

scAgeCom: a murine atlas of age-related changes in intercellular communication inferred with the package scDiffCom

Cyril Lager ^{1,*}, Eugen Ursu ^{2,*}, Anaïs Equey ³, Roberto A. Avelar ¹, Angela O. Pisco ⁴, Robi Tacutu ², João Pedro de Magalhães ^{1,#}

1. Integrative Genomics of Ageing Group, Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool, UK

2. Systems Biology of Aging Group, Institute of Biochemistry of the Romanian Academy, Bucharest, Romania

3. Department of Medicine, Karolinska Institutet, Stockholm, Sweden

4. Chan Zuckerberg Biohub, San Francisco, California, USA

*These authors contributed equally

#Corresponding author's email: jp@senescence.info

Abstract

Dysregulation of intercellular communication is a well-established hallmark of aging. To better understand how this process contributes to the aging phenotype, we built scAgeCom, a comprehensive atlas presenting how cell-type to cell-type interactions vary with age in 23 mouse tissues. We first created an R package, scDiffCom, designed to perform differential intercellular communication analysis between two conditions of interest in any mouse or human single-cell RNA-seq dataset. The package relies on its own list of curated ligand-receptor interactions compiled from seven established studies. We applied this tool to single-cell transcriptomics data from the Tabula Muris Senis consortium and the Calico murine aging cell atlas. All the results can be accessed online, using a user-friendly, interactive web application (<https://scagecom.org>). The most widespread changes we observed include upregulation of immune system processes, inflammation and lipid metabolism, and downregulation of extracellular matrix organization, growth, development and angiogenesis. More specific interpretations are also provided.

Introduction

Aging remains a poorly understood biological process despite affecting most organisms ¹. One of the difficult aspects to model is how the dynamics of OMICs and tissue homeostasis influence each other throughout the lifespan. To gain new 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 ^{2,3} and has recently been proposed as one of the causes leading to the cell-to-cell stochasticity arising with age ⁴. Well-known communication deregulations include inflammaging (a chronic low-grade age-associated inflammation) ⁵, impaired immune surveillance ⁶, increase in senescence-associated secretory phenotype (SASP) ⁷, altered communication between stem cells and their niche ^{8,9}, remodeling of the extracellular matrix ^{10,11} and changes in endocrine and neuronal communication ¹². 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 ^{13,14}, the use of anti-inflammatory compounds ¹⁵⁻¹⁷, heterochronic tissue transplants and heterochronic parabiosis ¹⁸⁻²⁰.

Direct measurement of intercellular communication is complicated and usually depends on the type of mediators considered, such as surface receptors, soluble factors, extracellular vesicles ²¹ or even mitochondria ²². However, recent studies have shown that specific aspects of ICC can be inferred from single-cell gene expression data ^{23,24}. Following the pioneer study that drafted the first comprehensive database of ligand-receptor interactions (LRIs) ²⁵, and based on the development of statistical tools dedicated to building cell-type to cell-type communication networks ²⁶⁻³³, it is now becoming standard to perform intercellular communication analyses alongside the workflow of single-cell transcriptomic studies. This notably includes recent aging articles on the mouse brain ³⁴, on the mouse mammary gland ²⁹, on several rat tissues ³⁵, on the primate cardiopulmonary system ³⁶ and on human skin fibroblasts ³⁷. 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 (e.g., young/old, 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, and 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 are due to noise or to a true biological effect.

To alleviate such limitations, we built a new statistical framework specifically designed to perform differential analysis in intercellular communication. Our resulting R package, scDiffCom, can be applied to any human or mouse scRNA-seq dataset to analyze changes in ICC between two given conditions in a given tissue. scDiffCom includes a collection of around 5000 curated ligand-receptor interactions that we have retrieved from seven publicly available resources ²⁶⁻³². 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 consortium³⁸ and the Calico murine aging cell atlas³⁹ to create scAgeCom, a large-scale atlas of age-related intercellular communication 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. 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 intercellular communication, depicting, 1) a widespread upregulation of immune system processes, inflammation and lipid metabolism, and 2) a downregulation of extracellular matrix organization, growth, development and angiogenesis. Despite significant differences across sex and experimental techniques, we were able to predict 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 believe that scAgeCom contains a large amount of generated data, waiting to be interpreted, which might provide the community with new therapeutic targets and new hypotheses regarding the relationship between aging and intercellular communication.

Results

Collection of ligand-receptor interactions from existing databases

As with other methods analyzing intercellular communication from scRNA-seq data, our approach first relied on the collection of ligand-receptor interactions (LRIs). In order to maximize the variety of interaction types, we retrieved LRIs from seven publicly available resources: CellChat²⁶, CellPhoneDB²⁷, CellTalkDB²⁸, NATMI/connectomeDB2020²⁹, ICELLNET³⁰, NicheNet³¹, and SingleCellSignalR³² (*Methods - Fetching ligand-receptor interactions*). Several processing steps were then necessary to obtain consistent human and mouse collections of curated LRIs (*Methods - Processing ligand-receptor interactions*). This means retaining only curated interactions, converting human genes to mouse orthologs for the human-only resources, and including both simple and complex LRIs. Simple LRIs are interactions involving a single-gene ligand with a single-gene receptor, e.g., *Apoe:Ldlr*. On the other hand, a complex interaction involves heteromeric ligands or receptors, e.g., *Col3a1:Itgal-Itgb1*.

Our approach resulted in 4741 mouse LRIs (simple: 3627, complex: 1114) and 4787 human LRIs (simple: 3650, complex: 1137) directly accessible from our R package scDiffCom (see also Supplemental Data 1). Fig.1a shows how those LRIs are distributed according to their database of origin. Whenever possible, we also included in Supplemental Data 1 the sources

used by each of the seven databases to curate their interactions, including references to PubMed identifiers, FANTOM5²⁵, HPMR⁴⁰, HPRD⁴¹, IUPHAR⁴², Reactome⁴³ or KEGG⁴⁴.

Functional annotation of ligand-receptor interactions

We annotated all LRIs with a standardized and consistent framework (*Methods - Annotating ligand-receptor interactions with GO terms and KEGG pathways*). We first associated gene ontology (GO) terms⁴⁵ 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 significant number of empty intersections (as most genes are annotated with specific GO terms and not all parent terms). Instead, as shown in Fig.1b, 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. Since 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. In addition to GO terms, we added KEGG pathways⁴⁴ to each LRI if all genes present in the interaction were part of the pathway. All annotations are directly accessible from scDiffCom.

Differential cell-cell communication analysis with scDiffCom

We designed a bioinformatics method, available within the R package scDiffCom, to detect cell-type to cell-type communication patterns that significantly change between two conditions in a given scRNA-seq dataset (Fig.2). Prior to the analysis, the dataset must be formatted as an R Seurat object⁴⁶⁻⁴⁸ 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 a cell-type pair, based on the ligand-receptor interactions described in the previous section. 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 expression of the genes in the respective cells (*Methods - CCI score based on the geometric mean*). A statistical procedure comprising two tests is finally performed to establish how likely each CCI is to correspond to an actual biological interaction and to be significantly differentially expressed between the two given conditions.

The first test consists in assessing the biological relevance of each CCI (in each condition independently) based on combining previously published approaches (*Methods - CCI detection and differential analyses*). To be considered detected, a CCI has to: 1) not be lowly expressed (based on the number and percentage of cells expressing each gene), 2) be *specific*, as originally defined by the authors of CellPhoneDB^{27,49} (this relies on a permutation test explained below), 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 ligand-receptor interaction 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 (*Methods - CCI detection and differential analyses*). This consists of randomly exchanging the condition label of each cell to see whether the test statistic, 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, i.e. they make no assumptions regarding the distributions of the underlying variables⁵⁰. 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" (*Methods - CCI classification*). It can either be up-regulated (UP), down-regulated (DOWN), stable (FLAT), or correspond to a non-significant change (NSC). As such, 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 for each condition.

We note that the most computationally demanding part of this workflow is to perform the permutation tests and therefore special care was taken to optimize them. Moreover, scDiffCom can easily be run in parallel. A toy model analysis (1000 permutations on a dataset of 1000 cells and 5 cell types) takes a couple of minutes if run sequentially on a single-core computer. A more realistic example (10000 permutations on 3107 cells and 16 cell types) was measured to take around 9 minutes when run in parallel on 30 cores.

Over-representation analysis and visualization

Every single detected CCI returned by our analysis corresponds to a communication signal whose change (or absence of change) between the two conditions can be interpreted on its own. However, 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 (*Methods - Fisher's exact test to find over-represented signals*). ORA measures the statistical association between CCI classes (UP, DOWN or FLAT) and CCI features of interest, e.g., gene, cell-type or functional annotation. For instance, it can evaluate if a given ligand-receptor interaction takes part in up-regulated CCIs more than expected by chance. Results can be sorted according to an over-representation score that combines the odd ratio (OR) and adjusted p-value returned by the test: $ORA\ score = \log_2 OR \cdot (-\log_{10} adj.\ pval)$.

ORA is performed on the following categories: LRIs, ligands, receptors, emitter-receiver cell-type pairs, emitter cell types, receiver cell types, GO terms and KEGG pathways. It is useful to analyze ligands and receptors alone to see if some ligands take part in various regulated CCIs involving different receptors (or vice-versa). The same logic is valid for the cell types. Moreover, scDiffCom offers the opportunity to perform ORA on additional user-defined attributes.

Finally, scDiffCom provides two tools to produce images of the over-represented results (*Methods - Visualization Tools*). First, the function PlotORA displays the top over-represented signals of a given category and regulation with their odds ratios and adjusted p-values. Second, the function BuildNetwork creates and displays a network representation of the over-represented cell-type pairs, as will be illustrated below. Other types of data visualization can be created from direct calls to standard R graphic packages and are therefore not implemented in scDiffCom.

Preparation and annotation of scRNA-seq aging datasets

We used the developed package to create a comprehensive murine atlas of age-related intercellular communication changes by applying scDiffCom to scRNA-seq datasets obtained from the Tabula Muris Senis (TMS) consortium³⁸ and the Calico murine aging cell atlas³⁹ (*Methods - Fetching and preparing scRNA-seq datasets*). These studies offer high-quality cell-type annotations: following standard cell clustering techniques⁵¹, TMS relied on experts to manually curate their cell-type classification, whereas Calico assigned cell types to their cells based on a semi-supervised neural network using the first Tabula Muris study⁵² as training data.

After data pre-processing, we retained 58 distinct datasets corresponding to 23 organs distributed across 5 categories: *TMS FACS (male)*, *TMS FACS (female)*, *TMS Droplet (male)*, *TMS Droplet (female)* and *Calico Droplet (male)* (Fig.3). *FACS* and *Droplet* refer to the two experimental techniques used by the original studies. We kept male and female samples separate to avoid possible confounding effects. Cells from different time points were classified as either young or old, as the two conditions on which scDiffCom performs the differential analysis. To integrate the 5 types of datasets and facilitate comparisons, we standardized the cell-type names across datasets, regrouped very specific categories into broader ones, and annotated them with cell-type families (*Methods - Cell-type characterization*).

The dataset preparation revealed that most tissues do not share the same cell types across the five experimental and sex categories. As shown in Fig.3, the number of cell types in the lung varies from 12 to 18. Moreover, some important cell types were not captured at all due to technical limitations (e.g., adipocytes in adipose tissues). As such, one must keep in mind that our results can only convey a partial representation of what is expected to happen *in vivo*.

scAgeCom: a mouse aging atlas of intercellular communication

We applied scDiffCom to each of the 58 aforementioned datasets, performing differential analysis between young and old cells using 10000 permutations and default parameters (as defined in *Methods*). To facilitate access to this large amount of data, we created a Shiny application (*Method - Building and deploying scAgeComShiny*). This website, available at <https://scagecom.org>, provides both a tissue-specific section, displaying results for each dataset independently and a global section summarizing the dominant signals shared across datasets.

Fig.4a-c illustrate the interactive tables and plots provided to explore detected CCIs and their age-regulation in each dataset. Several filtering options are available to restrict interactions to a subset of cell types, LRIs or functional annotations. Results of the over-representation analysis are available as graphs and tables for cell-type-centric, gene-centric and function-centric investigations (Fig.4d-f).

The high number of datasets analyzed allowed us to perform a global analysis to investigate age-related communication changes shared across multiple tissues. We used the signals returned by the over-representation analysis in each dataset. For each annotation, e.g., the GO term *T cell differentiation* or the LRI *B2m: Cd3g*, we counted the number of datasets in which it was over-represented (Fig.5). This allowed us to create tables with the keywords that were overrepresented in the largest number of tissues.

Aging dysregulates multiple aspects of intercellular communication

Over the 58 datasets considered, we detected 382522 CCIs (5% UP, 13% DOWN, 56% FLAT, 26% NSC), corresponding to an average of 98 detected LRIs (SD = 77) between any cell-type pair. The nature and age-regulation of the detected CCIs strongly vary across datasets, sex and experimental techniques, as shown in Fig. 6. The tissues from *TMS FACS (male)* clearly show a larger fraction of down-regulated CCIs compared to all other datasets including those from *TMS FACS (female)*. Unsurprisingly, we also observe more variability and noise (namely more NSC CCIs) in *FACS* compared to *Droplet*. Those observations are partially explained in the Discussion below.

The global section of the atlas reveals several broad mechanisms dysregulated with age. First, consistently with the literature, we observe a major up-regulation of inflammatory and immune system processes. GO terms over-represented in more than half of the tissues include *immune response*, *immune system process*, *defense response*, *T cell differentiation*, *cytokine/chemokine-mediated signaling pathway*, *lymphocyte migration*, *positive regulation of adaptive immune response*, *regulation of B cell activation*, *regulation of viral life cycle/process* and KEGG pathways such as *antigen processing and presentation*, *cytokine-cytokine receptor interaction*, *Epstein-Barr virus infection*. Analysis of cell-type families reveals that the interaction *leukocyte-leukocyte* is up-regulated in 16 tissues and that the interactions from *endothelial*, *connective* and *stem* cells towards *leukocytes* are each up-regulated in at least 5 tissues. From the gene perspective, the interactions that are overrepresented in the highest number of tissues (9 to 15 out of 23 tissues) include *B2m: Cd3g*, *B2m: Cd3d*, *Tnfsf12: Tnfrsf12a*, *H2-D1/K1/Q6: Cd8b1*, *Mif: Cd74*, *Hmgb1: Thbd*, and *Ccl5* interacting with different chemokine receptors such as *Ccr12*, *Ccr1* and *Ccr5* in 6, 4 and 4 tissues, respectively.

A second up-regulated process is related to lipid metabolism, with GO terms like *lipid metabolic process*, *cellular lipid metabolic process*, *lipid catabolic process*, *lipoprotein metabolic process*, *regulation of lipid localization*, *fatty acid biosynthetic process*. *Apoe* is over-represented as an up-regulated ligand in 16 tissues, mostly via its interactions with *Sdc4*, *Lrp1* and *Ldlr*. However, this pattern is mainly found in *TMS FACS (male)* and happens to be

reversed in four *TMS FACS (female)* tissues. Along the same line, *App:Lrp10*, is over-represented among down-regulated signals in 8 tissues from *TMS FACS (male)* but up-regulated in 4 tissues from *TMS FACS (female)* and 3 tissues from *TMS Droplet (female)*.

The most over-represented processes among signals down-regulated with age are related to the extracellular matrix (ECM) and adhesion. This includes the GO terms *biological adhesion*, *cell adhesion*, *extracellular matrix organization*, *extracellular structure organization*, *cell junction assembly*, *cell-matrix adhesion* and the KEGG pathways *ECM-receptor interaction*, *focal adhesion*, *adherens junction*. The main corresponding LRIs involve collagens, fibronectins and laminins interacting with integrins, mostly between connective tissue cells, epithelial cells and endothelial cells.

In addition to the changes undergone by the ECM, interactions related to growth, development, survival, differentiation and angiogenesis are also significantly down-regulated with age. Relevant GO terms include *anatomical structure morphogenesis/development*, *cell morphogenesis*, *developmental process and growth*, *response to growth factor*, *regulation of cell cycle*, *Notch signaling pathway*, *regulation of angiogenesis* and KEGG pathways include *PI3K-Akt signaling pathway*, *Wnt signalling pathway*. Genes such as *Fgf1*, *Wnt5a*, *Bmp2*, *Angpt1*, *Tgfb1*, *Tgfb3*, *Vegfa*, *Bmp6*, *Notch2*, *Fgfr1*, *Fgfr2*, *Tgfbr2*, *Egfr*, *Gpi1* are over-represented in down-regulated CCIs in 5 to 13 tissues. Cell-type pair over-representation also indicates down-regulation of communication among stem cells (6 tissues) and from stem cells towards endothelial cells (5 tissues).

The results reported so far only represent a fraction of the global signals present in our atlas. Tissue-specific results provide even more leads to be investigated, including age-regulated CCIs specific to an organ or cell type and therefore not showing up as over-represented. For instance, we noticed the down-regulation of multiple communication links towards intestinal crypt cells and *Lama2* in the large intestine; the up-regulation of *Sost* (sclerostin) and communication links towards podocytes in the kidney; the down-regulation of *Notch signalling* in the bone marrow; and the down-regulation of *Hgf* and *Tgfb1* in hepatic sinusoidal endothelial cells. Unfortunately, investigating all these signals is outside the scope of this article; however, we hope that scAgeCom will encourage other researchers to access our database for this purpose.

Discussion

In this work, we present a package to perform differential intercellular communication analysis as well as a comprehensive database of age-related mouse cell-cell interactions. From a technical point of view, our results illustrate the importance of performing a proper statistical analysis when comparing intercellular signals extracted from scRNA-seq data. Variability and noise were indeed responsible for the classification of 26% of the detected CCIs as NSC interactions (*non-significant change*), 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, as in some previous studies, 82.5% of those NSC signals (i.e., 82256 CCIs) would have been

falsely considered to be age-regulated. Moreover, such an approach would have missed all CCIs detected in both conditions but nevertheless showing a significant change with age, namely 16678 interactions in our case.

Fig.6 shows that there are significant disparities in the results depending on experimental techniques and sex. Previous comparisons of single-cell sequencing techniques⁵³ have claimed that Drop-seq, by using unique molecular identifiers (UMI), is less subject to amplification noise than Smart-seq2, potentially explaining why we observe less 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, as previously shown in Fig.3 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 the latter dataset (22540 vs. 11372) and in a different distribution of the percentage of UP/DOWN/FLAT/NCS CCIs (Fig.6). We also emphasize that the pronounced down-regulation observed in *TMS FACS (male)* datasets (Fig.6) was reported in a previous study performed on TMS datasets⁵⁴.

We now provide more specific interpretations of the main biological results reported above. Regarding the increase of immune response processes with age, we observed that the ligand *B2m* (β 2-microglobulin) is over-represented among up-regulated CCIs in 17 tissues. *B2m* has already been recognized as a gene consistently overexpressed with age⁵⁵ and as a pro-aging circulating factor whose elevated level negatively affects cognitive functions and neurogenesis in the mouse hippocampus⁵⁶. 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 on 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 suggested⁵⁷.

Regarding the increase in inflammation with age, our results show a global over-representation of the terms *cytokine-* and *chemokine-mediated pathways*, and of the ligands *Ccl5*, *Mif* and *Hmgbl*. Those ligands are known Senescence-Associated Secretory Phenotype (SASP) factors, but we cannot confirm from our analysis whether they are really emitted by senescent cells. Follow-up studies, such as for example a cross-analysis between scAgeCom and the SASP atlas⁵⁸, might allow us to explore such senescence-related hypotheses more precisely. Here, we only mention that the two most over-represented up-regulated LRIs that we reported earlier, *Mif:Cd74* and *Hmgbl: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^{59,60}. Along the same line, the high-mobility group box chromosomal protein 1 (Hmgbl) has proinflammatory effects when binding to RAGE/Ager⁶¹ but its sequestration by

thrombomodulin acts as an anti-inflammatory mechanism ⁶². According to our results, *Hmgbl:Thbd* is over-represented as up-regulated in 9 tissues and *Hmgbl:Ager* in 3 tissues, pointing towards a global over-emission of *Hmgbl* with age that tissues might try to compensate by over-expressing *Thbd*.

Changes in lipid metabolism with age are known to have an important impact on the lifespan and age-related diseases ⁶³. Our most intriguing results concern the widespread deregulation of *Apoe*, its receptor and *App*. We indeed observed a general over-emission of *Apoe* and an under-emission of *App* in most tissues at least in *TMS FACS (male)*. In the brain, those genes are known to play a role in the pathology of Alzheimer's disease (AD). *App* is the precursor of amyloid beta (A β) peptides ⁶⁴, *Apoe* is a known regulator of A β clearance ⁶⁵ and interactions between *App* and *Apoe* receptors influence A β metabolism and toxicity ^{66,67}. Much less research has been performed on the function of *App* outside the central nervous system ⁶⁸, but some studies point towards its role in several pathologies such as in obesity ^{69,70}, in the skin ⁷¹, in the intestine ⁷², where it is modulated by diet ⁷³, and in muscles ⁷⁴, particularly at neuromuscular junctions ⁷⁵. 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 down-regulated processes reported in our results, namely changes in the ECM 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 ⁷⁶. The multiple functions of these proteins and their deregulation that we observe across multiple organs (most notably when they interact with collagen) suggest they might play key roles in the structural decline of tissues with age. They are also consistently down-regulated in signaling involving stem cells, potentially impacting their maintenance and homeostasis, as suggested in previous articles ¹⁰.

A subpart of the development/growth/proliferation pathway concerns the global down-regulation of angiogenesis and qualitative changes occurring in vessels. Several studies have previously reported a decline with age in capillary density and in the formation of new blood vessels ^{77,78}, even leading some authors to postulate an “angiogenesis hypothesis of aging” ⁷⁹. Our analysis reveals that the pair *Gpil:Amfr*, mainly detected from or to endothelial cells, is over-represented among down-regulated CCIs in 11 tissues. Glucose-6-phosphate isomerase functions as an autocrine motility factor that stimulates endothelial cell motility ⁸⁰. This LRI could be an important regulator of microvascular aging and its therapeutic potential seems worth investigating. Moreover, we found down-regulation of angiopoietin-1 (*Angpt1*) in 7 tissues, which could lead to the formation of leaky vessels, as vessel stability relies on the balance of *Angpt1* and *Angpt2* ⁸¹.

Finally, we conclude with some of the limitations of using gene expression of ligands and receptors to infer intercellular communication 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 signalling, even though a fraction of them might be produced for intracellular processes. Second, it is not always clear how some interactions (e.g., *App:Lrp10*) are shared between cells, and which fraction of them are autocrine rather than paracrine. Third, by only looking at ligands and receptors, our method does not assess changes in other players acting in downstream signalling. Fourth, we do not consider all possible types of intercellular mediators and we do not investigate inter-tissue interactions such as endocrine signals. Despite these limitations, we hope that our atlas will be useful for the community and lead to new hypotheses on intercellular communication and aging to be further tested.

Methods

Fetching ligand-receptor interactions

We downloaded LRIs from the most recent versions (as of March 21, 2021) of seven publicly available databases. Data from CellChat, NicheNet, and SingleCellSignalR were directly accessed from their associated R packages. Data from CellPhoneDB, CellTalkDB, connectomeDB2020 and ICELLNET were fetched online 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* ³³, but we did not consider it further as all curated LRIs present in it were derived from the seven others mentioned above.

Processing ligand-receptor interactions

We analyzed the documentation and annotations of each resource to only keep 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 HGNC approved ⁸² and, when mouse data were available, MGI approved ⁸³. If mouse LRIs were not provided (namely for CellPhoneDB, connectomeDB2020, ICELLNET, NicheNet and SingleCellSignalR), we converted human LRIs to their mouse equivalent by retrieving high confidence orthology information from Ensembl version 102 ⁸⁴ accessed through the R package *biomaRt* ⁸⁵.

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

As some of the resources only provide 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, given 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 about 200 of them that we considered miss-curated, e.g., *Mapk1:Fgfr2*, *Calm1:Adcy8*, *Gnas:Adcy1*, *Hsp90aa1:Cftr*. To do so, we first annotated each gene with descriptions from *MyGene.Info* ^{86,87} and categories from *OmniPath* ⁸⁸. We then identified genes that seemed unlikely to participate in intercellular communication. Finally, we explored the interactions involving those genes and evaluated the evidence supporting their existence. Some of the removed LRIs relied on publications describing intracellular interactions that had been misinterpreted as intercellular, e.g., Loo et al. ⁸⁹ is presented as evidence for *Hsp90aa1:Cftr*, but the paper does not mention intercellular communication.

Annotating ligand-receptor interactions with GO terms and KEGG pathways

LRIs were annotated with GO terms by following the graph-based approach in Fig.1b. GO terms associated with each gene were retrieved from Ensembl version 102 ⁸⁴ accessed through the R package *biomaRt* ⁸⁵. For a given LRI, we used the R package *ontoProc* ⁹⁰ 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, e.g., 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, i.e., the intersection of those graphs at node-level.

To annotate LRIs with KEGG pathways, we used the R package *KEGGREST* ⁹¹ 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).

CCI score based on the geometric mean

For each cell-cell interaction in the form (emitter cell type, receiver cell type, ligand(s), receptor(s)), a score φ 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 intercellular communication from scRNA-seq data ^{23,24}, 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 towards 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 it would have been the

case with the arithmetic mean. Along the same line, although transcript counts or UMIs only give an indirect representation of protein levels⁹², molecular interactions are usually modeled by the law of mass action⁹³ 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,

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

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 φ_A and φ_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 Φ_A , 2) same for condition B , returning $\tilde{\varphi}_B^k$ to form the null distribution of Φ_B , and 3) keeping the original cell-type labels but shuffling the A and B condition labels, and returning the random score difference $\delta^k = \tilde{\varphi}_B^k - \tilde{\varphi}_A^k$ to form the null distribution of the random variable Δ . After iterating, the true values φ_A , φ_B and $\delta = \varphi_B - \varphi_A$ are compared to the three null distributions in order 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 procedure⁹⁴.

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 ($p_j^{adj.} \leq 0.05$), and 3) its score is among the top 80% of all the *specific* CCI scores of both conditions, namely $\varphi_j \geq q_{\varphi_{A,B}}(20)$, where $q_{\varphi_{A,B}}(x)$ is the x-th percentile of the scores. A CCI is called *differentially expressed* if $p_{DE}^{adj.} \leq 0.05$ and $|\log(\varphi_B/\varphi_A)| \geq \log(1.5)$.

CCI classification

We only kept CCIs *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 $\log fc \geq \log(1.5)$, 2) *DOWN* when $p_{DE}^{adj.} \leq 0.05$ and $\log fc \leq -\log(1.5)$, 3) *FLAT* when $|\log fc| < \log(1.5)$ and 4) *NSC* (non-significant change) when $p_{DE}^{adj.} > 0.05$ and $|\log fc| \geq \log(1.5)$.

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, i.e., CCIs that seem to appear or disappear because they fluctuate around the detection threshold, but which are in reality not differentially expressed. Using the aging datasets as benchmarking data, we considered all possible outcomes (Supplemental Data 2) and only noticed a marginal number of seemingly contradictory cases between the two tests, such as disappearing CCIs with positive log fold change. 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.

Fisher's exact test to find over-represented signals

Over-representation analysis (ORA) is used to evaluate frequent patterns in categorical data, for example, to find if a particular feature of CCIs, e.g., the annotation with the GO term *cell differentiation*, is more frequent in *up-regulated* CCIs compared to all other CCIs. This statistical association is measured by compiling the corresponding 2x2 contingency table (*up-regulated/not up-regulated* vs *annotated/not annotated*) and applying Fisher's exact test. We performed this procedure for every CCI feature (all GO Terms, KEGG pathways, LRIs, ligands, receptors, cell types) and classes (UP, DOWN, FLAT). It returns an odds ratio (OR) and a p-value adjusted for multiple testing according to the Benjamini-Hochberg procedure⁹⁴. In some instances (e.g., pattern ranking, 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 (π -value) used in differential gene expression analysis and GSEA⁹⁵: $ORA\ score = \log_2 OR \cdot (-\log_{10} pval)$.

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`⁹⁶. `scDiffCom::BuildNetwork` shows on a summary graph the over-represented cell types and cell-type pairs. It relies on the R package `igraph`⁹⁷ for internal computations and on the R package `visNetwork`⁹⁸ for the interactive rendering.

Fetching and preparing scRNA-seq datasets

We downloaded the latest version (as of March 21, 2021) of Tabula Muris Senis from the Amazon S3 `czb-tabula-muris-senis` repository and the Calico dataset from the calicolabs website. They had been preprocessed and annotated with the python toolkit `Scanpy`⁹⁹ prior to our work, and we converted the resulting `h5ad` files to R Seurat objects.

As stated in the original TMS article³⁸, 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 the mice age, TMS provided multiple time points that needed to be grouped into *young* and *old* categories. We removed 1-month-old and 30-month-old cells to avoid bias due to developmental or longevity-related processes. Therefore, we compared 3-month-old 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 (e.g., *TMS Droplet Fat* only contained 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 only contained two cell types (*macrophage* and *microglial cell*) and that merging the datasets allowed us to infer interactions with the other parts of the brain. We verified that this did not significantly 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 standards¹⁰⁰, e.g., *atrial myocyte* was renamed as *regular atrial cardiac myocyte*. We also regrouped some specialized cell clusters, e.g., CD4+ and CD8+ T cells, in order 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 worth keeping as distinct categories. Finally, we classified the cell types in 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. The list of all cell types with their new names and family annotations is available in Supplemental Data 3.

Building and deploying scAgeComShiny

We used the R package *golem*¹⁰¹ to build the Shiny App *scAgeComShiny* which contains all *scAgeCom* results. Interactive scatter plots were built with *plotly*¹⁰², which was also used to display the GO Terms treemaps. Those were internally computed with the R package *rrvgo*¹⁰³, according to the original method from REVIGO¹⁰⁴. 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.

Acknowledgements

We thank Jacob Kimmel for helpful information regarding the Calico murine cell atlas, Koki Tsuyuzaki for general discussion regarding intercellular communication at the early stage of this project, Zoya Abidi for her support to deploy the Shiny app and Ulysse Castet for his useful suggestions on the manuscript. Final figures were created with BioRender.com. CL 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 is 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 JPM. RT and EU are supported by the National Authority for Scientific Research and Innovation, and by the Ministry of European Funds, Romania, through the Competitiveness Operational Programme 2014-2020, POC-A.1-A.1.1.4-E-2015 [Grant number: 40/02.09.2016, ID: P_37_778, to RT].

Author contributions

CL: conceived the project, developed *scDiffCom/scAgeCom/scAgeComShiny*, compiled the ligand-receptor database, preprocessed scRNA-seq datasets, interpreted the results, wrote the manuscript

EU: developed *scDiffCom/scAgeCom/scAgeComShiny*, curated the ligand-receptor database, interpreted the results, wrote the manuscript

AE: contributed to *scDiffCom/scAgeComShiny*, annotated the ligand-receptor database, annotated the TMS/Calico cell types, interpreted the results, wrote the manuscript

RA: preprocessed scRNA-seq datasets, interpreted the results, revised the manuscript

AP: annotated the TMS/Calico cell types

RT: revised the manuscript, supervised the project

JPM: revised the manuscript, supervised the project

Competing interests

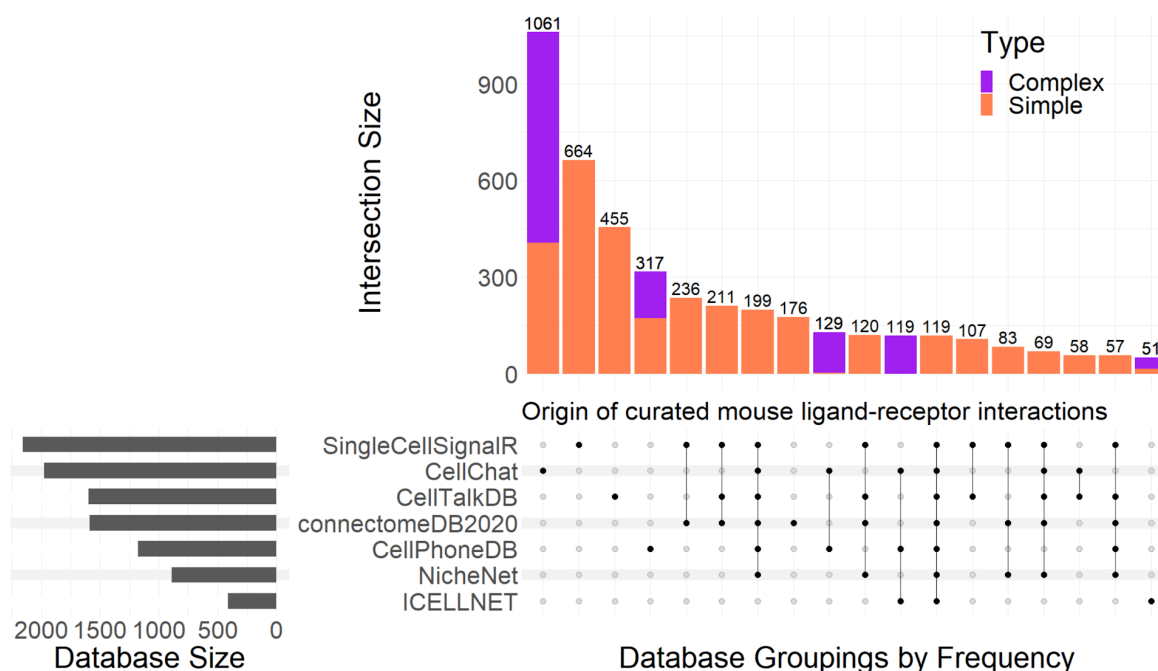
The authors declare no competing interests.

Code availability

- R package scDiffCom: <https://github.com/CyrilLagger/scDiffCom>
- R scripts for the aging analysis: <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>
- scAgeCom website: <http://scagecom.org/>

Fig.1: Distribution and annotation of ligand-receptor interactions

a



b

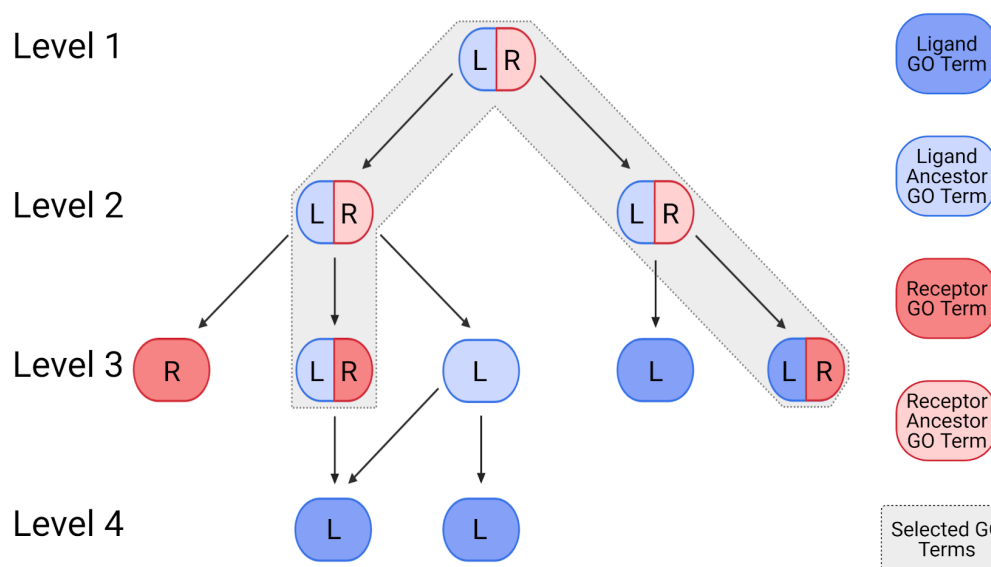


Fig.1: Repartition and annotation of ligand-receptor interactions in scDiffCom. a) UpSet plot showing the distribution of the 4741 curated mouse LRIs in terms of the seven databases of origin. Each column of the UpSet plot corresponds to one segment in an equivalent Venn diagram. Simplex/Complex distinguishes between LRIs composed of a single ligand gene and a single receptor gene from LRIs containing a heteromeric ligand or a heteromeric receptor. b) Representation of the method used to assign GO terms to an LRI. It consists of

intersecting the nodes of the two subgraphs made from the GO terms and ancestor GO terms of the ligand (blue) and receptor (red).

Fig.2: scDiffCom workflow

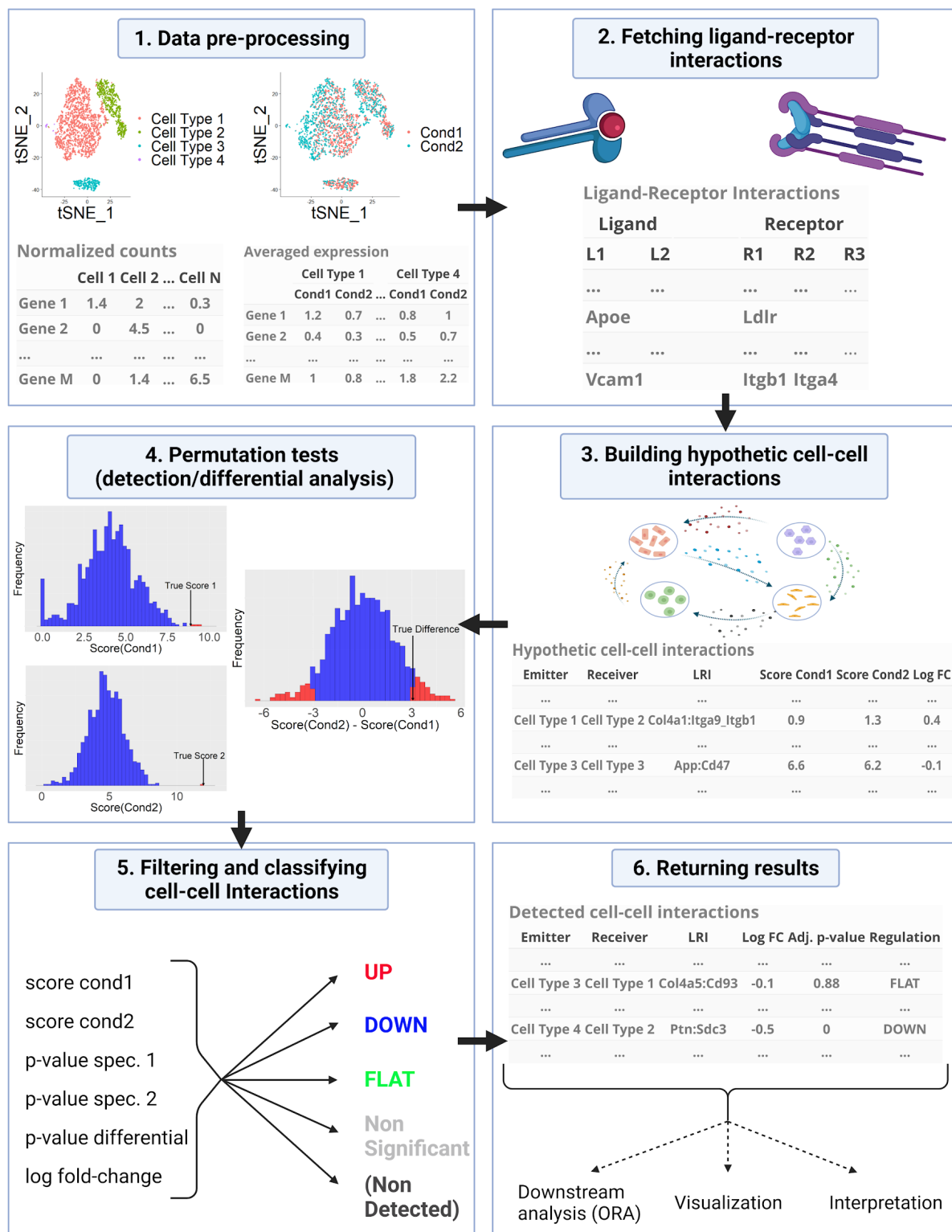


Fig.2: Workflow summary of scDiffCom. Read counts/UMIs from the single-cell dataset are aggregated by cell types and conditions (1). Genes are then joined with our database of ligand-receptor interactions (2) to build all the potential cell-cell interactions that can occur between cell types (3). Statistical 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).

Fig.3 The 58 aging scRNA-seq datasets used to build scAgeCom

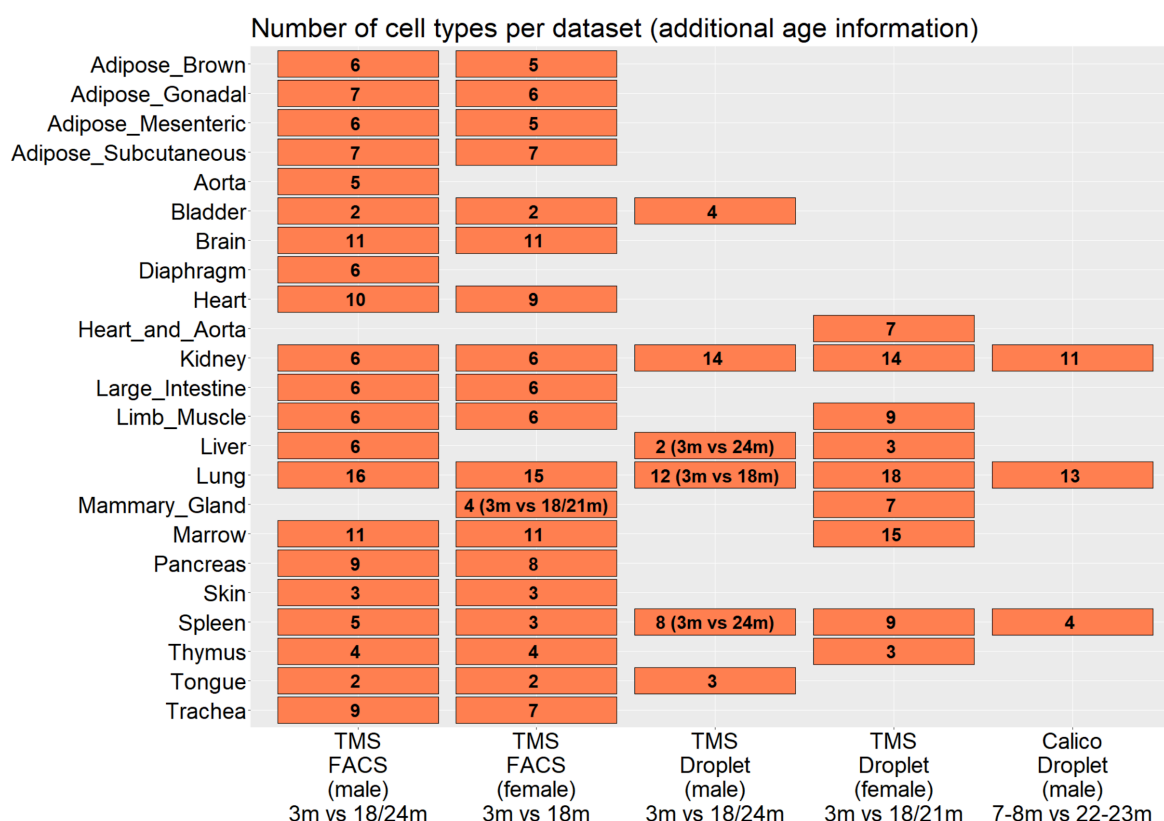


Fig.3: The 58 aging scRNA-seq datasets used to build scAgeCom. Repartition of the datasets according to their experimental/sex category (x-axis) and the tissue concerned (y-axis). Each rectangle represents a dataset and contains the number of detected cell types. The x-axis labels show information about the age of the mouse samples being compared.

scores, allowing for better visualization of the initial and final values that yield the absolute CCI score changes with age. For example, it can discriminate between CCIs undergoing similar change with age but having different young and old scores. This plot also highlights stable (FLAT) CCI with high scores, which may play an important role in intercellular communication regardless of age. c) The Ligand vs Receptor FC plot captures the log fold-change of the ligand and the log fold-change of the receptor. Our approach combines the ligand and receptor expressions into a single score, thereby losing some information, which this graph recovers *a posteriori*. It illustrates if the change with age in a given CCI is driven by the ligand or the receptor. d) Network representing which emitter cell types, receiver cell types and cell-type pairs are over-represented as either up-regulated, down-regulated, stable, or both up and down-regulated with age. The latter scenario is realized when both up and down-regulated LRIs are significantly present. e) GO Treemap is obtained by merging over-represented GO terms by semantic similarity and provides a useful high-level summary. Here, only up-regulated biological processes are shown, but the same plots are available for down-regulated terms as well as GO molecular functions and cellular components. f) Top over-represented LRIs, based on their ORA score. Here we show up-regulated interactions, but similar plots are available for the other cases, as well as for ligand and receptor genes alone. Marker size and color encode adjusted p-value and odds ratio respectively, which together determine the ORA Score.

Fig.5: examples of cross-tissue information available in scAgeCom

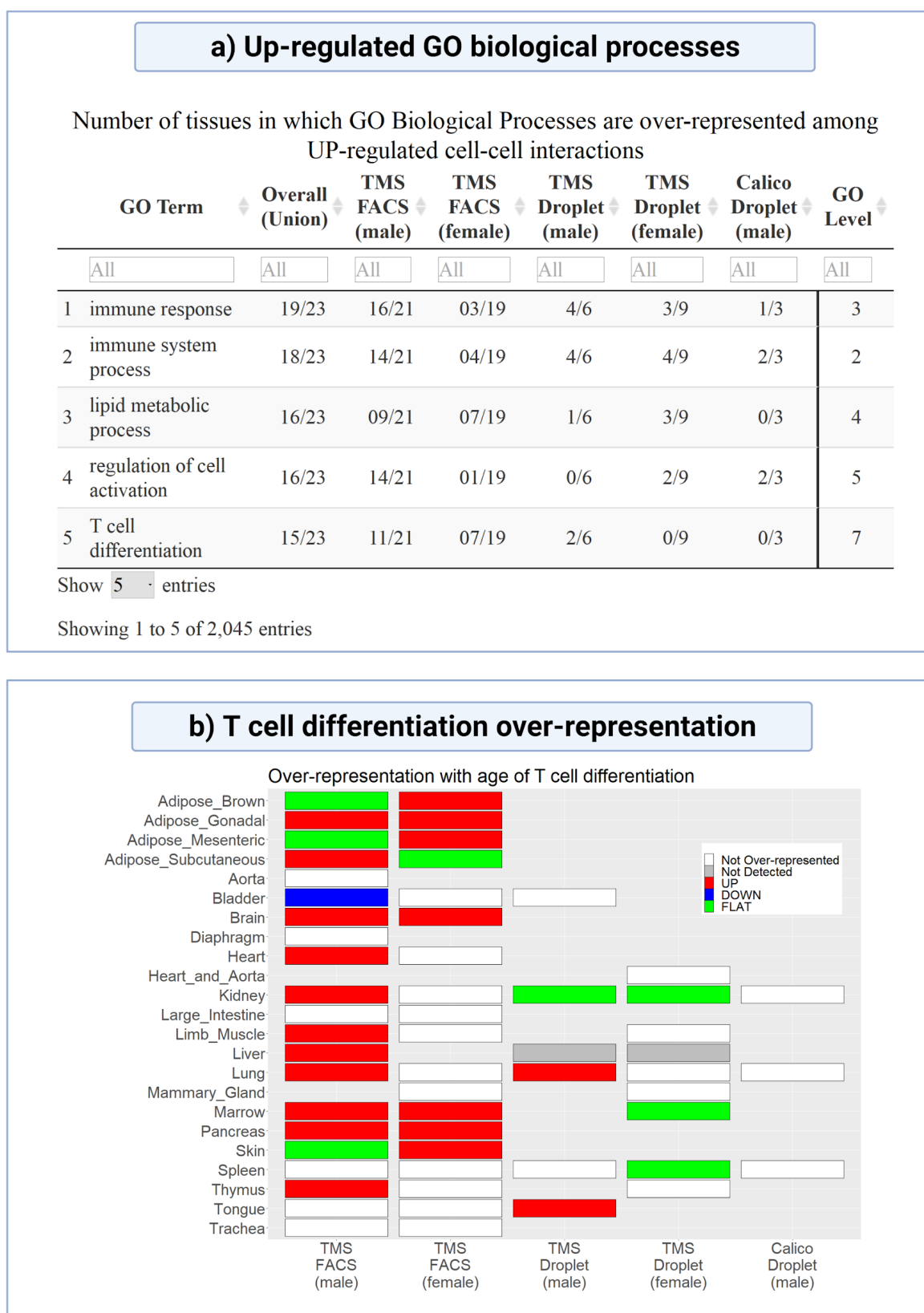


Fig.5: Presentation of the cross-tissue results available in scAgeCom. a) For each category on which over-representation analysis has been performed and for each regulation with age

(UP, DOWN, FLAT), we provide a table that summarizes the number of tissues in which a particular keyword is over-represented. An example is given here for the up-regulated GO biological processes. b) For each keyword of any category, we also provide a summary plot that shows how it is over-represented across the 58 datasets we analyzed. An example is given here for the GO term *T cell differentiation*.

Fig.6 Repartition of the age-regulated CCIs in each dataset

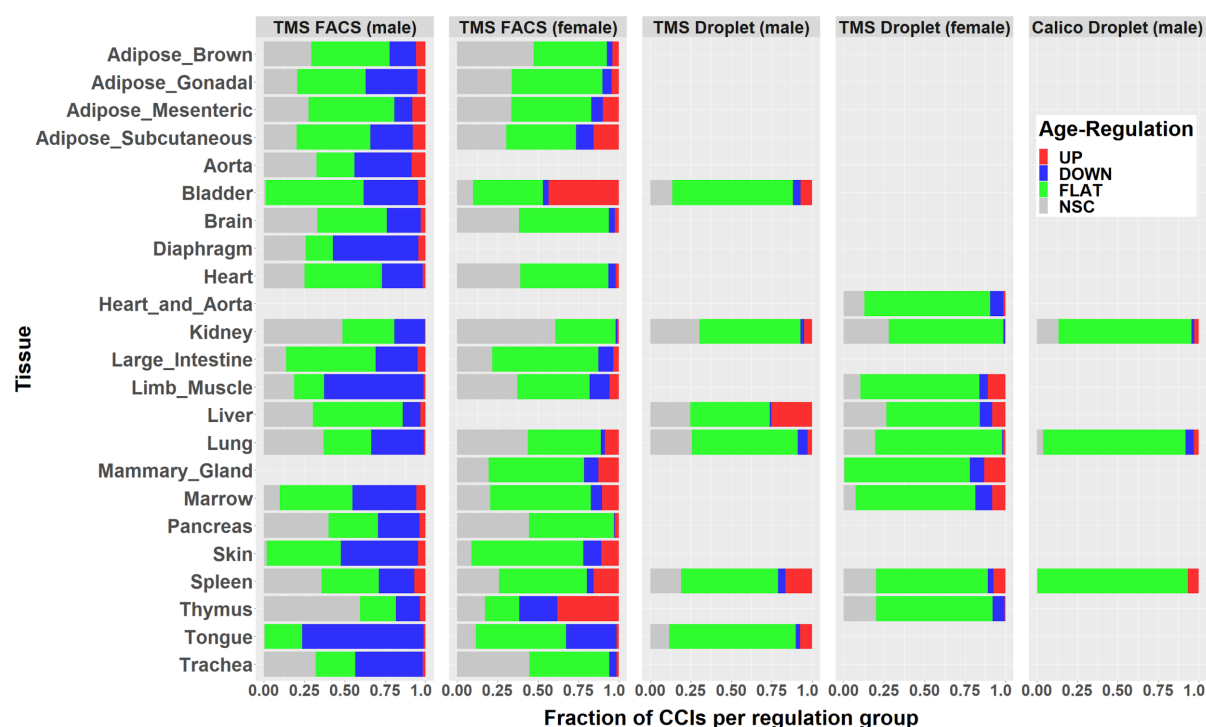


Fig.6: Stacked bar representation of the age-regulation of CCIs across tissues and data source. For each of the 58 datasets, the colored bar indicates the percentage of UP, DOWN, FLAT and NSC CCIs.

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-Otín, 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**, (2020).
4. Capp, J.-P. & Thomas, F. Tissue-disruption-induced cellular stochasticity and epigenetic drift: Common origins of aging and cancer? *Bioessays* e2000140 (2020) doi:10.1002/bies.202000140.
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’Loghlen, 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 Suppl**, S50-6 (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**, (2021).
 29. Hou, R., Denisenko, E., Ong, H. T., Ramilowski, J. A. & Forrest, A. R. 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. *BioRxiv* (2019) doi:10.1101/566182.
 34. Ximerakis, M. *et al.* Single-cell transcriptomic profiling of the aging mouse brain. *Nat.*

- Neurosci.* **22**, 1696–1708 (2019).
35. Ma, S. *et al.* Caloric Restriction Reprograms the Single-Cell Transcriptional Landscape of *Rattus Norvegicus* Aging. *Cell* **180**, 984-1001.e22 (2020).
 36. Ma, S. *et al.* Single-cell transcriptomic atlas of primate cardiopulmonary aging. *Cell Res.* (2020) doi:10.1038/s41422-020-00412-6.
 37. Zou, Z. *et al.* A Single-Cell Transcriptomic Atlas of Human Skin Aging. *Dev. Cell* (2020) doi:10.1016/j.devcel.2020.11.002.
 38. Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
 39. 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).
 40. 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).
 41. 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).
 42. 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).
 43. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
 44. 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).

45. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
46. 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).
47. 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).
48. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
49. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).
50. *Permutation, Parametric and Bootstrap Tests of Hypotheses*. (Springer-Verlag, 2005). doi:10.1007/b138696.
51. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
52. Tabula Muris Consortium *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
53. Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **65**, 631-643.e4 (2017).
54. 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**, (2021).
55. 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).
56. Smith, L. K. *et al.* β 2-microglobulin is a systemic pro-aging factor that impairs

- cognitive function and neurogenesis. *Nat. Med.* **21**, 932–937 (2015).
57. Castellano, J. M. Blood-Based Therapies to Combat Aging. *Gerontology* **65**, 84–89 (2019).
58. Basisty, N. *et al.* A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLoS Biol.* **18**, e3000599 (2020).
59. 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).
60. 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).
61. 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).
62. 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).
63. Johnson, A. A. & Stolzing, A. The role of lipid metabolism in aging, lifespan regulation, and age-related disease. *Aging Cell* **18**, e13048 (2019).
64. Chow, V. W., Mattson, M. P., Wong, P. C. & Gleichmann, M. An overview of APP processing enzymes and products. *Neuromolecular Med.* **12**, 1–12 (2010).
65. Yu, J.-T., Tan, L. & Hardy, J. Apolipoprotein E in Alzheimer’s disease: an update. *Annu. Rev. Neurosci.* **37**, 79–100 (2014).
66. 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).

67. Herz, J. & Beffert, U. Apolipoprotein E receptors: linking brain development and Alzheimer's disease. *Nat. Rev. Neurosci.* **1**, 51–58 (2000).
68. 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).
69. 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).
70. Lee, Y.-H. *et al.* Amyloid precursor protein expression is upregulated in adipocytes in obesity. *Obesity (Silver Spring)* **16**, 1493–1500 (2008).
71. Herzog, V., Kirfel, G., Siemes, C. & Schmitz, A. Biological roles of APP in the epidermis. *Eur. J. Cell Biol.* **83**, 613–624 (2004).
72. 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).
73. 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).
74. Askanas, V. & Engel, W. K. Inclusion-body myositis: a myodegenerative conformational disorder associated with Abeta, protein misfolding, and proteasome inhibition. *Neurology* **66**, S39-48 (2006).
75. 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).
76. Borghesan, M. & O'Loughlen, A. Integrins in senescence and aging. *Cell Cycle* **16**, 909–910 (2017).
77. Moriya, J. & Minamino, T. Angiogenesis, cancer, and vascular aging. *Front.*

- Cardiovasc. Med.* **4**, 65 (2017).
78. 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).
 79. Ambrose, C. T. Pro-Angiogenesis Therapy and Aging: A Mini-Review. *Gerontology* **63**, 393–400 (2017).
 80. 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).
 81. Hayashi, S.-I., Rakugi, H. & Morishita, R. Insight into the Role of Angiopoietins in Ageing-Associated Diseases. *Cells* **9**, (2020).
 82. Tweedie, S. *et al.* Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res.* **49**, D939–D946 (2021).
 83. Bult, C. J. *et al.* Mouse genome database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).
 84. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).
 85. 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).
 86. Xin, J. *et al.* High-performance web services for querying gene and variant annotation. *Genome Biol.* **17**, 91 (2016).
 87. Wu, C., Macleod, I. & Su, A. I. BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res.* **41**, D561–5 (2013).
 88. Türe, D. *et al.* Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol. Syst. Biol.* **17**, e9923 (2021).

89. 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).
90. Greene, D., Richardson, S. & Turro, E. ontologyX: a suite of R packages for working with ontological data. *Bioinformatics* **33**, 1104–1106 (2017).
91. Bioconductor - KEGGREST.
<http://bioconductor.org/packages/release/bioc/html/KEGGREST.html>.
92. 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).
93. Phillips. *Physical biology of the cell*. (Garland Science, 2012).
doi:10.1201/9781134111589.
94. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).
95. Xiao, Y. *et al.* A novel significance score for gene selection and ranking. *Bioinformatics* **30**, 801–807 (2014).
96. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
97. Csardi G & Nepusz T. *The igraph software package for complex network research*. (InterJournal, 2006).
98. Almende, B. V., Thieurmél, B. & Robert, T. *Network Visualization using “vis.js” Library [R package visNetwork version 2.0.9]*. (2019).
99. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
100. Diehl, A. D. *et al.* The Cell Ontology 2016: enhanced content, modularization, and

- ontology interoperability. *J. Biomed. Semantics* **7**, 44 (2016).
101. A Framework for Robust Shiny Applications [R package golem version 0.3.1].
<https://cran.r-project.org/web/packages/golem/index.html> (2021).
 102. Plotly Technologies Inc. *Collaborative data science Publisher: Plotly Technologies Inc.* (2015).
 103. Sayols S. *rrvgo: a Bioconductor package to reduce and visualize Gene Ontology term*. (Bioconductor, 2020).
 104. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).