Single cell landscape of sex differences in the progression of multiple sclerosis

Irene Soler-Sáez[1], Borja Gómez-Cabañes[1], Rubén Grillo-Risco[1], Cristina Galiana-Roselló[1], Lucas Barea-Moya[2], Héctor Carceller[3,4,5], María de la Iglesia-Vayá[5], Sara Gil-Perotin[2], Vanja Tepavcevic[6], Marta R. Hidalgo[1,7,*], Francisco García-García[1,*]

[1] Computational Biomedicine Laboratory, Principe Felipe Research Center (CIPF), 46012, Valencia, Spain

[2] Research Group in Immunotherapy and Biomodels for Autoimmunity, Instituto de Investigación Sanitaria La Fe, Valencia, Spain.

[3]Neurobiology Unit, Program in Neurosciences and Institute of Biotechnology and Biomedicine (BIOTECMED), Universitat de València, Burjassot, Spain.

[4]Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Spanish National Network for Research in Mental Health, Madrid, Spain.

[5]Joint Unit in Biomedical Imaging FISABIO-CIPF, Foundation for the Promotion of Health and Biomedical Research of Valencia Region, València, Spain.

[6]Achucarro Basque Center for Neurosciencie, Bº Sarriena s/n, 48940, Leioa, Spain.

[7]Area of Applied Mathematics, Department of Applied Mathematics, Universitat de València, Burjassot, Spain.

Corresponding authors:

Marta R. Hidalgo, marta.hidalgo@uv.es

Francisco García-García, fgarcia@cipf.es

Abstract

One of the major challenges in addressing multiple sclerosis is to understand the progression trajectory of patients. The pathological process evolves from acute phases predominantly driven by inflammation transitioning to progressive profiles where neurodegeneration takes precedence. It remains unresolved why this course is highly heterogeneous among patients. Currently we know that sex variable plays a crucial role in its understanding. Females are 2-3 times more likely to suffer from multiple sclerosis while males' progression is faster with greater severity. We investigate the potential molecular mechanisms underlying these sex-differential clinical traits analysing transcriptomic data at single cell resolution. 48,919 central nervous system and 336,934 peripheral immune cells, covering the multiple sclerosis spectrum, enabled us to provide the comprehensive landscape of sex differences by cell type. This includes signatures in gene expression patterns, functional profiling, pathways activation and cell-cell communication networks for females, males and their sex-differential profiles. Complete results can be explored in the user-friendly interactive webtool https://bioinfo.cipf.es/cbl-atlas-ms/. Among these findings, we unveiled that female neurons may exhibit protective mechanisms against excitotoxicity, glial cells dysregulated widely stress response genes in a sex-differential manner, and female oligodendrocytes increase expression of axon-myelin contact genes suggesting strong potential for myelin recovery. In the inflammatory-predominant forms, female immune cells present an inflammatory core driven by the AP-1 transcription factor, while male adaptive immune cells exhibit higher mitochondrial impairment. Conversely, larger differences are reported in CD8+ T cells, with females displaying homeostasis recovery patterns and males exhibiting cytolytic profiles. We consider that the molecular description of sex differences in multiple sclerosis progression may be a valuable resource for prevention and diagnosis through biomarker research and the development of personalised therapeutic strategies.

Background

Multiple sclerosis (MS) is a chronic, autoimmune and neurodegenerative condition of the central nervous system (CNS). This disease significantly impacts patients' quality of life, being the leading cause of non-traumatic neurological disability in young adults (25-30 years) (PMID: 29576504). Clinical manifestations are dependent on the damaged CNS areas and evolve throughout the individual's lifespan. They include sensory and motor symptoms, cognitive impairments, visual disturbances and autonomic system dysfunctions, among others (PMID: 27889190, PMID: 32949546). MS disability course is presented as a continuum of molecular and cellular events during ageing that accentuate and compensate for the clinical traits of the disease. These events are driven by interconnected neurodegenerative, inflammatory and reparative mechanisms during progression (PMID: 36410373). On this basis, immune cells infiltrate the CNS, where they range from actively constituting demyelinated plaques to accumulating in areas like the perivascular spaces. These defective autoimmune responses partly determined myelin destruction and axonal damage, which comprised MS hallmarks. Prolonged deviation from CNS physiological norms also promotes hyperexcitability through increased glutamate neuronal release and the generation of reactive oxygen species, enhancing neurodegeneration. Moreover, astrocytes, microglia, and oligodendrocytes play dual roles involving both protective (e.g., myelin recovery and maintenance of lipid homeostasis) and harmful (e.g., enhanced proinflammatory environment) activities (PMID: 30687321, PMID: 35190704). For the ease of communication, the spectrum of MS progression has been classified. The most common form is called Relapsing-Remitting MS (RRMS). It fluctuates among periods of relapses, driven by exacerbated inflammatory status, and followed by subsequent remissions. If this pattern evolves into a progressive course, it is renamed to Secondary Progressive MS (SPMS). Lastly, Primary Progressive MS (PPMS) is characterised by chronic neurological decline from the onset, predominantly driven by neurodegeneration and not preceded by relapses (PMID: 25398229, PMID: 24871874).

The sex variable contributes substantially to several epidemiological and clinical aspects of the disease (PMID: 34638664). Females tend to develop MS earlier, exhibit a higher prevalence (with a ratio of 2-3:1), and suffer more severe inflammatory patterns (PMID: 27364395). In contrast, males experience faster and drastic CNS deterioration, with greater atrophy in neuronal lesions (PMID: 37122747, PMID: 32859258). This is reflected in the clinical categorisation of the disease. Females are more prone to suffer RRMS (with more relapses) while the progressive forms SPMS and PPMS present a much more balanced ratio between the sexes (PMID: 26046348, PMID: 27870415). Studies in both human and model organisms have sought to unravel these sex-driven disparities,

primarily focusing on specific genes or molecular mechanisms (PMID: 37122747). Thus, a global unbiased understanding of sex differences in MS remains unexplored. We aim to uncover novel insights into how sex differences manifest at the single-cell resolution, thereby providing a deeper knowledge of the mechanisms driving these disparities in MS.

Here we report comprehensive atlases of sex differences in MS subtypes. We performed an in silico analysis of single-cell RNA sequencing (scRNA-seq) and single-nucleus RNA sequencing (snRNA-seq) data, compiling the state-of-the-art research on MS pathology. This allowed us to analyse 48,919 CNS cells and 336,934 peripheral blood immune cells covering RRMS, PPMS and SPMS. We examined sex-specific and sex-differential changes within each cell type in four areas: gene expression patterns, functional profiles, signalling pathways, and cell-cell communication interactions. Our results reveal significant changes that could explain sex differences in MS progression, such as potential compensatory mechanisms to cope with excitotoxicity in female neurons, sex-differential expression of neuroglial oxidative stress modulators, and sex differences in the activation status of hematopoietic lineages during progression. These findings may have important implications for understanding the molecular mechanisms underlying MS progression, and for designing novel sex-based therapeutic strategies to enhance neuroprotection; likely to efficiently control inflammation, promote myelin recovery and delay neurodegeneration.

Methods

Literature review and selection of studies

We conducted a systematic review in April 2022 following the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines (PMID: 19621072) to obtain a robust and standardised dataset search. Inclusion criteria were implemented by the keywords “multiple sclerosis”, “Homo sapiens”, “single cell” or “single nuclei” or “single nucleus” in the public databases GEO (PMID: 11752295), ArrayExpress (PMID: 31665515) and UCSC Cell Browser (PMID: 34244710). Web browser searching was also performed.

The identified studies were manually filtered by the following exclusion criteria: i) data type: not containing scRNA-seq or snRNA-seq data, ii) disease examined: studies not based on MS, iii) experimental design: required MS patients as cases and healthy individuals as controls, iv) sex: sex variable not registered or not available, v) sample

count: at least data from three different individuals at each condition and sex (control females, MS females, control males, MS males), and vi) unavailability of the gene expression matrix or metadata files. Finally, raw count matrices and sample metadata were downloaded for each of the selected studies.

Bioinformatics analysis

The bioinformatics analysis was performed individually for each of the selected studies. Firstly, we processed the data as follows: i) quality control filtering, ii) normalisation, iii) high variable genes selection, iv) dimensionality reduction, v) cell identity clustering, vi) cell type annotation, and vii) gene markers evaluation. Then, for each cell type, we characterised MS disease in females, in males and MS sex-differences by unveiling statistically changes in gene expression patterns, functional profiles, signalling pathways and cell-cell communication interactions. The whole bioinformatic code was developed using the programming language R (version 4.1.2) (R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. https://www.R-project.org/). The version of the corresponding packages can be consulted in Supplementary Table 1.

Data acquisition, preprocessing and cell type annotation

Gene expression matrix and metadata files were downloaded to develop the individual analyses. Firstly, we unified each set of data into a *SingleCellExperiment* object using the homonymous R package (PMID: 31792435). We also ensured the standardisation of the gene nomenclature, verifying that Gene Symbols were available in all datasets.

We performed afterwards a quality control analysis to incorporate solely data coming from viable cells and nuclei in the downstream processing. This step was adapted to suit the singularities of each dataset, maintaining homogeneity among studies as much as possible. The following parameters were calculated by sample: number of cells or nuclei, library size, number of expressed genes, and mitochondrial gene ratio. The graphical distribution of each parameter was represented before and after filtering to explore if some samples should be eliminated. Then, pertinent cutoffs were implemented to filter poor quality cells/nuclei. In detail, for UCSC-MS dataset we removed nuclei with less than 1000 library size, less than 500 gene expressed,

anomalous values (outliers) for mitochondrial gene ratio identified with *isOutlier* function from scuttle R package (PMID: 28088763), or nuclei aggregates detected with scDblFinder R package (Germain P (2021). scDblFinder: scDblFinder. R package version 1.6.0, https://github.com/plger/scDblFinder). We also performed the quality control analysis at gene level discarding genes expressed in less than 3 nuclei or genes other than protein-coding genes, microRNAs or lncRNAs. To identify these gene types we followed HGNC (HUGO Gene Nomenclature Committee) guidelines. On the other hand, the graphical distribution of parameters before filtering revealed that both GSE144744c1 and GSE144744c3 datasets were previously filtered by the authors (PMID: 33748804), corroborating that there were no outliers. Thus, we did not perform additional filtering steps.

Then, we normalised filtered data by implementing a deconvolution approach (https://doi.org/10.1186/s13059-016-0947-7) and subsequently transformed the obtained values into a log2 scale. We estimated for each gene the biological and technical variability considering batch effects reported in the metadata files. The 20% of the highest variable genes (HVG) based on biological heterogeneity were selected. All these steps were executed using the scran R package (PMID: 27909575). Expression levels of HVG were used as input data to implement dimensional reductions approaches that further compact expression profiles. In detail, summarization was assessed by PCA (*Principal Component Analysis*) using scater R package (PMID: 28088763). The relevant principal components were selected applying the elbow point method with PCAtools R package (Blighe K, Lun A (2021). PCAtools: PCAtools: Everything Principal Components Analysis. R package version 2.4.0, https://github.com/kevinblighe/PCAtools). Moreover, the non-linear dimensionality reduction strategies tSNE (*T-distributed Stochastic Neighbour Embedding*) and UMAP (Uniform Manifold Approximation and Projection) were computed for visualisation (PMID: 28088763).

Cells were classified in groups afterwards, based on similarity of the principal components previously selected. To this end, we constructed a graph built on the *Shared Nearest Neighbours* algorithm using scran R package (PMID: 27909575). We established groups as regions highly connected in the graph, considering each group as a different cell identity. Groups were defined with the *walktrap* algorithm with igraph R package (Csardi G, Nepusz T. The Igraph Software Package for Complex Network Research. InterJournal. 2005 Nov 30;Complex Systems:1695). We assigned the cell type to each cell and nucleus by contrasting its expression profile with public references. We used BRETIGEA R package (PMID: 29892006) for nervous tissue samples (UCSC-MS dataset) and SingleR package (PMID: 30643263) for peripheral blood mononuclear cells (PBMC) samples (GSE144744c1 and GSE144744c3 datasets). For the latter, we selected the *MonacoImmuneData* reference from celldex R package (PMID: 30643263), which is a

standard reference to analyse human PBMC samples. We discarded unassigned cells and neutrophils, basophils, progenitor cells, and T cells without classification in CD4+ or CD8+. Then, we screened for genes overexpressed in each cell type compared to the rest of the cell types. The statistical Wilcoxon test was implemented with scran R package (PMID: 27909575). Adjusted p-values were calculated by Benjamini-Hochberg (BH) method (PMID: 11682119) considering significant genes when False Discovery Rate (FDR) < 0.05. Finally, we corroborated that significantly overexpressed genes by cell type were marker genes described in scientific literature, increasing the robustness of the annotation.

Tested comparisons

We performed differential gene expression and several functional analyses to statistically identify differences in MS disease from a sex perspective (see following sections) testing three independent comparisons for each cell type analysed. Thoroughly, the comparison (MS.Female - Control.Female) assessed the impact of the disease in females (IDF), as it unveiled the differences among MS patients and healthy individuals being female. Likewise, the comparison (MS.Male - Control.Male) reported the impact of the disease in males (IDM), revealing the differences among MS patients and healthy individuals being male. Finally, SDID comparison ((MS.Female - Control.Female) - (MS.Male - Control.Male)) tested the sex-differential impact of the disease (SDID). With this comparison, we aim to identify the sex differences among MS patients taking into account the inherent sex variability in healthy individuals.

We calculated the corresponding statistics to find out whether the detected change is significant, and logFC values to establish the magnitude (logFC absolute value) and the direction (logFC sign) of this change. For further guidance on the interpretation of the logFC, please refer to the "Help" section of our website https://bioinfo.cipf.es/cbl-atlas-ms/. To delve deeper into the biological results, we have selected those characteristics that are significant in all three comparisons.

Differential gene expression analysis

We carried out differential gene expression (DGE) analysis using MAST R package (PMID: 26653891). To this end, a hurdle model was implemented by *zlm* function. A logistic

regression was considered to model if a gene would be expressed (discrete distribution), and a conditioned gaussian linear model to represent the level of expression (continuous distribution).

For every analysis, the model included a variable representing the condition and sex conforming to four groups (control females, MS females, control males, MS males), and another variable containing the scaled number of genes (MAST recommendation). Moreover, we added additional covariables to minimise the effect of confounders in the gene expression patterns. For UCSC-MS comparisons we incorporated: sample, lesion state, age, affected cerebral region, capture batch effect, sequencing batch effect and cell cycle state. On another note, for GSE144744c1 and GSE144c3 datasets we reported the variables sample, age, previous treatments, batch effect and cell cycle state. The cell cycle state was a source of variability not provided in the corresponding metadata files in any dataset, so it was calculated using the *cyclone* function (PMID: 26142758) from scran R package (PMID: 27909575).

After modelling, we performed a likelihood-ratio test to calculate the statistical parameters for each dataset, cell type and comparison. We also determined the logFC values to know the magnitude of the differential expression (logFC absolute value), and in which direction of the comparison the genes were more expressed (logFC sign). We adjusted p-values by BH method (PMID: 11682119), considering statistical significance when FDR < 0.05 and logFC absolute value > 0.5.

Functional profiling

We performed protein-protein interaction (PPI) analyses using STRINGdb R package (PMID: 30476243). Interaction networks were obtained keeping default parameters examining the total number of physical and functional interactions present in the database. For visualisation purposes, we represented weighted edges based on the confidence of the interaction (the greater the thickness, the greater the confidence) hiding disconnected proteins.

Additionally, we identified altered biological processes from the Gene Ontology (BP-GO) through a modular enrichment analysis implemented with the weight01 algorithm from topGO R package (PMID: 16606683). We worked with significant genes and BP-GO terms obtained from org.Hs.eg.db R package (Carlson, M. org.Hs.eg.db: Genome wide annotation for Human. R package version 3.8.2. (2019)) as input data. We calculated

p-value and logFC statistics. Significant terms were clustered, and a word cloud was generated for each cluster. This process was implemented with the simplifyEnrichment and simplifyGO R packages (PMID: 35680096). Semantic similarity matrices were obtained groupting terms by the Louvain clustering method. Word clouds were created, with the size of each word being proportional to its frequency. The variety of graphical representations generated to display functional results were elaborated with the R packages simplifyEnrichment (PMID: 35680096), ComplexHeatmap (PMID: 27207943) and ggplot2 (ggplot2: Elegant Graphics for Data Analysis (3e), https://ggplot2-book.org/).

We created visual atlases of the genes significantly dysregulated in each cell type. Genes were grouped into general functional categories based on BP-GO) terms and associated functional description from the literature.

Signalling pathways analysis

We spotted statistical differences in signalling pathways activity using hipathia R package (PMID: 28042959). Signal transduction was calculated for every effector subpathway of each pathway from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (PMID: 36300620). Differential effector activation (DEA) analysis was performed with MAST R package (PMID: 26653891), following the specifications described in the DGE analysis preceding section. Adjusted p-values were estimated by BH methodology (PMID: 11682119), determining statistical significance when FDR < 0.05. LogFC was also calculated, to determine the direction and the magnitude of the change. We selected the suitable pathways to be analysed . Criteria was established based on the type of tissue the samples came from (Supplementary Table 2).

Cell-cell communication analysis

Interactions between ligand-receptor pairs were inferred using the R package CellChat (PMID: 33597522). The CellChatDB.Human database was used to screen a total of 1939 ligand-receptor interaction pairs, involving 546 different ligands and 507 different receptors. We computed the communication probability for each interaction between pairs of cell types in each group of interest (control females, MS females, control males and MS males), being these values a measure of the interaction strength. The direction of interactions was assessed by establishing the cell type that provides the ligand and/or the receptor. Interactions were considered significant when FDR < 0.05. We then determined cellular communication networks by counting the total number of interactions, and calculated the differential number of interactions among females (MS

females vs. control females comparison) and males (MS males vs. control males comparison).

Web platform development

The web tool (https://bioinfo.cipf.es/cbl-atlas-ms/) has been developed under the structure of the R shiny package (Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B (2024). shiny: Web Application Framework for R. R package version 1.8.1.9000, https://github.com/rstudio/shiny, https://shiny.posit.co/.). This easy-to-use resource is divided into seven sections: 1) "Home" section, to provide a summary of the results available on the website; 2) "Gene expression" section, to explore changes in the expression of the genes of interest; 3) "Functional profiling" section, to delve into the functional profiling results; 4) "Signalling pathways" section, to identify the differential activation of protein effectors in signalling pathways of interest; 5) "Cell-cell communication networks" section, to inspect the ligand-receptor interactions strengths; 6) "Study overview" section, to review the outline of dataset selection and the implemented bioinformatic methods; and 7) "Help" section, to assist in the interpretation of the comparisons and the displayed results.

Results

Systematic review and exploratory analysis

The implementation of inclusion criteria in public databases and web browser searching resulted in the identification of 31 distinct studies. However, 29 studies were discarded due to the following exclusion criteria: i) data type (n=14), ii) disease examined (n=4), iii) experimental design (n=3), iv) sample count (n=6), and v) unavailability of data access (n=2). Finally, two studies were included in the analysis: UCSC-MS (from *Multiple sclerosis* identifier in UCSC Cell Browser database) and GSE144744 (from GEO database) (Figure 1). UCSC-MS study stores a snRNA-seq dataset coming from *postmortem* brain tissues in SPMS patients, whilst GSE144744 study contains three independent cohorts of scRNA-seq data from peripheral blood mononuclear cells (PBMC) (Supplementary Table 3). Of them, cohort 1 (GSE144744c1, RRMS subtype) was partially incorporated to the

analysis as we excluded samples from drug treated patients; cohort 2 (GSE144744c2, RRMS subtype) was completely discarded due to the absence of male data; and cohort 3 (GSE144744c3, PPMS subtype) was incorporated to the analysis for fulfilling the established criteria. Thus, three different datasets were ultimately included in the analysis: UCSC-MS, GSE144744c1 and GSE144744c3. Complete information about their sample distribution, and their clinical and technical characteristics is summarised in Table 1.

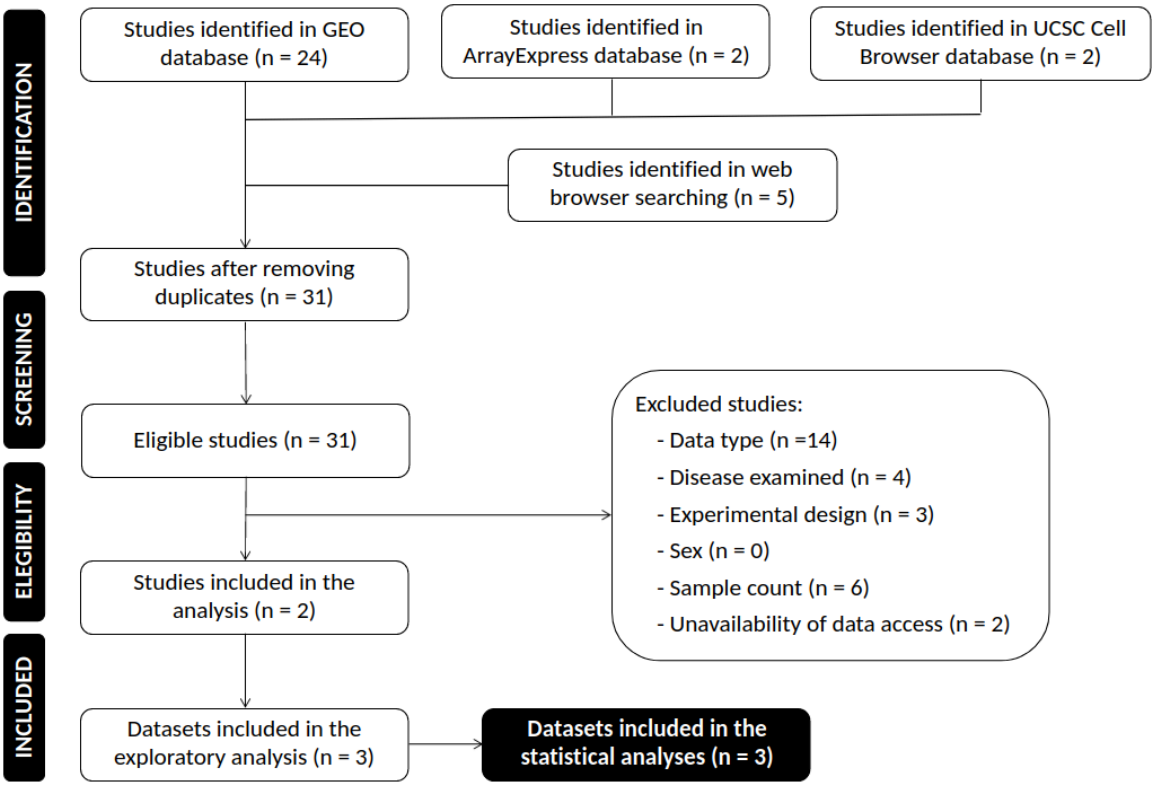

Figure 1. Systematic review workflow following PRISMA guidelines. Specification of the remaining studies number (n) through the identification, screening, eligibility and inclusion phases along with the corresponding exclusion justifications. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

	UCSC-MS	GSE144744c1	GSE144744c3
Database	UCSC Cell Browser	GEO	GEO

Sequencing type	snRNA-seq	scRNA-seq	scRNA-seq
Sequencing method	10x Genomics	10x Genomics	10x Genomics
Sequencing platform	Illumina HiSeq 2500	Illumina Nextseq 500	Illumina Nextseq 500
Sample type	*postmortem* brain tissue	PBMC	PBMC
MS subtype	SPMS	RRMS	PPMS
Patients by condition *	4 : 8 : 5 : 4	5 : 4 : 5 : 5	3 : 3 : 6 : 6
Status of the accessible data	unfiltered raw counts	filtered raw counts	filtered raw counts
Cell count	48.919	71.592	265.342
Gene count	65.217	15.354	15.354
PMID	31316211	33748804	33748804

Table 1. Description of the selected datasets. * Female control : female case : male control : male case. MS: multiple sclerosis; PBMC: peripheral blood mononuclear cells; PPMS: primary progressive MS; SPMS: secondary progressive MS; RRMS relapsing-remitting MS.

All datasets were processed independently. Their results are available in Supplementary Figures S1-S3 (UCSC-MS, GSE144744c1 and GSE144744c3, respectively). The outcome of the data processing and cellular annotation steps led to the identification of the major cell types in each tissue. For each cell type, we unveiled which genes have different expression patterns (DGE analysis), in which functions are these genes involved (functional profiling analysis) and which signalling pathways effectors are differentially activated (signalling pathways analysis); all of them tested in three different scenarios: IDF (MS.Female - Control.Female), IDM (MS.Male - Control.Male) and SDID ((MS.Female - Control.Female) - (MS.Male - Control.Male)). We also performed cell-cell communication analyses to quantitatively characterise cell-cell communication networks for each group (MS female, control female, MS male and control male). In this case, we evaluated the different number of interactions in IDF and IDM scenarios, obtaining homologous results to SDID comparison. As a result we provide precise landscapes of sex differences by MS subtype and cell type, which are further elaborated in the following sections and

can be explored in https://bioinfo.cipf.es/cbl-atlas-ms/. Detailed information on the proceedings can be consulted in the corresponding Methods paragraphs.

Atlas of sex differences in SPMS *postmortem* brain tissue: from glutamate excitotoxicity to myelin repair

We identified five cell types within brain samples: neurons, astrocytes, microglia, oligodendrocytes and oligodendrocyte precursor cells (OPCs) (Figure 2A). The expression pattern of the marker genes evaluated to support cell annotation results can be found in Supplementary Figure 4A, followed by the number of cells by cell type, condition and sex in Supplementary Figure 4B. Interestingly, sex differences have been revealed in all cell types (Figure 2B). It is worth noting the higher number of differentially expressed genes in IDM compared to IDF for astrocytes, oligodendrocytes and OPCs. This tendency is attenuated in the number of significant functions, although still noticeable (Figure 2B black box). Thus, the disease in males appears to have larger biological changes than in females. In addition, regardless of the analysis type, neurons have a greater number of significant features than the other cell types, especially in the signalling pathways analysis.

Although the majority of genes, functions and pathways are specific to one sex, there exist a considerable number of these which are altered at the same time in both sexes, creating a catalogue of different patterns (Figure 2C). We focus on the features significant in all three comparisons (IDF, IDM and SDID) (Figure 2C coloured combinations), that is, features that significantly change in the disease in both sexes but also present significant sex differences. They can present six different patterns, which in turn can be grouped into two by their logFC sign in SDID comparison: significant features increased in females (logFC > 0) and significant features increased in males (logFC < 0) (Figure 2D). For simplicity, henceforth this classification was used to illustrate results by sex. All cell types presented significant features, which are summarised in Supplementary Figures S5 to S9 and Supplementary Table S4. Most of them are cell type-specific, or at most found in two different cell types (Supplementary Figure S10).

We first explored the altered functions to obtain the general picture of the biological processes in which the dysregulated genes are involved (Figure 2E). Compared with males, females showed a higher proportion of functions related to *cell adhesion* (in microglia, neurons, oligodendrocytes and OPCs), and *nervous system development* (in astrocytes, neurons and oligodendrocytes). Conversely, males compared to females

exhibited a higher proportion of functions implicated in *cell death* (in neurons), *cellular homeostasis* (in astrocytes and microglia), and *cell growth* (in OPCs). Furthermore, except for oligodendrocytes, all cell types have a large proportion of terms classified as *Others*, approximately 30 - 50% of the total terms. This highlights the heterogeneity of alterations in specific functions by sex and cell type. With regard to significant pathway effectors, astrocytes most frequent terms were related to *Infection diseases* and *Cell growth and death* (increased in females) and *Signal transduction* (increased in males), while the neuron effectors are mostly related to *Signal transduction* (increased in females) and *Nervous system* (increased in males) (Supplementary Figure 11A).

We also conducted cell-cell communication analyses to complete the sex differences atlas. The number of inferred interactions between cell types for control females, MS females, control males and MS males is described in Supplementary Figure 11B. Notably, the disease patterns for each sex are practically opposite (Figure 2F): females mostly have a higher number of cell-cell pair interactions in disease whilst males in controls. Remarkably, neuron-neuron interactions exhibit an increase in the disease state in both sexes, with males showing a larger magnitude of change.

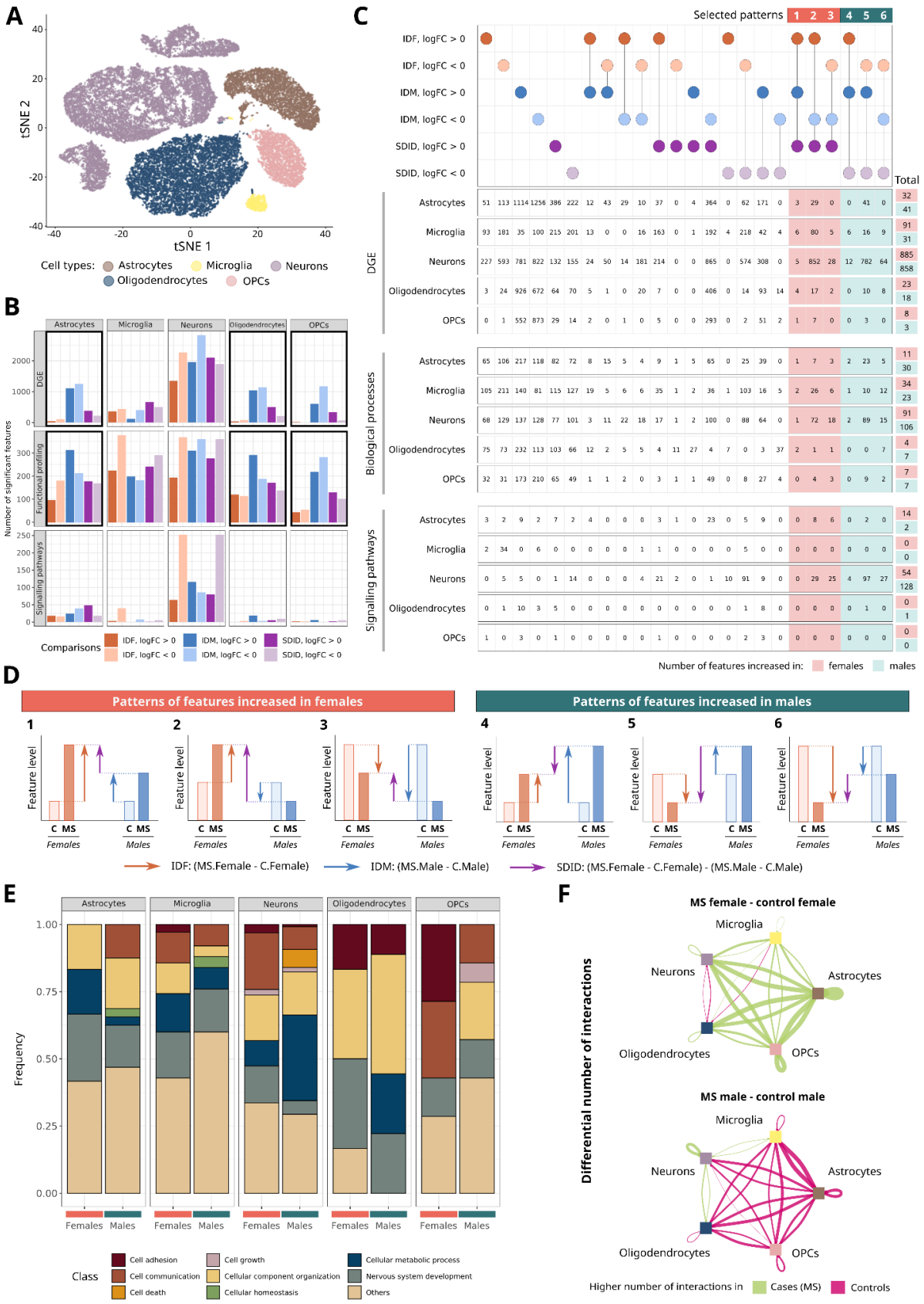
A
tSNE 2
tSNE 1
Cell types: Astrocytes Microglia Neurons Oligodendrocytes OPCs
B
Astrocytes
Microglia
Neurons
Oligodendrocytes
OPCs
DGE
Functional profiling
Signalling pathways
Number of significant features
Comparisons
IDF, logFC > 0
IDF, logFC < 0
IDM, logFC > 0
IDM, logFC < 0
SDID, logFC > 0
SDID, logFC < 0
C
Selected patterns
Total
DGE
Biological processes
Signalling pathways
Number of features increased in: females males
D
Patterns of features increased in females
Patterns of features increased in males
Feature level
C MS
Females
Males
IDF: (MS.Female - C.Female)
IDM: (MS.Male - C.Male)
SDID: (MS.Female - C.Female) - (MS.Male - C.Male)
E
Frequency
Females Males
Class
Cell adhesion
Cell communication
Cell death
Cell growth
Cellular component organization
Cellular homeostasis
Cellular metabolic process
Nervous system development
Others
F
MS female - control female
MS male - control male
Microglia
Neurons
Astrocytes
Oligodendrocytes
OPCs
Differential number of interactions
Higher number of interactions in Cases (MS) Controls

Figure 2. Transcriptomic landscape of sex differences in SPMS brain cell types. (A) Cell type distribution in tSNE dimensions. Each dot represents a cell, coloured by the annotated cell type. (B) Number of significant features by cell type, analysis and comparison tested. (C) Upset map of significant features for each cell type separated by comparison and direction of change (logFC). Dots display whether the feature is significant, with specific colours indicating the comparison and logFC sign. Each row of numbers indicates, for each cell type and specific analysis, the number of significant features corresponding to the evaluated column in the dot map. Coloured squares highlight the significant features in the three comparisons IDF, IDM and SDID; corresponding to six different patterns detailed qualitatively in (D). These features were selected to be described in the rest of the manuscript. (E) Relative frequency distribution of GO terms significantly overrepresented in females (orange) and males (green) by cell type. Individual terms have been classified into general categories (see legend). (F) Differential number of cell-cell interactions between MS females and control females (top) and MS males and control males (bottom). Colour indicates whether there are more interactions in MS (green) or controls (pink); the thickness of the interaction corresponds to the magnitude of change. SPMS: secondary progressive multiple sclerosis; tSNE: t-Distributed Stochastic Neighbour Embedding; OPCs: oligodendrocyte precursor cells; IDF: impact of disease in females (MS females vs control females); IDM: impact of disease in males (MS males vs control males); SDID: sex differential impact of disease ((MS females vs control females) - (MS males vs control males)); DGE: differential gene expression; GO: gene ontology. C: controls; MS: multiple sclerosis.

Sex differential alterations in the astrocyte-microglia-neuron triad implicates synaptic components and stress responses.

Semantic clustering of GO-BP functional signatures in astrocytes, neurons and microglia revealed which major functions are enriched in each sex (Figure 3A). The three cell types show differences in specific functions involving the same semantic areas, as no cell type specific clusters have been identified. Moreover, we found similarities comparing the results between sexes when inspecting the word clouds terms. In detail, RNA processing is altered in both sexes (female cluster 4 and male cluster 1), although males also present enriched functions related to protein degradation and apoptosis (male clusters 1 and 3). Detailed description of these functions can be found in Supplementary Table S5 (female cluster 4) and Supplementary Table S6 (male clusters 1 and 3). The remaining clusters for both female-enriched and male-enriched functions are associated with synapses, denoting the significance of alterations in different elements of the synaptic transmission. To further characterise these differences, Figure 3B summarises the

differentially expressed genes associated directly or indirectly with synaptic features. Neurons have a large number of sex-differential dysregulated presynaptic and postsynaptic terms. Some gene families are consistently increased in the same sex, as voltage-gated calcium channels in females (CACNA1C, CACNB2 and CACNB4), and ATPases subunits involved in maintaining ion gradients, vesicular acidification, and neurotransmitter packaging in males (ATP6V0E2, ATP1A3, ATP6V1F, ATP6V1G1, ATP6V1D, ATP6V1B2, ATP6V1E1 and ATP6V0D1). Neurons also exhibit sex differential activation of several types of chemical synapses. The majority of glutamatergic effectors, which would lead to increased excitability, are more activated in males (Figure 3 left); in spite of most GABAergic effectors, which would result in decreased excitability, more activated in females (Figure 3 right). Cholinergic, serotonergic and dopaminergic synapses display further significant differences (Supplementary Figure S12), sparking an intricate network of sex disparities in neuronal excitability.

Glial cells (e.g. astrocytes and microglia) mainly mediated responses to a wide spectrum of stressors derived from CNS tissue damage (Figure 3B *Oxidative stress and neuroprotection and Glutamate reuptake* boxes). Intriguingly, male astrocytes seem to counteract the enhanced glutamatergic neuronal excitotoxicity previously described with an increase of glutamate reuptake-related genes. Sex-differential stress response in astrocytes also ranges from increased genes related to calcium homeostasis (RYR3, TPD52L1 and TRPM3 in females and CALM3 in males), oxidative stress (REPS1 in females; ENC1, HSP90AB1, PRKCB and UCHL1 in males), hipoxia (HIF3A, EGLN3 and ZBTBW in females; EIF4A2 in males) and neuroprotection (NRG3, RORA, NEAT1, RANBP3L, MALAT1 and LINC-PINT in females; JUND, FOS, FLT1, FAIM2 and HES1 in males). Meanwhile, female microglia notably exhibit transmembrane transporters involved in metabolism and homeostasis maintenance, such as SLC24A2 (calcium/cation antiporter), SLC44A1 (choline transporter), TMEM144 (carbohydrate transporter), TMEM165 (cation/proton antiporter) and CACNA2D3 (calcium transporter), while male microglia display genes with high metal affinity as FTL and MT2A. Additionally, both female and male microglia show increased oxidative stress-related genes (e.g. COA1, CREB5 and FAF1 in females; SRGN and FGF12 in males). Glial cells also present dysregulated genes potentially involved in the myelin recovery process, which are illustrated in the following section.

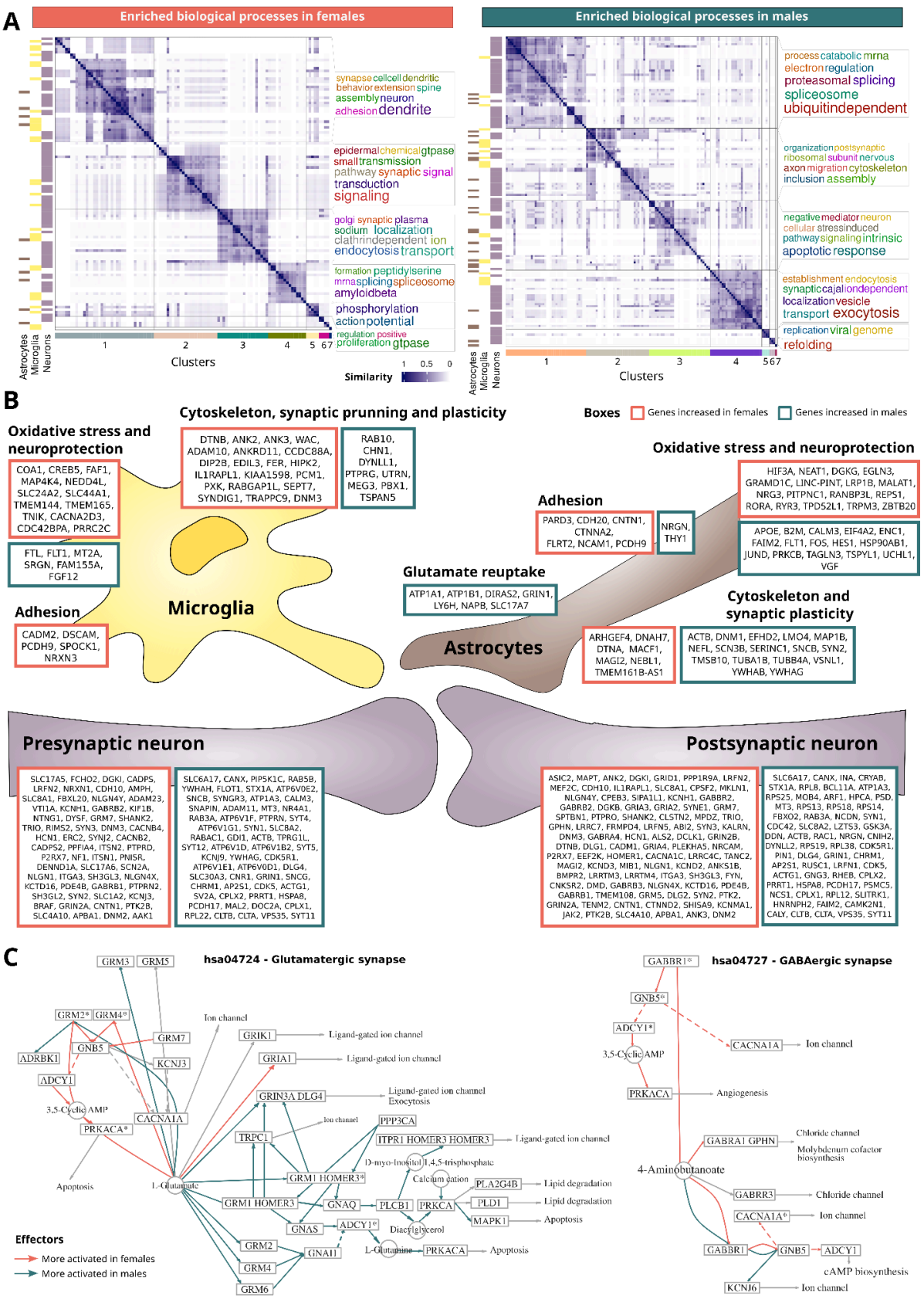

Figure 3. Sex differences in SPMS tetrapartite synapses. (A) Degree of similarity between GO Biological Processes overrepresented in females (left) and males (right) for

astrocytes, microglia and neurons. Each row and column corresponds to a significant GO term. Blue intensity indicates the degree of similarity. Left horizontal lines represent whether the term is significant in astrocytes, microglia and/or neurons. Clusters are shown at the bottom, with the associated word cloud on the right of the plot. (B) Atlas of sex differences in synapse-related genes. Significant genes involved directly or indirectly in synapses have been selected and classified in general categories based on their roles carried out in the cells. Colour boxes indicate the sex patterns: genes upregulated in females (orange) and in males (green). (C) Signalling pathways of glutamatergic (left) and GABAergic (right) synapses. Nodes represent proteins of the signalling pathway and edges the interactions between nodes. Effector proteins are the last node in each subpathway (arrow point). The effector nodes point to the biological functions they exerted. GO: Gene Ontology; BP: Biological Processes; GABA: Gamma-aminobutyric acid.

Sex differential alterations are also present in lipid metabolism and myelin recovery.

To gain insights into the sex differential potential for myelin repair and maintenance, we delved into the significant features of oligodendrocytes and OPCs. Enhancement of neuronal establishment and myelination are the major categories altered in both cell types, either in functions enriched in males or females (Supplementary Figure S13). However, dysregulated genes differ between sexes (Supplementary Table S7). In that way, oligodendrocytes differ in neuronal adhesion and synaptic molecules that support the maintenance of myelin targeting to axons as NCAM2, NLGN1, CADM2, CLDN11, ANK3, IL1RAPL1, CTNND2 and LPHN3 increased in females, while CAMK2B and NRGN in males. Females also display genes that boost myelin repair such as AMER2, a negative regulator of Wnt/ß-catenin pathway, and OMALINC, a lncRNA maker of oligodendrocyte maturation that regulates the expression genes associated with myelination. Interestingly, males also present markers for both oligodendrocyte differentiation (CAMK2B and MIR219A2) and myelinating activity (MAPB1, BCAS1 and MGAT5).

Among the genes upregulated in female OPCs are genes involved in the differentiation to oligodendrocytes and the myelin repair processes (GPNMB, PARD3, QKI, TNR) and with neuroprotective roles (HIF3A, VCAN), whereas males only display MAP1B related with this functions. Next, we delved into the characterization of PDGF (Platelet-derived Growth Factor) and FGF (Fibroblast growth factor) communication signalling in OPCs as the receiving cell type (Figure 4). The two pathways, which drive OPCs to growth and differentiate into oligodendrocytes promoting myelin repair, show higher interaction strength in MS than in controls, both in females and males (Figure 4A-B left). However, the contribution of the donor cells and the strength of ligand-receptor pairs differ according to the sex of the individuals, finding more pronounced changes in females

(Figure 4A-B right). In the PDGF signalling pathway, MS females PDGFC (OPCs) - PDGFRA (OPCs) interaction increases in proportion compared to control females, at the detriment of the interactions between the ligands provided by neurons. Moreover, in MS females, FGF ligands provided by oligodendrocytes increase in interaction strength proportion compared to female controls, whilst astrocyte ligands interactions are decreased. These patterns in PDGF and FGF signalling pathways are not observed when comparing MS males to male controls.

Of note, astrocytes and microglia show significant sex differences in genes related to lipid metabolism with potential to be associated with the myelin clearance process. We observed a higher number of genes increased in astrocytes females than males, such as DGKG, GRAMD1C, LRP1B and PITPNC1. However, males presented increased APOE which could potentiate the accumulation of lipid droplets that may impair its functionality. In regard to microglia, females enlisted genes as CELF2, CHST11, AGO3, ATG4C, GPM6B, LPAR1, LPGAT1, OSBPL1A, PIP4K2A, PLD1, QKI, SOX2-OT, TMEM131 and males RAB10, TNS3, COLEC12, PLSCR1, SCARB1, VIM. These gene sets unveiled the existence of adaptive differences in lipid metabolism and debris clearance from damaged myelin, being potential genes to evaluate the sex-differential impact of MS.

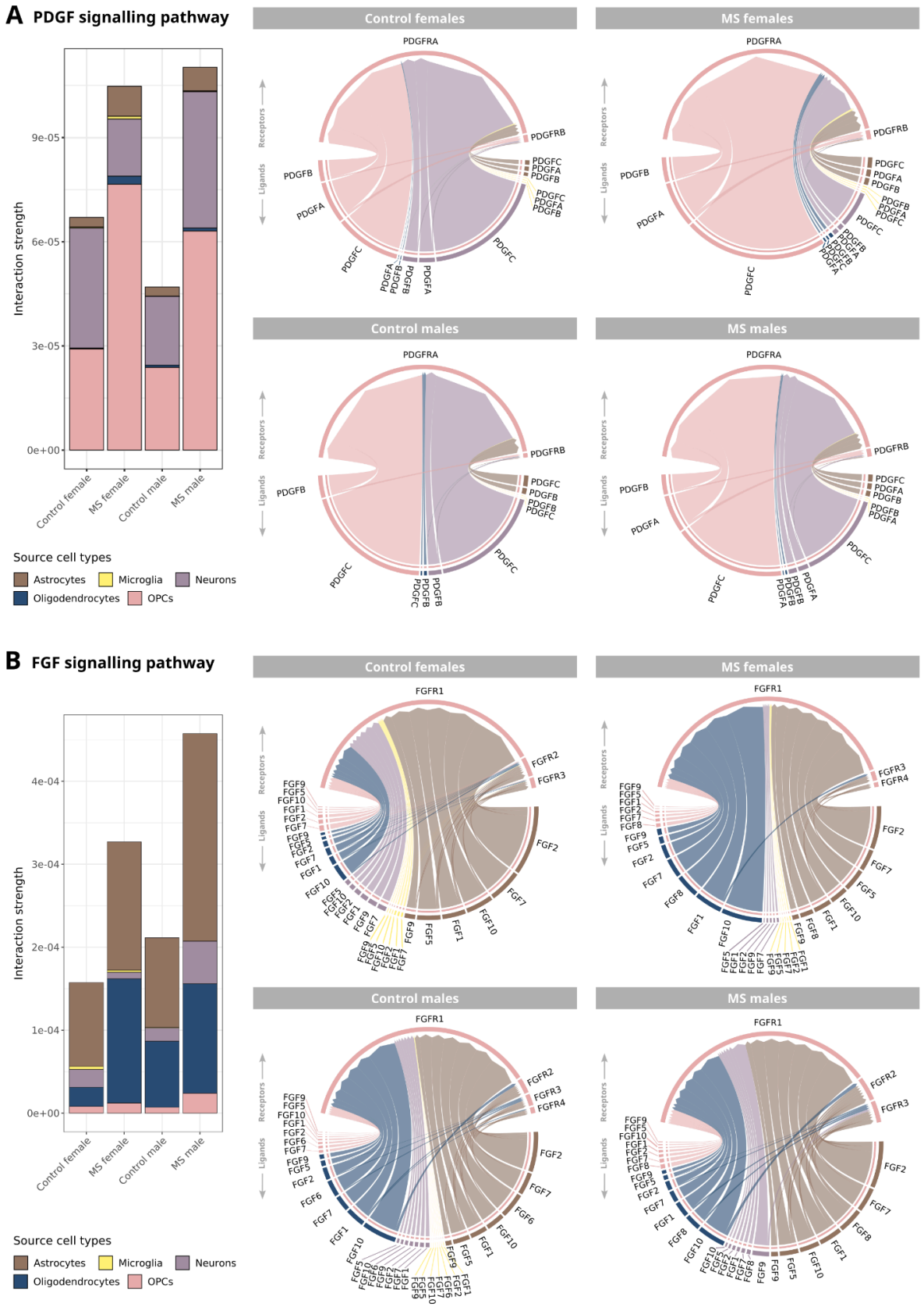

Figure 4. PDGF (A) and FGF (B) signals received by OPCs to promote proliferation and myelination. (Right) Total interaction strength by cell type for each group of

interest: control female, MS female, control male and MS male. (Left) Interaction strength frequency of ligand-receptors pairs. The outer circumference labels the cell type. The inner flux represents the interaction strength proportionally, coloured by the sender cell type which provides the ligand. Sender cells: astrocytes, microglia, neurons, oligodendrocytes and OPCs. Receiver cells: OPCs. PDGF: platelet-derived growth factor; FGF: fibroblast growth factor; MS: multiple sclerosis; OPCs: oligodendrocyte precursor cells.

Atlas of sex differences in RRMS peripheral blood mononuclear cells

B cells, CD4+ T cells, CD8+ T cells, NK cells, dendritic cells and monocytes were the cell types annotated in RRMS dataset (Figure 5A). Marker gene expression, cell distribution by group and sex, and the number of statistically significant features can be found in Supplementary Figure S14. In addition, an overview of the results for each cell type is available in Supplementary Figures S15 to S19. We now illustrate the significant features in the three comparisons as performed in the SPMS atlas (Figure 2D). Due to the limited number of results obtained for B cells and dendritic cells, we focused on CD4+ T cells, CD8+ T cells, NK cells and monocytes (Figure 5A coloured boxes). In all these cell types, females exhibit a higher number of significant functions than males except "Cellular component organisation" and "Others" categories (Figure 5B), in line with the higher number of significantly increased genes compared to males (Figure 5C). Noteworthy, females presented predominance in functional categories such as *Cell communication* and *Apoptosis and stress response* denoting multifaceted inflammatory responses. The exception is found in the category *Metabolism and bioenergetics* for male CD4+ cells, CD8+ cells and NK cells, being highly related with the mitochondrial electron transport. Lastly, the differential activation of signalling pathways and cell-cell communication interactions constitute an intricate framework of sex differences where, depending on the cell type, the feature assessed is increased in females than males or vice versa (Supplementary Figure S20).

Functional profiling reveals the immune signature core of females with relevance of AP-1 transcription factor

Female peripheral immune system displays a sophisticated set of genes that modulate MS response compared to males (Figure 5C). Noteworthy, we found 21 genes dysregulated in at least three of the four cell types evaluated providing a general insight into the state of the innate and adaptive immune system altogether (Figure 5D). These genes constitute a significantly connected protein-protein interaction network (PPI enrichment p-value: < 1.0e-16), where the prominent hub is composed by the AP-1

transcription factor complex (FOS, JUN, JUNB and JUND), a key regulator of several processes of differentiation and inflammatory response. We also identified the inhibitors NFKBIA and NFKBIZ, which could mitigate the characteristic activation of NF-kB pathway in MS response that led to over-inflammation profiles; and the activation biomarker CD69. Meanwhile, we have not identified a common gene core for males. C16orf54 gene, associated with immune infiltration, is the only significant gene in the three cell types CD4+ T cells, CD8+ T cells and monocytes (Supplementary Figure S21). These changes in the transcriptomic profile are also reflected in the differential interaction strength of ligand-receptors pairs in MS females compared to control females, where we detected increased interaction of some signalling pathways mediated by cytokines (e.g. IL6) and cell adhesion molecules (e.g. CD99 and cadherins) that are not reinforced in males (Supplementary Figure S20-B).

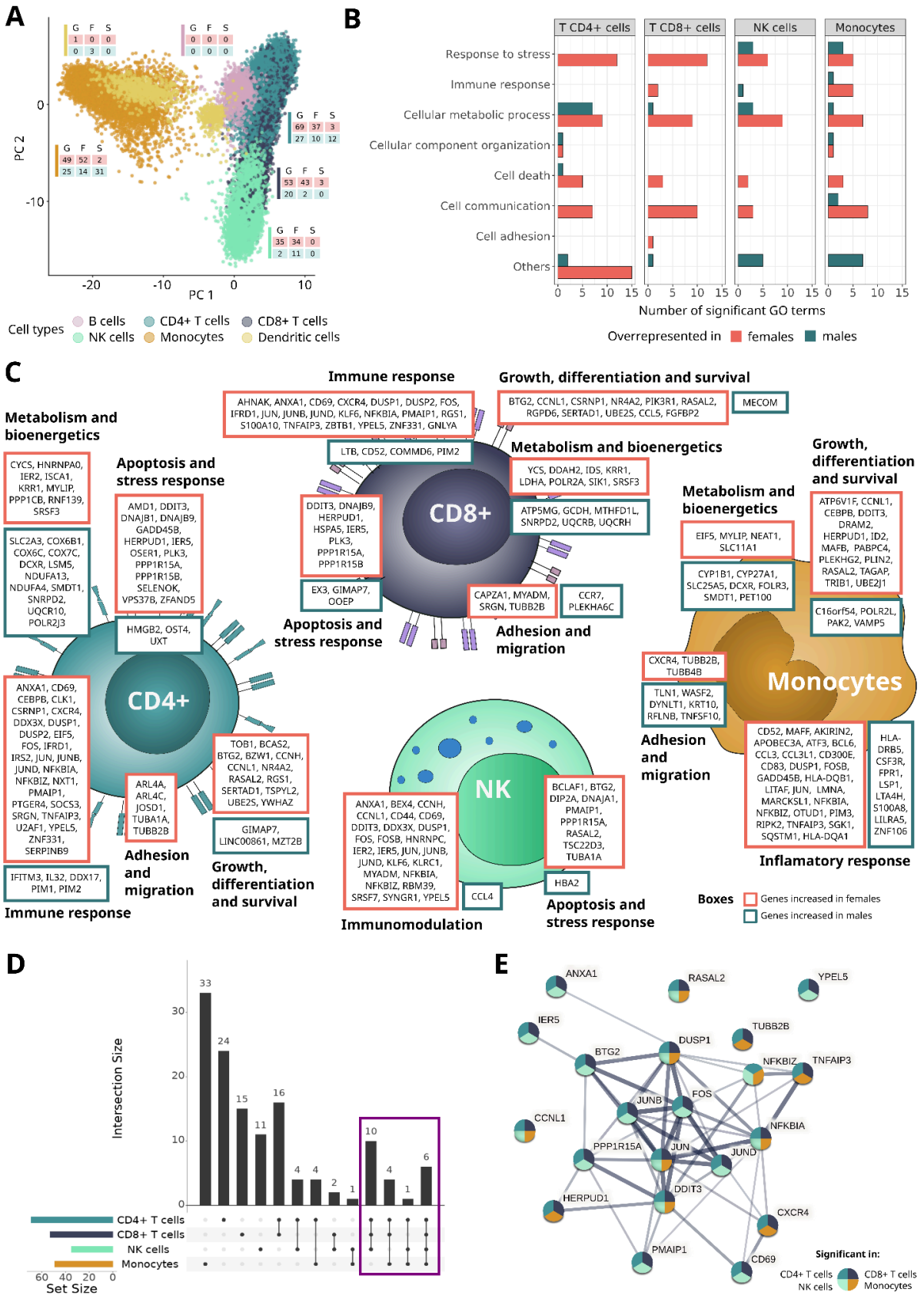

Figure 5. Transcriptomic landscape of sex differences in RRMS peripheral blood mononuclear cells. (A) Cell type distribution in PCA dimensions. Each dot represents a

cell, coloured by the annotated cell type. Coloured boxes indicate the number of genes (G), functions (F) and signalling pathways (S) significantly increased in females (orange) and males (blue). (B) Number of significant GO biological processes by cell type classified under broader biological categories. (C) Atlas of sex differences in gene expression patterns. Genes are classified in general biological terms. Coloured boxes indicate the sex patterns: genes upregulated in females (orange) and in males (green). (D) Upset plot of genes more expressed in females than males by cell type. Horizontal bars represent the number of significant genes in each cell type. Dots indicate the combinations of intersections tested, with the top vertical bars denoting the number of significant genes of the corresponding intersection. Purple square: significant genes in at least three of the four cell types evaluated. (E) Protein-protein interaction network of significant genes more expressed in females than males in at least three of the four cell types evaluated: CD4+ T cells, CD8+ T cells, NK cells and monocytes. Edge thickness indicates the structural and functional confidence of the interaction. NK: natural killer; GO: gene ontology.

Male adaptive immune response exhibits exacerbated mitochondrial dynamics compared to females

Previously, we denoted the absence of a central hub of increased genes in the immune system of males. Nevertheless, we found intriguing results by separating the innate and adaptive immune systems. The adaptive immune system presents different genes functionally related to the mitochondrial electron transport chain (Figure 5C). In detail, CD4+ T cells dispose genes from all steps: NADH to ubiquinone (NDUFA13 and NDUFA4), ubiquinol to cytochrome c (UQCR10) and cytochrome c to oxygen (COX6B1, COX6C and COX7C), while CD8+ T cells exhibit two additional ubiquinol-cytochrome C reductases (UQCRB and UQCRH) and ATP5MG, gene involved in ATP synthesis coupled with proton transport. Thus, males seem to be more susceptible to generate higher levels of reactive oxygen species during oxidative phosphorylation. Meanwhile, the sex-differential changes increased in the males' innate immune system of males are led by monocytes, which reveal a high difference in the number of increased pathways compared to females, mostly of them involved in cell adhesion and inflammatory modulation (Supplementary Figure 20-A). This pattern is also reflected in the increased expression of genes to favour shape-shifting infiltration as DYNLT1, TLN1 and RFLNB, and to mediate innate inflammation as LTA4H, S100A8 and LILRA2 (Figure 5C). We also detected an increase in the interaction of receptor-ligand pairs considering all cell types which involve the response modulation via semaphorins and the T-lymphocyte co-stimulatory molecules CD80 and CD86 (Supplementary Figure 20-C).

Atlas of sex differences in PPMS peripheral blood mononuclear cells

We next explored the PPMS sex-based transcriptomic profiles of B cells, CD4+ T cells, CD8+ T cells, NK cells, dendritic cells and monocytes (Figure 6A). Similarly to RRMS, we characterised the marker gene expression, cell distribution by group and sex, and the number of statistically significant features in Supplementary Figure 22, while the results for each cell type are summarised in Supplementary Figures 23-27. Centering on CD4+ T cells, CD8+ T cells, NK cells and monocytes, all of them dispose of significant functions in broad biological categories with similar proportions between females and males (Figure 6B). Thus, it reveals sex-dependent sex changes in the immune transcriptomic profile that impact several functional layers (Figure 6C). While the innate system demonstrates consistent proportions of changes in NK cells and monocytes, within the adaptive system there are notably more differences observed in CD8+ T cells compared to CD4+ T cells. Sex differences have also been found in the activation of different effectors of signalling pathways (Supplementary Figure 28), and in the strength of interaction between ligand-receptor pairs (Supplementary Figure 29). In the latter, it is important to note that there is almost no common pathway between those activated in MS males (vs. control males) and MS females (vs. control females).

Marked sex differences in PPMS CD8+ T cells comprised predominant cytolysis in males to restoring homeostasis in females

CD8+ T cells exhibit the most pronounced sex differences as previously described (Figure 6A-C), implying distinct disparities that could influence the physiological status between sexes. Both females and males increased genes in CD8+ T cells conformed significant interaction networks, with PPI enrichment p-values of 3.84e-14 and 1.0e-16, respectively (Figure 6D). From them we can infer that females tend to restore cellular homeostasis compared to males, as they increased genes related to the regulation of protein translation (red cluster), the differentiation and survival processes of T lymphocytes (ochre cluster) and energy production through oxidative phosphorylation and mitochondrial maintenance (yellow cluster). On the contrary, males appear to be in a more activated state than females by presenting higher cytolitic responses through granzyme and perforin-mediated apoptotic process (purple cluster), calcium regulation (light green cluster) and the formation of extracellular vesicles (park green cluster).

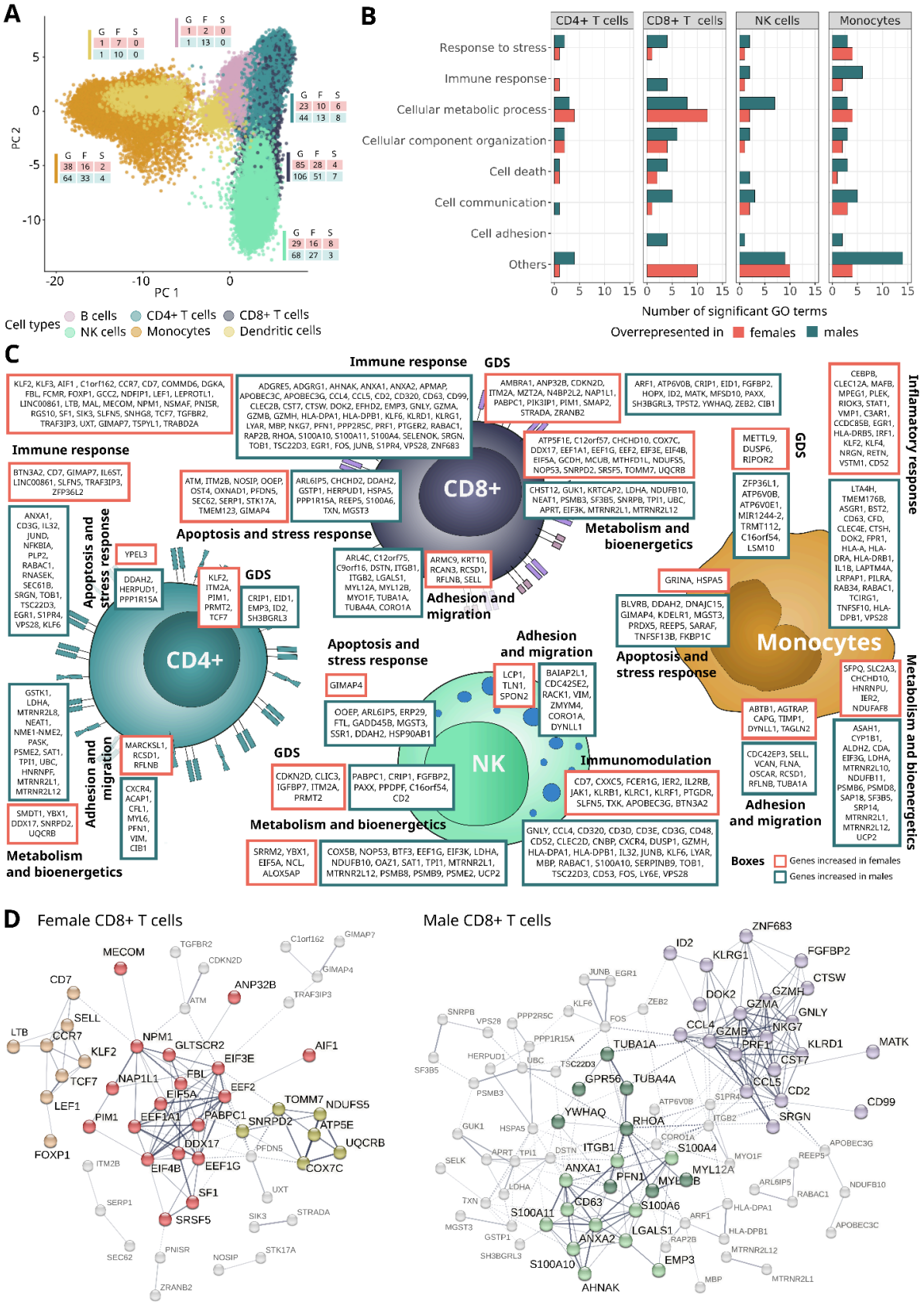

Figure 6. Transcriptomic landscape of sex differences in PPMS peripheral blood mononuclear cells. (A) Cell type distribution in PCA dimensions. Each dot represents a

cell, coloured by the annotated cell type. Coloured boxes indicate the number of genes (G), functions (F) and signalling pathways (S) significantly increased in females (orange) and males (blue).(B) Number of significant GO biological processes by cell type classified under broader biological categories. GO: gene ontology. (C) Atlas of sex differences in gene expression patterns. Genes are classified in general biological terms. Coloured boxes indicate the sex patterns: genes increased in females (orange) and in males (green). (D) Female and (E) male CD8+ T cells protein-protein interaction networks. Edge thickness indicates the structural and functional confidence of the interaction, representing those intra-group (solid line) and inter-group (dotted line) from MCL clustering with inflation parameter 2.2. Clusters of interest are highlighted with colours. PPMS: primary progressive multiple sclerosis.

Immune landscape by disease subtype unravelled sex differences between RRMS and PPMS

Sex differences in immune system status enable the clustering of RRMS and PPMS cell types

We previously delineated the sex-differential patterns in immune cells for both RRMS and PPMS independently. Next, we delve into the major disparities that vary across these MS subtypes in CD4+ T cells, CD8+ T cells, NK cells and monocytes. By clustering them based on genes with wider sex-differential patterns, we observed an initial separation by disease subtype (i.e., RRMS and PPMS), followed by further stratification based on their class (i.e., adaptive and innate immune system) (Figure 7). Importantly, these findings point to disparities in the immune system that may play a crucial role in the clinical variability observed between MS subtypes, and not just subtle variations within a specific cell type. The 67 genes that compose the clustering form a highly connected protein-protein interaction network (PPI enrichment p-value: < 1.0e-16, Supplementary Figure S30) mostly related to stimuli response such as reactive oxygen species, cytokines, lipids and to leukocyte differentiation. In depth, RRMS significant genes are more predominant in females, with a specific core not dysregulated in PPMS (Figure 7 grey box). This set of genes reflects an activated state (CD69) attempting to drive responses especially to stress and apoptosis (e.g. IER5, DDIT3 and BTG2). Conversely, the majority of significant PPMS genes are increased in males, with some of them increased in RRMS females. (Figure 7 yellow box). A high proportion is related to

inflammation modulation (e.g. CCL4, FOS and JUNB), in addition to processes not highlighted in RRMS female core such as glycolytic metabolism (LDHA and TPI1) and translation of mitochondrial proteins (MTRNR2L1 and MTRNR2L12). Interestingly, the adaptive immune system presents a gene set patterned inversely based on the disease subtype; scenario not observed in the innate immune system (Figure 7 orange box). Of note are ANXA1 and SRGN, relevant in the process of CNS infiltration, which are increased in RRMS females and PPMS males.

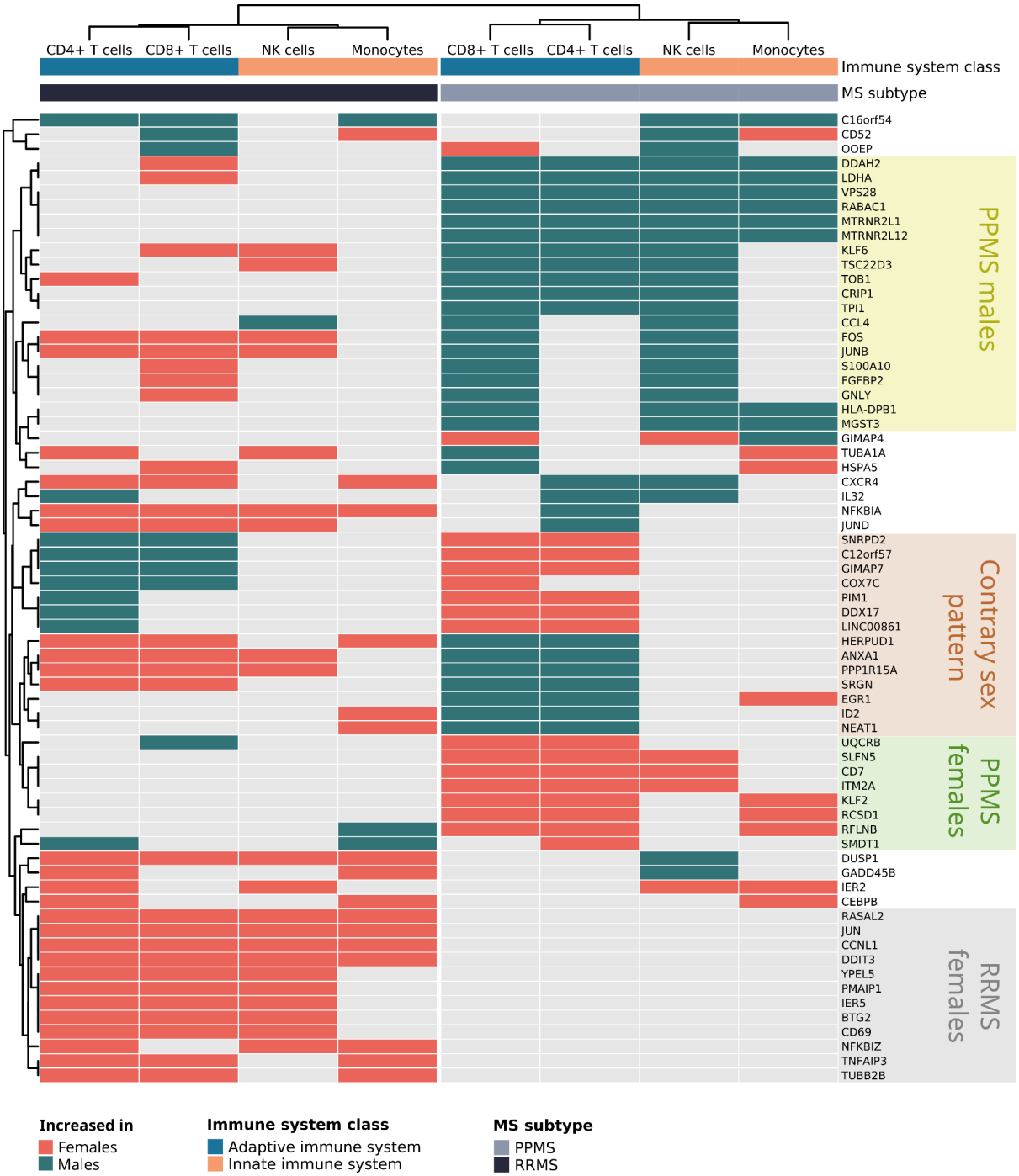

Figure 7. Clustering of RRMS and PPMS immune system cell types based on their sex differential profile. Heatmap showing the classification of CD4+ T cells, CD8+ T cells, NK cells and monocytes in RRMS and PPMS subtypes (columns) from genes with significant expression by sex in at least three of the eight cell types evaluated (rows). Disease subtypes, immune system classification (innate or adaptive) and the sex where the significant increment is detected (females or males) are specified. RRMS: relapsing-remitting multiple sclerosis; PPMS: primary progressive multiple sclerosis.

Unveiled the sex differential viral response and antigen presentation by MS subtype

Considering the extensive association of MS susceptibility with viral response and specifically to genetic variants in the human leukocyte antigens (HLAs), we aim to further investigate whether there are sex differences in this area. RRMS females compared to males present more significant functions related to viral infection, while the opposite pattern is defined in PPMS. In detail, RRMS females exhibit as significant functions *positive regulation by host of viral transcription* (GO:0043923) and *regulation of viral process* (GO:0050792) in CD4+ T cells, *cellular response to virus* (GO:0098586) in NK cells, and *negative regulation by host of viral transcription* (GO:0043922) in monocytes. No significant viral-related function has been found in males. Meanwhile, PPMS males displayed as significant functions p*ositive regulation of defense response to virus by host* (GO:0002230) in CD8+ T cells, *response to virus* (GO:0009615) in NK cells, and *defense response to virus* (GO:0051607) in monocytes. The latter is the only virus-related function overrepresented in PPMS females, also detected in monocytes. The same tendency is established when evaluating the expression pattern of the genes belonging to these functions, which can be explored in Figure 8A.

HLA genes play a critical role in the immune response by presenting foreign antigens to T cells. These genes can influence the susceptibility and protection to MS by regulating the immune system's response to self-antigens. We mainly identified sex differential expression of HLA class II genes, which are primarily expressed in antigen-presenting cells (monocytes, dendritic cells and activated B cells) to present the antigens to CD4+ T cells. Of them, it is remarkable the change of HLA-DRB5 pattern, being increased in male RRMS (monocytes) and female PPMS (B cells, monocytes and dendritic cells). While this gene is the only HLA family gene increased in PPMS females, several genes of the HLA-DR and HLA-DP families are increased in PPMS males (Supplementary Table S8). Major histocompatibility complexes (MHCs) are formed from the translation of HLA genes; where we have also found sex differences in their communication strengths (Figure 8B). The major sex difference in MHC-I is found for both MS subtypes with the HLA-E ligand with CD94:NKG2A and KLRC1 receptors (Figure 8B top). Interestingly, the interactions become significant for the MS female group in RRMS and the MS male

group in PPMS. Related to MHC-II, the interactions are stronger for control females than for control male at the baseline (Figure 8B bottom). In RRMS MS males increase the interaction strength to reach the same level as females, with both sexes showing significant interactions for almost all ligand-receptor pairs. Conversely, in PPMS ligand-receptor interactions disappear in MS females compared to control females, while MS males maintain a similar state to their respective controls.

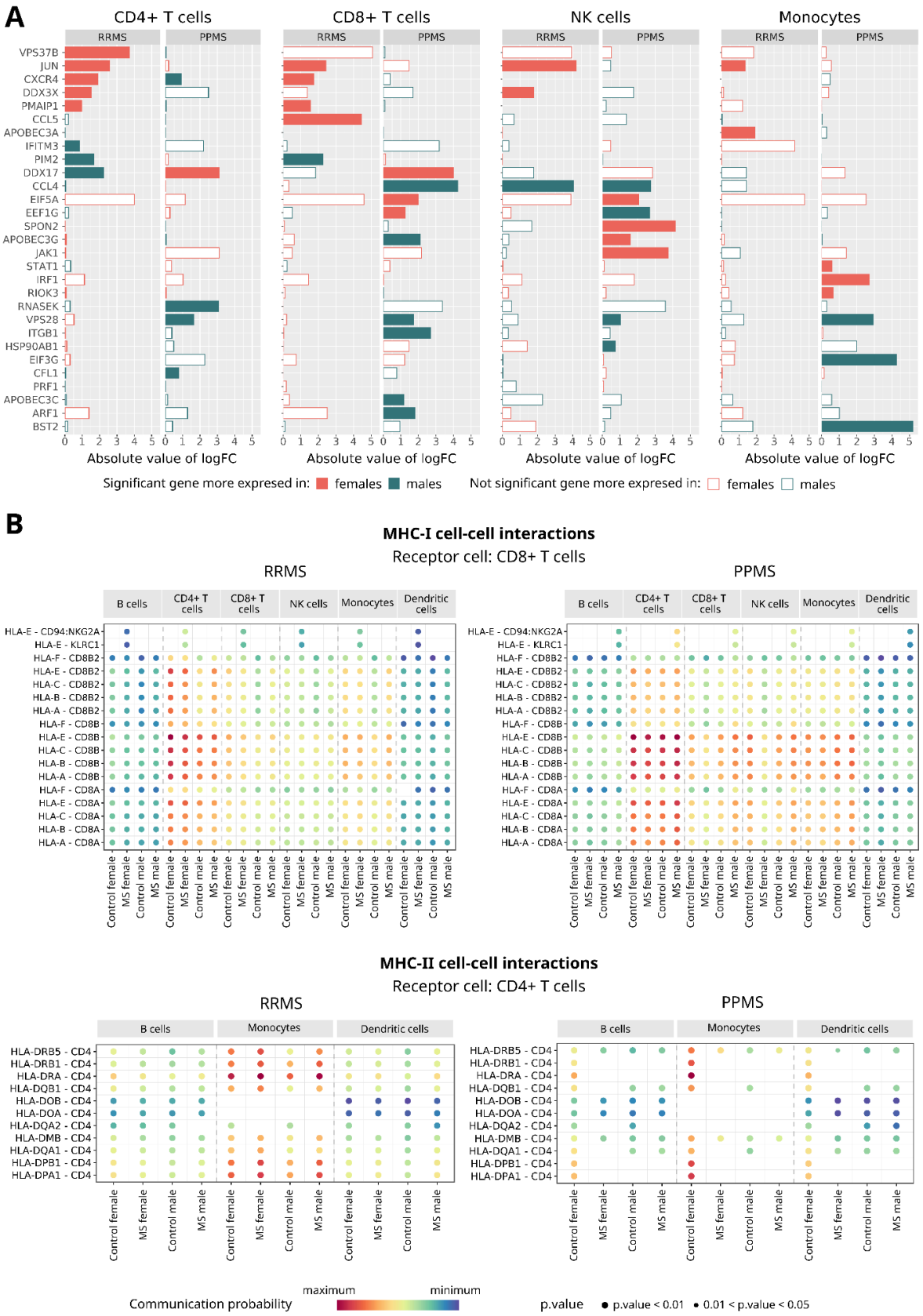

Figure 8. Sex differences in viral response and antigen presentation between RRMS and PPMS. (A) Pattern change of genes associated with viral response by cell type

and by disease subtype. Magnitude value of change in the SDID comparison (X-axis) for each gene extracted from the functions *response to virus* and *viral process* (Y-axis). Genes selected to be significant in at least one cell type. (B) Significant ligand-receptors pairs for MHC-I (top) and MHC-II (bottom) interactions. Dot presence indicates that the interaction has been detected as significant, while the colour reflects the interaction probability (i.e. interaction strength) of the ligand-receptor pair. RRMS: relapsing-remitting multiple sclerosis; PPMS: primary-progressive multiple sclerosis; NK: natural killer; MHC: Major histocompatibility complex.

Atlas-MS web platform

To encourage a wider coverage of this landscape, we developed the Shiny-based interactive web tool (https://bioinfo.cipf.es/cbl-atlas-ms/). This application offers a user-friendly interface for freely accessing and visualising the outcomes of this research.

Discussion

Despite the well-documented sex differences in MS pathophysiology, the specific molecular mechanisms underlying them remain underexplored, leaving a notable gap in the understanding of the disease. In this study, we novelly outline the intricate sex-differential landscape in MS subtypes at single-cell resolution. Thus, aiming to elucidate the molecular mechanisms that may be involved in MS progression between sexes within each cell type to potentiate biomarker discoveries and personalised therapeutic strategies.

Sex differences in MS have been widely characterised in the CNS. Males show greater brain atrophy driven by accelerated and more severe neurodegeneration than females, leading to acute demyelination and increased neuronal loss (PMID: 37122747). At epidemiology level it is reflected that males progress faster to SPMS and with more sequelaes (PMID: 26046348). To our knowledge, the molecular mechanisms driving sex-based differential neurodegeneration are poorly explored. In the present work, we propose potential molecular mechanisms that could explain these molecular differences in SPMS by characterising tetrapartite synapses involving astrocytes, neurons and microglia (Figure 3). Excitotoxicity, a phenomenon by which neurons are exposed to high levels of glutamate, is one of the major hallmarks of MS

neurodegeneration (PMID: 36675155). We found that female neurons presented increased genes that may act as compensatory mechanisms to cope with excitotoxicity, such as glutamate receptors (GRIA2, GRIA3, GRIA4, GRIN2A, GRIN2B, GRM5, GRM7, GRID1) and GABA postsynaptic receptors (GABRB1, GABBR2, GABRB2, GABRB3, GABRA4); cation transporters (SLC17A6, SLC1A2, SLC4A10, SLC17A5, SLC8A1, ASIC1, SCN2A, CACNA1C, CACNB4, CACNB2, KCNH1, KCNMA1, KCND2, KCND3, KCNJ3, HCN1, PTK2B, P2RX7, KCNH1) and neurotransmitter vesicle cycle mediators (CADPS, CADPS2, RIMS2, ERC2, BRAF, DNM2, DNM3, SYN2, SYN3, NLGN1, NLGN4X, NLGN4Y, NRXN1, SH3GL3, SYNJ2, SH3GL2, AMPH, ITSN1, ITSN2, FCHO2, DENND1A, TMEM108, ALS2, ANK2, ANK3, SPTBN1, PTK2, FYN, KALRN, MAGI2). Females also show an increased activation of GABA signalling effectors. In contrast, male neurons seem to be more susceptible to maintaining excitotoxicity by overexpressing ATPases subunits (ATP6V0E2, ATP1A3, ATP6V1F, ATP6V1G1, ATP6V1D, ATP6V1B2, ATP6V1E1 and ATP6V0D1) that provide energy for vesicle biogenesis, proteins involved in neurotransmitter secretion (DOC2A, SYT4, SYT5, SYT12, CPLX1, CPLX2, SYN1, SV2A, STX1A, SNAPIN, RAB3A, SYN1, NR4A1, PIP5K1C), and present a higher activation of glutamate signalling effectors. In addition, male neurons overexpress ribosomal proteins (RPS13, RPS14, RPS18, RPS19, RPS25, RPL8, RPL12, RPL38) compared to females; which could be a mechanism to counteract the increased axon loss enhancing synaptogenesis previously described in PMID: 31340150.

Glia-driven MS stress response is a complex set of processes to different stimuli that account for sex differences. However, studies currently differ in describing their severity between males and females (PMID: 37122747). Here, we highlight that male astrocytes have increased glutamate reuptake related genes, which could modulate the excitotoxicity in the synaptic cleft (PMID: 30791579). Conversely, female astrocytes dispose of potential key genes, such as HIF3A and EGLN3, to address the hypoxic stress characterised in MS demyelinating lesions (PMID: 33351069). By regulating the expression of multifaceted genes, astrocytes may modulate their profile towards neuroprotection or reactivity in MS. In this regard, females seem to balance their response via lncRNAs like NEAT1 - a negative regulator of neuronal excitability in neurodegenerative diseases (PMID: 30533572) and LINC-PINT - a potential regulator of oxidative stress in Parkinson’s disease substantia nigra which also it is accumulated in Alzheimer's and Huntington's diseases (PMID: 32080970). Meanwhile, we identified JUND and FOS increased in males compared to females. These genes, part of the activator protein 1 (AP-1) complex, promote transcription of several genes under stress conditions. Specifically, FOS activation has been described during experimental autoimmune encephalomyelitis (PMID: 30255127), and based on our knowledge, we propose JUND as a novel target to explore in this context. Another sex-based distinctive trademark uncovered in this work is that microglia transmembrane transporters of

inorganic and organic compounds are mostly increased in females (e. g. SLC24A2, SLC44A1, TMEM144, TMEM165 and CACNA2D3), which can be valuable to dechiper their role in the sexual dimorphism of microglia metabolism (https://doi.org/10.1038/s42003-021-02259-y).

It is highly relevant to unveil sex differences in MS myelin clearance and repair for the development of effective remyelination therapy strategies. Previous studies have demonstrated that remyelination is more effective in aged female rodents than males (PMID: 16797535) and sex-dependent regulation of myelin repair markers in EAE (PMID: 16797535). These differences in myelin regeneration are influenced by sexual hormones as estrogens and androgens (PMID: 36949062). Expanding on these findings, our research delves into sex disparities in human samples revealing that females have increased a greater number of neuronal-like genes than males, which would be relevant in establishing axon-myelin contacts (PMID: 31511515). Female OPCs also have increased a greater number of myelinating-related genes than males. Of note is QKI, a pivotal gene to promote oligodendrocyte differentiation (PMID: 17575274) and to maintain myelin sheaths by regulating myelin lipid homeostasis (PMID: 32202512). Among the signals received by OPCs to cope with myelin damage, notable pathways include PDGF and FGF (Figure 4). PDGF plays a crucial role by enhancing proliferation, differentiation, and survival of OPCs (PMID: 23383310). In MS males the interaction PDGFA (OPCs) - PDGFRA (OPCs) emerges compared to controls, which is crucial for recovery from chronic demyelination (PMID: 17984680). While this interaction is established in both control and MS females, the PDGFC (OPCs) - PDGFRA (OPCs) interaction intensifies, leading to a relative decrease in strength of neuron-OPCs interactions. Despite the known association of PDGFC with the reduction of neuroinflammation in MS (PMID: 33361796), we propose this interaction to further investigate its sex-dependent role in mediating communication between OPCs during the myelination process. Regarding the FGF signalling pathway, also involved in promoting myelination, females present a pronounced change in interactions than males. MS females increase the proportion of OPCs-neuron interactions that could support myelin sheath establishment. In this work, female microglia increased more genes related to phagocytosis and lipid metabolism, which could favour myelin clearance (PMID: 29301957). In this set CELF2 stands out for regulating the expression of TREM2, a key gene for eliminating myelin debris in MS (PMID: 33093587). Meanwhile, male microglia present increased SCARB1 which clears the deposition of amyloid-β in Alzheimer's disease (PMID: 22666621). Lastly, sex-differencial patterns have been also described in the lipid metabolism of astrocytes. Of note males presented increased APOE4, which potentially promotes lipid droplet accumulation, compromising their functionality and likely impeding their capacity to offer metabolic support to neurons (PMID: 33406436). Interestingly, APOE4 has also been associated with increased

tendency of progression (PMID: 16796581) and cognitive impairment (PMID: 17460153) in MS.

Previous efforts have been made to understand the role and status of the peripheral immune system given its relevance in the autoimmune response to CNS lesions. Recently, the German Competence Network Multiple Sclerosis underscored the importance of investigating the peripheral immune system in early MS by reporting three blood different endophenotypes linked to distinct disease outcomes (PMID: 38536936). However, the evidence related to sex differential disturbances so far is not conclusive, despite RRMS incidence is 2:3 times higher in females than males (PMID: 35467295). This work aims to expand knowledge in this area. We propose that females display a greater number of genes and functions increased in CD4+ T cells, CD8+ T cells, NK cells and monocytes (Figure 5), which could be involved in the exacerbated immune response. These features are involved in key processes such as inflammation, cell signalling and cell-cell adhesion that may comprise the higher female MS susceptibility compared to males. Of note, we identified a highly connected network of 21 genes in at least three of the four cell types evaluated (i.e. CD4+ T cells, CD8+ T cells, NK cells and monocytes) (Figure 5E). AP-1 transcription subunits FOS, JUN, JUNB and JUND conformed the major hub of the network, reflecting their importance as master regulators of a broad range of immune system processes such as proliferation, CNS infiltration and inflammation (PMID: 31340499). We also identified NFKBIA and NFKBIZ genes, inhibitors of NFkB - a pathway widely reported to be active in MS (PMID: 32265906); therefore, potentially counteracting the proinflammatory status. Interestingly, these findings are likely driven by sexual hormones, as oestrogen promotes both the activation of AP-1 (PMID: 10517669) and the mRNA and protein synthesis of NFKBIA (PMID: 22723832).

Our findings also suggest that RRMS males' adaptive immune system point to an enhanced activity of the mitochondrial electron transport chain (Figure 5C). Mitochondrial impairment corresponds to one hallmark of CNS active lesions (PMID: 19293237) that potentiates neuroinflammation (PMID: 38480879). mtDNA variants have been also proposed as susceptibility biomarkers of RRMS (PMID: 36534684). Regarding PBMCs, RRMS patients also exhibit mitochondrial dysfunction compared to controls, which is more pronounced in patients experiencing a more aggressive course of the disease (PMID: 36358334). These findings lead us to hypothesise that exploring the mitochondrial status of peripheral T lymphocytes that infiltrate the CNS may provide valuable insights into understanding the faster neurodegeneration observed in males (PMID: 37122747, PMID: 32859258).

Meanwhile, the role of the peripheral immune system in PPMS is less well documented, although this does not diminish its relevance or significance. Thus far, scientific community has observed that PPMS exhibits differences from other subtypes; treatments that are effective in RRMS show reduced efficacy in PPMS (PMID: 38554666), and the expression of the current blood neurofilament light marker is lower in PPMS compared to SPMS (PMID: 35379762). This does not imply that the immune system is not implicated in the disease's development, rather that a wider characterisation of PPMS immune profile is needed. In accordance with this goal, studies indicate meningeal inflammatory infiltration in progressive forms (PMID: 28620346). Additionally, Absinta M. *et al* (PMID: 34497421) revealed the significant diversity of glial and immune cells chronically active lesions. In contrast to the RRMS results where female incremental changes are predominant, we have found that in PPMS both sexes have a similar proportion of enriched functions involving all broad immunophenotypic features (Figure 6B). Interestingly, with a higher number of functional and gene expression sex differences in CD8+ T cells compared to the other cell types. They are of great relevance as intrathecal CD8+ T cells subtypes as key elements in the immunopathogenesis of PPMS, being associated with white matter injury and thalamic atrophy (PMID: 37369602).

The primary function of CD8+ T cells, also referred to as cytotoxic T cells, is to identify and eliminate infected, abnormal or cancerous cells. They are of particular interest in the context of PPMS, due to the reported greater decline in peripheral CD8+ T cells with age in patients compared to healthy controls. This scenario suggests an accelerated ageing effect in MS (PMID: 24842963). Additionally, the intracranial infiltration previously mentioned could drive autoreactive responses in the CNS. We reported sex-biassed CD8+ T cells responses (Figure 6D), with females showing tendencies to restore cellular homeostasis compared to males. Conversely, males exhibit a more activated state, characterised by heightened cytolytic responses. It may be a partly explanation to the more severe neurodegeneration in males triggering more rapid progression. Surprisingly, it is the opposite sex-differential response to cancer cells of CD8+ T cells (PMID: 36051310). It is therefore vital to characterise sex differences in each type of disease.

We have also identified a genetic signature composed of 67 genes that, based on their sex differential profile, classify the MS immune system by disease subtype (Figure 7). This genetic signature potentially enables the characterization of progression biomarkers and the development of tailored approaches to treatment by identifying the specific MS subtype of a patient. Specifically, 12 genes are increased in female RRMS and not significant in PPMS. They are mostly related to the process of modulating inflammation, a key differentiating factor between RRMS (characterised by high

inflammation) and PPMS (with lower inflammation levels). Of note is CD69, being significant on CD4+ T cells, CD8+ T cells and NK cells. CD69 protein has been detected in MS lesions in both brain and infiltrated immune cell types (PMID: 36753806, PMID: 29873694). Interestingly, it is also described as a master regulator of pro-inflammatory cytokines synthesis and autoimmune reactivity (PMID: 32291138). Therefore, its female increased expression holds promise for further investigation. Meanwhile, PPMS genes are mostly increased in males. Although many of them are also involved in inflammation, the glycolytic metabolism stand out, particularly with the genes TPI1 (that catalyses the isomerization of glyceraldehydes 3-phosphate and dihydroxy-acetone phosphate), and LDHA (that catalyses the reversible conversion of pyruvate to lactate). Peripheral cells of the immune system are known to exhibit metabolic dysfunction, partly due to their activation in the immune response and the mitochondrial damage they undergo (PMID: 38243312). This immune-metabolic phenomenon is a current field of interest to explore potentially as a therapeutic target. Given our findings, we emphasise the importance of elucidating the sex-specific metabolic profiles, which may offer valuable insights for targeted therapeutic interventions.

The strongest MS genetic associations are located in the HLA genes, responsible for encoding proteins crucial for the immune system's ability to distinguish between foreign organisms and our own cells (PMID: 21833088). This gene family has been implicated in several autoimmune conditions (https://doi.org/10.1016/j.yamp.2020.07.016). Furthermore, distinct sex-specific differences in HLA expression have been extensively documented across various contexts, including peripheral immune responses to lipopolysaccharide (PMID: 33441806) and generalised aggressive periodontitis (PMID: 12485327). MS is no exception, as we observe significant variations in both expression levels and interaction strengths within their antigen presentation roles (Figure 8). Particularly noteworthy is the pattern change in HLA-DRB5, which is increased in RRMS females and in PPMS males. HLA-DR increased expression is characterised from demyelinated lesions (PMID: 31882398). In detail, HLA-DRB5 presents an intricate role in MS as it is proposed to mitigate disease severity while also being effective at presenting antigens derived from myelin (PMID:34006391, PMID:18832704). Additionally, HLA-E protein interactions with CD8+ T cells became significant in RRMS females and PPMS males; which are proposed to present a pivotal role presenting antigens from Epstein Barr Virus — an inflammatory agent intricately linked to MS pathogenesis (PMID:38091993). Regarding HLA genes from class II, which also influence genetic susceptibility to MS (PMID:26343388), it is revealed that RRMS males exhibit heightened interaction strengths, aligning with levels observed in females. In contrast, the disappearance of ligand-receptor interactions among PPMS females compared to female controls is notable, while PPMS males maintain similar patterns to their respective controls. This loss of intensity in female PPMS connections could be an area

of further exploration to elucidate why female susceptibility is higher in RRMS, although not in PPMS.

We also acknowledge the strengths and limitations of our research. We performed an i*n silico* approach analysing publicly available data. Reanalyzing existing data to test new hypotheses facilitates scientific breakthroughs. This underscores the importance of promoting Open Science initiatives and adhering to FAIR (Findable, Accessible, Interoperable, Reusable) principles to encourage data reuse and enhancing the advance in research. Literature search yielded the final identification of three datasets -with a total of 385.853 cells comprising 19 cell types- to characterise the sex-differential transcriptomic profile of SPMS in the central nervous system, and RRMS and PPMS in PBMCs. Although this strategy captures the state of the art of MS scRNA-seq analysis in humans, we also had to deal with intrinsic technical and biological variability characteristic among the retrieved dataset. In depth, each study collects specific variables from each patient and we had lack of control over the execution of scRNA-seq technology. Moreover, further data will be needed to analyse novel combinations of tissue and MS subtypes, and to further enhance the robustness of our findings in the presented scenarios.

Previously we have characterised MS sex differences by analysing microarray and bulk RNA-seq transcriptomic data in brain and blood samples (PMID: 37023829). While in this research we identified sex-differential gene expression averaged across all cells within a sample, our current work of scRNA-seq/snRNA-seq data offers the ability to dissect cellular heterogeneity. Thus, shedding light on cell type-specific responses as evidenced in PPMS, where we unveiled that female CD8+ T cells tend to increase genes associated with cell maintenance. Contrarily, males exhibit increased expression of genes related to cytolysis. We have also identified common patterns among cell types that provide insights into the overall state of the immune system. A representative example is the increment of the AP-1 transcription factor in at least three cell types of the RRMS female immune system. A limitation to note is that we have focused on major cell types within each tissue. It would be of great interest to continue dissecting the differential expression profile in more populations, such as endothelial cells comprising the blood-brain barrier and the infiltration of the immune cells in the central nervous system. This, together with the characterisation of subpopulations of the current cell types, would further refine the understanding of the sex-differential molecular mechanisms.

This work contributes to generating an in-depth description of the transcriptional profile in different cell types, by not only performing differential gene expression analysis but also functional profiling, protein-protein interaction networks, signalling pathway

activation and cell-cell communication analyses. Although we have attempted to cover diverse areas of disease pathology in both the central nervous system and the immune system, it is unfeasible to address all encountered differences within a single manuscript. To provide broader and free access to our findings, we have developed a user-friendly website based on Shiny (https://bioinfo.cipf.es/cbl-atlas-ms/), enabling the interactive exploration of the complete results by any interested user.

Overall, we hope to contribute to the characterization of MS molecular pathology at single cell resolution, facilitating the search for potential biomarkers and the development of targeted therapies; consistently considering both the MS subtype and the sex of the individual. To our best knowledge, we have generated the first single-cell transcriptomic atlas of MS sex differences, deciphering results by tissue and disease subtype. Through the reanalysis of scRNA-seq and snRNA-seq transcriptomic data, we have delved into different bioinformatics approaches to cover as broadly as possible the understanding of sex differences in disease pathology. Additionally, our investigation has identified genes, functions and pathways worthy of consideration in the potential development of biomarkers and targeted therapies.

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

ACKNOWLEDGEMENTS

The authors thank the Principe Felipe Research Center (CIPF) for providing access to the cluster, co-funded by European Regional Development Funds (FEDER) to the Valencian Community 2014-2020.

This research was supported and partially funded by the Institute of Health Carlos III (project IMPaCT-Data, IMP/00019), co-funded by ERDF, "A way to make Europe", PID2021-124430OA-I00 funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe". Irene Soler-Sáez is supported by a predoctoral grant FPU20/03544 funded by the Spanish Ministry of Universities. Borja Gómez-Cabañes is supported by a PhD fellowship funded by the Spanish Association Against Cancer in Valencia (PRDVA234163GOME).

AUTHOR CONTRIBUTIONS STATEMENT (CRediT-compliant)

Conceptualization: ISS, MRH, FGG; Data Curation: ISS; Investigation: ISS, BGC, RGR, CGR, LBM, HC, MIV, SGP, VT, MRH, FGG; Bioinformatic Analysis: ISS, BGC, RGR; Supervision: MRH, FGG; Writing-Original Draft Preparation: ISS, BGC, RGR, CGR, LBM, HC, MIV, SGP, VT, MRH, FGG; all authors read and approved the final manuscript.