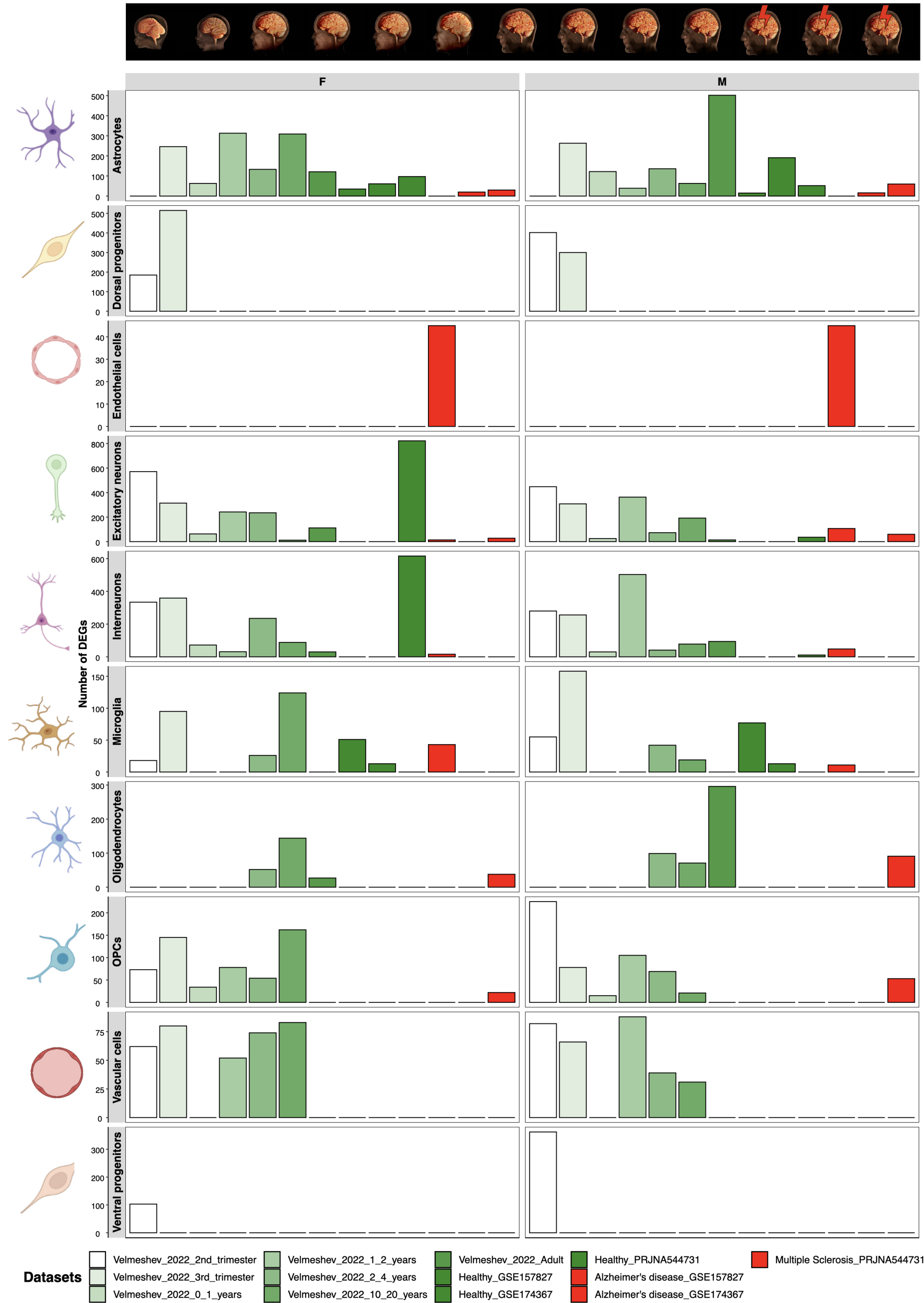


Figure 1

SG-biased genes are present in multiple brain cell types, and across multiple ages and pathologies. We confirmed the presence of SG-biased genes across cell types and datasets in both sexes but the number

of SG-DEGs var- ied greatly. Cell type icons were obtained from BioRender.com, and brain images, indicating the ages included in this study, were modified from [124].

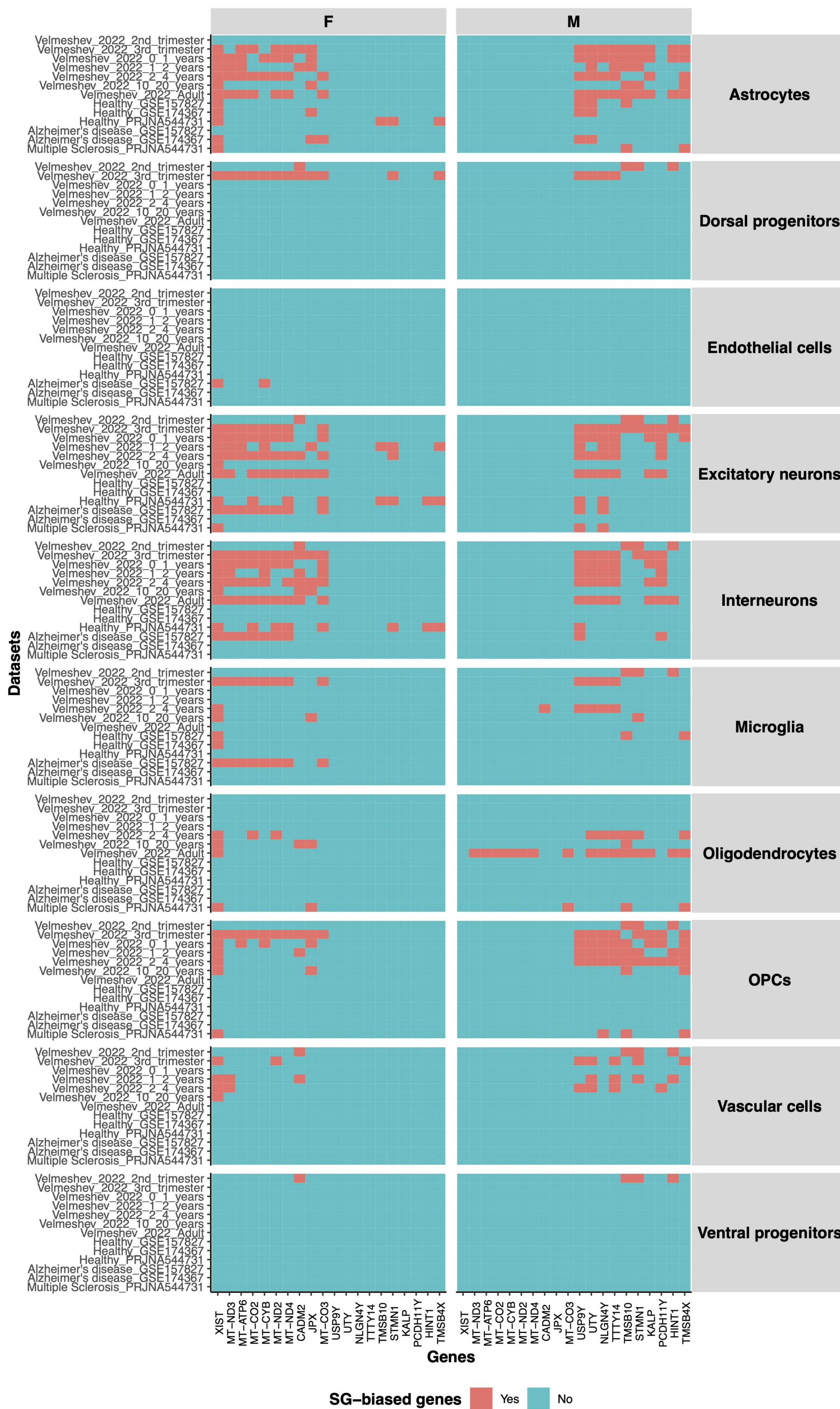


Figure 2

The most different SG-biased genes between sexes are mainly X- and Y-linked genes, with some exceptions. The heatmap displays the presence (not the actual expression) of the 20 most different

genes between sexes (10 per sex) in SG-biased genes lists from each cell type and dataset. Most of the genes belong to either the X or Y chromosome. However, among the genes most uniquely expressed by females, there are mitochondrial genes.

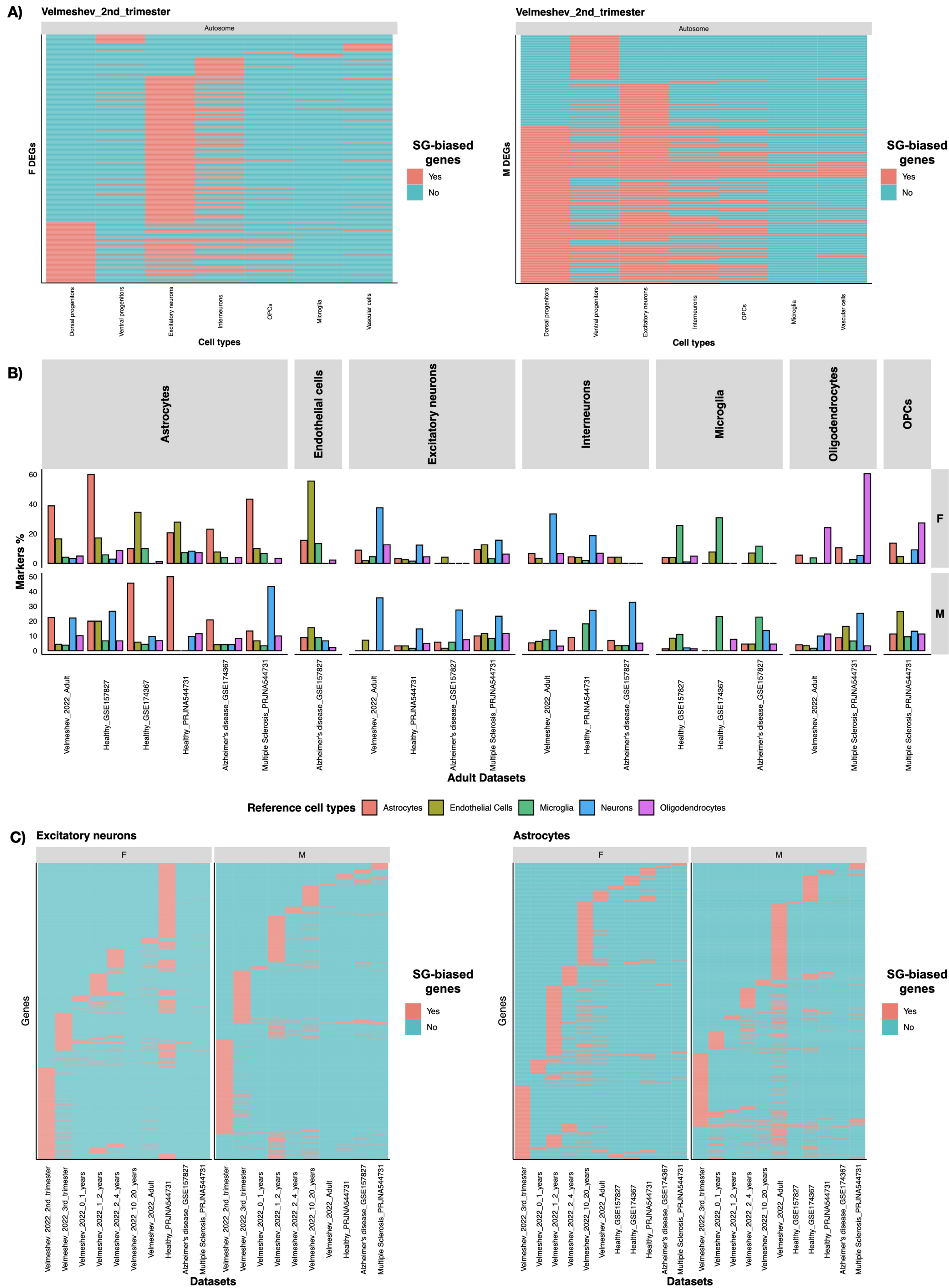
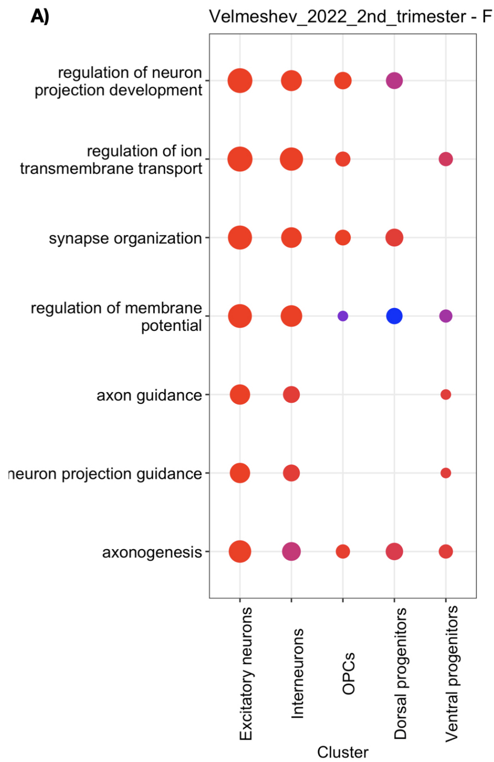
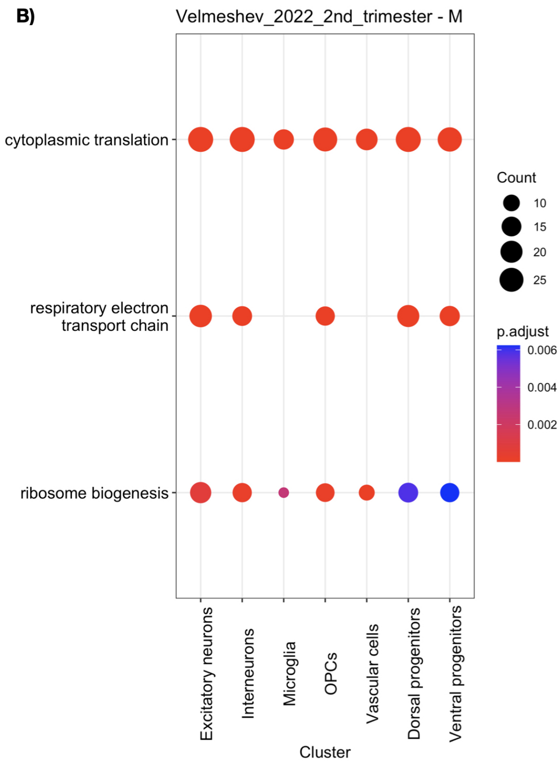
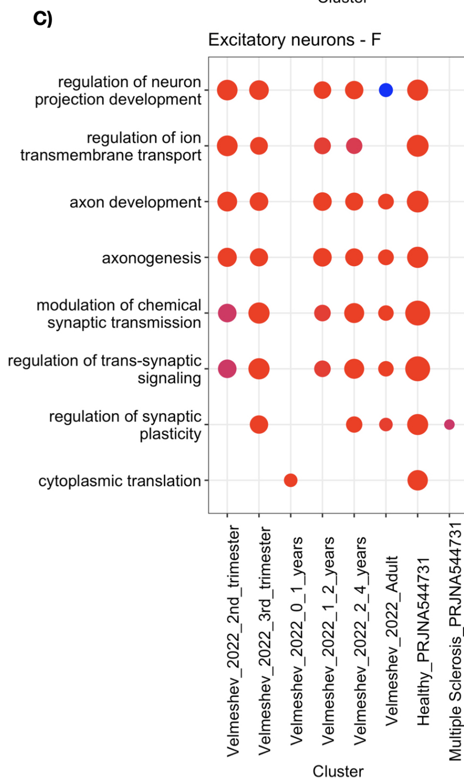


Figure 3

The SG-biased genes are cell type- and developmental stage- specific, and enriched for cell type specific markers. A) Heatmaps for SG-biased genes across cell types in the second trimester datasets. B) This bar plot shows the percentages of cell type markers found in the SG-biased genes in each cell type, by sex. For most of the cell types, the highest percentages corresponded to the expected cell type. For example, the excitatory neurons and interneurons have the highest per- centages of the neuronal markers. C) Heatmaps for SG-biased genes across datasets for excitatory neurons and astrocytes.







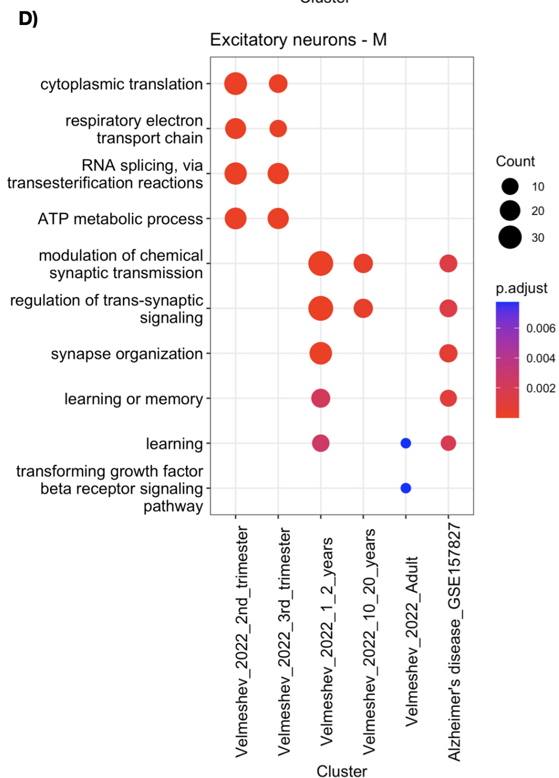


Figure 4

The SG-biased genes, despite being cell type- and develop- mental stage-specific, show overlap of enriched biological processes across both the cell types and the datasets. A,B) Both female- (A) and male-biased genes (B) showed overlap of enriched terms across different cell types, here as shown in the second trimester dataset. C, D) Similar findings were observed when instead of cell types, we compared multiple datasets in each cell type. Again, we observed an overlap in enriched biological processes both in female- (C) and male-biased genes (D) in excitatory neurons across datasets. 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).

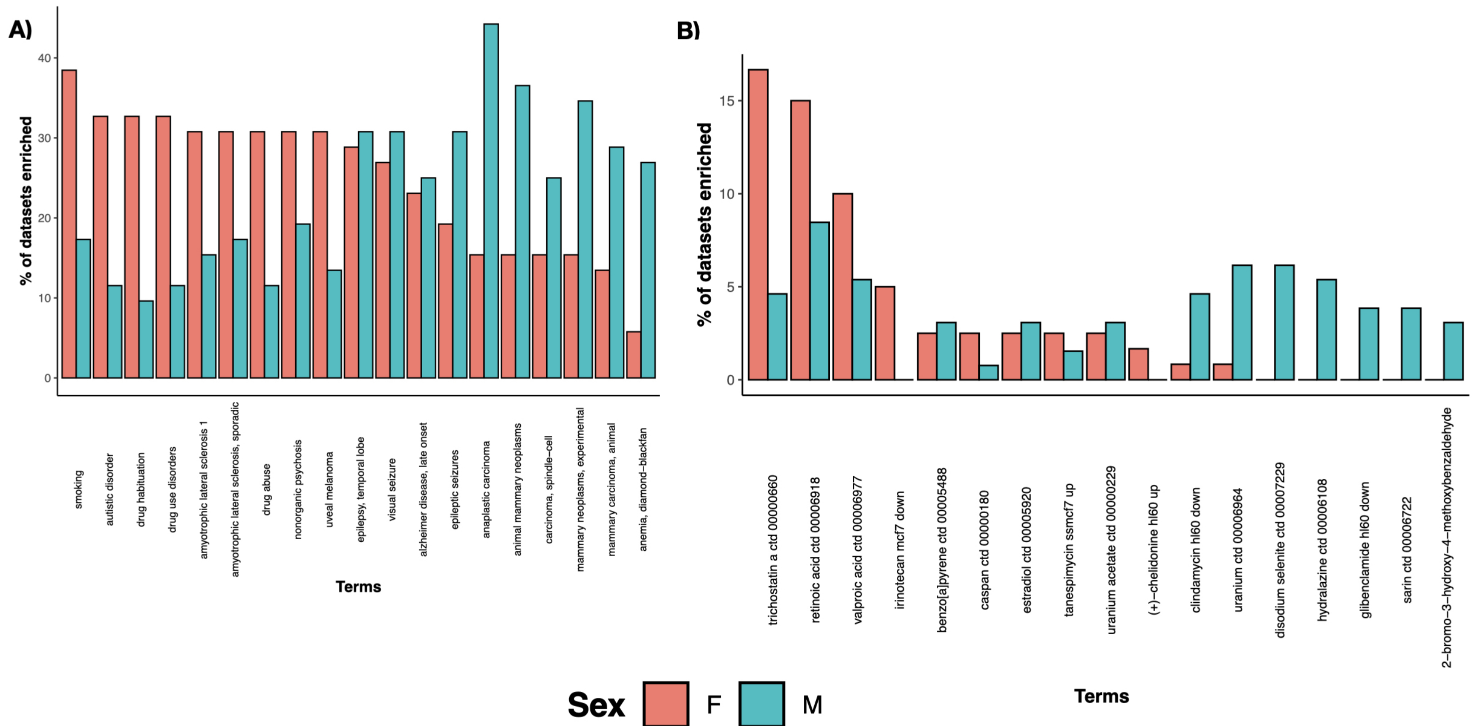


Figure 5

SG-biased genes show biased enrichment for diseases (includ- ing 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. Legend: NS: not significant; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$; ****: $p<0.0001$.

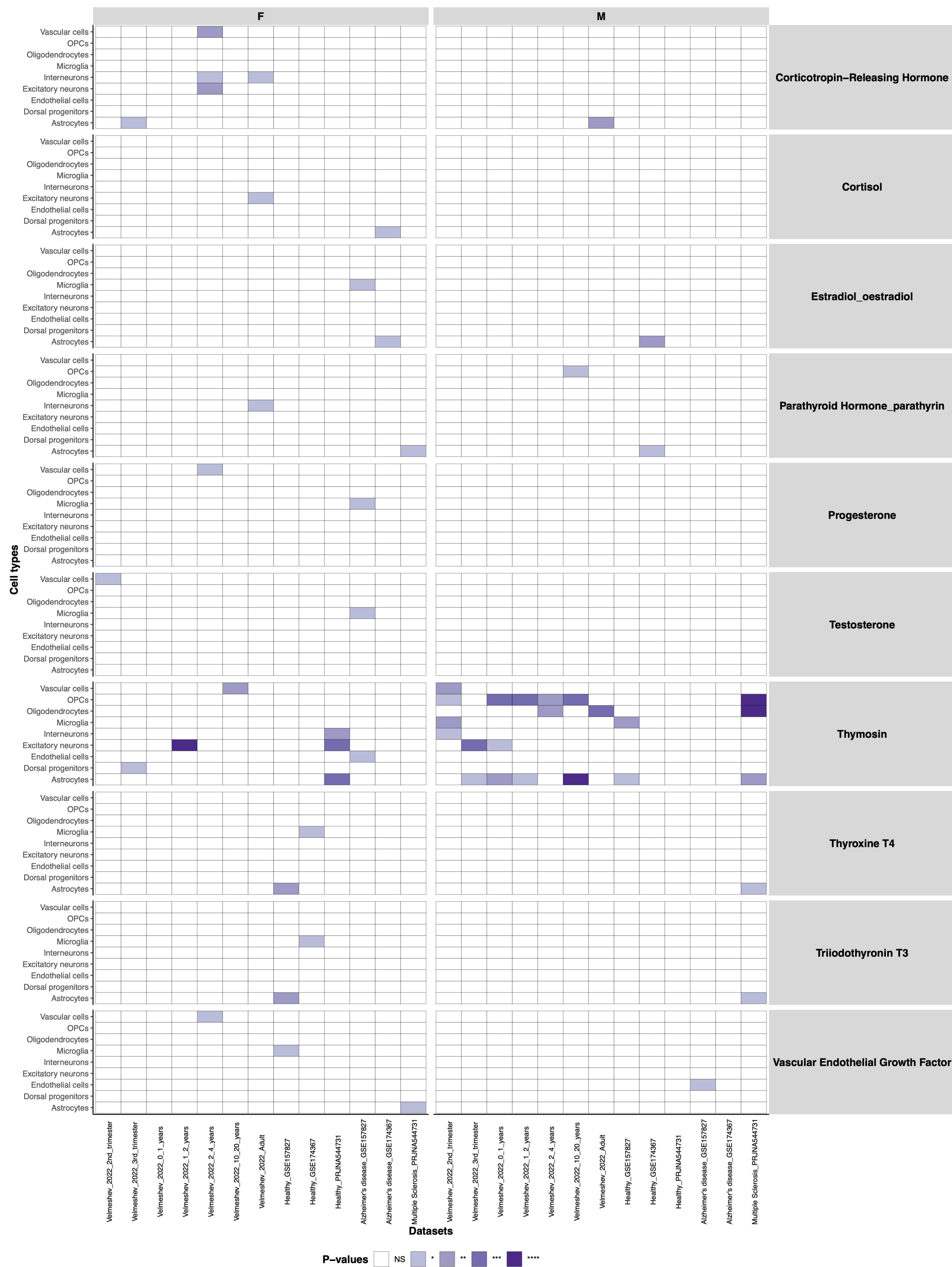


Figure 6

The SG-biased genes enrichment for hormone targets. This heatmap shows the p-values of enrichment for hormone targets from the literature in the SG-biased genes. Albeit most hormones are enriched in both sexes, there are some SG specific enrichments. For example, cortisol and progesterone are solely enriched in females. On the other hand, thymosin was strongly enriched in males, much more than in females,

across multiple cell types and datasets. Legend: NS: not significant; *: p<0.05; **: p<0.01; ***: p<0.001; ****: p<0.0001.

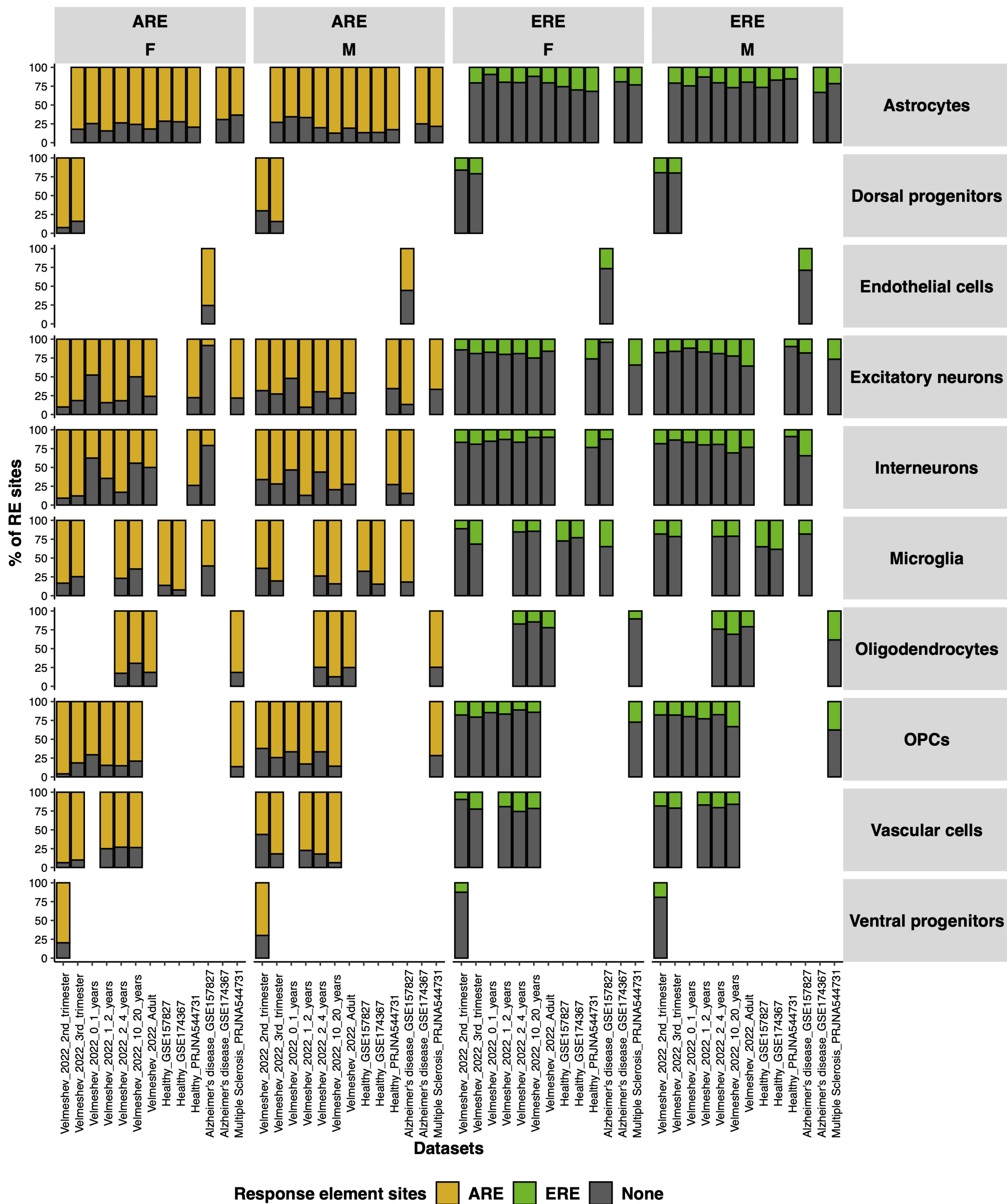


Figure 7

Androgen response element binding sites, but not estrogen, are consistently enriched in SG-biased genes across sexes, cell types and datasets. Androgen response element (ARE) and estrogen response element

(ERE) sites presence in the SG-biased genes, by sex, cell type and dataset. The majority of the SG-biased genes had ARE binding sites, in both sexes. ERE sites instead were present in less than one fourth of SG-biased genes overall. ”None” indicates SG-biased DEGs which did not have neither ARE nor ERE binding sites.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfile1.csv
- Supplementaryfile2.csv
- Supplementaryfile3.csv
- Supplementaryfile9.csv
- Supplementaryfile11.csv
- Supplementaryfile16.csv
- Supplementaryfile17.csv
- Supplementaryfile19.csv
- Supplementaryfile20.csv
- Supplementaryfile21.csv
- Supplementaryfile4.xlsx
- Supplementaryfile5.xlsx
- Supplementaryfile6.xlsx
- Supplementaryfile7.xlsx
- Supplementaryfile8.xlsx
- Supplementaryfile10.xlsx
- Supplementaryfile12.xlsx
- Supplementaryfile13.xlsx
- Supplementaryfile14.xlsx
- Supplementaryfile15.xlsx
- Supplementaryfile18.xlsx
- Supplementaryfigures.pdf