

## scDiffCom

Lagger, Cyril; Ursu, Eugen; Equey, Anaïs; Avelar, Roberto A.; Pisco, Angela Oliveira; Tacutu, Robi; de Magalhães, João Pedro

DOI:

[10.1038/s43587-023-00514-x](https://doi.org/10.1038/s43587-023-00514-x)

License:

Creative Commons: Attribution (CC BY)

### Document Version

Publisher's PDF, also known as Version of record

### Citation for published version (Harvard):

Lagger, C, Ursu, E, Equey, A, Avelar, RA, Pisco, AO, Tacutu, R & de Magalhães, JP 2023, 'scDiffCom: a tool for differential analysis of cell–cell interactions provides a mouse atlas of aging changes in intercellular communication', *Nature Aging*, vol. 3, no. 11, pp. 1446-1461. <https://doi.org/10.1038/s43587-023-00514-x>

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# scDiffCom: a tool for differential analysis of cell–cell interactions provides a mouse atlas of aging changes in intercellular communication

Received: 8 November 2021

Accepted: 27 September 2023

Published online: 2 November 2023

 Check for updates

Cyril Lager <sup>1,5,8</sup>, Eugen Ursu <sup>2,8</sup>, Anaïs Equey<sup>3</sup>, Roberto A. Avelar<sup>1</sup>, Angela Oliveira Pisco <sup>4,6</sup>, Robi Tacutu <sup>2</sup> & João Pedro de Magalhães <sup>1,7</sup> ✉

Dysregulation of intercellular communication is a hallmark of aging. To better quantify and explore changes in intercellular communication, we present scDiffCom and scAgeCom. scDiffCom is an R package, relying on approximately 5,000 curated ligand–receptor interactions, that performs differential intercellular communication analysis between two conditions from single-cell transcriptomics data. Built upon scDiffCom, scAgeCom is an atlas of age-related cell–cell communication changes covering 23 mouse tissues from 58 single-cell RNA sequencing datasets from Tabula Muris Senis and the Calico murine aging cell atlas. It offers a comprehensive resource of tissue-specific and sex-specific aging dysregulations and highlights age-related intercellular communication changes widespread across the whole body, such as the upregulation of immune system processes and inflammation, the downregulation of developmental processes, angiogenesis and extracellular matrix organization and the deregulation of lipid metabolism. Our analysis emphasizes the relevance of the specific ligands, receptors and cell types regulating these processes. The atlas is available online (<https://scagecom.org>).

Aging remains a poorly understood biological process despite affecting most organisms<sup>1</sup>. One of the difficult aspects to model is how the dynamics of molecular changes and tissue homeostasis influence each other throughout the lifespan. To gain further insights on how to bridge this gap, we focused our attention on intercellular communication (ICC). Dysregulation of ICC has been defined as a hallmark of aging<sup>2,3</sup> and has recently been proposed as one of the causes leading to the cell-to-cell stochasticity arising with age<sup>4</sup>. Well-known communication deregulations include inflammaging (a chronic low-grade

age-associated inflammation)<sup>5</sup>, impaired immune surveillance<sup>6</sup>, increase in senescence-associated secretory phenotype (SASP)<sup>7</sup>, altered communication between stem cells and their niche<sup>8,9</sup>, remodeling of the extracellular matrix<sup>10,11</sup> and changes in endocrine and neuronal communication<sup>12</sup>. Interestingly, interventions involving extracellular signals have been shown to partially reverse some of the aging phenotypes. This includes targeting endocrine mediators such as insulin-like peptides and growth hormones<sup>13,14</sup>, the use of anti-inflammatory compounds<sup>15–17</sup>, heterochronic tissue transplants and heterochronic parabiosis<sup>18–20</sup>.

<sup>1</sup>Integrative Genomics of Ageing Group, Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool, UK. <sup>2</sup>Systems Biology of Aging Group, Institute of Biochemistry of the Romanian Academy, Bucharest, Romania. <sup>3</sup>Department of Medicine, Karolinska Institutet, Stockholm, Sweden. <sup>4</sup>Chan Zuckerberg Biohub, San Francisco, CA, USA. <sup>5</sup>Present address: Altos Labs, San Diego, CA, USA. <sup>6</sup>Present address: Insitro, Inc., South San Francisco, USA. <sup>7</sup>Present address: Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK. <sup>8</sup>These authors contributed equally: Cyril Lager, Eugen Ursu. ✉e-mail: [jp@senescence.info](mailto:jp@senescence.info)

Direct measurement of ICC is complicated and usually depends on the type of mediators considered, such as surface receptors, soluble factors, extracellular vesicles<sup>21</sup> or even mitochondria<sup>22</sup>. However, recent studies have shown that specific aspects of ICC can be inferred from single-cell gene expression data<sup>23,24</sup>. After the pioneer study that drafted the first comprehensive database of ligand–receptor interactions (LRIs)<sup>25</sup>, and based on the development of statistical tools dedicated to building cell-type to cell-type communication networks<sup>26–34</sup>, it is now becoming standard to perform ICC analyses alongside the workflow of single-cell transcriptomic studies.

In the context of aging, single-cell omics is a recent but expanding field<sup>35,36</sup>, with already a few investigations reporting changes in ICC. This includes articles on the mouse brain<sup>37</sup>, on the mouse mammary gland<sup>29</sup>, on several rat tissues<sup>38</sup>, on the primate cardiopulmonary system<sup>39</sup> and on human skin fibroblasts<sup>40</sup>. However, the ICC analyses performed in some of those studies suffer from several limitations, as they rely on tools designed to detect interactions rather than to investigate how the interactions change between two biological conditions (for example, young/old and healthy/sick). Indeed, the main approach so far has been to detect interactions in young and old samples independently and then to focus on signals appearing or disappearing with age. As a result, this method does not account for interactions that are detected in both conditions but are, nevertheless, changing significantly; it also disregards the magnitude of the signal variation. More importantly, this approach lacks a statistical test to assess the significance of those changes and, thus, to evaluate if they occur due to noise or due to a true biological effect.

To alleviate such limitations, we built a statistical framework specifically designed to perform differential analysis in ICC. Our resulting R package, called *scDiffCom*, can be applied to any human or mouse single-cell RNA sequencing (scRNA-seq) dataset to analyze changes in ICC between two given conditions in a given tissue. *scDiffCom* includes a collection of approximately 5,000 curated LRIs that we retrieved from seven publicly available resources<sup>26–32</sup>. The typical output of the package is a table of detected cell-type to cell-type interactions indicating, in particular, their strength and how they are regulated between the two conditions of interest. To facilitate the interpretation of these results, we implemented an over-representation test to determine the dominant variations at the gene, cell-type or functional level. In addition, the package provides several visualization tools.

We used *scDiffCom* on several published scRNA-seq datasets from the Tabula Muris Senis (TMS) consortium<sup>41</sup> and the Calico murine aging cell atlas<sup>42</sup> to create *scAgeCom*, a large-scale atlas of age-related ICC changes across 23 mouse tissues. Samples obtained from male and female mice or from different experimental techniques were treated separately to avoid any confounding factors. By leveraging independent secretomics data from six cell lines, we confirmed that the interactions detected by *scDiffCom* using only gene expression are globally consistent with protein secretion profiles. The results are hosted and accessible via an online web application (<https://scagecom.org>) that contains both tissue-specific analyses and a global section summarizing changes shared across multiple tissues.

Our aging-related analysis supports previous knowledge regarding ICC, depicting a widespread upregulation of immune system processes and inflammation; a downregulation of extracellular matrix organization, growth, development and angiogenesis; and deregulation of lipid metabolism. We also report a generally complex sex-dependent regulation of ICC with age. Despite considerable differences across experimental techniques, we were able to predict some of the ligands, receptors and cell types that might play key roles in such dysregulation. Due to its generality and to the large number of tissues considered, we think that *scAgeCom* contains a large amount of valuable data waiting to be interpreted, which might provide the community with potential therapeutic targets and hypotheses regarding the relationship between aging and ICC.

## Results

### LRIs from existing databases

As with other methods analyzing ICC from scRNA-seq data, our approach first relied on the collection of LRIs. To maximize the variety of interaction types, we retrieved LRIs from seven publicly available resources: CellChat<sup>26</sup>, CellPhoneDB<sup>27</sup>, CellTalkDB<sup>28</sup>, NATMI/connectomeDB2020 (ref. 29), ICELLNET<sup>30</sup>, NicheNet<sup>31</sup> and SingleCellSignalR<sup>32</sup>. From them, we built both human and mouse databases, keeping only curated interactions, converting human genes to mouse orthologs for human-only resources and including both simple and complex LRIs. Simple LRIs are interactions involving a single-gene ligand with a single-gene receptor—for example, *ApoE:Ldlr*. On the other hand, a complex interaction involves heteromeric ligands or receptors—for example, *Col3a1:Itga1-Itgb1*.

Our approach resulted in 4,582 mouse LRIs (simple: 3,479, complex: 1,103) and 4,785 human LRIs (simple: 3,648, complex: 1,137) directly accessible from our R package *scDiffCom* (Fig. 1a and Supplementary Tables 1 and 2). Whenever possible, we also included the sources used by each of the seven databases to curate their interactions, including references to PubMed identifiers, FANTOM5 (ref. 25), HPMR<sup>43</sup>, HPRD<sup>44</sup>, IUPHAR<sup>45</sup>, Reactome<sup>46</sup> or KEGG<sup>47</sup>.

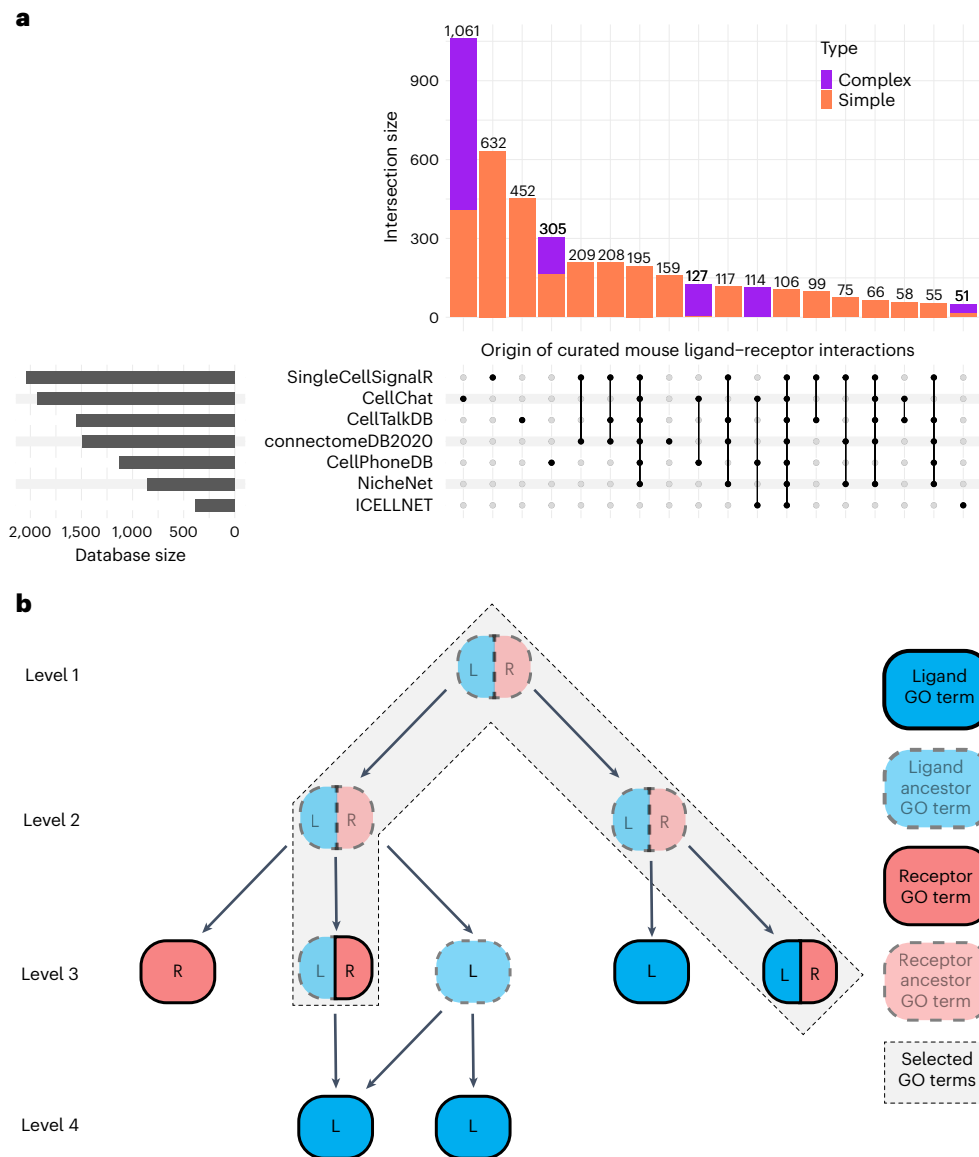
### Functional annotation of LRIs

We annotated all LRIs using a standardized and consistent framework. We first associated Gene Ontology (GO) terms<sup>48</sup> to each interaction in a way that conveys biological meaning related as much as possible to the interaction itself rather than to each gene independently. Simply taking the intersection between the ligand GO terms and receptor GO terms would have resulted in a considerable number of empty intersections (as most genes are annotated with specific GO terms and not all parent terms). Instead, we associated to each LRI all GO terms formed by the intersection of two sets of nodes: (1) the nodes of the GO graph made of the ligand GO terms with their ancestors and (2) the nodes of the corresponding receptor GO graph (Fig. 1b). Because this method is prone to attaching lowly informative terms (namely, those near the root of the GO graph), we also computed and indicated the level of each GO term (namely, its depth in the GO graph) to facilitate downstream analysis. Then, we added KEGG pathways<sup>47</sup> to each LRI if all genes present in the interaction were part of the pathway. All these annotations are directly accessible from *scDiffCom* (Supplementary Tables 3–6).

Anticipating our aging analysis, we also annotated the genes from the mouse LRI database with the number of PubMed articles associating each of these genes (or its human homolog) with aging or age-related diseases (excluding cancers). We also indicated if each gene (or its human homolog) is referenced in the GenAge<sup>49</sup>, CellAge<sup>50,51</sup>, LongevityMap<sup>52</sup> or Gene Expression<sup>53</sup> database of the Human Ageing Genomic Resources (HAGR)<sup>49</sup>. According to our search, the top 10 ligands/receptors associated the most with the aging literature are *ApoE\**, *App\**, *Snca*, *Psen1\**, *Ldlr*, *Tnf\**, *Il6\**, *Crp*, *Tlr4\** and *Il1b*, where a star indicates genes that are also referenced in the HAGR (Supplementary Table 7).

### Differential cell–cell communication analysis with *scDiffCom*

We designed a bioinformatics method, available within the R package *scDiffCom* (<https://github.com/CyrilLagger/scDiffCom>), to detect cell-type to cell-type communication patterns that significantly change between two conditions in a given scRNA-seq dataset (Fig. 2). The dataset must be formatted as an R Seurat object<sup>54–56</sup> and contain cells annotated by cell types and by the two conditions on which the differential analysis will be performed. The package then considers all possible interactions between each cell-type pair, based on the LRIs described in the previous subsection. We call each of those potential signals a cell–cell interaction (CCI). Each CCI is then assigned a score (independently in each condition) based on the average expression of the genes across the respective cells. As such, the CCI score should be



**Fig. 1 | Origin and annotation of LRIs used by scDiffCom. a**, Distribution of the 4,582 curated mouse LRIs classified according to the seven databases that they originated from. Each column of the UpSet plot corresponds to one intersection in an equivalent Venn diagram. Simple LRIs are composed of a single ligand gene and a single receptor gene, whereas complex LRIs contain a heteromeric ligand

and/or a heteromeric receptor. Data are available in Supplementary Table 2.

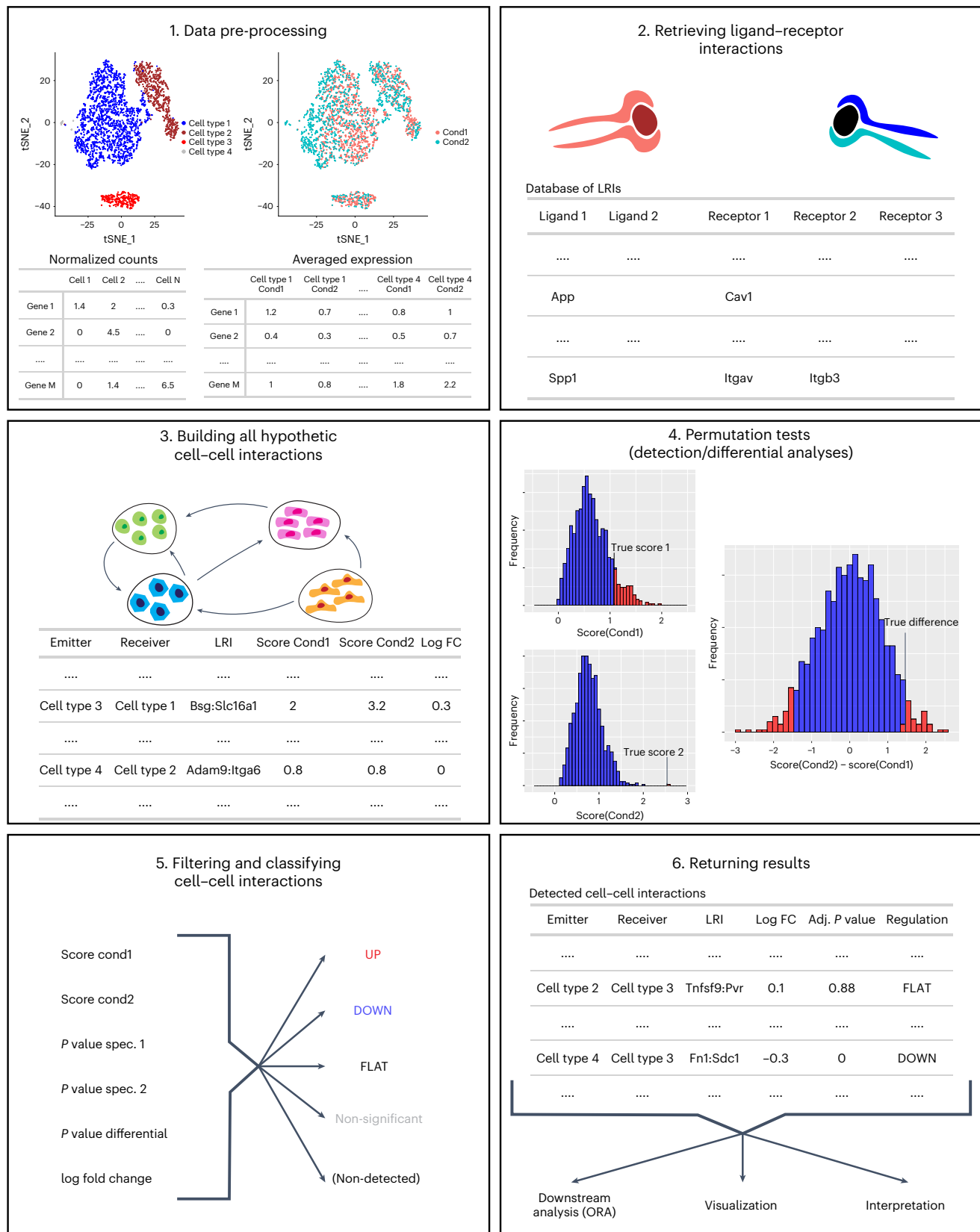
**b**, Method used to assign GO terms to a given LRI. Selected terms correspond to the intersection of the nodes of the two GO subgraphs made from the terms and ancestor terms of the ligand (blue) and receptor (red).

interpreted as the typical interaction that would occur between two cells randomly picked from the pool of emitter, respectively receiver, cells. This score has the advantage of not depending on the number of cells captured, a quantity that is usually biased by current experimental protocols<sup>57</sup>. However, a limitation of this score is its inability to account for absolute communication across all cells and, consequently, to account for potential changes in cell composition.

Based on the CCI score, we first assess the biological relevance of each CCI (in each condition independently) by combining previously proposed approaches. To be considered detected, a CCI has to (1) not be lowly expressed; (2) be ‘specific’, as originally defined by the authors of CellPhoneDB<sup>27,58</sup>, and (3) have a large enough score (compared to the other remaining CCIs). The permutation test shuffles the cell-type annotation attached to each cell to estimate how specific to a given emitter–receiver cell-type pair a particular LRI is. This allows the removal of non-specific CCIs that are likely not biologically relevant.

Alongside the detection analysis, we implemented a second permutation test to assess if the score of each CCI significantly changes between the two conditions of interest. This consists of randomly exchanging the condition label of each cell to see whether the test statistic, namely the difference between the CCI scores of each condition, is different from zero. Choosing permutation tests was motivated by the fact that they are non-parametric—that is, they make no assumptions regarding the distributions of the underlying variables<sup>59</sup>. This was particularly useful in this context, as there is no obvious way to model the CCI score distribution. In addition, permutation tests can be applied to unbalanced and low sample size scenarios, which frequently appear in single-cell studies.

Finally, based on both the detection and differential analysis, each CCI is classified into one of four possible categories, further referred to as ‘regulation’. It can be upregulated (UP), downregulated (DOWN), stable (FLAT) or correspond to a non-significant change (NSC). As such,



**Fig. 2 | Workflow summary of scDiffCom.** Read counts/UMIs from the single-cell dataset are aggregated by cell types and conditions (1). Genes are then joined with our database of LRIs (2) to build all the potential CCI that can occur between cell types (3). Statistical permutation tests are then performed to evaluate the biological significance of each CCI and its differential expression (4). They are

then classified based on several computed variables, such as their scores, *P* values and log fold change (5). Results are returned in a convenient format for downstream analyses and interpretation (6). FC, fold change; tSNE, t-distributed stochastic neighbor embedding.



the main output of scDiffCom is a table that contains all detected CCIs with relevant information, including their regulation, log fold changes, adjusted *P* values and scores.

### Over-representation analysis to extract dominant changes

As it is typical to detect several thousands of CCIs in a single tissue, scDiffCom also performs an over-representation analysis (ORA) to extract the dominant differential patterns. Our method differs from usual enrichment analyses that typically infer GO terms associated with a list of genes of interest and, therefore, requires defining a background set of genes<sup>60</sup>. Here, the detected CCIs (as opposed to the genes) are classified into groups of interest (namely, UP, DOWN or FLAT) and a background (all other CCIs). Then, ORA measures the statistical association between the CCIs of interest and any feature that can be associated with them, such as GO terms and KEGG pathways, but also genes and cell types. As will be confirmed below, this approach also has the advantage of not being biased toward generic and lowly informative GO terms, despite the way that we annotated LRIs in the first place. Indeed, as being generic, these terms tend to be attached to most CCIs and, therefore, to have similar odds in both the group of interest and the background, making them rarely over-represented.

In practice, scDiffCom performs ORA on the following features by default: LRIs, individual ligands, individual receptors, emitter–receiver cell-type pairs, emitter cell types, receiver cell types, GO terms and KEGG pathways. Analyzing individual ligands is useful to detect those that may take part in various regulated CCIs involving different receptors (or vice versa). The same logic is valid for considering the cell types individually. Moreover, scDiffCom offers the opportunity to perform ORA on additional user-defined attributes. To facilitate the interpretation of these results, over-represented features can be sorted according to a score that combines the usual odds ratio (OR) and adjusted *P* value:  $ORA\ score = -\log_2(OR) \cdot \log_{10}(adj.pval)$ . Finally, scDiffCom includes several built-in visualization functions that allow the user to easily take advantage of these statistical scores.

### scAgeCom: a mouse aging atlas of ICC

We built a mouse atlas of age-related changes in ICC by applying scDiffCom (using 10,000 permutations and default parameters) to 58 scRNA-seq datasets from the TMS consortium<sup>41</sup> and the Calico murine aging cell atlas<sup>42</sup>. The datasets cover 23 organs distributed across five categories (TMS FACS (male), TMS FACS (female), TMS Droplet (male), TMS Droplet (female) and Calico Droplet (male)) and contain both young and old samples (Fig. 3). When necessary, cell types were renamed or regrouped to facilitate the comparison across datasets. We also noticed that some important cell types were occasionally not captured in specific tissues (for example, adipocytes in adipose tissues) due to technical limitations. Notably, male and female samples were treated independently to avoid confounding effects that would otherwise be likely to arise, as we observed that cells were not uniformly distributed across sex, cell types and tissues (Supplementary Table 8). Keeping male and female samples separate also allowed us to perform a sex-differential analysis (Supplementary Text 1 and Supplementary Fig. 1) with the potential to reveal the effects of sex dimorphism on ICC aging.

Our aging atlas can be accessed from an online application (<https://scagecom.org/>) that provides either results for each dataset independently or results summarizing the dominant age-related signals shared across tissues (Fig. 4). Over the 58 datasets, scDiffCom detected 393,035 CCIs regulated as follows with age: 5% UP, 13% DOWN, 56% FLAT and 26% NSC. The number of returned CCIs corresponds to an average of 100 detected LRIs (s.d. = 81) between any cell-type pair. Interestingly, 1,135 LRIs (out of 4,555) were never observed as part of any detected CCI (Supplementary Table 9), although their genes were individually detected in the scRNA-seq datasets. This illustrates the importance of the detection procedure to filter out interactions and minimize false discoveries.

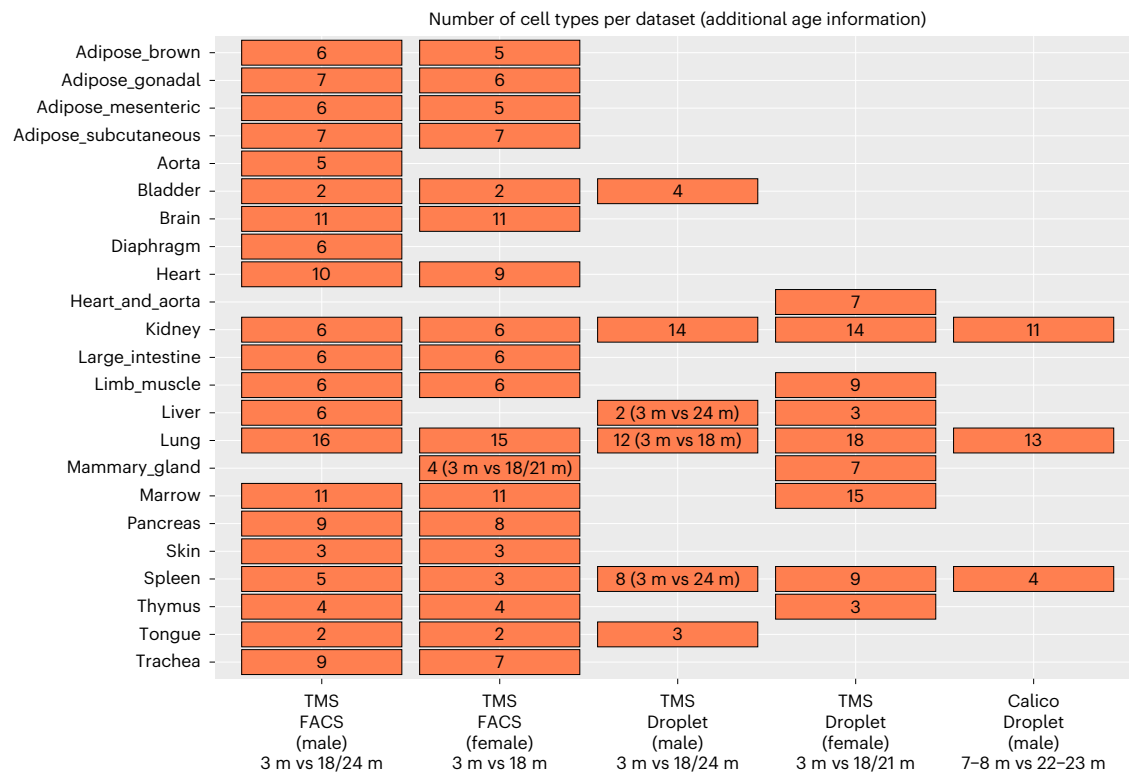
As scDiffCom is based on mRNA expression rather than protein expression, we employed multiple secretomics datasets to indirectly validate the biological significance and cell-type specificity of the CCIs identified in scAgeCom. We analyzed six proteomics datasets that characterize the secretion profiles of the following cell types cultured in vitro: mouse bone-marrow-derived macrophages (mBMMs)<sup>61</sup>, mouse neurons (mNeurons)<sup>62</sup>, mouse mesenchymal stem cells of adipose tissues (mMSC-ATs)<sup>63</sup>, rat cardiomyocytes (rCMs)<sup>64</sup>, human umbilical vein endothelial cells (hUVECs)<sup>65</sup> and human pancreatic ductal epithelial (hPDE) cells<sup>66</sup>. We assessed the validity of scDiffCom by computing the association between these cell lines and the emitter cells in scAgeCom, based on if the ligands of the identified CCIs were secreted by the given cell line. Despite the inherent limitations of comparing proteomics to transcriptomics data, comparing data across different species and comparing cells from in vivo samples to cells cultured in vitro, we observed a significant level of consistency. For example, when considering CCIs whose ligand has been detected in the hUVEC secretome, the odds of their emitter cells being endothelial cells are 1.68 times greater (adjusted  $P < 1 \times 10^{-15}$ , Fisher's exact test), in contrast to other cell types. Conversely, the odds of their emitter cells being lymphocytes are 1.98 times lower (adjusted  $P < 1 \times 10^{-15}$ , Fisher's exact test) (Fig. 5a). A similar consistent association is confirmed for the other secretomics datasets and scAgeCom emitter cell-type families (Fig. 5b–f). This pattern can also generally be observed at the tissue-specific level when considering cell types rather than families (Extended Data Fig. 1). Overall, this gives further support that CCIs predicted by scDiffCom using only scRNA-seq likely correspond to actual protein-mediated extracellular signaling, even though transcriptome–proteome correlations might be modest.

In addition to its detection capabilities, we also assessed several other aspects of scDiffCom using scAgeCom results as a benchmark. First, we confirmed the relevance of performing the differential analysis test on the CCI score that combines the expression of the ligand and receptor rather than performing a standard differential test on each gene separately. By comparing these two approaches on all scAgeCom CCIs, we observed that the regulation of a considerable fraction of them was determined without ambiguity by scDiffCom but not by standard differential analysis (Extended Data Fig. 2). Second, we confirmed that scDiffCom does not introduce a bias toward generic GO terms when performing ORA (Supplementary Text 2 and Supplementary Figs. 2 and 3). Finally, we confirmed the importance of the prior knowledge contained in LRI databases to characterize CCIs from scRNA-seq data (Supplementary Text 3 and Supplementary Fig. 4).

### Aging dysregulates several aspects of cell–cell communication

The regulation of the CCIs with age strongly varies across datasets, sex and experimental techniques (Fig. 6). The tissues from TMS FACS (male) clearly show a larger fraction of downregulated CCIs compared to all other datasets, including those from TMS FACS (female). We also observed more variability and noise (namely, more NSC CCIs) in FACS compared to Droplet datasets. These general observations are partially explained in the discussion below. Then, as it was not possible to present all results contained in scAgeCom, we focused our attention on age-related changes that we considered of primary interest because they were shared across several tissues, involving genes not previously associated with aging, involving genes detected in secretomics datasets or showing interesting sex-dependent patterns.

Consistent with the literature, we first observed a major upregulation of inflammatory, immune system and viral processes, as revealed by the numerous related GO terms, KEGG pathways and immune cell types over-represented across several tissues (Extended Data Fig. 3a). From the gene perspective, the interactions that are over-represented in the highest number of tissues include *B2m:Cd3g*, *B2m:Cd3d*, *Tnfsf12:Tnfrsf12a*, *H2-D1/K1/Q6:Cd8b1*, *Mif:Cd74*, *Hmgbl:Thbd*, *Slpi:Plscr1* and *Ccl5* interacting with different chemokine



**Fig. 3 | The 58 scRNA-seq aging datasets used to build scAgeCom.** scAgeCom was built by applying scDiffCom to 58 public scRNA-seq aging datasets encompassing 23 tissues (y axis) from both male and female mice and different experimental techniques (x axis). The number of cell types detected per dataset

is indicated in each rectangle. Information regarding the ages of the mice is indicated along the x axis and applies to all datasets of a given column unless specified otherwise in specific rectangles. Empty areas indicate unavailable data. Data are available in Supplementary Table 8. m, months.

receptors, such as *Ccr2*, *Ccr1* and *Ccr5* (Extended Data Fig. 3b). As a protein, B2M was observed in the secretome of five of the six cell lines that we analyzed, further supporting a tissue and cell-type-independent role. Leveraging our annotation of LRIs with PubMed articles on aging, we also noticed that *Sipi:Plscr1*, although being over-represented as upregulated in eight tissues, is an interaction whose genes have almost not been studied in the context of aging. We, therefore, provide further interpretation regarding its potential role in aging in our discussion below.

Lipid metabolism is another process dysregulated in many tissues with age. Several related GO terms and KEGG pathways are globally over-represented as upregulated (Extended Data Fig. 4a). However, looking at over-represented LRIs revealed more complex patterns, such as the upregulation of CCIs involving *ApoE* in male tissues and *App* in female tissues but the downregulation of CCIs involving *ApoE* in female tissues and *App* in male tissues (Extended Data Fig. 4b). Understanding such patterns in detail requires going beyond the over-representation results and exploring the specific CCIs of each tissue. For example, focusing on male brain samples, we observed that changes in CCIs involving *ApoE* are cell-type dependent (Extended Data Fig. 4c), typically upregulated when emitted from microglial cells but downregulated when emitted from ependymal cells. Both *App* and *ApoE* are associated with a large number of PubMed articles on aging, essentially for their relevance in Alzheimer's disease. However, our results suggest that these two proteins might play important roles in aging and ICC outside of the central nervous system as well, as they appear in dysregulated CCIs across many tissues and as they are detected in the secretome of most of the cell lines that we analyzed (Extended Data Fig. 4b).

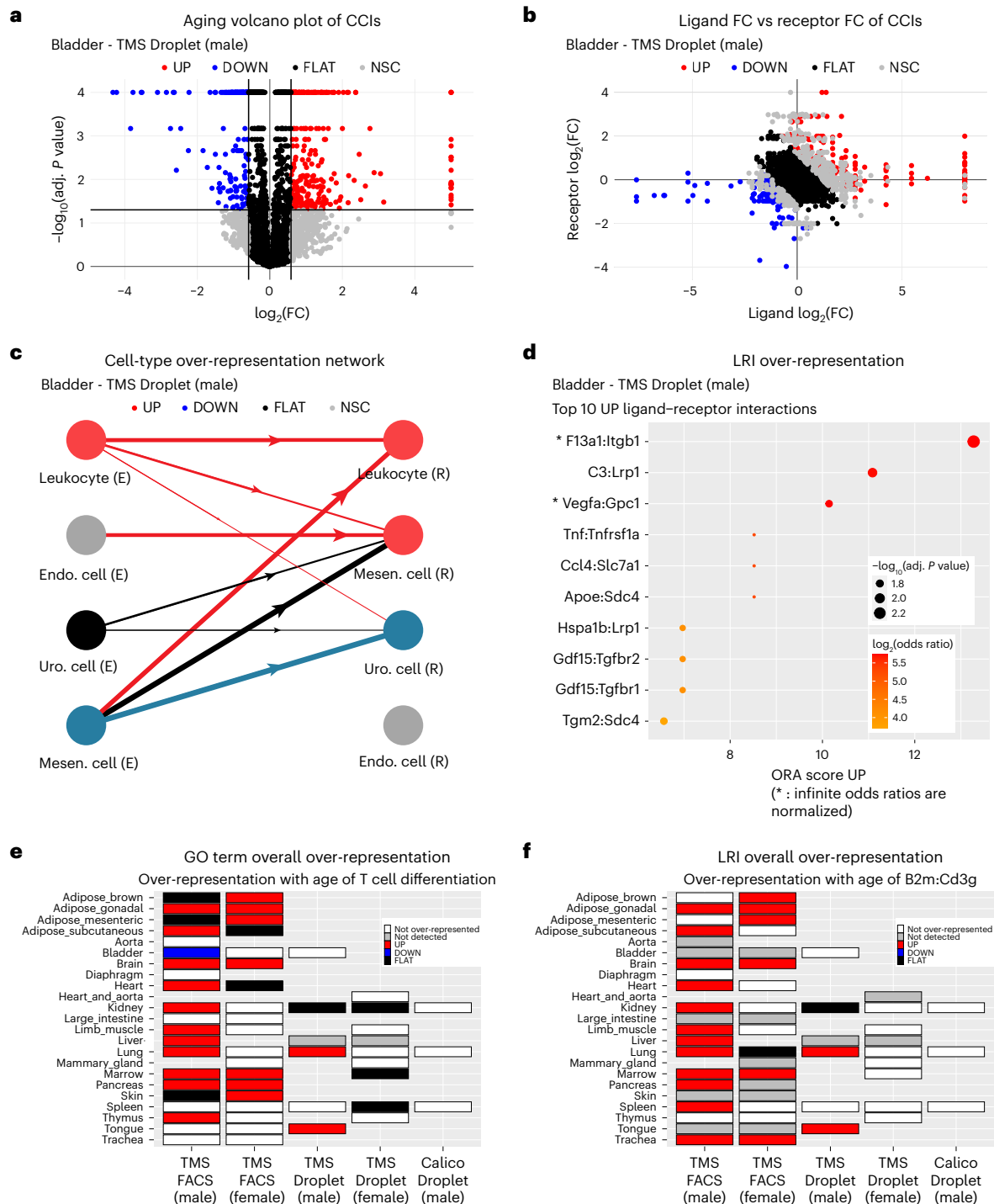
The most over-represented processes among signals down-regulated with age are related to the extracellular matrix and adhesion (Extended Data Fig. 5a). The main corresponding LRIs involve

collagens, cadherins and metalloproteinases interacting mainly with integrins (Extended Data Fig. 5b), mostly among connective tissue cells, epithelial cells and endothelial cells (Extended Data Fig. 5c). In addition to the changes undergone by the extracellular matrix, interactions related to growth, development, survival, differentiation and angiogenesis are also significantly downregulated with age (Extended Data Fig. 6). Cell-type pair over-representation also indicates the down-regulation of communication among stem cells and from stem cells toward endothelial cells (Extended Data Fig. 5c).

In addition to these widespread age-related changes, scAgeCom contains much more specific information that cannot be detailed here but is available to the community from our atlas. Several large-scale downstream analyses are further possible to perform, such as combining scAgeCom with the results from the sex-differential analysis (Supplementary Text 1) to reveal patterns of sex dimorphism in ICC aging. As an illustration, we observed from the TMS FACS Lung data that 13% of the differentially expressed CCIs had a complex pattern, namely being more expressed in young males than in young females, decreasing with age in males and increasing with age in females. Involved in these CCIs are bronchial smooth muscle cells (64%), *App* (21%), *Pecam1* (4%), *Tgm2* (4%), *Itgb1* (22%), *Lrp10* (4%), *Mcam* (4%) and *Nrp1* (4%). Sex differences in airway remodeling and inflammation have been recently reviewed, highlighting sex-related and age-related lung modifications, such as surfactant secretion<sup>67</sup>. We expect that other similar complex patterns could be mined from scAgeCom in the future.

## Discussion

We provide a package to perform differential ICC analysis as well as a comprehensive database of age-related mouse CCIs. From a technical point of view, our results illustrate the importance of performing a proper statistical analysis when comparing intercellular signals



**Fig. 4 | Typical figures available in scAgeCom.** Various types of figures are available as part of our atlas scAgeCom and accessible online at <https://scagecom.org/>. A non-exhaustive list is provided here as an illustration.

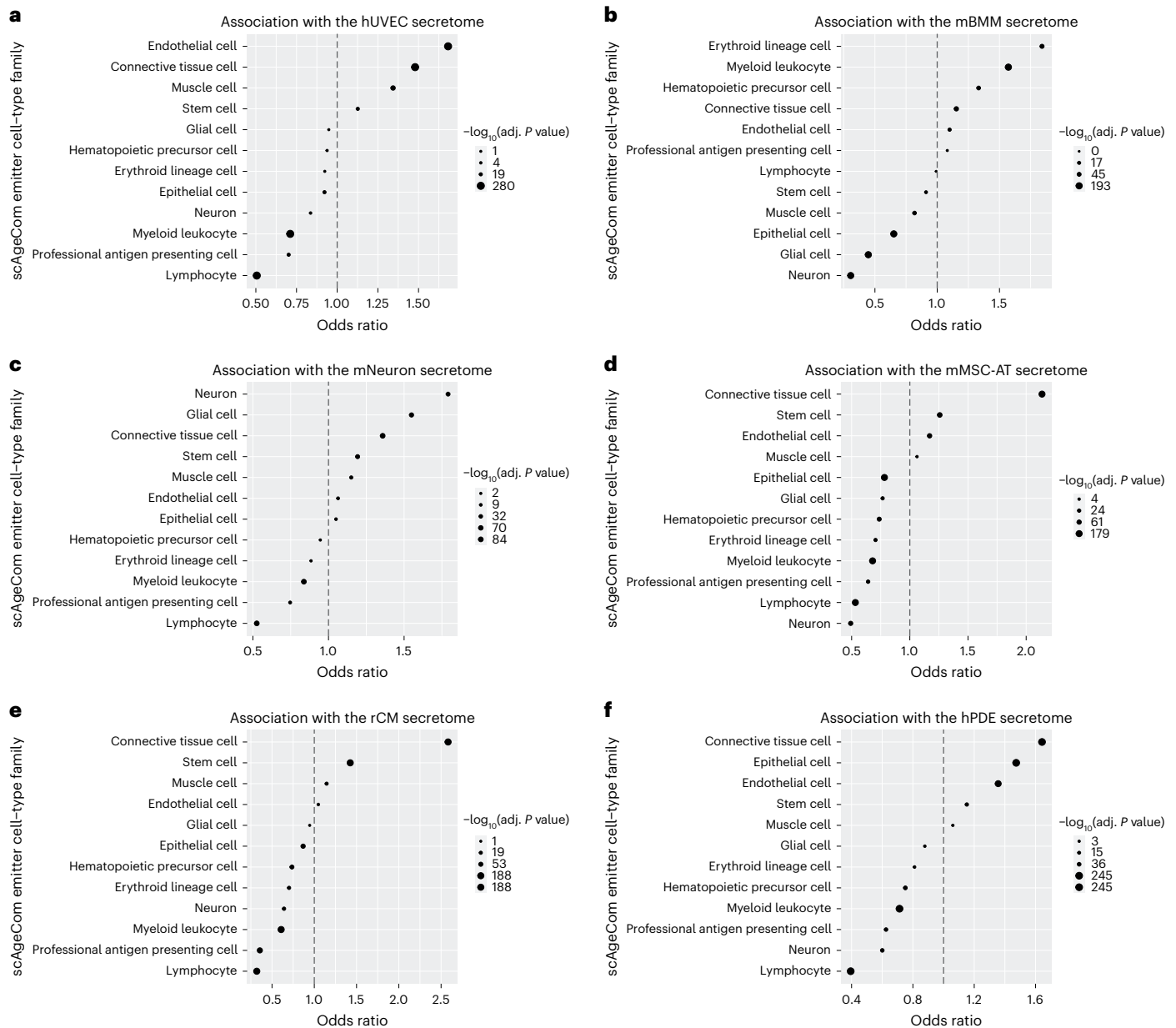
**a**, Volcano plots show how CCIs are distributed in terms of their aging log fold change, adjusted  $P$  value and regulation class (UP, DOWN, FLAT and NCS) for each specific dataset. Adjusted  $P$  values are computed by the scDiffCom two-sided permutation-based differential expression test. **b**, Ligand FC versus receptor FC plots indicate if the change with age of a given CCI score is driven by a change in the expression of the ligand, the receptor or both. **c**, Cell-type networks show which emitter cell types (E), receiver cell types (R) and cell-type pairs are

over-represented as upregulated, downregulated, stable or both upregulated and downregulated with age. The latter scenario might occur when both upregulated and downregulated LRIs are significantly present between two cell types. **d**, ORA plots display the top over-represented terms of a given category (here, LRIs upregulated with age in the bladder) in terms of their OR, adjusted  $P$  value and ORA score computed from the two-sided Fisher's exact test implemented in scDiffCom. **e, f**, Overall ORA plots summarize over-representation results of a given term from a category (here, **e**, the GO term 'T cell differentiation', and **f**, the LRI B2m: Cd3g) across all datasets. FC, fold change.

extracted from scRNA-seq data. Using scAgeCom as a benchmark, variability and noise were indeed responsible for the classification of 26% of the detected CCIs as NSC interactions, namely those with

a fold change larger than 1.5 but a non-significant adjusted  $P$  value. Had we not used a statistical test and based our analysis solely on the appearance and disappearance of CCIs between the two conditions,





**Fig. 5 | scDiffCom predicts CCIs globally consistent with secretomics data.**

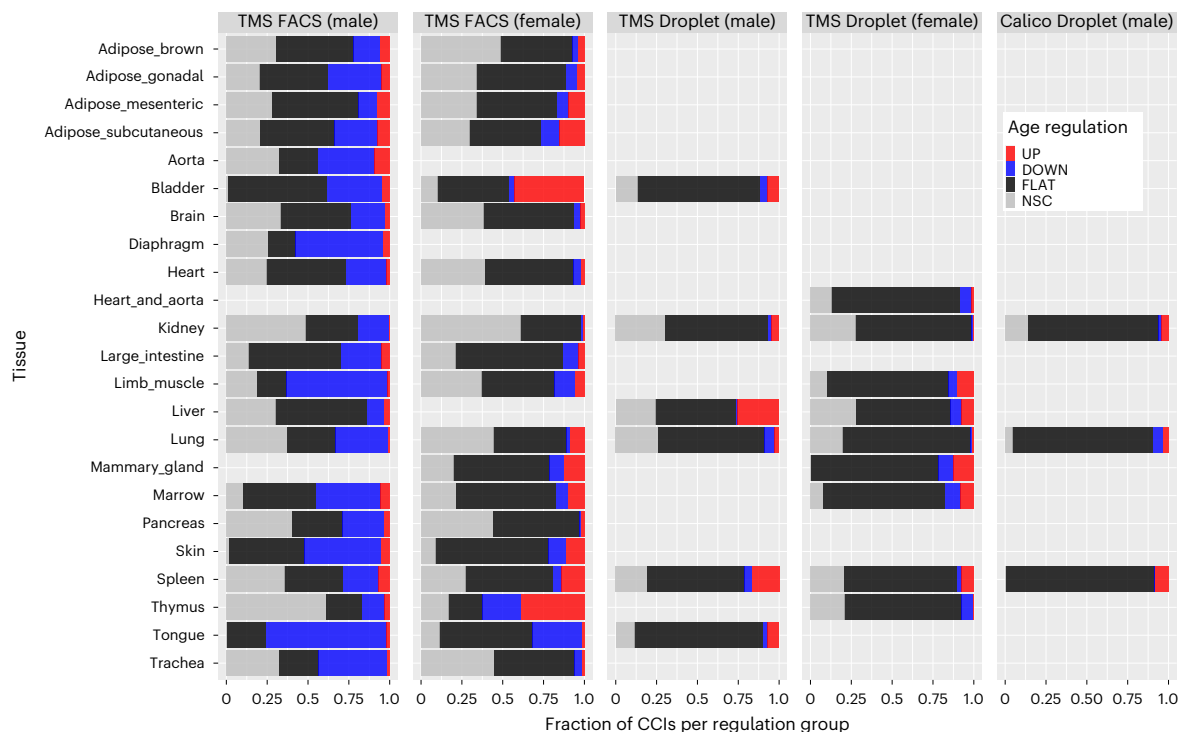
CCIs detected by scDiffCom from transcriptomics data are globally consistent with the proteomics secretion profile of six specific cell lines. Each cell line has a strong and positive association with related cell-type families from scAgeCom as well as a strong and negative association with unrelated families. ORs and  $P$  values were computed from a two-sided Fisher's exact test, and  $P$  values were further adjusted for multiple comparisons (Benjamini–Hochberg procedure). **a**, CCIs with a ligand that is part of the secretome from hUVECs are more likely

emitted from endothelial or connective tissue cells, as opposed to immune cells. **b**, The secretome from mBMMs is positively associated with CCIs emitted from hematopoietic lineage. **c**, The secretome of mNeurons is mostly associated with CCIs emitted from brain cells. **d**, The secretome from mMSC-ATs is associated with CCIs emitted from connective tissue cells and stem cells. **e**, The secretome from rCMs is mostly associated with CCIs emitted from connective tissue cells and stem cells. **f**, The secretome from hPDE cells is mainly associated with CCIs emitted from connective tissue cells, epithelial cells and endothelial cells.

as in some previous studies, 82.4% of those NSC signals (that is, 84,998 CCIs) would have been falsely considered to be age regulated. Moreover, such an approach would have missed all CCIs detected in both conditions but showing a significant change with age–18,146 interactions, in our case.

We observed considerable disparities in the results depending on experimental techniques and sex (Fig. 6). Previous comparisons of single-cell sequencing techniques<sup>68</sup> have claimed that droplet-based methods using unique molecular identifiers (UMIs) are less subject to amplification noise than Smart-seq2, potentially explaining why we observed fewer NSC CCIs in TMS Droplet and Calico Droplet than in TMS

FACS. The other differences can be explained by several factors. First, different datasets sometimes compare different age groups (Fig. 3). Second, due to experimental limitations, captured cell types are rarely the same between two datasets of the same tissue. For example, for the Lung, there are 16 cell types in TMS FACS (male) against 12 cell types in TMS Droplet (male), resulting in more detected CCIs in the former than in the latter dataset (23,200 versus 11,711) and with a different distribution of the percentage of UP/DOWN/FLAT/NCS CCIs (Fig. 6). We also emphasize that the pronounced downregulation observed in TMS FACS (male) datasets (Fig. 6) was reported in a previous study performed on TMS datasets<sup>69</sup>.



**Fig. 6 | Distribution of the scAgeCom CClis in terms of age regulation per dataset.** The percentage of CClis being upregulated (UP), downregulated (DOWN), stable (FLAT) or changing non-significantly (NSC) with age varies across datasets in functions of the tissue, sex and experimental techniques. We observed that TMS FACS datasets are generally more noisy (that is, have a higher

percentage of NSC CClis) than Droplet datasets. TMS FACS (male) datasets also admit a higher percentage of downregulated CClis with age compared to all other conditions, in agreement with the general decrease in gene expression previously reported in these samples<sup>41</sup>.

We now provide more specific interpretations of the main biological results reported above. Regarding the increase of immune response processes with age, we observed a widespread upregulation of CClis involving the ligand *B2m* ( $\beta$ 2-microglobulin). *B2m* has already been recognized as a gene consistently overexpressed with age<sup>53</sup> and as a pro-aging circulating factor whose elevated level negatively affects cognitive functions and neurogenesis in the mouse hippocampus<sup>70</sup>. Our results indicate that the increased secretion of the ligand *B2m* with age is systemic and appears to target T cells via their receptors *Cd3g*, *Cd3d* and, to a lesser extent, *Cd247* (explaining the occurrence of the GO term ‘T cell differentiation’). This could be a sign of increased antigen presentation in major histocompatibility complex (MHC) class I. Such a widespread pattern might also indicate detrimental effects of *B2m* not limited to the brain and reinforce the idea that this protein might be a potential therapeutic target, as previously suggested<sup>71</sup>.

Regarding the increase in inflammation with age, our results show a global over-representation of the terms ‘cytokine-mediated pathway’ and ‘chemokine-mediated pathway’ and of the ligands *Ccl5*, *Mif* and *Hmgb1*. Those ligands are known SASP factors, but we cannot say from our analysis whether they originate in senescent cells. The original analysis performed on the Calico data found very few cells expressing markers of senescence and no significant changes in aged tissues<sup>42</sup>. The original TMS study reported a higher fraction of cells expressing senescence markers in some (but not all) older age groups<sup>41</sup>, although the number of such cells was also very small (unpublished data). Follow-up studies, such as a cross-analysis between scAgeCom and the SASP atlas<sup>72</sup>, might allow us to explore such senescence-related hypotheses more precisely. Here, we only mention that the two most over-represented upregulated LRIs that we reported earlier, *Mif:Cd74* and *Hmgb1:Thbd*, actually seem to play compensatory roles in the context of senescence and inflammation. Indeed, it has been reported that macrophage migration inhibitory factor (Mif), a pleiotropic cytokine, can

prevent cellular senescence and rejuvenate mesenchymal stem cells from age-induced senescence via CD74/AMPK/FOXO3a and autophagy in both rats and humans<sup>73,74</sup>. Along the same line, the high-mobility group box chromosomal protein 1 (Hmgb1) has pro-inflammatory effects when binding to RAGE/Ager<sup>75</sup>, but its sequestration by thrombomodulin acts as an anti-inflammatory mechanism<sup>76</sup>. According to our results, *Hmgb1:Thbd* is over-represented as upregulated in nine tissues and *Hmgb1:Ager* in three tissues (Extended Data Fig. 3b), pointing toward a global over-emission of *Hmgb1* with age that tissues might try to compensate by over-expressing *Thbd*.

Our analysis also revealed widespread upregulation of the LRI *Slpi:Plscr1* whose genes have received little attention in the context of aging. Secretory leukocyte peptidase inhibitor (SLPI) has anti-inflammatory, anti-bacterial, anti-fungal and anti-viral activities<sup>77</sup>, and its interaction with phospholipid scramblase 1 (PLSCR1) was suggested to inhibit HIV-1 infection of T cells<sup>78</sup>. In scAgeCom, *Slpi:Plscr1* is notably upregulated with age in all three bone marrow datasets (Extended Data Fig. 3c). Therefore, the increase with age in such communication patterns might represent a response to the latent and persistent invasion of the immune system by pathogens such as cytomegaloviruses, which can use the bone marrow as a reservoir of latency<sup>79</sup>.

Changes in lipid metabolism with age are known to have an important impact on the lifespan and age-related diseases<sup>80</sup>. Our most intriguing results concern the widespread sex-dependent deregulation of *ApoE*, its receptors and *App*. We indeed observed a general over-emission of *ApoE* and under-emission of *App* in most TMS FACS (male) tissues but an opposite trend in TMS FACS (female) tissues (Extended Data Fig. 4b). In the brain, these genes are known to play a role in Alzheimer’s disease, for which there are well-documented and important sex differences<sup>81</sup>. *App* is the precursor of amyloid beta (A $\beta$ ) peptides<sup>82</sup>; *ApoE* is a known regulator of A $\beta$  clearance<sup>83</sup>; and interactions between *App* and *ApoE* receptors influence A $\beta$  metabolism and toxicity<sup>84,85</sup>.

Much less research has been performed on the function of *App* outside the central nervous system<sup>86</sup>, but some studies point toward its role in other pathologies, such as in obesity<sup>87,88</sup>, and in different tissues, such as skin<sup>89</sup>, intestine<sup>90</sup> (where it is modulated by diet<sup>91</sup>) and muscles<sup>92</sup>, particularly at neuromuscular junctions<sup>93</sup>. Taken together, those studies and our results suggest that changes with age in *App* intercellular trafficking might lead to a variety of tissue-specific diseases.

The two main downregulated processes reported in our results—namely, changes in the extracellular matrix and development/growth/proliferation—are partially interdependent. They have in common the under-expression of integrins, which typically act as bidirectional mediators between the cytoplasm and the extracellular space and regulate mechanisms such as cell migration, adhesion, proliferation, apoptosis, tumor progression and senescence<sup>94</sup>. The multiple functions of these proteins and their deregulation that we observed across multiple organs (most notably when they interact with collagen) suggest that they might play key roles in the structural decline of tissues with age. They are also consistently downregulated in signaling involving stem cells, potentially impacting their maintenance and homeostasis, as suggested in previous articles<sup>10</sup>. Changes in processes related to growth and development also fit the hypotheses that the run-on from developmental mechanisms and programs may later impact on aging<sup>95</sup>.

A subpart of the development/growth/proliferation pathway concerns the global downregulation of angiogenesis and qualitative changes occurring in vessels. Several studies previously reported a decline with age in capillary density and in the formation of new blood vessels<sup>96,97</sup>, even leading some authors to postulate an ‘angiogenesis hypothesis of aging’<sup>98</sup>. Our analysis reveals that the pair *Gpi1:Amfr*, mainly detected from or to endothelial cells, is over-represented among downregulated CCIs in 10 male tissues (Extended Data Fig. 6b). Glucose-6-phosphate isomerase functions as an autocrine motility factor that stimulates endothelial cell motility<sup>99</sup>. This LRI could be an important regulator of microvascular aging, and its therapeutic potential seems worth investigating. Moreover, we found downregulation of *Angpt1* (angiopoietin-1) in several tissues (Extended Data Fig. 6b), which could lead to the formation of leaky vessels, as vessel stability relies on the balance of *Angpt1* and *Angpt2* (ref. 100).

We conclude with some of the limitations of using gene expression of ligands and receptors to infer ICC activity. First, using mRNA counts as a proxy for the actual level of secreted proteins may lead to an overestimation of some intercellular interactions, as we assume that all transcripts participate in signaling, even though a fraction of them might be produced for intracellular processes. Despite this concern, scDiffCom was still able to predict CCIs that were globally consistent with secretomics data. Second, it is not always clear how some interactions (for example, *App:Lrp10*) are shared between cells and which fraction of them are autocrine rather than paracrine. Third, by looking only at ligands and receptors, our method does not assess changes in other players acting in downstream signaling. Fourth, we did not consider all possible types of intercellular mediators, and we did not investigate inter-tissue interactions, such as endocrine signals. Fifth, the CCI score does not account for potential changes in cell composition. Despite these limitations, we think that our atlas will be useful for the community and lead to hypotheses on ICC and aging to be further validated.

## Methods

### Retrieving LRIs

We downloaded LRIs from seven publicly available databases. Data from CellChat (version 1.1.3), NicheNet (version 1.1.1) and SingleCellSignalR (version 1.10.0) were directly accessed from their associated R packages. Data from CellPhoneDB, CellTalkDB, connectomeDB2020 and ICELLNET were retrieved online (as of 16 April 2022) from their respective websites. All details regarding download dates and links

are directly accessible from our package scDiffCom. We initially also retrieved LRIs from an eighth database, LRBaseDb from scTensor<sup>33</sup>, but we did not consider it further, as all curated LRIs present in it were derived from the seven others mentioned above.

### Processing LRIs

We analyzed the documentation and annotations of each resource to keep only their curated LRIs. We removed the interactions that were only bioinformatically predicted, such as from protein–protein interaction networks. We checked that gene symbols were approved by the HUGO Gene Nomenclature Committee (HGNC)<sup>101</sup> and, when mouse data were available, approved by Mouse Genome Informatics (MGI)<sup>102</sup>. If mouse LRIs were not provided (namely, for CellPhoneDB, connectomeDB2020, ICELLNET, NicheNet and SingleCellSignalR), we converted human LRIs to their mouse equivalent by retrieving orthology information from Ensembl version 102 (ref. 103), accessed through the R package biomaRt (version 2.50.2)<sup>104</sup>. We only kept mouse LRIs whose genes had high-confidence homolog pairs (setting the parameter `mmusculus_homolog_orthology_confidence` to 1) or whose genes were already included in the two databases providing mouse LRIs directly (CellChat and CellTalkDB).

LRIs from each resource were combined in a single list. Special care was taken to avoid duplicates arising from the same interactions available for both directions (for example, G1:G2 versus G2:G1), typically for juxtacrine signaling where the notion of ligand or receptor can sometimes be arbitrary.

As some of the resources provide only simple interactions, some of their LRIs could be incomplete. To partially correct this effect, we removed simple LRIs present in such databases if they were also found in complex databases but only in a complex form. For instance, we removed *Col3a1:Itgb1*, existing in SingleCellSignalR, as it always appears in a complex form in CellPhoneDB, such as in *Col3a1:Itga1-Itgb1*.

We manually verified the combined list of LRIs and removed approximately 200 records that we considered mis-curated—for example, *Mapk1:Fgfr2*, *Calml1:Adcy8*, *Gnas:Adcy1* and *Hsp90aa1:Cftr*. To do so, we first annotated each gene with descriptions from MyGene.Info (version 3.2.2)<sup>105,106</sup> and categories from OmniPath (version 3.3.13)<sup>107</sup>. We then identified genes that seemed unlikely to participate in ICC because they were not annotated as being potentially secreted and observed extracellularly. Finally, we explored the interactions involving those genes and evaluated the evidence supporting their existence. Some of the removed LRIs were initially included based on publications describing intracellular interactions that had been misinterpreted as intercellular—for example, Loo et al.<sup>108</sup> is presented as evidence for *Hsp90aa1:Cftr*, but the article does not mention ICC.

### Annotating LRIs with GO terms and KEGG pathways

LRIs were annotated with GO terms by following a graph-based approach (Fig. 1b). GO terms associated with each gene were retrieved from Ensembl version 102 (ref. 103), accessed through the R package biomaRt (version 2.50.2)<sup>104</sup>. For a given LRI, we used the R package ontoProc (version 1.16.0)<sup>109</sup> to build the ligand and receptor ontology subgraphs, whose nodes included their respective GO terms and ancestors up to the root node. For complex LRIs including, for example, multiple ligand genes, we considered the union of the terms associated with each ligand gene. The final LRI GO terms were those present in both the ligand and the receptor subgraphs—that is, the intersection of those graphs at the node level. scDiffCom can provide other GO term annotation methods (Supplementary Text 2), but we do not recommend them outside benchmarking purposes as we consider them less biologically relevant.

To annotate LRIs with KEGG pathways, we used the R package KEGGREST (version 1.34.0)<sup>110</sup> to retrieve all pathways associated with a given gene. For each LRI, we then only retained the pathways that included both the ligand gene(s) and receptor gene(s).



### Annotating ligands and receptors with aging resources

To annotate the genes from the mouse LRI database with the number of PubMed articles associating them with aging, we first downloaded the gene2pubmed table from <ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2pubmed.gz> on 16 April 2022. We filtered the table by keeping only mouse and human genes and by removing articles associated with more than 50 genes. Second, we performed a PubMed search with the R package rentrez (version 1.2.3)<sup>111</sup> for all articles containing at least one of the following keywords in their title or abstract (TIAB search): aging, longevity, senescence, age-related, dementia, alzheimer, parkinson, atherosclerosis, stroke, arthritis, osteoporosis and cataract. Diseases without a very direct and strong connection with aging, as well as cancers that have a too-broad scope, were excluded. Finally, we merged the gene2pubmed table with our aging search to count how many PubMed articles mentioning a given mouse ligand or receptor (or its human homolog) were related to aging.

To further determine if a gene of the mouse LRI database (or its human homolog) was previously associated with aging, we downloaded GenAge (human and mouse genes, build 20), LongevityMap (build 3) and the Microarray meta-analysis of Ageing Gene Expression database from the HAGR website (<https://genomics.senescence.info/download.html>) on 16 April 2022. On the same date, we retrieved the CellAge database of cellular senescence genes (build 2) from the supplementary material of Tejada-Martinez et al.<sup>51</sup>

### CCI score based on the geometric mean

For each CCI in the form (emitter cell type, receiver cell type, ligand(s) and receptor(s)), a score  $\varphi$  is computed in each condition as the geometric mean  $\varphi = \sqrt{e_L \cdot e_R}$  between the averaged expression  $e_L$  of the ligand gene in the emitter cells and the averaged expression  $e_R$  of the receptor gene in the receiver cells (based on normalized non-log-transformed read counts/UMIs). In the case of complex LRIs with multiple ligand genes (or receptor genes) involved,  $e_L$  (or  $e_R$ ) is given by the minimum value from the set of average expressions of those genes.

Defining a CCI score is a standard approach when investigating ICC from scRNA-seq data<sup>23,24</sup>, although the way of computing the score varies between studies. Here, the choice of the geometric mean (similar to SingleCellSignalR) rather than the arithmetic mean (as used by CellPhoneDB) is motivated by several advantages. The first one is that the geometric mean tends toward zero if either  $e_L$  or  $e_R$  tends to zero. This implies that when a highly expressed ligand is combined with a lowly expressed receptor (or vice versa), the score is not dominated by the large ligand value, as would have been the case with the arithmetic mean. Along the same line, although transcript counts or UMIs give only an indirect representation of protein levels<sup>112</sup>, molecular interactions are usually modeled by the law of mass action<sup>113</sup>, which is by essence multiplicative and not additive in protein concentrations. Finally, the geometric mean provides a clear interpretation of the log fold change of the scores between the two conditions of interest. Indeed, we see that the log fold change of the CCI score across two hypothetical conditions  $A$  and  $B$  corresponds to the (arithmetic) average between the respective ligand and receptor log fold changes,

$$\log\left(\frac{\varphi_B}{\varphi_A}\right) = \log\left(\frac{\sqrt{e_{L,B}e_{R,B}}}{\sqrt{e_{L,A}e_{R,A}}}\right) = \frac{1}{2}\left(\log\left(\frac{e_{L,B}}{e_{L,A}}\right) + \log\left(\frac{e_{R,B}}{e_{R,A}}\right)\right).$$

At present, scDiffCom calculates CCI scores exclusively from gene expression and relies only on the LRI database as a prior source of information. However, additional forms of prior knowledge, such as binding kinetics parameters, may become available soon and enable more precise score calculations. Supplementary Text 4 provides some concrete examples.

### CCI detection and differential analyses

Our approach relies on three permutation tests to assess if a CCI is (1) cell-type pair specific in condition  $A$ , (2) cell-type pair specific in condition  $B$  and (3) differentially expressed between  $A$  and  $B$ . To be

computationally more efficient, the three tests are done together as part of a single iteration loop. All threshold parameters described below can be adjusted by the users.

Given  $m$  cell types and  $l$  LRIs that are found in the scRNA-seq dataset, scDiffCom builds a table of  $m^2 \cdot l$  hypothetical CCIs. For each CCI, we compute the CCI scores  $\varphi_A$  and  $\varphi_B$ , the log fold change  $\log(\varphi_B/\varphi_A)$  and the variables  $n_{i,j}$  and  $d_{i,j}$  corresponding to the number and fraction of emitter cells expressing the ligand ( $i = L$ ) or receiver cells expressing the receptor ( $i = R$ ) in either condition ( $j \in \{A, B\}$ ). A CCI is deemed 'not expressed' in condition  $j$  if ( $n_{L,j} < 5$  or  $n_{R,j} < 5$ ) and ( $d_{L,j} < 0.1$  or  $d_{R,j} < 0.1$ ).

Only the CCIs 'expressed' in at least  $A$  or  $B$  are passed to the iteration loop. At each iteration  $k$ , three independent operations are done: (1) shuffling the cell-type labels of cells from condition  $A$  and returning the random score  $\tilde{\varphi}_A^k$  as the  $k$ -th element of the null distribution representing the random variable  $\Phi_A$ ; (2) same for condition  $B$ , returning  $\tilde{\varphi}_B^k$  to form the null distribution of  $\Phi_B$ ; and (3) keeping the original cell-type labels but shuffling the  $A$  and  $B$  condition labels and returning the random score difference  $\delta^k = \varphi_B^k - \varphi_A^k$  to form the null distribution of the random variable  $\Delta$ . After iterating, the true values  $\varphi_A$ ,  $\varphi_B$  and  $\delta = \varphi_B - \varphi_A$  are compared to the three null distributions to compute the two one-sided specificity  $P$  values  $p_A = P(\Phi_A > \varphi_A)$  and  $p_B = P(\Phi_B > \varphi_B)$  and the differential two-sided  $P$  value  $p_{DE} = P(|\Delta| > |\delta|)$ . Those  $P$  values are then adjusted for false discovery rate according to the Benjamini-Hochberg procedure<sup>114</sup>.

A CCI is considered 'detected' in condition  $j \in \{A, B\}$  if (1) it is 'expressed'; (2) it is 'specific', based on the specificity  $P$  values ( $p_{DE}^{adj.} \leq 0.05$ ); and (3) its score is among the top 80% of all the 'specific' CCI scores of both conditions, namely  $\varphi_j \geq q_{\varphi_{AB}}(20)$ , where  $q_{\varphi_{AB}}(x)$  is the  $x$ -th percentile of the scores. A CCI is called 'differentially expressed' if  $p_{DE}^{adj.} \leq 0.05$  and  $|\log(\varphi_B/\varphi_A)| \geq \log(1.5)$ .

As performing permutation tests can be computationally demanding, special care was taken to optimize the above computations. In particular, scDiffCom relies on the R package data.table (version 1.14.8)<sup>115</sup> for fast manipulations of large data. Along the same line, we leveraged the R package future (version 1.32.0)<sup>116</sup> allowing scDiffCom to perform permutation iterations in parallel. A toy model analysis (1,000 iterations on a dataset of 1,000 cells and five cell types) took a couple of minutes when run sequentially on a single-core computer. A more realistic example (10,000 permutations on 3,107 cells and 16 cell types) was measured to take around 9 min when run in parallel on 30 CPUs.

### CCI classification

We only kept CCIs that were 'detected' in at least one of the two conditions. They are then classified into four categories: (1) UP when  $p_{DE}^{adj.} \leq 0.05$  and  $\logfc \geq \log(1.5)$ ; (2) DOWN when  $p_{DE}^{adj.} \leq 0.05$  and  $\logfc \leq -\log(1.5)$ ; (3) FLAT when  $|\logfc| < \log(1.5)$ ; and (4) NSC when  $p_{DE}^{adj.} > 0.05$  and  $|\logfc| \geq \log(1.5)$ .

The detection analysis was used to remove biologically irrelevant interactions but not to predict actual changes. Using the detection test for this purpose was indeed prone to return false-positive varying signals—that is, CCIs that seem to appear or disappear because they fluctuate around the detection threshold but that are, in reality, not differentially expressed. Using the aging datasets as benchmarking data, we considered all possible outcomes and noticed only a marginal number of seemingly contradictory cases between the two tests, such as disappearing CCIs with positive log fold changes (Supplementary Table 11). Those are, for instance, due to a reduction of the fraction of expressing cells, despite the increase of the signal, hence our decision to prioritize the classification of the CCIs based on the differential test that shuffles the  $A$  and  $B$  condition labels as explained in the previous subsection.

For benchmarking purposes, scDiffCom also performs standard differential expression analysis on individual genes in their respective cell type. Following the same notation as above, scDiffCom builds a

distribution for the difference in expression of the ligand in emitter cells,  $E_{L,B} - E_{L,A}$ , alongside the same permutation loop as for the CCI scores (similarly for the receptor  $R$  in receiver cells). The  $P$  value for the ligand to be differentially expressed is computed as  $p_{L,DE} = P(|E_{L,B} - E_{L,A}| > |e_{L,B} - e_{L,A}|)$  and then adjusted according to Benjamini–Hochberg (similarly for the receptor  $R$ ). There are more direct methods than a permutation test (for example, a Wilcoxon test) to obtain such differential  $P$  values. However, our approach ensures that comparing these  $P$  values to the ‘CCI score  $P$  value’ does not suffer from technical biases as they are computed from the same permutations. In practice, we performed such a comparison using the 393,035 CCIs from scAgeCom. Using  $p_{L,DE}^{adj.} < 0.05$  and  $|\log_2 f_{C,L}| \geq \log(1.5)$  as thresholds for differential expression of the ligands (respectively, receptors), we saw that the classification of a considerable fraction of CCIs was determined without ambiguity by scDiffCom but not by standard differential expression (Extended Data Fig. 2).

### Fisher’s exact test to find over-represented signals

ORA is used to evaluate frequent patterns in categorical data—for example, to find if a particular feature of CCIs—for example, the annotation with the GO term ‘T cell differentiation’ is more frequent in upregulated CCIs compared to all other CCIs. This statistical association is measured by compiling the corresponding  $2 \times 2$  contingency table (upregulated/not upregulated versus annotated/not annotated) and applying two-sided Fisher’s exact test. We performed this procedure for every CCI feature (all GO terms, KEGG pathways, LRIs, ligands, receptors and cell types) and classes (UP, DOWN and FLAT). It returns an OR and a  $P$  value adjusted for multiple testing according to the Benjamini–Hochberg procedure<sup>114</sup>. In some instances (for example, pattern ranking and plots), to sort the results based on a single value, we combined the OR and the  $P$  value to create an ORA score, by adapting the gene significance score ( $\pi$  value) used in differential gene expression analysis and gene set enrichment analysis<sup>117</sup>:  $ORA\ score = \log_2 OR \cdot (-\log_{10} pval)$ .

### Visualization tools in scDiffCom

We implemented two functions in scDiffCom to visualize the over-representation results. scDiffCom::PlotORA displays the top over-represented keywords of a given category and regulation. It is implemented on top of the R package ggplot2 (version 3.4.2)<sup>118</sup>. scDiffCom::BuildNetwork shows on a summary graph the over-represented cell types and cell-type pairs. It relies on the R package igraph (version 1.4.2)<sup>119</sup> for internal computations and on the R package visNetwork (version 2.1.2)<sup>120</sup> for the interactive rendering.

### Retrieving and preparing scRNA-seq datasets

We downloaded the latest version (as of 21 March 2021) of TMS from the Amazon S3 czb-tabula-muris-senis repository and the Calico dataset from the calicolabs website. They had been pre-processed and annotated with the Python toolkit Scanpy<sup>121</sup> before our work, and we converted the resulting h5ad files to R Seurat objects.

As stated in the original TMS article<sup>41</sup>, FACS and Droplet refer to the technique used to capture the cells, namely (1) cell sorting in microtiter well plates followed by Smart-seq2 library preparation and full-length sequencing and (2) cell capture by microfluidic droplets as per the 10x Genomics protocol followed by 3’ end counting. The Calico data were exclusively obtained using the Droplet technique. Regarding mice age, TMS provided multiple timepoints that needed to be grouped into ‘young’ and ‘old’ categories. We removed 1-month-old cells and 30-month-old cells to avoid bias due to developmental or longevity-related processes. Therefore, we compared 3-month-old cells to 18/24-month-old cells from TMS and 7/8-month-old cells to 22/23-month-old cells from Calico. Finally, we filtered out tissues that were missing one age group (for example, TMS Droplet Fat contained only old cells).

The cells from each dataset were sequenced together. However, we decided to regroup TMS FACS Brain\_Myeloid and Brain\_Non-myeloid as the former contained only two cell types (macrophage and microglial cell), and merging the datasets allowed us to infer interactions with the other parts of the brain. We verified that this did not considerably alter the interactions detected in each dataset independently.

### Cell-type characterization

In each dataset, we standardized the names of the cell types based on Cell Ontology standards<sup>122</sup>—for example, ‘atrial myocyte’ was renamed as ‘regular atrial cardiac myocyte’. We also regrouped some specialized cell clusters—for example, CD4<sup>+</sup> and CD8<sup>+</sup> T cells—to increase sample size and avoid overlapping cell types. These overlaps were exceptionally kept in some tissues—for instance, the ‘undetermined myeloid leukocytes’ in the Lung dataset from Calico Droplet (male) overlap with some specialized cell types, such as ‘classical monocytes’, but were kept as distinct categories. Finally, we classified the cell types into 10 families to facilitate downstream analyses: ‘endothelial cells’, ‘epithelial cells’, ‘connective tissue cells’, ‘leukocytes’, ‘stem cells’, ‘neurons’, ‘glial cells’, ‘muscle cells’, ‘erythroid lineage cells’ and ‘hematopoietic precursor cells’ (Supplementary Table 12).

### Building and deploying scAgeComShiny

We used the R package golem (version 0.4.0)<sup>123</sup> to build the Shiny app scAgeComShiny, which contains all scAgeCom results. Interactive scatter plots were built with plotly (version 4.10.1)<sup>124</sup>, which was also used to display the GO terms tree maps. Those were internally computed with the R package rrvgo (version 1.10.0)<sup>125</sup>, according to the original method from REVIGO<sup>126</sup>. To deploy the application, we first used golem to create a Docker image of the scAgeComShiny app and then serve it with the containerized version of ShinyProxy open-source middleware (version 3.0.1).

### Comparing scDiffCom to secretomics data

We leveraged the supplementary data of six recent studies to retrieve datasets of proteins secreted by six different cell types: mBMM<sup>61</sup>, mNeuron<sup>62</sup>, mMSC-AT<sup>63</sup>, rCM<sup>64</sup>, hUVEC<sup>65</sup> and hPDE<sup>66</sup>. When necessary, we only kept the proteins secreted by control cells and discarded results related to specific conditions, such as cancer cells. Protein names were converted to MGI gene symbols (using orthology conversion if necessary) and intersected with the genes from our mouse LRI database. For each secretomics dataset, remaining gene symbols were used as input to an over-representation analysis assessing the cell-type specificity of the CCIs returned by scDiffCom. More precisely, each CCI was categorized into one of four groups depending on whether its ligand gene was part of the detected proteins and if its emitter cell type (or cell type family) was part of the cell type of interest. A two-sided Fisher’s exact test was applied to each such contingency table to compute corresponding  $P$  values and ORs.  $P$  values were further adjusted for multiple comparisons (Benjamini–Hochberg procedure). We focused our attention on the ligand and emitter parts of each CCI, as secretomics data do not directly provide information regarding the targets of secreted proteins.

### Statistics and reproducibility

No statistical method was used to pre-determine sample size. Sample size selection was informed and restricted by the availability of the public scRNA-seq data used in this study<sup>41,42</sup>. Randomization and blinding were not performed in this study, as all data collected were public, and the authors performing the analyses had already worked on most of the data before starting this study. Few groups of cells were excluded from the original scRNA-seq datasets to avoid biases, as explained in Methods. Both statistical tests implemented in scDiffCom (differential expression and over-representation) are non-parametric and do not assume the data to follow specific distributions. Other statistical tests are mentioned in the figures with additional details provided in



Methods. Adjustments for multiple testing were always performed when relevant, as reported.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Source data are provided with this paper

Datasets used in this study:

- scRNA-seq Tabula Muris Senis: <https://s3.console.aws.amazon.com/s3/buckets/czb-tabula-muris-senis>
  - scRNA-seq Calico murine cell atlas: <https://mca.research.calicolabs.com/>
  - LRIs from CellChat: available from their R package
  - LRIs from NicheNet: available from their R package
  - LRIs from SingleCellSignalR: available from their R package
  - LRIs from CellPhoneDB: <https://www.cellphonedb.org/>
  - LRIs from CellTalkDB: <https://github.com/ZJUFanLab/CellTalkDB>
  - LRIs from connectomeDB2020: <https://asrhou.github.io/NATMI/>
  - LRIs from ICELLNET: <https://github.com/soumelis-lab/ICELLNET>
  - General gene information: <https://mygene.info/> and <https://omnipathdb.org/>
  - GO terms: Ensembl (version 102), <https://useast.ensembl.org/index.html>
  - KEGG: accessed via the R package KEGGREST
  - gene2pubmed table: <ftp://ftp.ncbi.nlm.nih.gov/gene/DATE/gene2pubmed.gz>
  - GenAge (build 20): <https://genomics.senescence.info/genes/index.html>
  - LongevityMap (build 3): <https://genomics.senescence.info/longevity/>
  - Microarray meta-analysis of Ageing Gene Expression database: [https://genomics.senescence.info/gene\\_expression/](https://genomics.senescence.info/gene_expression/)
  - CellAge (build 2): Supplementary Material of <https://doi.org/10.1093/molbev/msab369>
  - Secretomics data for mBMM: Supplementary Material of <https://doi.org/10.1126/science.1232578>
  - Secretomics data for mNeuron: Supplementary Material of <https://doi.org/10.15252/embj.2020105693>
  - Secretomics data for mMSC-AT: Supplementary Material of <https://doi.org/10.18632/aging.202423>
  - Secretomics data for rCM: Supplementary Material of <https://doi.org/10.1161/circulationaha.119.044914>
  - Secretomics data for hUVEC: Supplementary Material of <https://doi.org/10.1016/j.ajpath.2019.10.007>
  - Secretomics data for hPDE: Supplementary Material of <https://doi.org/10.1002/pmic.202100320>
- Datasets created in this study:
- scDiffCom LRIs: available from the R package scDiffCom (<https://github.com/CyrilLagger/scDiffCom/tree/master/data>)
  - Aging and sex scAgeCom results: <https://doi.org/10.6084/m9.figshare.17074964>
  - Data to run the app scAgeComShiny and scagecom.org: [https://figshare.com/articles/dataset/scAgeComShiny\\_data/17075375](https://figshare.com/articles/dataset/scAgeComShiny_data/17075375)

### Code availability

All code used in this study is available from the following repositories:

- R package scDiffCom: <https://github.com/CyrilLagger/scDiffCom>
- R scripts for the aging/sex analyses: <https://github.com/CyrilLagger/scAgeCom>
- Golem package for the Shiny app: <https://github.com/CyrilLagger/scAgeComShiny>
- Docker image of the Shiny app: <https://hub.docker.com/r/ursueugen/scagecom>

Although scagecom.org is available online, we note that the content of the website can always be reproduced locally by installing our Shiny app from Ras follows: `devtools::install_github('CyrilLagger/scAgeComShiny')`. The app can then be launched after having downloaded the necessary data in the app directory ([https://figshare.com/articles/dataset/scAgeComShiny\\_data/17075375](https://figshare.com/articles/dataset/scAgeComShiny_data/17075375)).

### References

1. de Magalhães, J. P. & Costa, J. A database of vertebrate longevity records and their relation to other life-history traits. *J. Evol. Biol.* **22**, 1770–1774 (2009).
2. López-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
3. Miller, H. A., Dean, E. S., Pletcher, S. D. & Leiser, S. F. Cell non-autonomous regulation of health and longevity. *eLife* **9**, e26259 (2020).
4. Capp, J.-P. & Thomas, F. Tissue-disruption-induced cellular stochasticity and epigenetic drift: common origins of aging and cancer? *Bioessays* **43**, e2000140 (2021).
5. Franceschi, C., Garagnani, P., Parini, P., Giuliani, C. & Santoro, A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat. Rev. Endocrinol.* **14**, 576–590 (2018).
6. Ovadya, Y. et al. Impaired immune surveillance accelerates accumulation of senescent cells and aging. *Nat. Commun.* **9**, 5435 (2018).
7. Fafián-Labora, J. A. & O’Loughlen, A. Classical and nonclassical intercellular communication in senescence and ageing. *Trends Cell Biol.* **30**, 628–639 (2020).
8. Signer, R. A. J. & Morrison, S. J. Mechanisms that regulate stem cell aging and life span. *Cell Stem Cell* **12**, 152–165 (2013).
9. Oh, J., Lee, Y. D. & Wagers, A. J. Stem cell aging: mechanisms, regulators and therapeutic opportunities. *Nat. Med.* **20**, 870–880 (2014).
10. Kurtz, A. & Oh, S.-J. Age related changes of the extracellular matrix and stem cell maintenance. *Prev. Med.* **54**, S50–S56 (2012).
11. Kehlet, S. N. et al. Age-related collagen turnover of the interstitial matrix and basement membrane: implications of age- and sex-dependent remodeling of the extracellular matrix. *PLoS ONE* **13**, e0194458 (2018).
12. Russell, S. J. & Kahn, C. R. Endocrine regulation of ageing. *Nat. Rev. Mol. Cell Biol.* **8**, 681–691 (2007).
13. Tatar, M., Bartke, A. & Antebi, A. The endocrine regulation of aging by insulin-like signals. *Science* **299**, 1346–1351 (2003).
14. Bartke, A. Growth hormone and aging: updated review. *World J. Mens Health* **37**, 19–30 (2019).
15. Rothwell, P. M. et al. Short-term effects of daily aspirin on cancer incidence, mortality, and non-vascular death: analysis of the time course of risks and benefits in 51 randomised controlled trials. *Lancet* **379**, 1602–1612 (2012).
16. Strong, R. et al. Nordihydroguaiaretic acid and aspirin increase lifespan of genetically heterogeneous male mice. *Aging Cell* **7**, 641–650 (2008).
17. Rothwell, P. M. et al. Effect of daily aspirin on long-term risk of death due to cancer: analysis of individual patient data from randomised trials. *Lancet* **377**, 31–41 (2011).
18. Conboy, I. M. et al. Rejuvenation of aged progenitor cells by exposure to a young systemic environment. *Nature* **433**, 760–764 (2005).
19. Conboy, I. M. & Rando, T. A. Heterochronic parabiosis for the study of the effects of aging on stem cells and their niches. *Cell Cycle* **11**, 2260–2267 (2012).
20. Villeda, S. A. et al. Young blood reverses age-related impairments in cognitive function and synaptic plasticity in mice. *Nat. Med.* **20**, 659–663 (2014).

21. Raposo, G. & Stahl, P. D. Extracellular vesicles: a new communication paradigm? *Nat. Rev. Mol. Cell Biol.* **20**, 509–510 (2019).
22. Al Amir Dache, Z. et al. Blood contains circulating cell-free respiratory competent mitochondria. *FASEB J.* **34**, 3616–3630 (2020).
23. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2021).
24. Shao, X., Lu, X., Liao, J., Chen, H. & Fan, X. New avenues for systematically inferring cell–cell communication: through single-cell transcriptomics data. *Protein Cell* **11**, 866–880 (2020).
25. Ramilowski, J. A. et al. A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).
26. Jin, S. et al. Inference and analysis of cell–cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
27. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
28. Shao, X. et al. CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice. *Brief. Bioinformatics* **22**, bbaa269 (2021).
29. Hou, R., Denisenko, E., Ong, H. T., Ramilowski, J. A. & Forrest, A. R. Predicting cell-to-cell communication networks using NATMI. *Nat. Commun.* **11**, 5011 (2020).
30. Noël, F. et al. Dissection of intercellular communication using the transcriptome-based framework ICELLNET. *Nat. Commun.* **12**, 1089 (2021).
31. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
32. Cabello-Aguilar, S. et al. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2020).
33. Tsuyuzaki, K., Ishii, M. & Nikaido, I. Uncovering hypergraphs of cell–cell interaction from single cell RNA-sequencing data. Preprint at *bioRxiv* <https://doi.org/10.1101/566182> (2019).
34. Solovey, M. & Scialdone, A. COMUNET: a tool to explore and visualize intercellular communication. *Bioinformatics* **36**, 4296–4300 (2020).
35. He, X., Memczak, S., Qu, J., Belmonte, J. C. I. & Liu, G.-H. Single-cell omics in ageing: a young and growing field. *Nat. Metab.* **2**, 293–302 (2020).
36. Uyar, B. et al. Single-cell analyses of aging, inflammation and senescence. *Ageing Res. Rev.* **64**, 101156 (2020).
37. Ximerakis, M. et al. Single-cell transcriptomic profiling of the aging mouse brain. *Nat. Neurosci.* **22**, 1696–1708 (2019).
38. Ma, S. et al. Caloric restriction reprograms the single-cell transcriptional landscape of *Rattus norvegicus* aging. *Cell* **180**, 984–1001 (2020).
39. Ma, S. et al. Single-cell transcriptomic atlas of primate cardiopulmonary aging. *Cell Res.* **31**, 415–432 (2021).
40. Zou, Z. et al. A single-cell transcriptomic atlas of human skin aging. *Dev. Cell* **56**, 383–397.e8 (2021).
41. Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
42. Kimmel, J. C. et al. Murine single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging. *Genome Res.* **29**, 2088–2103 (2019).
43. Ben-Shlomo, I., Yu Hsu, S., Rauch, R., Kowalski, H. W. & Hsueh, A. J. W. Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. *Sci STKE* **2003**, RE9 (2003).
44. Prasad, T. S. K., Kandasamy, K. & Pandey, A. Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods Mol. Biol.* **577**, 67–79 (2009).
45. Harding, S. D. et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY. *Nucleic Acids Res.* **46**, D1091–D1106 (2018).
46. Fabregat, A. et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
47. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
48. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
49. Tacutu, R. et al. Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res.* **46**, D1083–D1090 (2018).
50. Avelar, R. A. et al. A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biol.* **21**, 91 (2020).
51. Tejada-Martinez, D. et al. Positive selection and enhancer evolution shaped lifespan and body mass in great apes. *Mol. Biol. Evol.* **39**, msab369 (2022).
52. Budovsky, A. et al. LongevityMap: a database of human genetic variants associated with longevity. *Trends Genet.* **29**, 559–560 (2013).
53. de Magalhães, J. P., Curado, J. & Church, G. M. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* **25**, 875–881 (2009).
54. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
55. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
56. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
57. Büttner, M., Ostner, J., Müller, C. L., Theis, F. J. & Schubert, B. scCODA is a Bayesian model for compositional single-cell data analysis. *Nat. Commun.* **12**, 6876 (2021).
58. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 (2018).
59. Good, P. *Permutation, Parametric and Bootstrap Tests of Hypotheses* (Springer, 2005).
60. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb.)* **2**, 100141 (2021).
61. Meissner, F., Scheltema, R. A., Mollenkopf, H.-J. & Mann, M. Direct proteomic quantification of the secretome of activated immune cells. *Science* **340**, 475–478 (2013).
62. Tüshaus, J. et al. An optimized quantitative proteomics method establishes the cell type-resolved mouse brain secretome. *EMBO J.* **39**, e105693 (2020).
63. Acar, M. B. et al. Obesity induced by high-fat diet is associated with critical changes in biological and molecular functions of mesenchymal stromal cells present in visceral adipose tissue. *Ageing (Albany, NY)* **12**, 24894–24913 (2020).
64. Kuhn, T. C. et al. Secretome analysis of cardiomyocytes identifies PCSK6 (proprotein convertase subtilisin/kexin type 6) as a novel player in cardiac remodeling after myocardial infarction. *Circulation* **141**, 1628–1644 (2020).
65. Zhao, Y. et al. Quantitative proteomics of the endothelial secretome identifies RC0497 as diagnostic of acute rickettsial spotted fever infections. *Am. J. Pathol.* **190**, 306–322 (2020).
66. Li, X. et al. Proteome and secretome analysis of pancreatic cancer cells. *Proteomics* **22**, e2100320 (2022).

67. Ekpruke, C. D. & Silveyra, P. Sex differences in airway remodeling and inflammation: clinical and biological factors. *Front. Allergy* **3**, 875295 (2022).
68. Ziegenhain, C. et al. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **65**, 631–643 (2017).
69. Zhang, M. J., Pisco, A. O., Darmanis, S. & Zou, J. Mouse aging cell atlas analysis reveals global and cell type-specific aging signatures. *eLife* **10**, e62293 (2021).
70. Smith, L. K. et al.  $\beta$ 2-microglobulin is a systemic pro-aging factor that impairs cognitive function and neurogenesis. *Nat. Med.* **21**, 932–937 (2015).
71. Castellano, J. M. Blood-based therapies to combat aging. *Gerontology* **65**, 84–89 (2019).
72. Basisty, N. et al. A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLoS Biol.* **18**, e3000599 (2020).
73. Zhang, Y. et al. Macrophage migration inhibitory factor rejuvenates aged human mesenchymal stem cells and improves myocardial repair. *Aging (Albany, NY)* **11**, 12641–12660 (2019).
74. Xia, W., Zhang, F., Xie, C., Jiang, M. & Hou, M. Macrophage migration inhibitory factor confers resistance to senescence through CD74-dependent AMPK-FOXO3a signaling in mesenchymal stem cells. *Stem Cell Res. Ther.* **6**, 82 (2015).
75. Kokkola, R. et al. RAGE is the major receptor for the proinflammatory activity of HMGB1 in rodent macrophages. *Scand. J. Immunol.* **61**, 1–9 (2005).
76. Abeyama, K. et al. The N-terminal domain of thrombomodulin sequesters high-mobility group-B1 protein, a novel antiinflammatory mechanism. *J. Clin. Invest.* **115**, 1267–1274 (2005).
77. Dumas, S., Kolokotronis, A. & Stefanopoulos, P. Anti-inflammatory and antimicrobial roles of secretory leukocyte protease inhibitor. *Infect. Immun.* **73**, 1271–1274 (2005).
78. Py, B. et al. The phospholipid scramblases 1 and 4 are cellular receptors for the secretory leukocyte protease inhibitor and interact with CD4 at the plasma membrane. *PLoS ONE* **4**, e5006 (2009).
79. Jergović, M., Contreras, N. A. & Nikolich-Žugich, J. Impact of CMV upon immune aging: facts and fiction. *Med. Microbiol. Immunol.* **208**, 263–269 (2019).
80. Johnson, A. A. & Stolzing, A. The role of lipid metabolism in aging, lifespan regulation, and age-related disease. *Aging Cell* **18**, e13048 (2019).
81. Riedel, B. C., Thompson, P. M. & Brinton, R. D. Age, APOE and sex: triad of risk of Alzheimer's disease. *J. Steroid Biochem. Mol. Biol.* **160**, 134–147 (2016).
82. Chow, V. W., Mattson, M. P., Wong, P. C. & Gleichmann, M. An overview of APP processing enzymes and products. *Neuromolecul. Med.* **12**, 1–12 (2010).
83. Yu, J.-T., Tan, L. & Hardy, J. Apolipoprotein E in Alzheimer's disease: an update. *Annu. Rev. Neurosci.* **37**, 79–100 (2014).
84. Holtzman, D. M., Herz, J. & Bu, G. Apolipoprotein E and apolipoprotein E receptors: normal biology and roles in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **2**, a006312 (2012).
85. Herz, J. & Beffert, U. Apolipoprotein E receptors: linking brain development and Alzheimer's disease. *Nat. Rev. Neurosci.* **1**, 51–58 (2000).
86. Puig, K. L. & Combs, C. K. Expression and function of APP and its metabolites outside the central nervous system. *Exp. Gerontol.* **48**, 608–611 (2013).
87. Lee, Y.-H., Martin, J. M., Maple, R. L., Tharp, W. G. & Pratley, R. E. Plasma amyloid- $\beta$  peptide levels correlate with adipocyte amyloid precursor protein gene expression in obese individuals. *Neuroendocrinology* **90**, 383–390 (2009).
88. Lee, Y.-H. et al. Amyloid precursor protein expression is upregulated in adipocytes in obesity. *Obesity (Silver Spring)* **16**, 1493–1500 (2008).
89. Herzog, V., Kirfel, G., Siemes, C. & Schmitz, A. Biological roles of APP in the epidermis. *Eur. J. Cell Biol.* **83**, 613–624 (2004).
90. Puig, K. L., Swigost, A. J., Zhou, X., Sens, M. A. & Combs, C. K. Amyloid precursor protein expression modulates intestine immune phenotype. *J. Neuroimmune Pharmacol.* **7**, 215–230 (2012).
91. Galloway, S., Jian, L., Johnsen, R., Chew, S. & Mamo, J. C. L.  $\beta$ -Amyloid or its precursor protein is found in epithelial cells of the small intestine and is stimulated by high-fat feeding. *J. Nutr. Biochem.* **18**, 279–284 (2007).
92. Askanas, V. & Engel, W. K. Inclusion-body myositis: a myodegenerative conformational disorder associated with A $\beta$ , protein misfolding, and proteasome inhibition. *Neurology* **66**, S39–S48 (2006).
93. Wang, P. et al. Defective neuromuscular synapses in mice lacking amyloid precursor protein (APP) and APP-like protein 2. *J. Neurosci.* **25**, 1219–1225 (2005).
94. Borghesan, M. & O'Loughlin, A. Integrins in senescence and aging. *Cell Cycle* **16**, 909–910 (2017).
95. de Magalhães, J. P. Ageing as a software design flaw. *Genome Biol.* **24**, 51 (2023).
96. Moriya, J. & Minamino, T. Angiogenesis, cancer, and vascular aging. *Front. Cardiovasc. Med.* **4**, 65 (2017).
97. Hodges, N. A., Suarez-Martinez, A. D. & Murfee, W. L. Understanding angiogenesis during aging: opportunities for discoveries and new models. *J. Appl. Physiol.* **125**, 1843–1850 (2018).
98. Ambrose, C. T. Pro-angiogenesis therapy and aging: a mini-review. *Gerontology* **63**, 393–400 (2017).
99. Funasaka, T., Haga, A., Raz, A. & Nagase, H. Tumor autocrine motility factor is an angiogenic factor that stimulates endothelial cell motility. *Biochem. Biophys. Res. Commun.* **285**, 118–128 (2001).
100. Hayashi, S.-I., Rakugi, H. & Morishita, R. Insight into the role of angiopoietins in ageing-associated diseases. *Cells* **9**, 2636 (2020).
101. Tweedie, S. et al. Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res.* **49**, D939–D946 (2021).
102. Bult, C. J. et al. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).
103. Yates, A. D. et al. Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).
104. 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**, 1184–1191 (2009).
105. Xin, J. et al. High-performance web services for querying gene and variant annotation. *Genome Biol.* **17**, 91 (2016).
106. Wu, C., Macleod, I. & Su, A. I. BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res.* **41**, D561–D565 (2013).
107. Türei, D. et al. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol. Syst. Biol.* **17**, e9923 (2021).
108. Loo, M. A. et al. Perturbation of Hsp90 interaction with nascent CFTR prevents its maturation and accelerates its degradation by the proteasome. *EMBO J.* **17**, 6879–6887 (1998).
109. Greene, D., Richardson, S. & Turro, E. ontologyX: a suite of R packages for working with ontological data. *Bioinformatics* **33**, 1104–1106 (2017).
110. Bioconductor. KEGGREST. <https://doi.org/10.18129/B9.bioc.KEGGREST>
111. Winter, D. J. rentrez: an R package for the NCBI eUtils API. *R J.* **9**, 520–526 (2017).



112. Quinn, T. P., Erb, I., Richardson, M. F. & Crowley, T. M. Understanding sequencing data as compositions: an outlook and review. *Bioinformatics* **34**, 2870–2878 (2018).
113. Phillips, R., Kondev, J., Theriot, J. & Garcia, H. *Physical Biology of the Cell* (Garland Science, 2012).
114. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
115. Extension of ‘data.frame’. R package data.table version 1.14.8. <https://cran.r-project.org/web/packages/data.table/index.html> (2023).
116. Bengtsson, H. Unified parallel and distributed processing in R for everyone. R package future version 1.32.0 (2023).
117. Xiao, Y. et al. A novel significance score for gene selection and ranking. *Bioinformatics* **30**, 801–807 (2014).
118. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
119. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **1695**, 1–9 (2006).
120. Almende, B. V., Thieurmel, B. & Robert, T. Network visualization using ‘vis.js’ library. R package visNetwork version 2.0.9 (2019).
121. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
122. Diehl, A. D. et al. The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *J. Biomed. Semantics* **7**, 44 (2016).
123. golem: a framework for robust Shiny applications. R package golem version 0.3.1. <https://cran.r-project.org/web/packages/golem/index.html> (2021).
124. Plotly Technologies, Inc. *Collaborative Data Science* (Plotly Technologies, Inc., 2015).
125. Sayols, S. rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms. *microPubl. Biol.* <https://doi.org/doi:10.17912/micropub.biology.000811> (2023).
126. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).

## Acknowledgements

We thank J. Kimmel for helpful information regarding the Calico murine cell atlas, K. Tsuyuzaki for general discussion regarding intercellular communication at the early stage of this project, Z. Abidi for support to deploy the Shiny app and U. Castet for useful suggestions on the manuscript. C.L. is grateful for the funding provided by the Human Frontier Science Program (fellowship LT000741/2019-C). Our work in the Integrative Genomics of Ageing Group was further supported by grants from the Wellcome Trust (208375/Z/17/Z), LongeCity and the Biotechnology and Biological Sciences Research Council (BB/R014949/1) to J.P.d.M. R.T. and E.U. were supported by the National Authority for Scientific Research and Innovation and by the Ministry of European Funds, Romania, through the Competitiveness Operational Program 2014–2020, POC-A1-A.1.1.4-E-2015 (grant no.: 40/02.09.2016, ID: P\_37\_778, to R.T.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Author contributions

C.L. conceived the project, developed scDiffCom/scAgeCom/scAgeComShiny, compiled the ligand–receptor database, pre-processed scRNA-seq datasets, pre-processed and analyzed the secretomics data, interpreted the results and wrote the manuscript. E.U. developed scDiffCom/scAgeCom/scAgeComShiny, curated and analyzed the ligand–receptor database, performed supplementary analyses, interpreted the results and wrote the manuscript. A.E. contributed to scDiffCom/scAgeComShiny, annotated the ligand–receptor database, annotated the TMS/Calico cell types, interpreted the results and wrote the manuscript. R.A.A. pre-processed scRNA-seq datasets, interpreted the results and revised the manuscript. A.O.P. annotated the TMS/Calico cell types. R.T. revised the manuscript and supervised the project. J.P.d.M. revised the manuscript and supervised the project.

## Competing interests

J.P.d.M. is Chief Scientific Officer of YouthBio Therapeutics, an advisor/consultant for the Longevity Vision Fund, 199 Biotechnologies and NOVOS and the founder of Magellan Science Ltd., a company providing consulting services in longevity science. The remaining authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s43587-023-00514-x>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43587-023-00514-x>.

**Correspondence and requests for materials** should be addressed to João Pedro de Magalhães.

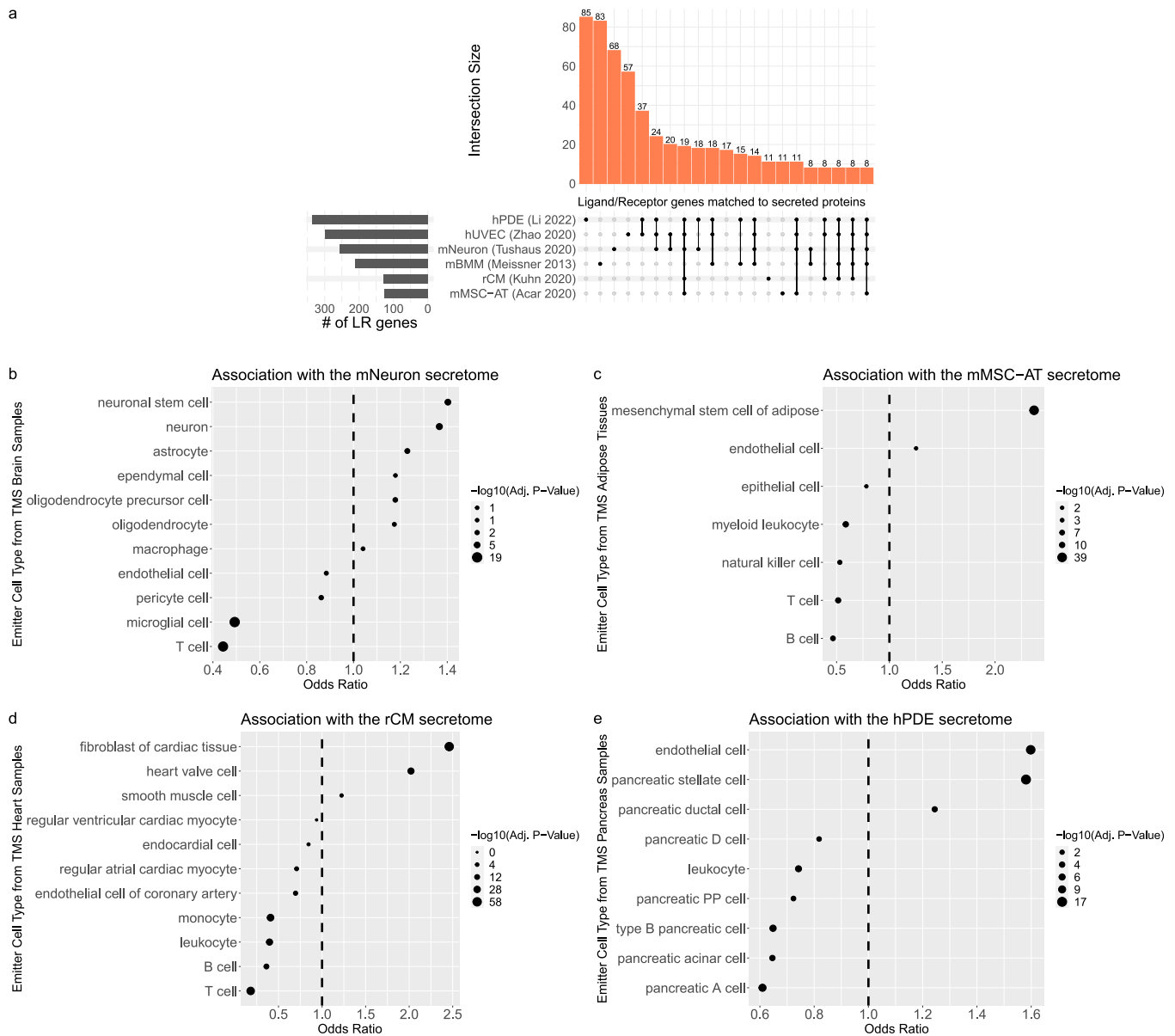
**Peer review information** *Nature Aging* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

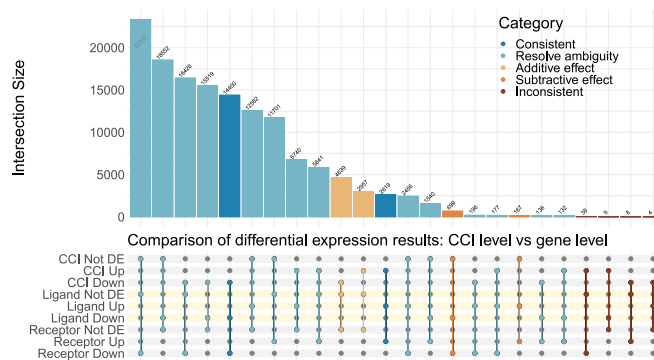
© The Author(s) 2023



**Extended Data Fig. 1 | Consistency between secretomics data and tissue-specific scDiffCom detection results.** **a**, After keeping the secreted proteins that matched to ligand and receptor genes present in the scDiffCom LRI database, mBMM, hPDE, mNeuron, and hUVEC cells have the highest number of specific proteins (that is only secreted by them), whereas rCM and mMSC-AT cells have most of their secreted proteins in common with other cell types; **b-e**, several secretomics datasets have their cell line originating from a tissue present in scAgeCom, allowing us to perform tissue-specific association analyses. Odds ratios and p-values were computed from a two-sided Fischer's exact test and p-values were further adjusted for multiple comparisons (Benjamini-Hochberg procedure); **b**, for scAgeCom TMS Brain results, the secretome from mNeuron is

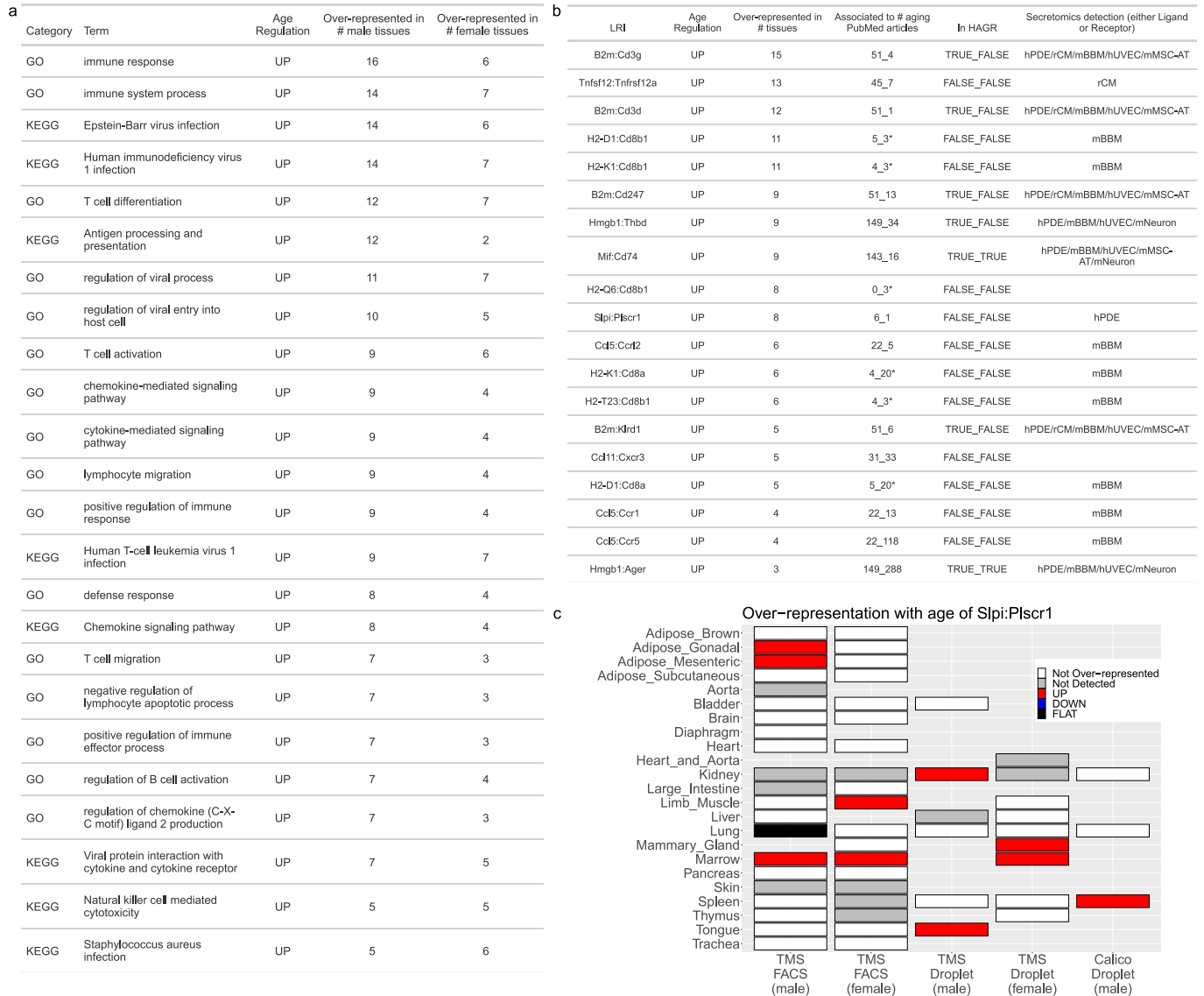
predominantly associated with CCI's emitted from consistent cell types, namely neuronal stem cells and neurons; **c**, for scAgeCom TMS Adipose Tissue results, the secretome of mMSC-AT is predominantly associated with CCI's emitted from the corresponding cell type, namely mesenchymal stem cells; **d**, for scAgeCom TMS Heart results, the secretome of rCM is predominantly associated with CCI's emitted from heart-specific cell types (fibroblast of cardiac tissue and heart valve cell) but is neither positively nor negatively associated with regular ventricular cardiac myocytes; **e**, for scAgeCom TMS Pancreas results, the secretome of hPDE is predominantly associated with CCI's emitted from endothelial cells and pancreatic stellate cells and to a lesser extent with its corresponding cell type, namely pancreatic ductal cells.





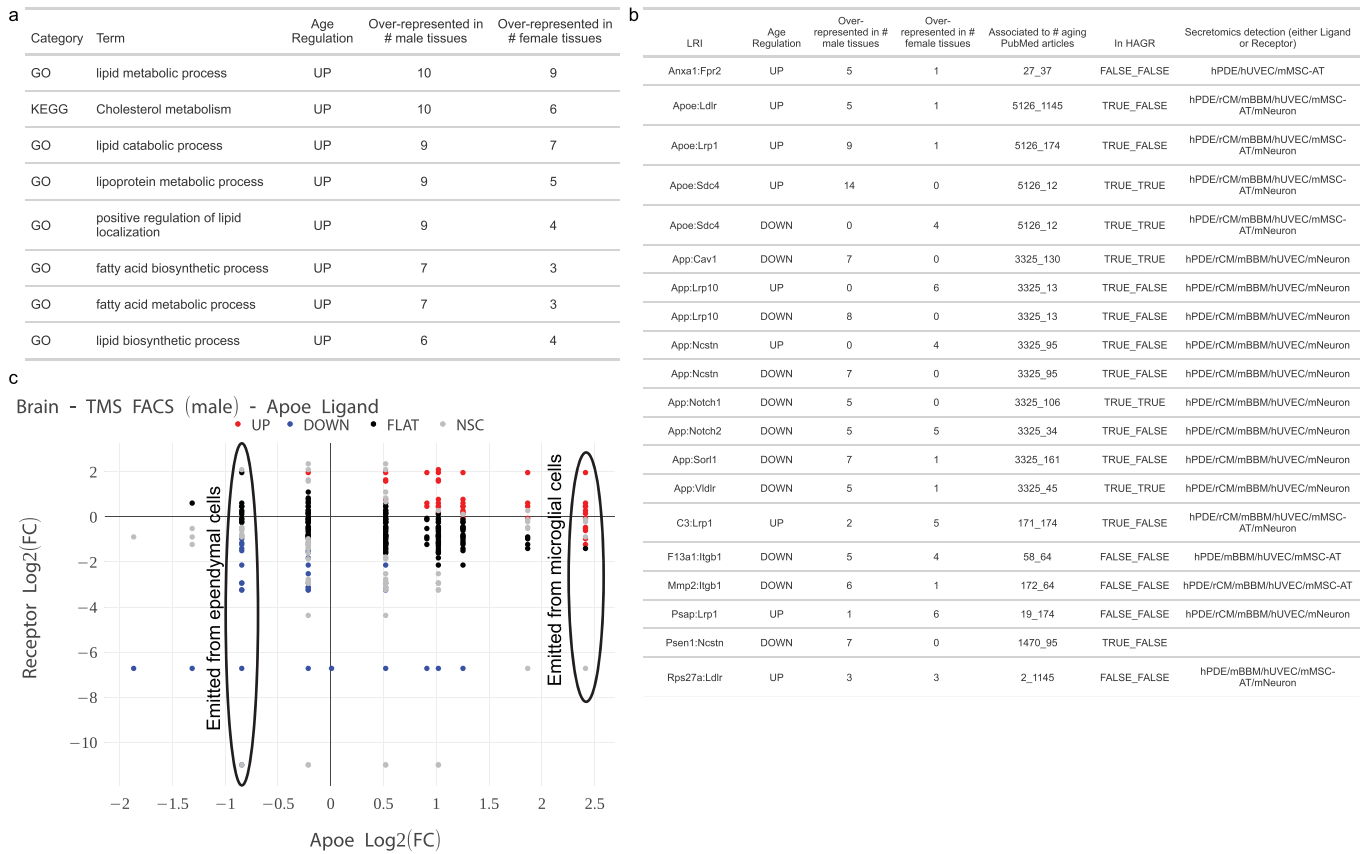
**Extended Data Fig. 2 | Comparing scDiffCom differential test to standard differential gene expression analysis.** Using the 393,035 scAgeCom CCIs as a benchmark, the UpsetPlot shows how they are distributed in functions of scDiffCom differential analysis (rows 1-3), standard differential analysis of the ligand (rows 4-6), and standard differential analysis of the receptor (rows 7-9). For clarity, we removed the column counting CCIs not differentially expressed (Not DE) in all three cases. For most of the remaining CCIs, scDiffCom resolves the ambiguity that appears when performing differential analysis on the ligand and receptor individually. For example, scDiffCom considers that among the CCIs with a non-differentially expressed ligand but a down-regulated receptor, 23,337

of those are not differentially expressed whereas 16,426 are down-regulated. We also observed additive effects (CCI differentially expressed while both the ligand and receptor are not) or subtractive effects (CCI not DE while both the ligand and receptor are DE in the same direction). Additive effects can be easily explained by the CCI having a fold change larger than the threshold while this is not the case for its genes individually. Subtractive effects are rare and mainly due to the p-value of the CCI being not significant, despite ligand and receptor p-values being significant. Inconsistent cases only represent a marginal fraction of all cases and can be explained by CCIs and genes having fold-changes and p-values close to the thresholds.



**Extended Data Fig. 3 | Processes and LRIs related to inflammation, the immune system, and viral processes are globally over-represented as up-regulated with age across tissues. a**, List of selected GO terms and KEGG pathways over-represented as up-regulated with age in several male and female tissues. **b**, List of selected LRIs over-represented as up-regulated with age in several tissues. We indicate how many aging PubMed articles mention individual genes and if each gene is part of the Human Ageing Genomic Resources (HAGR). Of note, the fact that histocompatibility 2 (H2) genes were

only associated with a low number of aging PubMed articles (as indicated with a star) is technical rather than biological and because most of these genes are not often discussed individually but rather as a family in the literature. The column Secretomics detection indicates if at least one of the genes of a given LRI has been detected in some of the six secretomics datasets we analyzed. **c**, Over-representation results for the LRI Slpi:Plscr1 across the 58 scAgeCom datasets. It is over-represented as up-regulated with age in 10 datasets and consistently in the three bone marrow datasets.



**Extended Data Fig. 4 | Processes and LRIs related to lipid metabolism are over-represented as globally dysregulated with age across tissues. a,** List of selected GO terms and KEGG pathways over-represented as up-regulated with age in several male and female tissues. **b,** List of selected LRIs over-represented as either up or down-regulated with age in several tissues. The column Secretomics detection indicates if at least one of the genes of a given LRI has been detected in

some of the six secretomics datasets we analyzed. Several LRIs display opposite aging patterns in male compared to female datasets. **c,** Comparison between the fold changes of the ligand and the receptor for the CCI from the TMS FACS (male) Brain dataset that have Apoe as ligand. Changes in Apoe expression with age (x-axis) are strongly cell-type-specific. For example, Apoe is most strongly up-regulated in microglial cells while it is down-regulated in endopymal cells.

a				b							
Category	Term	Age Regulation	Over-represented in # male tissues	Over-represented in # female tissues	LRI	Age Regulation	Over-represented in # male tissues	Over-represented in # female tissues	Associated to # aging PubMed articles	In HAGR	Secretomics detection (either Ligand or Receptor)
KEGG	ECM-receptor interaction	DOWN	18	5	Adam17:Itgb1	DOWN	8	2	68_64	FALSE_FALSE	hPDE/mBBM/hUVEC/mMSC-AT
GO	basement membrane	DOWN	17	4	Adam9:Itgb1	DOWN	7	3	11_64	FALSE_FALSE	hPDE/mBBM/hUVEC/mMSC-AT/mNeuron
GO	extracellular matrix	DOWN	17	5	Alcam:Nrp1	DOWN	4	1	6_25	FALSE_FALSE	hPDE/mBBM/hUVEC/mMSC-AT/mNeuron
GO	collagen-containing extracellular matrix	DOWN	16	5	Cdh1:Cdh1	DOWN	2	2	35_35	TRUE_TRUE	hPDE
GO	cell adhesion	DOWN	14	4	Cdh1:Itga1:Itgb1	DOWN	3	1	35_14_64	TRUE_TRUE_FALSE	hPDE/mBBM/hUVEC/mMSC-AT
KEGG	Focal adhesion	DOWN	14	7	Ceacam1:Ceacam1	DOWN	0	4	12_12	TRUE_TRUE	hPDE
GO	cell-substrate adhesion	DOWN	13	4	Col15a1:Itgb1:Itga1	DOWN	4	1	4_64_14	TRUE_FALSE_TRUE	hPDE/mBBM/hUVEC/mMSC-AT
GO	collagen binding	DOWN	13	5	Col1a1:Cd44	DOWN	3	3	173_59	TRUE_FALSE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
GO	cell junction	DOWN	10	4	Col1a2:Sdc4	DOWN	2	3	36_12	FALSE_TRUE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
GO	cell-matrix adhesion	DOWN	10	4	Col3a1:Ddr2	DOWN	3	2	14_14	TRUE_FALSE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
					Col4a1:Cd44	DOWN	4	1	46_59	FALSE_FALSE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
c											
Emitter-Receiver Cell Families		Age Regulation	Over-represented in # male tissues	Over-represented in # female tissues	Col4a2:Itgb1:Itga1	DOWN	4	2	26_64_14	FALSE_FALSE_TRUE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
connective tissue cell_connective tissue cell		DOWN	5	4	Col4a2:Itgb5	DOWN	5	1	26_8	FALSE_FALSE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
connective tissue cell_endothelial cell		DOWN	5	2	Col4a2:Sdc4	DOWN	1	4	26_12	FALSE_TRUE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
endothelial cell_endothelial cell		DOWN	8	0	Col6a2:Itgb1:Itga1	DOWN	3	1	0_64_14	FALSE_FALSE_TRUE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
endothelial cell_epithelial cell		DOWN	3	2	Fbln1:Itgb1	DOWN	4	1	7_64	FALSE_FALSE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
epithelial cell_endothelial cell		DOWN	4	3	Gcg:Dpp4	DOWN	3	3	24_47	FALSE_TRUE	
epithelial cell_epithelial cell		DOWN	5	5	Hspg2:Itgb1	DOWN	6	1	24_64	FALSE_FALSE	hPDE/mBBM/hUVEC/mMSC-AT/mNeuron
epithelial cell_leukocyte		DOWN	4	2	Mmp2:Itgb1	DOWN	6	1	172_64	FALSE_FALSE	hPDE/rCM/mBBM/hUVEC/mMSC-AT
leukocyte_epithelial cell		DOWN	3	4	Nid1:Itgb1	DOWN	6	4	4_64	FALSE_FALSE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
leukocyte_leukocyte		DOWN	2	3	Postn:Itgb1	DOWN	4	3	34_64	FALSE_FALSE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
stem cell_endothelial cell		DOWN	5	0	Sell:Cd34	DOWN	5	7	29_19	FALSE_TRUE	mBBM
stem cell_stem cell		DOWN	6	2							

**Extended Data Fig. 5 | Processes, LRIs, and cell-type families related to the extracellular matrix and adhesion processes are over-represented as globally down-regulated with age across tissues. a.** List of selected GO terms and KEGG pathways over-represented as down-regulated with age in several male and female tissues. **b.** List of selected LRIs over-represented as down-regulated

with age in several tissues. The column Secretomics detection indicates if at least one of the genes of a given LRI has been detected in some of the six secretomics datasets we analyzed. **c.** Emitter-Receiver cell-type family pairs that are mainly over-represented as down-regulated with age in several tissues.

**a**

Category	Term	Age Regulation	Over-represented in # male tissues	Over-represented in # female tissues
GO	anatomical structure morphogenesis	DOWN	17	4
GO	cell morphogenesis	DOWN	12	3
GO	regulation of cell cycle	DOWN	12	4
GO	response to growth factor	DOWN	12	3
KEGG	PI3K-Akt signaling pathway	DOWN	12	6
GO	cell morphogenesis involved in differentiation	DOWN	10	3
GO	Notch signaling pathway	DOWN	9	5
GO	animal organ morphogenesis	DOWN	8	5
GO	morphogenesis of an epithelium	DOWN	6	3
GO	regulation of growth	DOWN	4	3
GO	response to fibroblast growth factor	DOWN	4	5
GO	regulation of angiogenesis	DOWN	3	6

**b**

LRI	Age Regulation	Over-represented in # male tissues	Over-represented in # female tissues	Associated to # aging PubMed articles	In HAGR	Secretomics detection (either Ligand or Receptor)
Angpt1.Igfb1	DOWN	5	3	34_64	FALSE_FALSE	hPDE/mBBM/hUVEC/mMSC-AT
App.Notch2	DOWN	5	5	3325_34	TRUE_FALSE	hPDE/rCM/mBBM/hUVEC/mNeuron
Fgfr2.Cd44	DOWN	2	3	24_59	TRUE_FALSE	hPDE/mBBM/hUVEC/mNeuron
Gpi1.Amfr	DOWN	10	1	23_6	FALSE_FALSE	hPDE/hUVEC/mMSC-AT/mNeuron
Jag1.Notch2	DOWN	1	6	26_34	FALSE_FALSE	hPDE/mBBM/hUVEC/mNeuron
Postn.Igfb1	DOWN	4	3	34_64	FALSE_FALSE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron
Psen1.Ncstn	DOWN	7	0	1470_95	TRUE_FALSE	
Psen1.Notch2	DOWN	4	7	1470_34	TRUE_FALSE	mNeuron
Tgfb1.Igfb1	DOWN	4	0	371_64	TRUE_FALSE	hPDE/mBBM/hUVEC/mMSC-AT
Tgfb3.Igfb1	DOWN	4	2	10_64	FALSE_FALSE	hPDE/rCM/mBBM/hUVEC/mMSC-AT
Thbs1.Igfa4	DOWN	4	3	56_17	FALSE_FALSE	hPDE/rCM/mBBM/hUVEC/mNeuron
Vegfa.Igfb1	DOWN	2	3	421_64	TRUE_FALSE	hPDE/rCM/mBBM/hUVEC/mMSC-AT/mNeuron

**Extended Data Fig. 6 | Processes and LRIs related to morphogenesis, development, and angiogenesis are over-represented as globally down-regulated with age across tissues. a.** List of selected GO terms and KEGG pathways over-represented as down-regulated with age in several male and

female tissues. **b.** List of selected LRIs over-represented as down-regulated with age in several tissues. The column Secretomics detection indicates if at least one of the genes of a given LRI has been detected in some of the six secretomics datasets we analyzed.



## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

Some data were collected with the following software and packages: R(v4.1.0), rentrez (v1.2.3, R package), CellChat (v1.1.3, R package), NicheNet (v1.1.1, R package), SingleCellSignalR (v.1.10.0, R package).

Multiple datasets were also collected online from their respective websites (see Data section below).

The way each data type was collected is fully described in detail in our Methods section and/or in our github repositories: <https://github.com/CyrilLagger/scDiffCom> and <https://github.com/CyrilLagger/scAgeCom>.

#### Data analysis

Data analyses were performed with the following software/code: R(v4.1.0), scDiffCom (v0.2.4, R package), scAgeComShiny (v0.2, R package), data.table (v1.14.2, R package), biomaRt (v2.50.2, R package), rentrez (v1.2.3, R package), Seurat (v4.1.0, R package), org.Mm.eg.db (v3.14.0, R package), future (v1.25.0, R package), rrvgo (v1.6.0, R package), GOsemSim (v2.20.0, R package), ggplot2 (v3.3.6, R package), ComplexUpset (v1.3.3, R package), plotly (v4.10.0, R package), gt (v0.9.0, R package), ontoProc (v1.16.0, R package), KEGGREST (v1.34.0, R package), igraph (v1.2.11, R package), visNetwork (v2.1.0, R package).

scDiffCom and scAgeComShiny are two R packages we built for this study. They are available at <https://github.com/CyrilLagger/scDiffCom>, and <https://github.com/CyrilLagger/scAgeComShiny>, which also provide the full lists of their dependencies.

All scripts are available on github at: <https://github.com/CyrilLagger/scAgeCom> and <https://github.com/CyrilLagger/scAgeComShiny>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

### Datasets used in this study:

- scRNA-seq Tabula Muris Senis: <https://s3.console.aws.amazon.com/s3/buckets/czb-tabula-muris-senis>
- scRNA-seq Calico murine cell atlas: <https://mca.research.calicolabs.com/>
- LRIs from CellChat: available from the R package (see data collection)
- LRIs from NicheNet: available from the R package (see data collection)
- LRIs from SingleCellSignalR: available from the R package (see data collection)
- LRIs from CellPhoneDB: <https://www.cellphonedb.org/>
- LRIs from CellTalkDB: <https://github.com/ZJUFanLab/CellTalkDB>
- LRIs from connectomeDB2020: <https://asrhou.github.io/NATMI/>
- LRIs from ICELLNET: <https://github.com/soumelis-lab/ICELNET>
- general gene information: <https://mygene.info/> and <https://omnipathdb.org/>
- GO terms: Ensembl (v102) <https://useast.ensembl.org/index.html>
- KEGG: accessed via the R package KEGGREST (v1.34.0)
- gene2pubmed table: <ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2pubmed.gz>
- GenAge (build 20): <https://genomics.senescence.info/genes/index.html>
- LongevityMap (build 3): <https://genomics.senescence.info/longevity/>
- Microarray meta-analysis of Ageing Gene Expression database: [https://genomics.senescence.info/gene\\_expression/](https://genomics.senescence.info/gene_expression/)
- CellAge (build 2): Supplementary Material of <https://doi.org/10.1093/molbev/msab369>
- secretomics data for mBMM: Supplementary Material of <https://doi.org/10.1126/science.1232578>
- secretomics data for mNeuron: Supplementary Material of <https://doi.org/10.15252/embj.2020105693>
- secretomics data for mMSC-AT: Supplementary Material of <https://doi.org/10.18632/aging.202423>
- secretomics data for rCM: Supplementary Material of <https://doi.org/10.1161/circulationaha.119.044914>
- secretomics data for hUVEC: Supplementary Material of <https://doi.org/10.1016/j.ajpath.2019.10.007>
- secretomics data for hPDE: Supplementary Material of <https://doi.org/10.1002/pmic.202100320>

### Datasets created in this study:

- scDiffCom LRIs: available from the R package scDiffCom (<https://github.com/CyrilLagger/scDiffCom/tree/master/data>)
- aging and sex scAgeCom results: <http://doi.org/10.6084/m9.figshare.17074964>
- data to run the app scAgeComShiny and [scagecom.org](https://scagecom.org/): [https://figshare.com/articles/dataset/scAgeComShiny\\_data/17075375](https://figshare.com/articles/dataset/scAgeComShiny_data/17075375)

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

### Reporting on sex and gender

Our analysis studied murine age-related changes in intercellular communication from public scRNA-seq datasets that contain both male and female mouse samples. Therefore, we performed first an aging-differential analysis on male and female samples separately and second a sex-differential analysis on young and old samples separately. This allowed us to create both an age-related database and a sex-related database of intercellular communication changes that are publicly available (see data availability). As aging is the main focus of this study, we mainly focused on interpreting our age-related results but we also used our sex-related results to reveal sex-specific age-related changes in extracellular signalling.

### Population characteristics

N/A

### Recruitment

N/A

### Ethics oversight

N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used to predetermine sample size. Sample size selection was informed and restricted by the availability of the public scRNA-seq data used in this study (ref 41,42).
Data exclusions	Few groups of cells were excluded from the original scRNA-seq datasets to avoid biases as explained in the Methods.
Replication	Experimental replication was not performed in this study as all data collected were public and no new experimental data have been generated. However, some of the results in our atlas have been obtained from several independent public datasets (TMS FACS, TMS Droplet and the Calico murine aging cell atlas) allowing to compare them.
Randomization	We controlled for sex in our aging analysis by treating male and female datasets separately. Likewise, we control for age in the sex analysis by treating young and old datasets separately.
Blinding	Blinding was not performed in this study as all data collected were public and the authors performing the analyses had already worked on most of the data before starting this study.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging