

METHOD

Uncovering hypergraphs of cell-cell interaction from single cell RNA-sequencing data

Koki Tsuyuzaki¹, Manabu Ishii¹ and Itoshi Nikaido^{1,2*}

*Correspondence:

itoshi.nikaido@riken.jp

²Bioinformatics Course,

Master's/Doctoral Program in Life

Science Innovation (T-LSI),

School of Integrative and Global

Majors (SIGMA), University of

Tsukuba, Wako, 351-0198,

Saitama, Japan

Full list of author information is

available at the end of the article

Abstract

Complex biological systems can be described as a multitude of cell-cell interactions (CCIs). Recent single-cell RNA-sequencing technologies have enabled the detection of CCIs and related ligand-receptor (L-R) gene expression simultaneously. However, previous data analysis methods have focused on only one-to-one CCIs between two cell types. To also detect many-to-many CCIs, we propose *scTensor*, a novel method for extracting representative triadic relationships (hypergraphs), which include (i) ligand-expression, (ii) receptor-expression, and (iii) L-R pairs. When applied to simulated and empirical datasets, *scTensor* was able to detect some hypergraphs including paracrine/autocrine CCI patterns, which cannot be detected by previous methods.

Keywords: Single-cell RNA-sequencing; Cell-cell interaction; Hypergraph; Dimension Reduction; Tensor Decomposition; Non-negative Tucker Decomposition; R/Bioconductor

1

2

3 Background

4 Complex biological systems such as tissue homeostasis [1, 2], neurotransmission [3,
5 4], immune response [5], ontogenesis [6], and stem cells niche [7, 8] are composed by
6 cell-cell interaction (CCI). Many molecular biology studies have been decomposed
7 the system into constituent parts (e.g., genes, proteins, and metabolites) to clarify

1 their functions. Nevertheless, more sophisticated methodologies are still required,
2 because CCI is the difference between the whole system and sum of their parts.

3 Previous studies have investigated CCIs using technologies such as fluorescence
4 microscopy [9–13], microdevice-based methods such as microwells, micropatterns,
5 single-cell traps, droplet microfluidics, and micropillars [14–22], and transcriptome-
6 based method [23–53]. In particular, the recent development of single-cell RNA-
7 sequencing (scRNA-seq) technologies has enabled the detection of exhaustive cell-
8 type-level CCIs based on ligand and receptor (L-R) coexpression.

9 To assess the coexpression of known L-R genes, circle plots [24, 33, 36, 44, 45, 50],
10 bigraph/Sankey diagrams [29, 30, 34, 49, 52, 53], network diagrams [26–28, 31, 37,
11 40, 41, 44, 46, 54] and heatmaps [30, 32, 34, 39, 46, 47, 49] are often drawn. Some
12 studies have also introduced more systematical approaches for quantifying the de-
13 gree of CCIs based on the L-R coexpression, such as the number of coexpressed
14 L-R pairs [23, 24, 26, 27], Spearman correlation coefficients between L-R expres-
15 sion profiles [26, 31], original interaction scores between L-R coexpression [39], or
16 hypothetical test based on random cell-type label permutation [29, 32, 37, 55].

17 All the approaches described above implicitly suppose that CCIs are the one-to-
18 one relationships between two cell types and that the corresponding L-R coexpres-
19 sion is observed as the cell-type-specific manner. In real empirical data, however,
20 the situation can be more complex; CCIs often exhibit many-to-many relationships
21 involving many cell types, and an particular L-R pair can also function across mul-
22 tiple cell-type pairs. Therefore, in this work, we propose **scTensor**, which is a novel
23 method based on a tensor decomposition algorithm. Our method regards CCIs as
24 hypergraphs and extracts some representative triadic relationship from the data
25 tensor, which includes (i) ligand-expressing cell types, (ii) receptor-expressing cell
26 types, and (iii) L-R pairs.

1 CCI as hypergraph (*CaH*) and CCI-tensor

2 The simplest CCI representation is perhaps a directed graph, where each node
3 represents a cell type and each edge represents the coexpression of all L-R pairs
4 (Figure 1a, left). The direction of the edge is set as ligand \rightarrow receptor. Such a data
5 structure can also be described as an asymmetric adjacency matrix, in which each
6 row and column represents a ligand and receptor, respectively. If some combinations
7 of cell types are regarded as interacting, corresponding elements of the matrix are
8 filled with 1 and otherwise 0. If the degree of CCI is not a binary relationship,
9 weighted graphs and corresponding weighted adjacent matrices may also be used.
10 The previous analytical methods are categorized within this approach [23, 24, 26–
11 34, 36, 37, 39–41, 44–47, 49, 50, 52, 53, 55].

12 In contrast, this work describes CCIs as directed hypergraphs (CCI as hypergraph;
13 *CaH*), where each node is a cell type but the edges are distinguished from each other
14 by the different related L-R pair sets (Figure 1a, right). Such a context-aware edge
15 is called a hyperedge and is described as multiple different adjacency matrices and
16 the set of matrices is called a higher-order matrix or tensor. In contrast with the
17 simple adjacency matrix, tensor contains considerable high-resolution information
18 owing to its higher-order.

19 Prediction of many-to-many CCIs using tensor decomposition

20 Tensor data are constructed through the following steps (Figure 1b). Here, a scRNA-
21 seq matrix and the cellular label specifying cell types are supposed to be provided
22 by users. Firstly, Freeman-Tukey transformation [56] (FTT, $\sqrt{x} + \sqrt{x+1}$), which
23 is a variance-stabilizing transformation, is performed to the data matrix. Next,
24 the matrix is converted to a cell-type-level average matrix according to the label.
25 Combined with a L-R database, two corresponding row-vectors of an L-R pair are
26 extracted from the matrix. We originally developed the databases for 12 organisms
27 (for more details, see Additional Files 1, 2, Table 2 and 3). The outer product (direct

1 product) of the two vectors is then calculated, and a matrix is generated. The matrix
2 can be considered as the similarity matrix of all possible cell-type combinations
3 using the L-R pair. Finally, for each L-R pair, the matrix is calculated, and a tensor
4 is generated as the combined matrices. In this work, this is called the CCI-tensor.

5 After the construction of a CCI-tensor, we perform non-negative Tucker decompo-
6 sition (NTD [57, 58]), which is known as a tensor decomposition algorithm. We orig-
7 inally implemented this algorithm and confirmed its convergence within a realistic
8 number of iterations (for more details, see Additional File 3). NTD assumes that the
9 CCI-tensor can be approximated by the summation of some representative *CaH*s.
10 NTD has the three rank parameters ($R1$, $R2$, and $R3$) and a *CaH* is calculated as
11 the outer product of the column vectors of three factor matrices $\mathbf{A}^{(1)} \in \mathbb{R}^{J \times R1}$,
12 $\mathbf{A}^{(2)} \in \mathbb{R}^{J \times R2}$, and $\mathbf{A}^{(3)} \in \mathbb{R}^{K \times R3}$ calculated by NTD (Figure 1c). Each *CaH*-
13 strength is calculated by the core tensor $\mathcal{G}(r1, r2, r3) \in \mathbb{R}^{R1 \times R2 \times R3}$ of NTD. In this
14 work, each *CaH* is termed $CaH(r1, r2, r3) = \mathbf{A}_{:r1}^{(1)} \circ \mathbf{A}_{:r2}^{(2)} \circ \mathbf{A}_{:r3}^{(3)} \in \mathbb{R}^{J \times J \times K}$, where
15 $r1$, $r2$, and $r3$ are the indexes of the columns of three factor matrices. All *CaH*s are
16 ordered by the size of elements of the core tensor, and the patterns explaining the
17 top 40 % of cumulative core tensor value are reserved as representative *CaH*s. For
18 more details on *CaH*, see the Materials and Methods. The *CaH*s are extracted in a
19 data-driven way without the assumption of one-to-one CCIs. Therefore, it can also
20 detect many-to-many CCIs according to the data complexity.

21 **Results and discussion**

22 Evaluation of multiple CCI prediction

23 *Accuracy of the detection of CCIs and the related L-R pairs*

24 Here, we demonstrate the efficacy of **scTensor** by using the two simulation datasets
25 (Figure 2a). Three different cell types are indicated as "A", "B", and "C". In the case
26 I dataset, all CCIs represent the one-to-one relationships between two cell types.
27 CCIs corresponding to $A \rightarrow B$, $B \rightarrow C$, and $C \rightarrow A$ are colored by red, blue, and

1 green, respectively. In contrast, the CCIs in the case II dataset represent many-to-
2 many relationships involving many cell types such as $A \rightarrow B/C$ (red), $C \rightarrow A/B/C$
3 (blue), and $A/B/C \rightarrow B/C$ (green). To evaluate whether such ground truth combi-
4 nation of cell types and their related L-R pairs are enriched by **scTensor**, receiver
5 operating characteristic (ROC) curves and their corresponding area under the curve
6 (AUC) values were calculated (Figure2b). In the analysis of each L-R set, only the
7 *CaHs* with the maximum AUC value for each ground truth L-R set were regarded
8 as the corresponding *CaH* being accurately detected by **scTensor** (Figure2c, for
9 more details on the simulation datasets, see the Materials and Methods).

10 Again, note that the *CaHs* detected by **scTensor** are not just CCIs, but sets of
11 CCIs and their related L-R pairs. To the best of our knowledge, the label permu-
12 tation method implemented in CellPhoneDB [37] is the only previous method that
13 can detect CCIs and their related L-R pairs simultaneously. To demonstrate the
14 efficacy of **scTensor** in terms of the detection of many-to-many CCIs, we also orig-
15 inally implemented this method and compared it with **scTensor** (for more details
16 on the algorithm, see the Materials and Methods). Note that each combination of
17 a CCI and its related L-R pair is separately extracted by NTD of **scTensor**, but
18 under the label permutation method, the combinations are not separated and are
19 just sorted in ascending order of their *P*-values. Combinations with low *P*-values
20 indicate significant triadic relationships.

21 In the case I dataset, ground truth L-R sets are highly enriched according to the
22 measures of both methods, and the AUC values show that there is no difference
23 in their performance (Figure2b and c, left). On the other hand, in the case II
24 dataset, the label permutation method cannot correctly detect blue or green L-R
25 sets, and the AUC value becomes lower (Figure2b, right). In the **scTensor** analysis,
26 red, blue, and green L-R sets are separately extracted as three *CaHs* (Figure2c,
27 right), and the AUC values are still high. This is because, the label permutation

1 is implicitly hypothesized the CCI as a one-to-one relationship. Therefore, in the
2 case II dataset, many-to-many CCIs such as the CCIs corresponding to green L-R
3 sets, are hard to detect by the method. This is because for each L-R pair, mean
4 values for any combination of cell types are basically high in such situations, and a
5 *P*-value corresponding to a one-to-one CCI tends to be large (i.e., not significant);
6 accordingly, the observed L-R coexpression and the null distribution calculated
7 are hard to distinguish. In the analysis of real datasets presented later, however,
8 the L-R gene expression pairs are not always the cell-type specific, and it is more
9 natural that the CCI corresponding to the L-R has a many-to-many relationship.
10 This simulation shows that **scTensor** is a more general method for detecting CCIs
11 and their related L-R pairs at once, irrespective of whether a particular CCI is
12 one-to-one or many-to-many.

13 *Biological interpretation of real datasets*

14 To demonstrate the efficacy of **scTensor** in the analysis of empirical datasets, we
15 applied **scTensor** to four real scRNA-seq datasets (Table 1). First, we used the
16 scRNA-seq data derived from fetal germ cells (FGCs (φ)) and their gonadal niche
17 cells (Soma) from female human embryos (Germline.Female [25]). As a known *CaH*
18 pattern, CCIs of Soma with FGCs (φ) involving the BMP signaling pathway are
19 reported, and **scTensor** accurately detects the CCIs and CCI-related L-R pairs
20 as *CaH*(4,2,5) (Figure 3). Moreover, **scTensor** was able to extract some putative
21 *CaH*s such as *CaH*(3,4,14) and *CaH*(1,4,22). *CaH*(3,4,14) is the autocrine-type CCI
22 within FGCs (φ) and L-R pairs such as Wnt (WNT5A/WNT6) and some growth
23 factor genes (NGF/IGF/FGFR/VEGF), suggesting that this *CaH* is related to the
24 proliferation and differentiation of Soma. *CaH*(1,4,22) is the CCI corresponding to
25 FGCs (φ) and Soma, and most of the receptors are related to G protein-coupled
26 receptor (GPCR). The conjugated ligands are some neuropeptides, and this suggests

1 that the peptides are related to the activation of GPCR in FGCs (φ) by some
2 mechanism.

3 We also used the scRNA-seq data derived from FGCs (σ) and their gonadal
4 niche cells from male human embryos (Germline_Male [25]). In contrast to the
5 Germline_Female data, the original study reported that FGCs (σ) interact with
6 Soma through AMH-BMPR interactions, and **scTensor** also detected the triadic
7 relationship as *CaH*(1,1,10) (Figure 4). The conjugate receptor AMHR2 is also de-
8 tected in this *CaH*. In this dataset, like in Germline_Female, in this data, autocrine-
9 type CCIs such as *CaH*(3,3,9) and CCIs corresponding to FGCs (σ) and Soma
10 (*CaH*(1,3,16)) were detected.

11 Next, we used the scRNA-seq data derived from immune cells isolated from the
12 metastatic melanoma patients (Melanoma [59]). The meta-analysis of scRNA-seq
13 and TCGA datasets in the original published study claimed that T cell abundance
14 and the expression of complement system-related genes in cancer-associated fibrob-
15 lasts (CAFs) are correlated, and CCIs between T cells and CAFs were therefore
16 inferred. **scTensor** detected these CCIs as *CaH*(1,1,3)/*CaH*(1,1,9) and comple-
17 ment system-related genes such as C3, CXCL12, CFB, and C4A were also detected
18 (Figure 5). **scTensor** was also able to capture a well-known CCI between T cells
19 and B cells as *CaH*(2,2,6), although this was not a focus of the original study; this
20 CCI is the antigen-presenting from B cells to T cells by class II major histocom-
21 patibility complex (MHC) through the coexpression of CD8 (T cell receptor) and
22 human leukocyte antigen (HLA) genes. **scTensor** also detected *CaH*(4,1,13) and
23 *CaH*(1,4,8), which are the CCIs of macrophages with T cells and an autocrine-type
24 CCI of known chemokine ligands with their receptors.

25 Finally, we used data derived from non-myocyte cells isolated from mouse heart
26 (NonMyocyte [27]). The original study focused on the CCIs of pericytes/fibroblasts
27 with macrophages through the L-R pairs *Il34/Csfl* and *Csflr*, and the correspond-

1 ing CCIs were detected as $CaH(2,2,19)$ by `scTensor` (Figure 6). `scTensor` also
2 detected $CaH(1,2,17)$ and $CaH(3,2,21)$, which are the autocrine-type CCI among
3 macrophages and the CCI of NK cells with macrophages by known chemokine lig-
4 ands and their receptors.

5 Application to minor organism

6 To demonstrate the applicability of using `scTensor` in the species that is not
7 mouse or human, we used a scRNA-seq data derived from zebrafish habenular
8 neurons (Habenular_Larva [60]). Although the original study did not focus on
9 the CCIs among the neuronal cell types, `scTensor` detected some triadic relation-
10 ships as $CaH(3,3,4)$ (La.Hb01/03/07 with La.Hb02/08), $CaH(2,1,3)$ (La.Hb09 with
11 La.Hb02/08), and $CaH(1,3,1)$ (La.Hb02/04-06/08/11-15/Olf with La.Hb02/08)
12 (Figure 7). The spatial distribution of the cell types measured by RNA-fluorescence
13 *in situ* hybridization (FISH) shows that the cell-type pairs detected as $CaH(3,3,4)$
14 and $CaH(2,1,3)$ are dorsally located and proximal to each other in the habenula.
15 However, $CaH(1,3,1)$ was a more global interaction related to the entire habenula
16 regions. Although the spatial distribution of rare cell types La.Hb03/05/14 could
17 not be determined by RNA-FISH in the original study, `scTensor` was able to as-
18 sign the conjugated cell types of La.Hb03 as La.Hb02/08 in the dorsal region. This
19 result suggests that `scTensor` may also be useful in spatial transcriptomics [61, 62].

20 `scTensor` and L-R database implementations as R/Bioconductor packages

21 All the algorithms and L-R lists are available as R/Bioconductor packages and a
22 web application described below.

23 `nnTensor` and `scTensor` packages

24 NTD is implemented as the function of `nnTensor` R/CRAN package and internally
25 imported in `scTensor`. `scTensor` constructs the CCI-tensor, decomposes the tensor
26 by NTD, and generates an HTML report. `scTensor` is assumed to be used with

1 LRB`base.XXX.eg.db`, which are the L-R databases for multiple organisms. To en-
2 hance the biological interpretation of *CaHs*, a wide variety of gene information is
3 assigned to the L-R lists through the other R/Bioconductor packages (Figure 8).
4 For example, gene annotation is assigned by `biomaRt` [63] (Gene Name, Description,
5 Gene Ontology (GO), STRING, and UniProtKB), `reactome.db` [64] (Reactome)
6 and `MeSH.XXX.eg.db` [65] (Medical Subject Headings; MeSH), while the enrich-
7 ment analysis (also known as over-representative analysis; ORA) is performed by
8 `G0stats` [66] (GO-ORA), `meshr` [65] (MeSH-ORA), `ReactomePA` [67] (Reactome-
9 ORA), and `DOSE` [68] (Disease Ontology; DO, Network of Cancer Gene; NCG,
10 and DisGeNET-ORA). To validate that the detected gene expression of L-R gene
11 pair is also consistently detected in the other data with tissue- or cell-type-level
12 transcriptome data, the hyperlinks to RefEx [69], Expression Atlas [70], Single-
13 Cell Expression Atlas [71], scRNASeqDB [72], PanglaoDB [73] are embedded in
14 the HTML report, facilitating comparisons of the L-R expression with the data
15 from large-scale genomics projects such as GTEx [74], FANTOM5 [75], NIH Epige-
16 nomics Roadmap [76], ENCODE [77], and Human Protein Atlas [78]. Additionally,
17 in consideration of users who might want to experimentally investigate detected
18 CCIs, we embedded the hyperlinks to Connectivity Map (CMap [79]), which pro-
19 vides the relationships between perturbation by the addition of particular chemical
20 compounds/genetic reagents and succeeding gene expression change.

21 LRB`base.XXX.eg.db`-type packages

22 For data sustainability and the extension to the wide range of organisms, in
23 this work, we originally constructed L-R databases as R/Bioconductor packages
24 named LRB`base.XXX.eg.db` (where XXX represents the abbreviation for an organ-
25 ism, such as “Hsa” for Homo sapiens). LRB`base.XXX.eg.db` currently provides the
26 L-R databases for 12 organisms (Table 2 and 3). The data process pipeline is almost
27 the same as that of the FANTOM5 project for constructing the putative L-R lists.

1 Precise differences between `LRBase.XXX.eg.db` and FANTOM5 are summarized in
2 Table S4 in Additional File 1.

3 `LRBaseDbi` package

4 All the `LRBase.XXX.eg.db` packages are generated by `LRBaseDbi`, which is the an-
5 other R/Bioconductor package. `LRBaseDbi` generates the `LRBase.XXX.eg.db` pack-
6 ages from the CSV files, in which NCBI Gene IDs are saved as two columns de-
7 scribing the L-R relationship (we call this function as “meta”-packaging).

8 In addition to the `LRBase.XXX.eg.db` packages we summarized, the users may
9 want to specify the user’s original L-R list. For example, there are some other L-R
10 databases such as IUPHAR [80], DLRP [81], FANTOM5 [75], CellPhoneDB [28],
11 and Cell-Cell Interaction Database summarized by Gary Bader *et. al.* ([http://
12 baderlab.org/CellCellInteractions](http://baderlab.org/CellCellInteractions)) (Additional File 1). Besides, if the users
13 want to apply the `scTensor` to minor species, the L-R list may be constructed by the
14 orthologous relationship with major species (e.g., human) [27]. The corresponding
15 `LRBase.XXX.eg.db` is easily generated from the original L-R list by `LRBaseDbi` and
16 can be used with `scTensor`.

17 `CellCelldb`

18 All analytical results `scTensor` are outputted as HTML reports. Hence, com-
19 bined with a cloud web service such as Amazon Simple Storage Service (Ama-
20 zon S3), reports can be used as simple web applications, enabling the user to
21 share their results with collaborators or to develop an exhaustive CCI database.
22 We have already performed `scTensor` analyses using a wide-variety of scRNA-
23 seq datasets, including the five empirical datasets examined in this study ([https:
24 //q-brain2.riken.jp/CellCelldb/](https://q-brain2.riken.jp/CellCelldb/)).

1 **Conclusions**

2 In this work, CCIs were regarded as *CaHs*, which are hypergraphs that represent
3 triadic relationships, and a novel algorithm `scTensor` for detecting such *CaH* was
4 developed. In evaluations with empirical datasets from previous CCI studies, pre-
5 viously reported CCIs were also detected by `scTensor`. Moreover, some CCIs were
6 detected only by `scTensor`, suggesting that the previous studies may have over-
7 looked some CCIs. To extend the use of `scTensor` to a wide range of organisms,
8 we also developed multiple L-R databases as `LRBase.XXX.eg.db`-type packages.
9 When combined with `LRBase.XXX.eg.db`, `scTensor` can currently be applied to 12
10 organisms.

11 There are still some plans for improving both `LRBase.XXX.eg.db` and `scTensor`
12 to build on the advantages of this current framework. For example, the range of
13 corresponding organisms and the L-R lists can be extended with the spread of
14 genome-wide researches. Additionally, the algorithm can be improved, for exam-
15 ple, by utilizing acceleration techniques such as randomized algorithm/sketching
16 methods [82], incremental algorithm/stochastic optimization [83, 84], or distributed
17 computing with MapReduce/Hadoop on large-scale memory machines [85] for NTD,
18 which is now available. Tensors are a very flexible way to represent heterogeneous
19 biological data [86], and easily integrate the side information about genes or cell
20 types with semi-supervised manner. Such information will improve the accuracy
21 and extend the scope of the data.

22 We aim to tackle such problems and develop the framework further through
23 updates of the R/Bioconductor packages. In the package registration process
24 for R/Bioconductor, package source code is peer-reviewed via the Bioconductor
25 single package builder system and assigned to a curator ([https://github.com/
26 Bioconductor/packagebuilder](https://github.com/Bioconductor/packagebuilder)), and even after the package is accepted, the daily
27 package builder tests the source code every day (<https://www.bioconductor.org/>

1 [checkResults/](#)). Furthermore, biannual updates of Bioconductor require that the
2 internal data of all data packages are updated (<https://www.bioconductor.org/>
3 [developers/release-schedule/](#)). Such strict check systems for source code and
4 internal data improve the sustainability and usability of packages. Our team has
5 maintained over one hundred R/Bioconductor packages since 2015 [65], and we are
6 still organizing a system for the maintenance of the combined `LRBase.XXX.eg.db`
7 and `scTensor` framework.

8 **Materials and methods**

9 Construction of L-R list

10 *Public databases*

11 To compare our database with other databases (Additional File 1 and 2), the data
12 from FANTOM5 ([http://fantom.gsc.riken.jp/5/suppl/Ramilowski\\$_set\\$_al\\$_2015/](http://fantom.gsc.riken.jp/5/suppl/Ramilowski$_set$_al$_2015/)
13 [data/PairsLigRec.txt](#)), DLRP (<http://dip.doe-mbi.ucla.edu/dip/dlrp/dlrp.txt>),
14 IUPHAR (<http://www.guidetopharmacology.org/DATA/interactions.csv>), and
15 HPRD ([ftp://ftp.ebi.ac.uk/pub/databases/genenames/new/tsv/locus\\$_groups/](ftp://ftp.ebi.ac.uk/pub/databases/genenames/new/tsv/locus$_groups/)
16 [protein-coding\\$_gene.txt](#)) were downloaded. The subcellular localization data
17 from SWISSPROT and TrEMBL were downloaded from UniProtKB ([https://](https://www.uniprot.org/downloads)
18 [www.uniprot.org/downloads](#)). Protein-protein interaction (PPI) data of STRING
19 (v-10.5) were downloaded from <https://stringdb-static.org/cgi/download.pl>.
20 To unify the gene identifier as NCBI Gene ID, we retrieved the correspond-
21 ing table from Biomart (Ensembl release 92). All the data were downloaded by
22 RESTful access using the wget command and query.xml <http://www.biomart.org/>
23 [martservice.html](#).

24 *Simulation datasets*

25 The simulated single-cell gene expression data were sampled from the negative bi-
26 nomial distribution $NB(FC_{gc}m_g, \phi_g)$, where FC_{gc} is the fold-change (FC) for gene
27 g and cell type c , and m_g and ϕ_g are the average gene expression and the dispersion

1 parameter of gene g , respectively. For the setting of differentially expressed genes
2 (DEGs) and non-DEGs, FC_{gc} values were calculated based on the non-linear rela-
3 tionship of FC and the gene expression level $\log_{10} FC_{gc} = a \exp(-b \log_{10}(m_g + 1))$
4 as follows:

$$FC_{gc} = \begin{cases} 10^{4.42 \exp(-0.81 \log_{10}(m_g + 1))} & \text{(DEG)} \\ 1 & \text{(non-DEG)}. \end{cases}$$

5 The m_g and gene-wise variance v_g were calculated from the scRNA-seq dataset
6 of human embryonic stem cells (hESCs) measured by Quartz-Seq [87], and the
7 gene-wise dispersion parameter ϕ_g was estimated as $\phi_g = (v_g - m_g)/m_g^2$. The
8 NB distribution reduces to Poisson when $\phi_g = 0$. To simulate the “dropout”
9 phenomena of scRNA-seq experiments, we introduced the dropout probability
10 $p_{dropout_{gc}} = \exp(-cFC_{gc}m_g^2)$, which is used in ZIFA [88] (default: $c=1$), and the
11 expression values were randomly converted to 0 based on this dropout probability.

12 For the setting of the case I datasets, $150 \times 150 \times 500$ CCI-tensor was constructed.
13 For each cell type, 50 cells were established, and in total, three cell types were set.
14 For L-R set 1 (red), 50 L-R pairs were established, and the cell-type-wise ligand and
15 receptor patterns were (1,0,0) and (0,1,0), respectively (0, non-DEG; 1, DEG). For
16 L-R set 2 (blue), 50 L-R pairs were established, and the cell-type-wise ligand and
17 receptor patterns were (0,1,0) and (0,0,1), respectively. For L-R set 3 (green), 50
18 L-R pairs were established, and the cell-type-wise ligand and receptor patterns were
19 (0,0,1) and (1,0,0), respectively. The other 350 L-R pairs were sampled randomly
20 as non-DEGs.

21 For the setting of the case II datasets, $150 \times 150 \times 500$ CCI-tensor were con-
22 structed. For each cell type, 50 cells were established, and in total, three cell types
23 were established. For L-R set 1 (red), 50 L-R pairs were established, and the cell-

1 type-wise ligand and receptor patterns were (1,0,0) and (0,1,1), respectively. For
2 L-R set 2 (blue), 50 L-R pairs were established, and the cell-type-wise ligand and
3 receptor patterns were (0,0,1) and (1,1,1), respectively. For L-R set 3 (green), 50
4 L-R pairs were established, and the cell-type-wise ligand and receptor patterns were
5 (1,1,1) and (1,0,1), respectively. The other 350 L-R pairs were sampled randomly
6 as non-DEGs.

7 *Real datasets*

8 The gene expression matrix and cellular labels for Germline_Female and Germline_Male
9 scRNA-seq data were retrieved from the GEO database (GSE86146), and only
10 highly variable genes (HVGs : http://pklab.med.harvard.edu/scw2014/subpop_tutorial.html)
11 with low P -values ($\leq 1E-7$) were extracted. The gene expression matrix and cel-
12 lular label of Melanoma scRNA-seq data were retrieved from the GEO database
13 (GSE72056), and only HVGs with low P -values ($\leq 1E-10$) were extracted. The gene
14 symbols were converted to NCBI GeneIDs using the R/Bioconductor *Homo.sapiens*
15 package. The gene expression matrix and cellular labels for NonMyocyte scRNA-
16 seq data were retrieved from the ArrayExpress database (E-MTAB-6173), and only
17 HVGs with low P -values ($\leq 1E-10$) were extracted. The gene symbols are con-
18 verted to NCBI GeneIDs using the R/Bioconductor *Mus.musculus* package. For
19 each dataset, t-Distributed Stochastic Neighbor Embedding (t-SNE) with 40 per-
20 plexity was performed using the Rtsne R package.

21 *scTensor algorithm details*

22 *Construction of CCI-tensor*

Here we assume that a data matrix $\mathbf{Y} \in \mathbb{R}^{I \times H}$ is the gene expression matrix of
scRNA-seq, where I is the number of genes and H is the number of cells. Next, the
matrix \mathbf{Y} is converted to cell-type mean matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$, where J is the number
of mean vectors for each cell type. The cell-type label is supposed to be specified by
user's prior analysis such as clustering or confirmation of marker gene expression.

The relationship between the \mathbf{X} and \mathbf{Y} is described as below:

$$\mathbf{X} = \mathbf{Y}\mathbf{A}, \quad (1)$$

where the matrix $\mathbf{A} \in \mathbb{R}^{H \times J}$ converts cellular-level matrix \mathbf{Y} to cell-type-level matrix \mathbf{X} and each element of \mathbf{A} is

$$\mathbf{A}_{hj} = \begin{cases} 1/n_j & (\text{h-th cell belongs to j-th cell type}) \\ 0 & (\text{otherwise}), \end{cases}$$

1 where n_j is the number of cells belonging to j 's cell type.

2 Next, NCBI gene IDs of each L-R pair stored in `LRBase.XXX.eg.db` are searched
3 in the row names of matrix \mathbf{X} and if both IDs are found, corresponding J -length
4 row-vectors of the ligand and receptor genes (\mathbf{x}_L and \mathbf{x}_R) are extracted.

5 Finally, a $J \times J$ matrix is calculated as the outer product of \mathbf{x}_L and \mathbf{x}_R and
6 incrementally stored as a sub-tensor (frontal slice) of the CCI-tensor $\chi \in \mathbb{R}^{J \times J \times K}$
7 as below:

$$\chi_{::k} = \mathbf{x}_{L(k)} \circ \mathbf{x}_{R(k)} \quad (2)$$

8 where K is the number of L-R pairs found in the row names of matrix \mathbf{X} .

9 *CANDECOMP/PARAFAC and Tucker decomposition*

10 Here, we suppose that the CCI-tensor has some representative triadic relationship.

11 To extract the triadic relationships from a CCI-tensor, here we consider perform-

12 ing some tensor decomposition algorithms. There are two typical decomposition

13 methods; CANDECOMP/PARAFAC (CP) and Tucker decomposition [57, 58].

In CP decomposition, CCI-tensor χ is decomposed as follows:

$$\begin{aligned}\chi &= \Lambda \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} \\ &= \sum_{r=1}^R \lambda_r \mathbf{A}_{:r}^{(1)} \circ \mathbf{A}_{:r}^{(2)} \circ \mathbf{A}_{:r}^{(3)} \\ &\text{subject to } \|\mathbf{A}_{:r}^{(1)}\| = \|\mathbf{A}_{:r}^{(2)}\| = \|\mathbf{A}_{:r}^{(3)}\| = 1,\end{aligned}\tag{3}$$

1 where \times_n is mode- n product, R is the rank of χ , and Λ is diagonal cubical tensor,
2 in which the element λ_r on the superdiagonal can be non-zero. $\mathbf{A}^{(1)} \in \mathbb{R}^{J \times R}$,
3 $\mathbf{A}^{(2)} \in \mathbb{R}^{J \times R}$, and $\mathbf{A}^{(3)} \in \mathbb{R}^{K \times R}$ are factor matrices. $\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$ is rank-1 tensor,
4 and the scalar λ_r is the size of rank-1 tensor. The rank-1 tensor indicates the
5 triadic relationship described above, and CP model suppose that CCI-tensor is
6 approximated by the summation of R rank-1 tensor. There are some algorithms for
7 optimizing CP decomposition problem such as alternative least squares (ALS) or
8 power method [58]. Despite its wide use, CP decomposition has a drawback when
9 using the problem in this work; the number of columns of three factor matrices must
10 be a common number R , and the correspondence of $\mathbf{A}_{:r}^{(1)}$, $\mathbf{A}_{:r}^{(2)}$, and $\mathbf{A}_{:r}^{(3)}$ in each r
11 is one-to-one. This constraint is sometimes too strict and unnatural for biological
12 applications. For example, in the CCI-tensor case, the number of ligand expression,
13 receptor expression, and L-R-pair patterns are commonly R , and all of them must
14 correspond to each other in each r . Thus, if an L-R pair assigned in a CCI, this L-R
15 pair cannot be part of other CCIs.

To deal with this problem, the application of Tucker decomposition can be considered next. In Tucker decomposition, a CCI-tensor is decomposed as follows:

$$\begin{aligned}\chi &= \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} \\ &= \sum_{r1=1}^{R1} \sum_{r2=1}^{R2} \sum_{r3=1}^{R3} \mathcal{G}(r1, r2, r3) \mathbf{A}_{:r1}^{(1)} \circ \mathbf{A}_{:r2}^{(2)} \circ \mathbf{A}_{:r3}^{(3)} \\ &\text{subject to } \|\mathbf{A}_{:r1}^{(1)}\| = \|\mathbf{A}_{:r2}^{(2)}\| = \|\mathbf{A}_{:r3}^{(3)}\| = 1,\end{aligned}\tag{4}$$

1 where R_1 , R_2 , and R_3 are the rank of mode-1,2, and 3, respectively. Unlike the
2 CP model, the constraint conditions of the Tucker model are relaxed, that is, three
3 factor matrices $\mathbf{A}^{(1)} \in \mathbb{R}^{J \times R_1}$, $\mathbf{A}^{(2)} \in \mathbb{R}^{J \times R_2}$, and $\mathbf{A}^{(3)} \in \mathbb{R}^{K \times R_3}$ can differ in their
4 numbers of columns and any combination of $\mathbf{A}_{:,r_1}^{(1)}$, $\mathbf{A}_{:,r_2}^{(2)}$, and $\mathbf{A}_{:,r_3}^{(3)}$ can be considered.
5 This is because $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is a dense core tensor and any element, including
6 non-diagonal elements, can have a non-zero value. There are some algorithms for
7 optimizing Tucker decomposition, such as higher order singular value decomposition
8 (HOSVD) or higher orthogonal iteration of tensors (HOOI) [58].

9 *Non-negative Tucker decomposition*

10 Despite its effectiveness, Tucker decomposition cannot be directly applied to the
11 extraction of *CaHs*. This is because the factor matrices of the Tucker model can
12 have negative value elements, and these make interpretation difficult. For example,
13 if $\mathbf{A}_{:,r_1}^{(1)}$ contains very large positive elements and very small negative elements, we
14 cannot determine which cell type is highly related to a ligand expression pattern
15 and which cell type is not.

For the above reason, here we utilize NTD. Unlike Tucker decomposition based on
singular value decomposition (SVD), NTD is based on non-negative matrix factor-
ization (NMF), which is another matrix decomposition method. NMF is formalized
as follows:

$$\mathbf{X} = \mathbf{W}\mathbf{H} \tag{5}$$

subject to $\mathbf{W} \geq 0$, $\mathbf{H} \geq 0$.

The typical algorithm for optimizing the NMF problem is multiplicative updating
(MU) [58]. Two widely used forms are considered. The first form is minimization
problem of Euclidean distance ($\min \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_{Euclid}$), where \mathbf{H} and \mathbf{W} are iter-

actively updated by considering Gaussian noise:

$$\begin{aligned} \mathbf{H} &\leftarrow \mathbf{H} * \frac{\mathbf{W}^T \mathbf{X}}{\mathbf{W}^T \mathbf{W} \mathbf{H}} \\ \mathbf{W} &\leftarrow \mathbf{W} * \frac{\mathbf{X} \mathbf{H}^T}{\mathbf{W} \mathbf{H} \mathbf{H}^T}, \end{aligned} \quad (6)$$

where $*$ is the element-wise (Hadamard) product. The second form is a minimization problem of Kullback-Leibler (KL) divergence ($\min \|\mathbf{X} - \mathbf{W} \mathbf{H}\|_{KL}$), where \mathbf{H} and \mathbf{W} are iteratively updated by considering Poisson noise:

$$\begin{aligned} \mathbf{H} &\leftarrow \mathbf{H} * \frac{\mathbf{W}^T \frac{\mathbf{X}}{\mathbf{W} \mathbf{H}}}{\mathbf{W}^T \mathbf{1}} \\ \mathbf{W} &\leftarrow \mathbf{W} * \frac{\frac{\mathbf{X}}{\mathbf{W} \mathbf{H}} \mathbf{H}^T}{\mathbf{1} \mathbf{H}^T}. \end{aligned} \quad (7)$$

- 1 These update rules are derived from the element-wise gradient descent with the
- 2 spatial form of the learning rate [58]. Starting with a random non-negative initial
- 3 value, update of \mathbf{W} and \mathbf{H} are updated iteratively until convergence. In this work,
- 4 MU with the KL-form, which shows a stable convergence with simulation data, is
- 5 used for following initialization step of NTD (for more details, see Additional File
- 6 3).

To extend the KL form of MU to NTD, we consider iterative updating $\mathbf{A}^{(n)} \mathcal{G}_n \mathbf{A}^{(-n)T}$, which is the matricized expression of Tucker decomposition. Here $\mathbf{A}^{(-n)}$ represents the factor matrices without $\mathbf{A}^{(n)}$. For example, if $n=1$, this part becomes $\mathbf{A}^{(2)T} \mathbf{A}^{(3)T}$. By considering a part of $\mathbf{A}^{(n)} \mathcal{G}_n \mathbf{A}^{(-n)T}$ as a variable and fixing other parts as constants, the KL form of MU can be performed to the matricized tensor, such as $\mathbf{A}^{(1)} \rightarrow \mathbf{A}^{(2)} \rightarrow \mathbf{A}^{(3)} \rightarrow \mathcal{G} \rightarrow \dots$. Each updating rule for $\mathbf{A}^{(n)}$ is as follows:

$$\mathbf{A}^{(n)} \leftarrow \mathbf{A}^{(n)} * \frac{\frac{\mathbf{X}_{(n)}}{\mathbf{A}^{(n)} \mathcal{G}_A^{(n)}} \mathcal{G}_A^{(n)T}}{\mathbf{1} \mathbf{z}^T}. \quad (8)$$

Additionally, the updating rule for core tensor \mathcal{G} is:

$$\begin{aligned} \mathcal{G}^{(n)} &\leftarrow \max\{\chi \times_1 \mathbf{A}^{(1)T} \times_2 \mathbf{A}^{(2)T} \times_3 \mathbf{A}^{(3)T}, \epsilon\} \\ \mathcal{G}^{(n)} &\leftarrow \frac{\chi \times_1 \mathbf{A}^{(1)T} \times_2 \mathbf{A}^{(2)T} \times_3 \mathbf{A}^{(3)T}}{\mathcal{G} \times_1 \mathbf{A}^{(1)T} \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)T} \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)T} \mathbf{A}^{(3)}}, \end{aligned} \quad (9)$$

1 where ϵ is a small value included to avoid generating negative values. In the
2 `nnTensor`, 1e-10 is used.

3 *Extraction of CCIs as hypergraphs*

4 To extract the representative *CaHs*, `scTensor` estimates the NTD ranks by SVD
5 performed for each matricized CCI-tensor ($X^{(n)}$, $n=1,2$, and 3). The eigenvalues
6 and eigenvectors that explain the top 80% to 90% of variance are selected. With
7 the estimated ranks of NTD ($\hat{R}1, \hat{R}2, \hat{R}3$), NTD is performed, and only the triads
8 $(r1,r2,r3)$ with large core tensor values are selected as representative *CaHs*. In its
9 default mode, `scTensor` selects the *CaHs* that explain the top 40% of cumulative
10 core tensor values. For each *CaH*, corresponding column vectors of factor matrices
11 were selected as $CaH(r1, r2, r3) = A_{:r1}^{(1)} \circ A_{:r2}^{(2)} \circ A_{:r3}^{(3)}$.

12 To enhance the interpretation, each column vector is binarized in advance by
13 two-class hierarchical clustering using Ward's minimum variance method, and only
14 large values are converted to 1, with other values becoming 0.

15 CCI-strength (cf. Figure 3, 4, 5, 6, 7) is calculated as the summation of mode-3
16 of reconstructed tensor from all *CaHs*. With the selected indexes in each mode
17 $(r1' \in \{1..\hat{R}1\}, r2' \in \{1..\hat{R}2\}, r3' \in \{1..\hat{R}3\})$, CCI-strength is defined as follows:

$$CCI\text{-strength}(i, j) = \sum_{k=1}^K \left[\sum_{r1=1}^{\hat{R}1} \sum_{r2=1}^{\hat{R}2} \sum_{r3=1}^{\hat{R}3} \mathcal{G}(r1, r2, r3) A_{:r1}^{(1)} \circ A_{:r2}^{(2)} \circ A_{:r3}^{(3)} \right]_{ijk} \quad (10)$$

1 *Label permutation method*

2 In this method, the cluster labels of all cells are randomly permuted 1000 times, and
3 the average ligand expression level of a cluster and the average receptor expression
4 level of a cluster are calculated [37]. For each L-R pair, the mean values of the
5 averaged L-R expression level are calculated in all possible combinations of the cell
6 types. This process generates 1,000 of synthetic L-R coexpression matrices and these
7 are used to generate the null distribution, that is, in a combination of cell types, the
8 proportion of the means which are “as or more extreme” than the observed mean
9 is the calculated as *P*-value.

10 **Availability and requirements**

- 11 • scTensor: <https://bioconductor.org/packages/devel/bioc/html/scTensor.html>
- 12 • nnTensor: <https://cran.r-project.org/web/packages/nnTensor/index.html>
- 13 • LRBase.Hsa.eg.db: [https://bioconductor.org/packages/release/bioc/](https://bioconductor.org/packages/release/bioc/html/LRBase.Hsa.eg.db.html)
14 [html/LRBase.Hsa.eg.db.html](https://bioconductor.org/packages/release/bioc/html/LRBase.Hsa.eg.db.html)
- 15 • LRBase.Mmu.eg.db: [https://bioconductor.org/packages/release/bioc/](https://bioconductor.org/packages/release/bioc/html/LRBase.Mmu.eg.db.html)
16 [html/LRBase.Mmu.eg.db.html](https://bioconductor.org/packages/release/bioc/html/LRBase.Mmu.eg.db.html)
- 17 • LRBase.Ath.eg.db: [https://bioconductor.org/packages/release/bioc/](https://bioconductor.org/packages/release/bioc/html/LRBase.Ath.eg.db.html)
18 [html/LRBase.Ath.eg.db.html](https://bioconductor.org/packages/release/bioc/html/LRBase.Ath.eg.db.html)
- 19 • LRBase.Rno.eg.db: [https://bioconductor.org/packages/release/bioc/](https://bioconductor.org/packages/release/bioc/html/LRBase.Rno.eg.db.html)
20 [html/LRBase.Rno.eg.db.html](https://bioconductor.org/packages/release/bioc/html/LRBase.Rno.eg.db.html)
- 21 • LRBase.Bta.eg.db: [https://bioconductor.org/packages/release/bioc/](https://bioconductor.org/packages/release/bioc/html/LRBase.Bta.eg.db.html)
22 [html/LRBase.Bta.eg.db.html](https://bioconductor.org/packages/release/bioc/html/LRBase.Bta.eg.db.html)
- 23 • LRBase.Cel.eg.db: [https://bioconductor.org/packages/release/bioc/](https://bioconductor.org/packages/release/bioc/html/LRBase.Cel.eg.db.html)
24 [html/LRBase.Cel.eg.db.html](https://bioconductor.org/packages/release/bioc/html/LRBase.Cel.eg.db.html)
- 25 • LRBase.Dme.eg.db: [https://bioconductor.org/packages/release/bioc/](https://bioconductor.org/packages/release/bioc/html/LRBase.Dme.eg.db.html)
26 [html/LRBase.Dme.eg.db.html](https://bioconductor.org/packages/release/bioc/html/LRBase.Dme.eg.db.html)

- 1 • LRBase.Dre.eg.db: <https://bioconductor.org/packages/release/bioc/html/LRBase.Dre.eg.db.html>
- 2
- 3 • LRBase.Gga.eg.db: <https://bioconductor.org/packages/release/bioc/html/LRBase.Gga.eg.db.html>
- 4
- 5 • LRBase.Pab.eg.db: <https://bioconductor.org/packages/release/bioc/html/LRBase.Pab.eg.db.html>
- 6
- 7 • LRBase.Xtr.eg.db: <https://bioconductor.org/packages/release/bioc/html/LRBase.Xtr.eg.db.html>
- 8
- 9 • LRBase.Ssc.eg.db: <https://bioconductor.org/packages/release/bioc/html/LRBase.Ssc.eg.db.html>
- 10
- 11 • LRBaseDbi: <https://bioconductor.org/packages/release/bioc/html/LRBaseDbi.html>
- 12
- 13 • Operating system: Linux, Mac OS X, Windows
- 14
- 15 • Programming language: R (v-3.5.0 or higher)
- 16
- 17 • License: Artistic-2.0
- 18
- 19 • Any restrictions to use by non-academics: For non-profit use only

16 **Abbreviations**

17 CCI: cell-cell interaction; scRNA-seq: single-cell RNA sequencing; L-R: ligand and receptor; *CaH*: CCI as
18 hypergraph; FTT: Freeman-Tukey transformation; NTD: non-negative Tucker decomposition; ROC: receiver
19 operating characteristic; AUC: area under the curve; FGCs: fetal germ cells; Soma: gonadal niche cells; GPCR: G
20 protein-coupled receptor; CAFs: cancer-associated fibroblasts; MHC: major histocompatibility complex; HLA: human
21 leukocyte antigen; RNA-FISH: RNA-fluorescence *in situ* hybridization; PM: plasma membrane; PPI:
22 protein-protein interaction; GO: Gene Ontology; MeSH: medical subject headings; ORA: over-representative
23 analysis; DO: Disease Ontology; NCG: network of cancer gene; CMap: Connectivity Map; Amazon S3: Amazon
24 simple storage service; FC: fold-change; DEGs: differentially expressed genes; human embryonic stem cells: hESCs;
25 HVGs: highly variable genes; t-SNE: t-distributed stochastic neighbor embedding; CP: CANDECOMP/PARAFAC;
26 ALS: alternative least squares; HOSVD: higher order singular value decomposition; HOOI: higher orthogonal
27 iteration of tensors; NMF: non-negative matrix factorization; MU: multiplicative updating; KL: Kullback-Leibler

28 **Competing interests**

29 The authors declare that they have no competing interests.

30 **Funding**

31 This work was supported by MEXT KAKENHI Grant Number 16K16152 and by JST CREST grant number
32 JPMJCR16G3, Japan.

1 **Authors' contributions**

2 KT and IN designed the study. KT designed the algorithm and benchmark, retrieved and preprocessed the test data
3 to evaluate the proposed method, implemented the source code, and performed all analyses. MI implemented the
4 pipeline for bi-annual automatic update of the R/Bioconductor packages. All authors have read and approved the
5 manuscript.

6 **Acknowledgements**

7 Some cell images used in Figure 1 were previously presented by © 2016 DBCLS TogoTV. We thank Dr. Yoshihiro
8 Taguchi for discussions about the algorithms. We thank Mr. Akihiro Matsushima for their assistance with the IT
9 infrastructure for the data analysis. We are also grateful to all members of the Laboratory for Bioinformatics
10 Research, RIKEN Center for Biosystems Dynamics Research for their helpful advice.

11 **Author details**

12 ¹Laboratory for Bioinformatics Research RIKEN Center for Biosystems Dynamics Research, Wako, 351-0198,
13 Saitama, Japan. ²Bioinformatics Course, Master's/Doctoral Program in Life Science Innovation (T-LSI), School of
14 Integrative and Global Majors (SIGMA), University of Tsukuba, Wako, 351-0198, Saitama, Japan.

15 **References**

- 16 1. Yu, Y., Elble, R.C.: Homeostatic signaling by cell–cell junctions and its dysregulation during cancer progression.
17 *Journal of Clinical Medicine* **5**(2), 26 (2016)
- 18 2. Livshits, G., Kobiela, A., Fuchs, E.: Governing epidermal homeostasis by coupling cell–cell adhesion to integrin
19 and growth factor signaling, proliferation, and apoptosis. *PNAS* **109**(3), 4886–4891 (2012)
- 20 3. Chao, D.L., Ma, L., Shen, K.: Transient cell-cell interactions in neural circuit formation. *Nature Review*
21 *Neuroscience*, 262–271 (2009)
- 22 4. Kasukawa, T., Masumoto, K., Nikaido, I., Nagano, M., Uno, K.D., Tsujino, K., Hanashima, C., Shigeyoshi, Y.,
23 Ueda, H.R.: Quantitative expression profile of distinct functional regions in the adult mouse brain. *PLOS ONE*,
24 23228 (2011)
- 25 5. Miller, J.F.A.P., Mitchell, G.F.: Cell to cell interaction in the immune response v. target cells for tolerance
26 induction. *Journal of Experimental Medicine* **131**(4), 675–699 (1970)
- 27 6. Pieters, T., Roy, V.F.: Role of cell–cell adhesion complexes in embryonic stem cell biology. *Journal of Cell*
28 *Science* **127**, 2603–2613 (2014)
- 29 7. Tweedell, K.S.: The adaptability of somatic stem cells: A review. *Journal of Stem Cells and Regenerative*
30 *Medicine* **13**(1), 3–13 (2017)
- 31 8. Plaks, V., Kong, N., Werb, Z.: The cancer stem cell niche: How essential is the niche in regulating stemness of
32 tumor cells? *Cell Stem Cell* **16**, 225–238 (2015)
- 33 9. Hegerfeldt, Y., Tusch, M., Brocker, E.B., Friedl, P.: Collective cell movement in primary melanoma explants:
34 Plasticity of cell-cell interaction, beta1-integrin function, and migration strategies. *Cancer Research* **62**,
35 2125–2130 (2002)
- 36 10. Hofschroer, V., Koch, K.A., Ludwig, F.T., Friedl, P., Oberleithner, H., Stock, C., Schwab, A.: Extracellular
37 protonation modulates cell-cell interaction mechanics and tissue invasion in human melanoma cells. *Scientific*
38 *Reports* **7**(42369) (2017)
- 39 11. Stein, J.V., Gonzalez, S.F.: Dynamic intravital imaging of cell-cell interactions in the lymph node. *Mechanisms*
40 *of allergic diseases* **139**(1), 12–20 (2016)
- 41 12. Reinhar-King, C.A., Dembo, M., Hammer, D.A.: Cell-cell mechanical communication through compliant
42 substrates. *Biophysical Journal* **95**, 6044–6051 (2008)

- 1 13. Dewji, N.N., Mukhopadhyay, D., Singer, S.J.: An early specific cell–cell interaction occurs in the production of
2 beta-amyloid in cell cultures. *PNAS* **103(5)**, 1540–1545 (2006)
- 3 14. Konry, T., Sarkar, S., Sabhachandani, P., Cohen, N.: Innovative tools and technology for analysis of single cells
4 and cell–cell interaction. *The Annual Review of Biomedical Engineering* **18**, 259–284 (2016)
- 5 15. Rothbauer, M., Zirath, H., Ertl, P.: Recent advances in microfluidic technologies for cell-to-cell interaction
6 studies. *Lab on Chip* **18(2)**, 249–270 (2018)
- 7 16. Li, R., Lv, X., Zhang, X., Saeed, O., Deng, Y.: Microfluidics for cell-cell interactions: A review. *10(1)* **10(1)**,
8 90–98 (2016)
- 9 17. Wiklund, M., Christakou, A.E., Ohlin, M., Iranmanesh, I., Frisk, T., Vanherberghen, V., Onfelt, B.:
10 Ultrasound-induced cell–cell interaction studies in a multi-well microplate. *Micromachines* **5**, 27–49 (2014)
- 11 18. Tauriainen, J., Gustafsson, K., Gothlin, M., Gertow, J., Buggert, M., Frisk, T.W., Karlsson, A.C., Uhlin, M.,
12 Onfelt, B.: Single-cell characterization of in vitro migration and interaction dynamics of t cells expanded with
13 il-2 and il-7. *Frontiers in Immunology* **6**, 196 (2015)
- 14 19. Merouane, A., Rey-Villamizar, N., Lu, Y., Liadi, I., Romain, G., Lu, J., Singh, H., Cooper, L.J.N., Varadarajan,
15 N., Roysam, B.: Automated profiling of individual cell–cell interactions from high-throughput time-lapse
16 imaging microscopy in nanowell grids (timing). *Bioinformatics* **31(19)**, 3189–3197 (2015)
- 17 20. Espulgar, W., Yamaguchi, Y., Aoki, W., Mita, D., Saito, M., Lee, J.K., Tamiya, E.: Single cell trapping and
18 cell–cell interaction monitoring of cardiomyocytes in a designed microfluidic chip. *Sensors and Actuators B:*
19 *Chemical* **207**, 43–50 (2015)
- 20 21. Sarkar, S., Sabhachandani, P., Stroopinsky, D., Palmer, K., Cohen, N., Rosenblatt, J., Avigan, D., Konry, T.:
21 Dynamic analysis of immune and cancer cell interactions at single cell level in microfluidic droplets.
22 *Biomicrofluidics* **10(5)**, 054115 (2016)
- 23 22. Dura, B., Dougan, S.K., Barisa, M., Hoehl, M.M., Lo, C.T., Ploegh, H.L., Voldman, J.: Profiling lymphocyte
24 interactions at the single-cell level by microfluidic cell pairing. *Nature Communication* **6(5940)** (2015)
- 25 23. Ramiłowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V.P., Itoh, M., Kawaji,
26 H., Carninci, P., Rost, B., Forrest, A.R.R.: A draft network of ligand–receptor-mediated multicellular signalling
27 in human. *Nature Communication* **22(6)**, 7866 (2015)
- 28 24. Camp, J.G., Sekin, K., Gerber, T., Loeffler-Wirth, H., Binder, H., Gac, M., Kanton, S., Kageyama, J., Damm,
29 G., Seehofer, D., Belicova, L., Bickle, M., Barsacchi, R., Okuda, R., Yoshizawa, E., Kimura, M., Ayabe, H.,
30 Taniguchi, H., Takebe, T., Treutlein, B.: Multilineage communication regulates human liver bud development
31 from pluripotency. *Nature* **546(7659)**, 533–538 (2017)
- 32 25. Li, L., Dong, J., Yan, L., Yong, J., Liu, X., Hu, Y., Fan, X., Wu, X., Guo, H., Wang, X., Zhu, X., Li, R., Yan,
33 J., Wei, Y., Zhao, Y., Wang, W., Ren, Y., Yuan, P., Yan, Z., Hu, B., Guo, F., Wen, L., Tang, F., Qiao, J.:
34 Single-cell rna-seq analysis maps development of human germline cells and gonadal niche interactions. *Cell*
35 *Stem Cell* **20**, 858–873 (2017)
- 36 26. Zhou, J.X., Taramelli, R., Pedrini, E., Knijnenburg, T., Hunag, S.: Extracting intercellular signaling network of
37 cancer tissues using ligand-receptor expression patterns from whole-tumor and single-cell transcriptomes.
38 *Scientific Reports* **7(1)**, 8815 (2017)
- 39 27. A, S.D., Squiers, G.T., McLellan, M.A., Bolisetty, M.T., Robson, P., Rosenthal, N.A., Pinto, A.R.: Single-cell
40 transcriptional profiling reveals cellular diversity and intercommunication in the mouse heart. *Cell Reports*
41 **22(3)**, 600–610 (2018)
- 42 28. Pavlicev, M., Wagner, G.P., Chavan, A.R., Owens, K., Maziarz, J., Dunn-Fletcher, C., Lallapur, S.G., Muglia,
43 L., Jones, H.: Single-cell transcriptomics of the human placenta: inferring the cell communication network of

- 1 the maternal-fetal interface. *Genome Research* **27**, 349–361 (2017)
- 2 29. Joost, S., Jacob, T., Sun, X., Annusver, K., La Manno, G., Sur, I., Kasper, M.: Single-cell transcriptomics of
3 traced epidermal and hair follicle stem cells reveals rapid adaptations during wound healing. *Cell Reports* **25(3)**,
4 585–597 (2018)
- 5 30. Kramann, R., Machado, F., Wu, H., Kusaba, T., Hoeft, K., Schneider, R.K., Humphreys, B.D.: Parabiosis and
6 single-cell rna sequencing reveal a limited contribution of monocytes to myofibroblasts in kidney fibrosis. *JCI*
7 *insight* **3(9)**, 99561 (2018)
- 8 31. Cohen, M., Giladi, A., Gorki, A.D., Solodkin, D.G., Zada, M., Hladik, A., Miklosi, A., Salame, T.M., Halpern,
9 K.B., David, E., Itzkovitz, S., Harkany, T., Knapp, S., Amit, I.: Lung single-cell signaling interaction map
10 reveals basophil role in macrophage imprinting. *Cell* **175(4)**, 1031–1044 (2018)
- 11 32. Davidson, S., Efremova, M., Riedel, A., Mahata, B., Pramanik, J., Huuhtanen, J., Kar, G., Vento-Tormo, R.,
12 Hagai, T., Chen, X., Haniffa, M.A., Shields, J.D., Teichmann, S.A.: Single-cell rna sequencing reveals a
13 dynamic stromal niche within the evolving tumour microenvironment. *bioRxiv* (2018). doi:[10.1101/467225](https://doi.org/10.1101/467225)
- 14 33. Potter, S.S., Mucenski, M.L., Mahoney, R., Adam, M., Potter, A.S.: Single cell rna-seq study of wild type and
15 *hox9,10,11* mutant developing uterus. *bioRxiv* (2018). doi:[10.1101/395574](https://doi.org/10.1101/395574)
- 16 34. Wu, H., Uchimura, K., Donnelly, E.L., Kirita, Y., Morris, S.A., Humphreys, B.D.: Comparative analysis and
17 refinement of human psc-derived kidney organoid differentiation with single-cell transcriptomics. *Cell Stem Cell*
18 **23(6)**, 869–881 (2018)
- 19 35. Chen, L., Lee, J.W., Chou, C.L., Nair, A.V., Battistone, M.A., Paunescu, T.G., Merkulova, M., Breton, S.,
20 Verlander, J.W., Wall, S.M., Brown, D., Burg, M.B., Knepper, M.A.: Transcriptomes of major renal collecting
21 duct cell types in mouse identified by single-cell rna-seq. *PNAS* **114(46)**, 9989–9998 (2017)
- 22 36. Menon, R., Otto, E.A., Kokoruda, A., Zhou, J., Zhang, Z., Yoon, E., Chen, Y.C., Troyanskaya, O., Spence,
23 J.R., Kretzler, M., Cebrian, C.: Single-cell analysis of progenitor cell dynamics and lineage specification in the
24 human fetal kidney. *Development* **145(16)** (2018)
- 25 37. Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.E.,
26 Stephenson, E., Polanski, K., Goncalves, A., Gardner, L., Holmqvist, S., Henriksson, J., Zou, A., Sharkey, A.M.,
27 Millar, B., Innes, B., Wood, L., Wilbrey-Clark, A., Payne, R.P., Ivarsson, M.A., Lisgo, S., Filby, A., Rowitch,
28 D.H., Bulmer, J.N., Wright, G.J., Stubbington, M.J.T., Haniffa, M., Moffett, A., Teichmann, S.A.: Single-cell
29 reconstruction of the early maternal–fetal interface in humans. *Nature* **563(7731)**, 347–353 (2018)
- 30 38. Biton, M., Haber, A.L., Rogel, N., Burgin, G., Beyaz, S., Schnell, A., Ashenberg, O., Su, C.W., Smillie, C.,
31 Shekhar, K., Chen, Z., Wu, C., Ordovas-Montanes, J., Alvarez, D., Herbst, R.H., Zhang, M., Tirosh, I.,
32 Dionne, D., Nguyen, L.T., Xifaras, M.E., Shalek, A.K., von Andrian, U.H., Graham, D.B., Rozenblatt-Rosen,
33 O., Shi, H.N., Kuchroo, V., Yilmaz, O.H., Regev, A., Xavier, R.J.: T helper cell cytokines modulate intestinal
34 stem cell renewal and differentiation. *Cell* **175(5)**, 1307–1320 (2018)
- 35 39. Kumar, M.P., Du, J., Lagoudas, G., Jiao, Y., Sawyer, A., Drummond, D.C., Lauffenburger, D.A., Raue, A.:
36 Analysis of single-cell rna-seq identifies cell-cell communication associated with tumor characteristics. *Cell*
37 *Reports* **25(6)**, 1458–1468 (2018)
- 38 40. Single cell analyses of the effects of amyloid-beta42 and interleukin-4 on neural stem/progenitor cell plasticity
39 in adult zebrafish brain **XX(X)**, (20XX)
- 40 41. Verma, M., Asakura, Y., Murakonda, B.S.R., Pengo, T., Latroche, C., Chazaud, B., McLoon, L.K., Asakura,
41 A.: Muscle satellite cell cross-talk with a vascular niche maintains quiescence via vegf and notch signaling. *Cell*
42 *Stem Cell* **23(4)**, 530–543 (2018)
- 43 42. Jerby-Arnon, L., Shah, P., Cuoco, M.S., Rodman, C., Su, M.J., Melms, J.C., Leeson, R., Kanodia, A., Mei, S.,

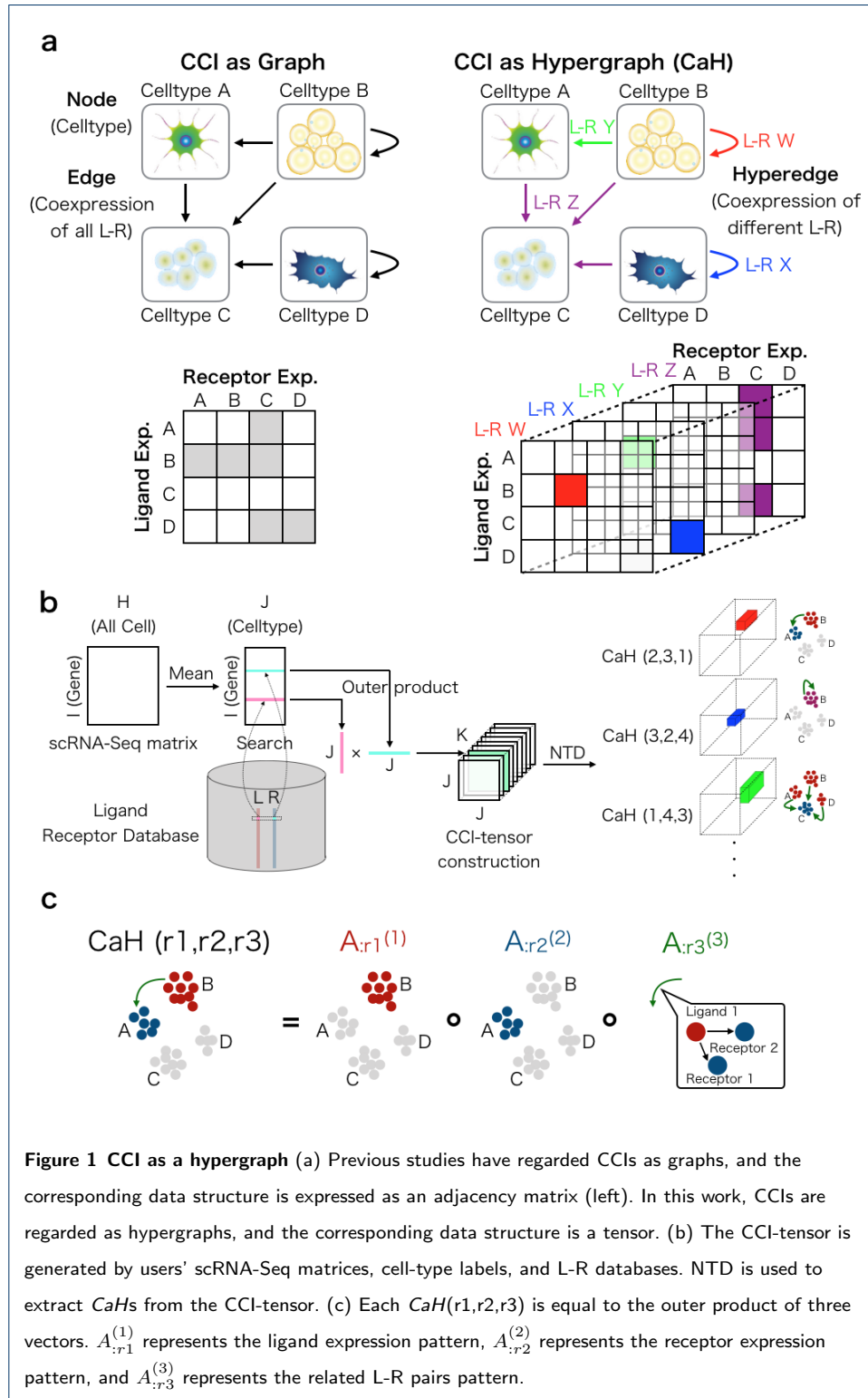
- 1 Lin, J.R., Wang, S., Rabasha, B., Liu, D., Zhang, G., Margolis, C., Ashenberg, O., Ott, P.A., Buchbinder, E.I.,
2 Haq, R., Hodi, F.S., Boland, G.M., Sullivan, R.J., Frederick, D.T., Miao, B., Moll, T., Flaherty, K.T., Herlyn,
3 M., Jenkins, R.W., Thummalapalli, R., Kowalczyk, M.S., Canadas, I., Schilling, B., Cartwright, A.N.R., Luoma,
4 A.M., Malu, S., Hwu, P., Bernatchez, C., Forget, M.A., Barbie, D.A., Shalek, A.K., Tirosh, I., Sorger, P.K.,
5 Wucherpennig, K., Van Allen, E.M., Schadendorf, D., Johnson, B.E., Rotem, A., Rozenblatt-Rosen, O.,
6 Garraway, L.A., Yoon, C.H., Izar, B., Regev, A.: A cancer cell program promotes t cell exclusion and resistance
7 to checkpoint blockade. *Cell* **175(4)**, 984–997 (2018)
- 8 43. Kelleher, A.M., Milano-Foster, J., Behura, S.K., Spencer, T.E.: Uterine glands coordinate on-time embryo
9 implantation and impact endometrial decidualization for pregnancy success. *Nature Communications* **9(1)**, 2435
10 (2018)
- 11 44. Yin, J., Li, Z., Yan, C., Fang, E., Wang, T., Zhou, H., Luo, W., Zhou, Q., Zhang, J., Hu, J., Jin, H., Wang, L.,
12 Zhao, X., Li, J., Qi, X., Zhou, W., Huang, C., He, C., Yang, H., Kristiansen, K., Hou, Y., Zhu, S., Zhou, D.,
13 Wang, L., Dean, M., Wu, K., Hu, H., Li, G.: Comprehensive analysis of immune evasion in breast cancer by
14 single-cell rna-seq. *bioRxiv* (2018). doi:[10.1101/368605](https://doi.org/10.1101/368605)
- 15 45. Biase, F.H., Kimble, K.M.: Functional signaling and gene regulatory networks between the oocyte and the
16 surrounding cumulus cells. *BMC Genomics* **19(1)**, 351 (2018)
- 17 46. Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F.,
18 Plaisier, C.L., Eddy, J.A., Ziv, E., Culhane, A.C., Paull, E.O., Sivakumar, I.K.A., Gentles, A.J., Malhotra, R.,
19 Farshidfar, F., Colaprico, A., Parker, J.S., Mose, L.E., Vo, N.S., Liu, J., Liu, Y., Rader, J., Dhankani, V.,
20 Reynolds, S.M., Bowlby, R., Califano, A., Cherniack, A.D., Anastassiou, D., Bedognetti, D., Rao, A., Chen, K.,
21 Krasnitz, A., Hu, H., Malta, T.M., Noushmehr, H., Peadarallu, C.S., Bullman, S., Ojesina, A.I., Lamb, A.,
22 Zhou, W., Shen, H., Choueiri, T.K., Weinstein, J.N., Guinney, J., Saltz, J., Holt, R.A., Rabkin, C.E., Network,
23 C.G.A.R., Lazar, A.J., Serody, J.S., Demicco, E.G., Disis, M.L., Vincent, B.G., Shmulevich, L.: The immune
24 landscape of cancer. *Immunity* **48(4)**, 812–830 (2018)
- 25 47. Han, X., Chen, H., Huang, D., Chen, H., Fei, L., Cheng, C., Huang, H., Yuan, G.C., Guo, G.: Mapping human
26 pluripotent stem cell differentiation pathways using high throughput single-cell rna-sequencing open access.
27 *BMC Genome Biology* **19(1)**, 47 (2018)
- 28 48. Costa, A., Kieffer, Y., Scholer-Dahirel, A., Pelon, F., Bourachot, B., Cardon, M., Sirven, P., Magagna, I.,
29 Fuhrmann, L., Bernard, C., Bonneau, C., Kondratova, M., Kuperstein, I., Zinovyev, A., Givel, A.M., Parrini,
30 M.C., Soumelis, V., Vincent-Salomon, A., Mechta-Grigoriou, F.: Fibroblast heterogeneity and
31 immunosuppressive environment in human breast cancer. *Cancer Cells* **33(3)**, 463–479 (2018)
- 32 49. Hrvatin, S., Hochbaum, D.R., Nagy, M.A., Cicconet, M., Robertson, K., Cheadle, L., Zilionis, R., Ratner, A.,
33 Borges-Monroy, R., Klein, A.M., Sabatini, B.L., Greenberg, M.E.: Single-cell analysis of experience-dependent
34 transcriptomic states in the mouse visual cortex. *Nature Neuroscience* **21(1)**, 120–129 (2018)
- 35 50. Suryawanshi, H., Morozov, P., Straus, A., Sahasrabudhe, N., Max, K.E.A., Garzia, A., Kustagi, M., Tuschl, T.,
36 Z, W.: A single-cell survey of the human first-trimester placenta and decidua. *Science Advances* **4(10)**, 4788
37 (2018)
- 38 51. Puram, S.V., Tirosh, I., Parikh, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C.L., Mroz, E.A.,
39 Emerick, K.S., Deschler, D.G., Varvares, M.A., Mylvaganam, R., Rozenblatt-Rosen, O., Rocco, J.W., Faquin,
40 W.C., Lin, D.T., Regev, A., Bernstein, B.E.: Single-cell transcriptomic analysis of primary and metastatic tumor
41 ecosystems in head and neck cancer. *Cell* **171(7)**, 1611–1624 (2017)
- 42 52. Ximerakis, M., Lipnick, S.L., Simmons, S.K., Adiconis, X., Innes, B.T., Dionne, D., Nguyen, L., Mayweather,
43 B.A., Ozek, C., Niziolek, Z., Butty, V.L., Isserlin, R., Buchanan, S.M., Levine, S.R., Regev, A., Bader, G.D.,

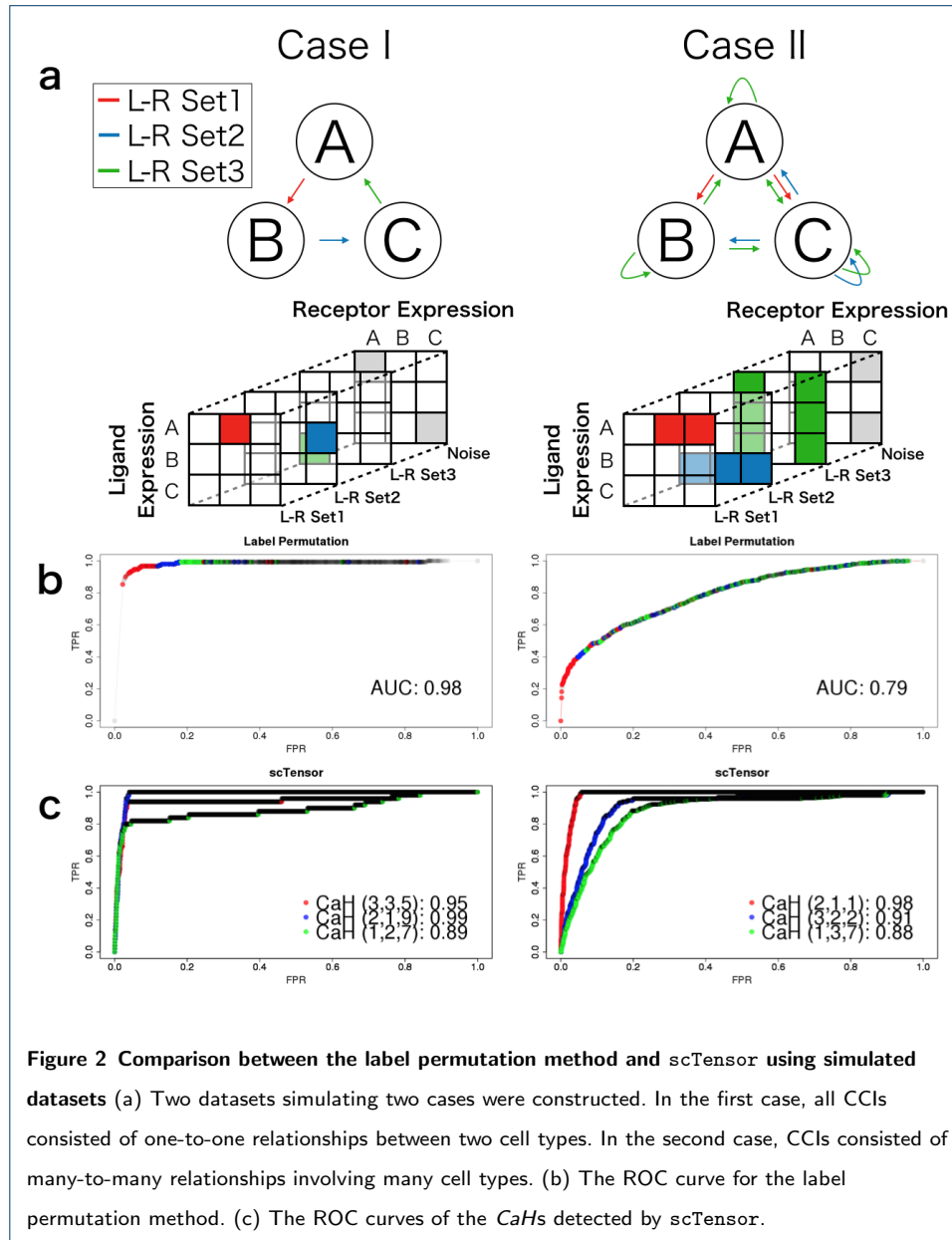
- 1 Levin, J.Z., Rubin, L.L.: Single-cell transcriptomics of the aged mouse brain reveals convergent, divergent and
- 2 unique aging signatures. *bioRxiv* (2018). doi:[10.1101/440032](https://doi.org/10.1101/440032)
- 3 53. Sivakamasundari, V., Bolisetty, M., Sivajothi, S., Bessonett, S., Ruan, D., Robson, P.: Comprehensive cell type
- 4 specific transcriptomics of the kidney. *bioRxiv* (2017). doi:[10.1101/238063](https://doi.org/10.1101/238063)
- 5 54. Yuanxin, W., Ruiping, W., Shaojun, Z., Shumei, S., Changying, J., Guangchun, H., Michael, W., Jaffer, A.,
- 6 Andy, F., Wang, L.: italk: an r package to characterize and illustrate intercellular communication. *bioRxiv*
- 7 (2019). doi:[10.1101/507871](https://doi.org/10.1101/507871)
- 8 55. Boisset, J.C., Vivie, J., Murano, M.J., Lyubimova, A., van-Oudenaarden, A.: Mapping the physical network of
- 9 cellular interactions. *Nature Methods* (2018)
- 10 56. Freeman, M.F., Tukey, J.W.: Transformations related to the angular and the square root. *The Annals of*
- 11 *Mathematical Statistics* **4(21)**, 607–611 (1950)
- 12 57. Kim, Y.-D., Choi, S.: Nonnegative tucker decomposition. In: *In IEEE Conference on Computer Vision and*
- 13 *Pattern Recognition* (2007)
- 14 58. Cichocki, A., Zdunek, R., Amari, S.: Nonnegative Matrix and Tensor Factorizations. *IEEE Signal Processing*
- 15 *Magazine*, ??? (2008)
- 16 59. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H. II, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C.,
- 17 Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., , L. Jia-Ren, Cohen, O., Shah, P., Lu, D.,
- 18 Genshaft, A.S., Hughes, T.K., Ziegler, C.G.K., Kazer, S.W., Gaillard, A., Kolb, K.E., Villani, A.C.,
- 19 Johannessen, C.M., Andreev, A.Y., Van Allen, E.M., Bertagnolli, M., Sorger, P.K., Sullivan, R.J., Flaherty,
- 20 K.T., Frederick, D.T., Jane-Valbuena, J., Yoon, C.H., ROzenblatt-Rosen, O., Shalek, A.K., Regev, A.,
- 21 Garraway, L.A.: Dissecting the multicellular ecosystem of metastatic melanoma by single-cell rna-seq. *Science*
- 22 **352(6282)**, 189–196 (2016)
- 23 60. Pandey, S., Shekhar, K., Regev, A., Schier, A.F.: Comprehensive identification and spatial mapping of
- 24 habenular neuronal types using single-cell rna-seq. *Current Biology* **28**, 1052–1065 (2018)
- 25 61. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., Regev, A.: Spatial reconstruction of single-cell gene
- 26 expression data. *Nature Computational Biology* **33(5)**, 495–502 (2015)
- 27 62. Svensson, V., Teichmann, S.A., Stegle, O.: Spatialde: identification of spatially variable genes. *Nature methods*
- 28 **15**, 343–346 (2018)
- 29 63. Durinck, S., Spellman, P., Birney, E., Huber, W.: Mapping identifiers for the integration of genomic datasets
- 30 with the r/bioconductor package biomart. *Nature Protocols* **4**, 1184–1191 (2009)
- 31 64. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B.,
- 32 Korninger, F., May, B., Milacic, M., Roca, C.D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai,
- 33 T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P.: The reactome pathway
- 34 knowledgebase. *Nucleic Acids Research* **46(D1)**, 649–655 (2018)
- 35 65. Tsuyuzaki, K., Morota, G., Ishii, M., Nakazato, T., Miyazaki, S., Nikaido, I.: Mesh ora framework:
- 36 R/bioconductor packages to support mesh over-representation analysis. *BMC Bioinformatics* **16(45)** (2015)
- 37 66. Falcon, S., Gentleman, R.: Using gostats to test gene lists for go term association. *Bioinformatics* **23(2)**,
- 38 257–258 (2007)
- 39 67. Yu, G., He, Q.: Reactomepa: an r/bioconductor package for reactome pathway analysis and visualization.
- 40 *Molecular BioSystems* **12(12)**, 477–479 (2016)
- 41 68. Yu, G., Wang, L., Yan, G., He, Q.: Dose: an r/bioconductor package for disease ontology semantic and
- 42 enrichment analysis. *Bioinformatics* **31(4)**, 608–609 (2015)
- 43 69. Ono, H., Ogasawara, O., Okubo, K., Bono, H.: Refex, a reference gene expression dataset as a web tool for the

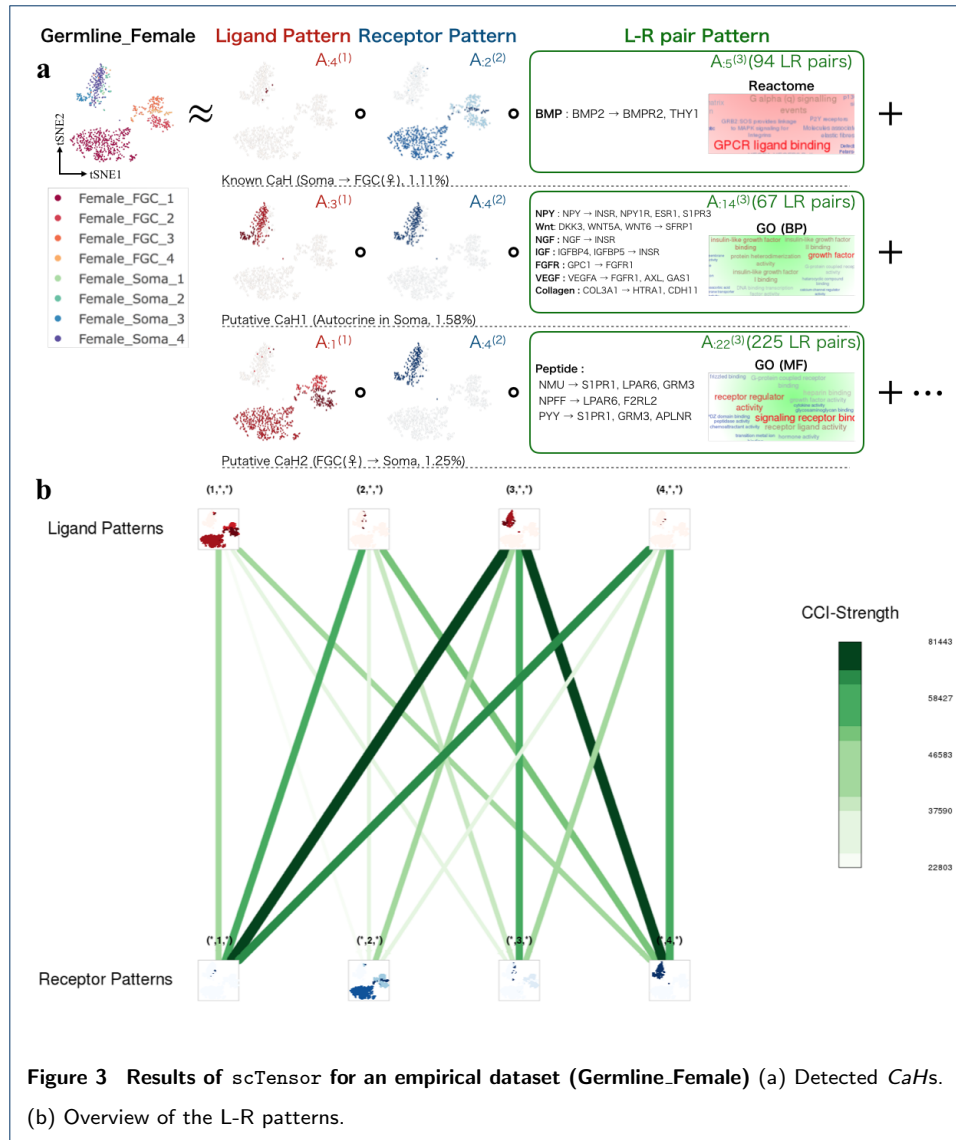
- 1 functional analysis of genes. *Scientific Data* **4**, 170105 (2017)
- 2 70. Papatheodorou, I., Fonseca, N.A., Keays, M., Tang, Y.A., Barrera, E., Bazant, W., Burke, M., Fullgrabe, A.,
3 Fuentes, A.M., George, N., Huerta, L., Koskinen, S., Mohammed, S., Geniza, M., Preece, J., Jaiswal, P.,
4 Jarnuczak, A.F., Huber, W., Stegle, O., Vizcaino, J.A., Brazma, A., Petryszak, R.: Expression atlas: gene and
5 protein expression across multiple studies and organisms. *Nucleic Acids Research* **46(Database issue)**, 246–51
6 (2018)
- 7 71. Single Cell Expression Atlas Single cell gene expression across species. <https://www.ebi.ac.uk/gxa/sc/home>
- 8 72. Cao, Y., Zhu, J., Jia, P., Zhao, Z.: scrnaseqdb: A database for rna-seq based gene expression profiles in human
9 single cells. *Genes (Basel)* **8(12)**, 368 (2017)
- 10 73. PanglaoDB. <https://panglaoDB.se/index.html>
- 11 74. Carithers, L.J., Ardlie, K., Barcus, M., Branton, P.A., Britton, A., Buia, S.A., Compton, C.C., DeLuca, D.S.,
12 Peter-Demchok, J., Gelfand, E.T., Guan, P., Korzeniewski, G.E., Lockhart, N.C., Rabiner, C.A., Rao, A.K.,
13 Robinson, K.L., Roche, N.V., Sawyer, S.J., Segre, A.V., Shive, C.E., Smith, A.M., Sobin, L.H., Undale, A.H.,
14 Valentino, K.M., Vaught, J., Young, T.R., Moore, H.M., Consortium., G.: A novel approach to high-quality
15 postmortem tissue procurement: The gtex project. *Biopreservation and Biobanking* **13(5)**, 311–319 (2015)
- 16 75. Collection — 29 August 2017 The FANTOM5 project. <https://www.nature.com/collections/jcxdjndxy>
- 17 76. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M.,
18 Marra, M.A., Beaudet, A.L., Ecker, J.R., Farnham, P.J., Hirst, M., Lander, E.S., Mikkelsen, T.S., Thomson,
19 J.A.: The nih roadmap epigenomics mapping consortium. *Nature Biotechnology* **28(10)**, 1045–1048 (2010)
- 20 77. Consotium, E.P.: An integrated encyclopedia of dna elements in the human genome. *Nature* **489(7414)**, 57–74
21 (2012)
- 22 78. Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf,
23 C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szgyarto, C.A., Odeberg, J.,
24 Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg,
25 J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F.,
26 Zwahlen, M., von Heijne, G., Nielsen, J., F, P.: Proteomics. tissue-based map of the human proteome. *Science*
27 **347(6220)**, 1260419 (2015)
- 28 79. Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P.,
29 Subramanian, A., Ross, K.N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S.A., Haggarty, S.J., Clemons,
30 P.A., Wei, R., Carr, S.A., Lander, E.S., Golub, T.R.: The connectivity map: using gene-expression signatures to
31 connect small molecules, genes, and disease. *Science* **313(5795)**, 1929–1935 (2006)
- 32 80. Harding, S.D., Sharman, J.L., Faccenda, E., Southan, C., Pawson, A.J., Ireland, S., Gray, A.J.G., Bruce, L.,
33 Alexander, S.P.H., Anderton, S., Bryant, C., Davenport, A.P., Doerig, C., Fabbro, D., Levi-Schaffer, F.,
34 Spedding, M., Davies, N.-I. J A: The iuphar/bps guide to pharmacology in 2018: updates and expansion to
35 encompass the new guide to immunopharmacology. *Nucleic Acids Research* **46(D1)**, 1091–1106 (2018)
- 36 81. Graeber, T.G., Eisenberg, D.: Bioinformatic identification of potential autocrine signaling loops in cancers from
37 gene expression profiles. *Nature Genetics* **29(3)**, 295–300 (2001)
- 38 82. Wang, Y., Tung, H.-Y., Smola, A., Anandkumar, A.: Fast and guaranteed tensor decomposition via sketching.
39 In: *In NIPS*, vol. 1, pp. 991–999 (2015)
- 40 83. Maehara, T., Hayashi, K., Kawarabayashi, K.: Expected tensor decomposition with stochastic gradient descent.
41 In: *AAAI'16*, pp. 1919–1925 (2016)
- 42 84. Smith, S., Park, J., Karypis, G.: An exploration of optimization algorithms for high performance tensor
43 completion. In: *SC '16 Proceedings of the International Conference for High Performance Computing*,

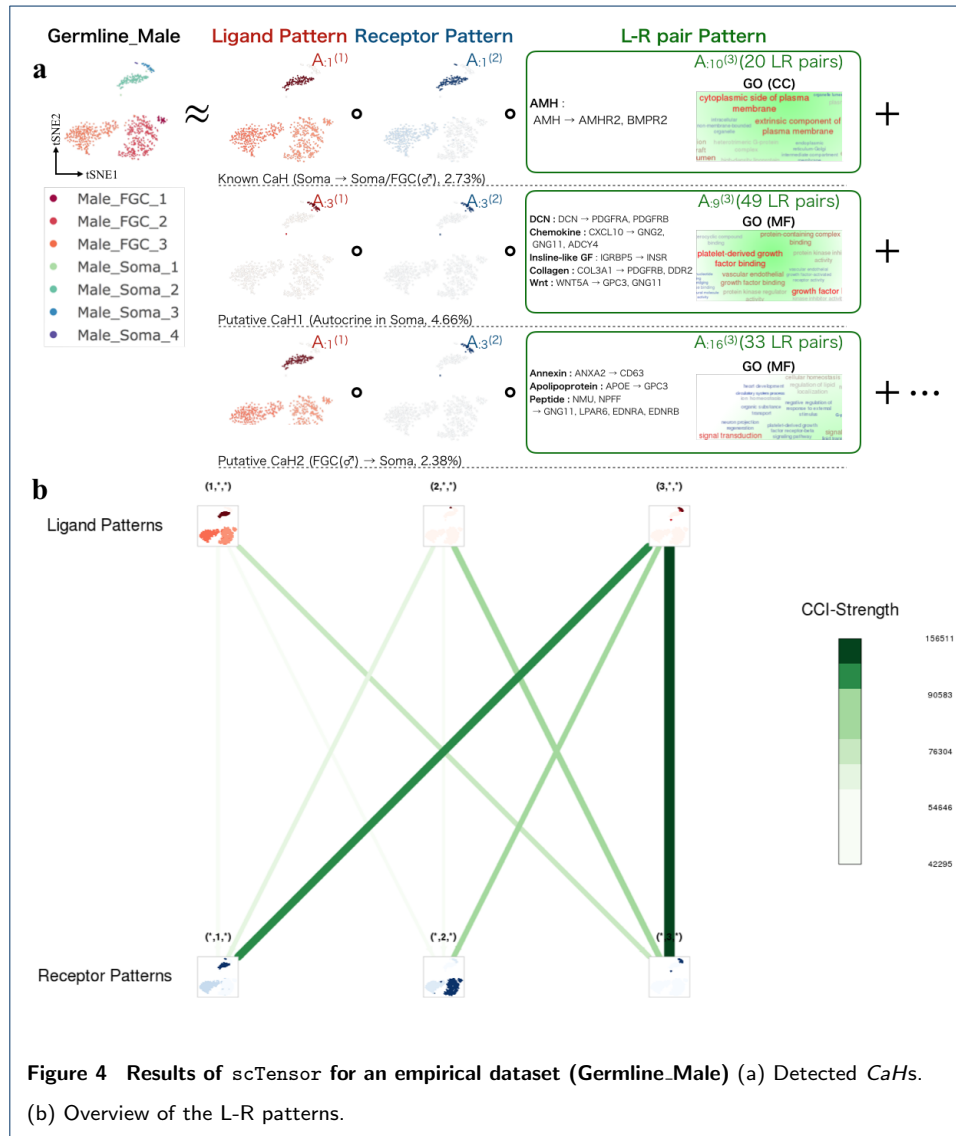
- 1 Networking, Storage and Analysis, vol. 31 (2016)
- 2 85. Shin, K., Sael, L., Kang, U.: Fully scalable methods for distributed tensor factorization. *IEEE Transactions on*
- 3 Knowledge and Data Engineering **29(1)**, 100–113 (2017)
- 4 86. Tsuyuzaki, K., Nikaïdo, I.: Biological systems as heterogeneous information networks: A mini-review and
- 5 perspectives. *HeteroNAM'18* (2018)
- 6 87. Sasagawa, Y., Nikaïdo, I., Hayashi, T., Danno, H., Uno, K.D., Imai, T., Ueda, H.R.: Quartz-seq: a highly
- 7 reproducible and sensitive single-cell rna sequencing method, reveals non-genetic gene-expression heterogeneity.
- 8 *BMC Genome Biology* **14(4)**, 31 (2013)
- 9 88. Pierson, E., Yau, C.: Zifa: Dimensionality reduction for zero-inflated single-cell gene expression analysis. *BMC*
- 10 *Genome Biology* **16(241)** (2015)

1 Figures









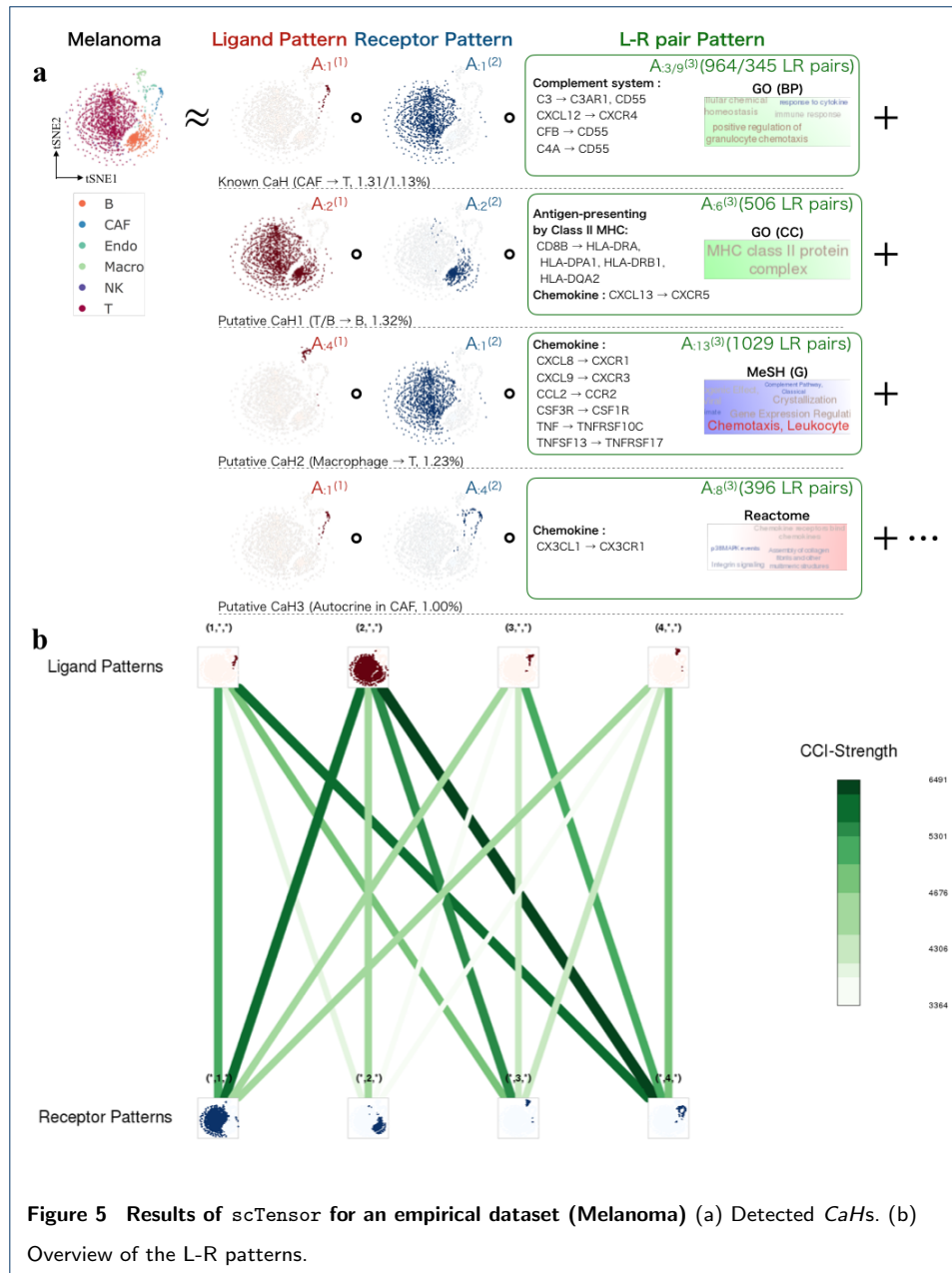
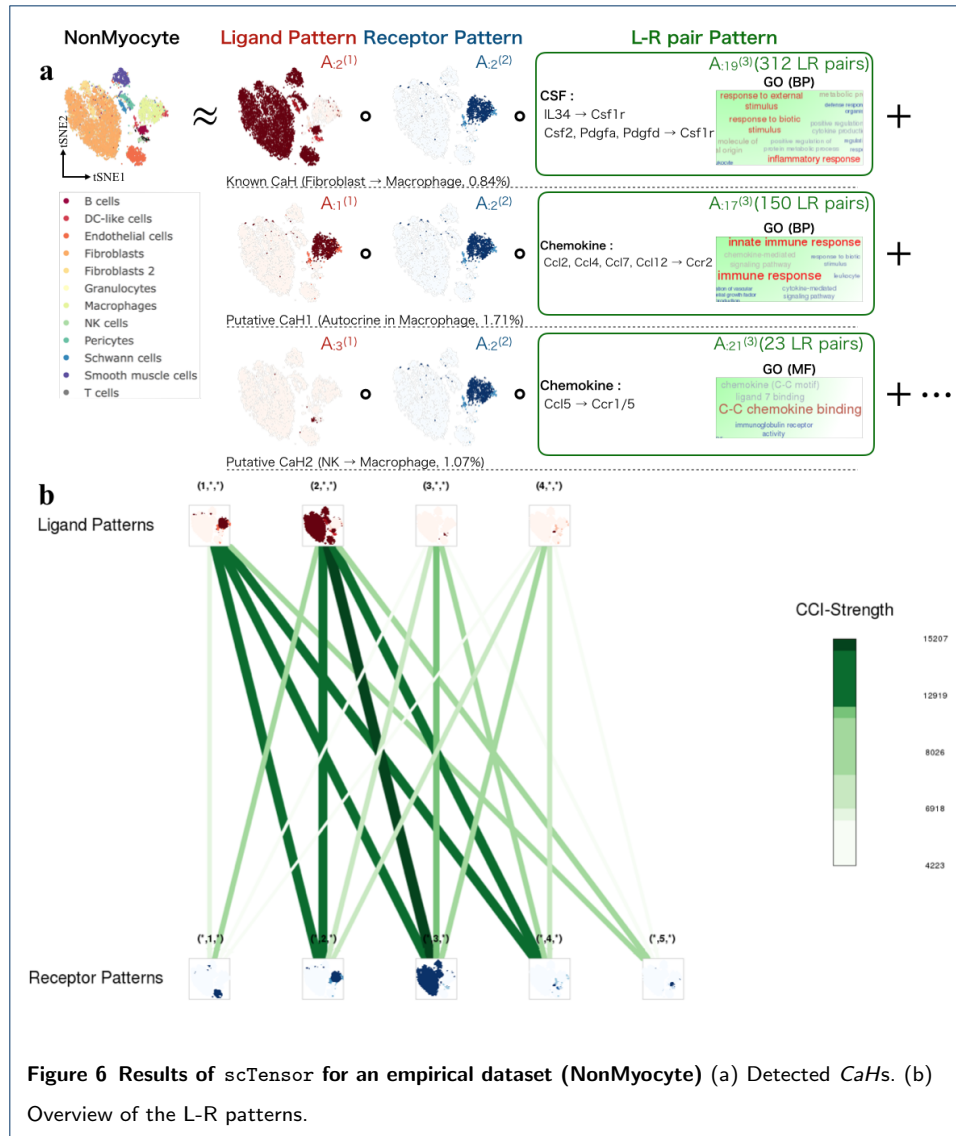
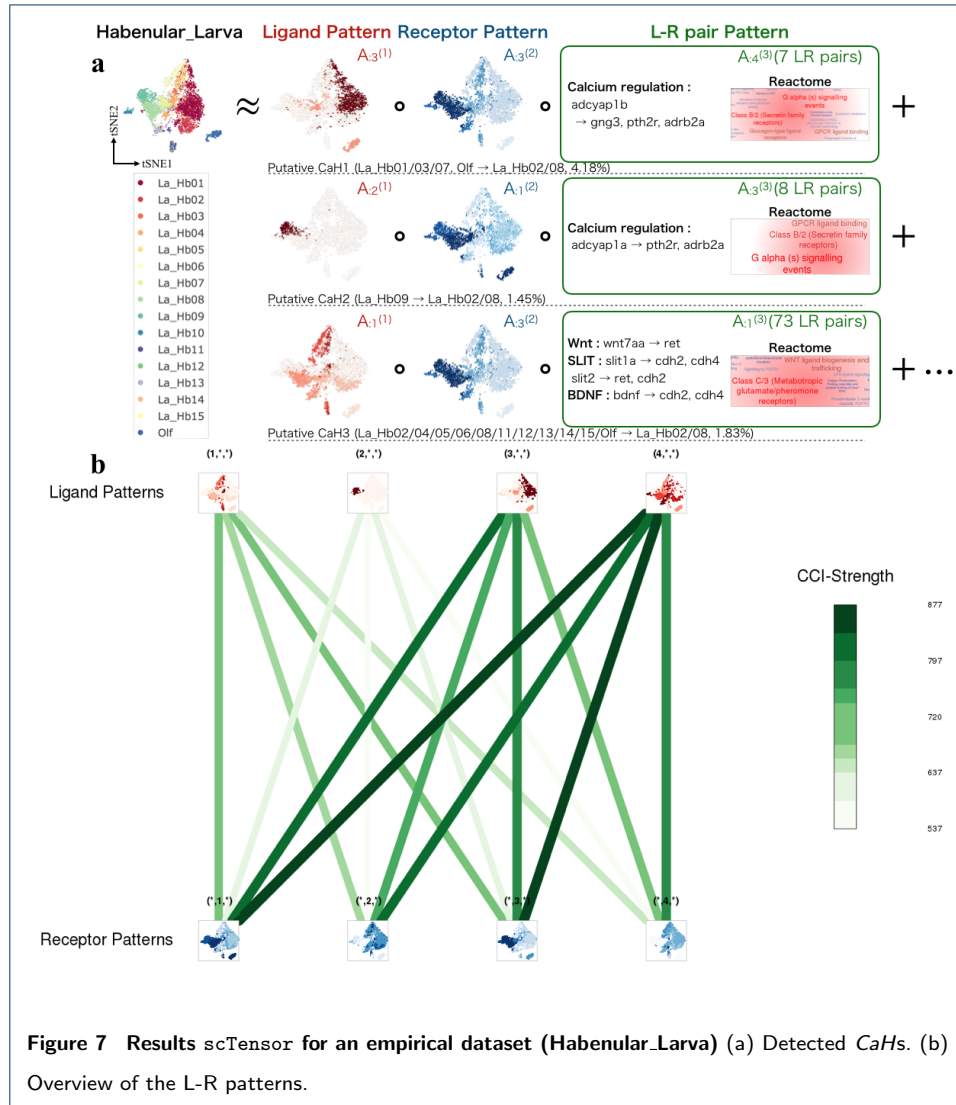


Figure 5 Results of scTensor for an empirical dataset (Melanoma) (a) Detected CaHs. (b) Overview of the L-R patterns.





	Gene Name	Description	GO	STRING	UniProtKB	Reactome	MeSH	GO	MeSH	Reactome	DO/NCI/DisGeNET	RefTex	Expression Atlas	SingleCell Expression Atlas	scRNA-SeqDB	PangaoDB	CMap
	Annotation						Enrichment			Tissue / Celltype				Chem			
Homo sapiens (9606)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mus musculus (10090)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Arabidopsis thaliana (3702)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rattus norvegicus (10116)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Bos taurus (9913)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Caenorhabditis elegans (6239)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Drosophila melanogaster (7227)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Danio rerio (7955)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gallus gallus (9031)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pongo abelli (9601)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Xenopus (Silurana) tropicalis (8364)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sus scrofa (9823)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Figure 8 Correspondence table containing the available L-R gene information for each organism For each detected *CaH*, gene annotation, enrichment analysis, tissue- or cell-type-specific gene expression, and chemical-gene expression relationships are assigned.

1 Tables

Table 1 Empirical datasets

Name	Organisms	ID	# Gene	# Cell	# Cell type	Unit
Germline.Female [25]	<i>Homo sapiens</i>	GSE86146	2717	992	8	TPM
Germline.Male [25]	<i>Homo sapiens</i>	GSE86146	2790	852	7	TPM
Melanoma [59]	<i>Homo sapiens</i>	GSE72056	6603	2250	6	TPM
NonMyocyte [27]	<i>Mus musculus</i>	E-MTAB-6173	2587	10519	12	UMI count
Habenular.Larva [60]	<i>Danio rerio</i>	GSE109158	15206	4233	16	UMI count

Table 2 Summary of LRBase.XXX.eg.db for 12 organisms (1/2)

XXX	Organisms	SWISSPROT (Secreted / Membrane)	TrEMBL (Secreted / Membrane)	STRING (PPI)
Hsa	<i>Homo sapiens</i>	1592 / 2269	176 / 334	18838
Mmu	<i>Mus musculus</i>	1309 / 1806	325 / 1555	19715
Ath	<i>Arabidopsis thaliana</i>	1260 / 1001	244 / 80	24174
Rno	<i>Rattus norvegicus</i>	643 / 983	232 / 1229	19963
Bta	<i>Bos taurus</i>	517 / 448	192 / 390	18349
Cel	<i>Caenorhabditis elegans</i>	198 / 247	28 / 60	13545
Dme	<i>Drosophila melanogaster</i>	249 / 333	89 / 148	11903
Dre	<i>Danio rerio</i>	119 / 169	318 / 376	21746
Gga	<i>Gallus gallus</i>	175 / 173	185 / 154	13084
Pab	<i>Pongo abelii</i>	80 / 134	212 / 211	16691
Xtr	<i>Xenopus Silurana tropicalis</i>	57 / 83	141 / 114	15338
Ssc	<i>Sus scrofa</i>	223 / 153	202 / 445	18683

Table 3 Summary of LRBase.XXX.eg.db for 12 organisms (2/2)

XXX	# L-R Pairs (SWISSPROT × STRING)	# L-R Pairs (TrEMBL × STRING)
Hsa	21882	472
Mmu	16386	476
Ath	8697	94
Rno	5270	65
Bta	2220	237
Cel	106	1
Dme	384	9
Dre	99	432
Gga	140	105
Pab	34	184
Xtr	19	107
Ssc	277	130

1 **Additional Files**

- 2 Additional file 1 — Development of L-R databases for multiple organisms (PDF 2.2 MB)
- 3 Additional file 2 — Distributions of and correlations among 8 STRING-scores (ZIP 18.5 MB)
- 4 Additional file 3 — Convergence of NTD with toy model and empirical data (PDF 9.5 MB)