ISMB Tutorial 2024, Montreal, Canada

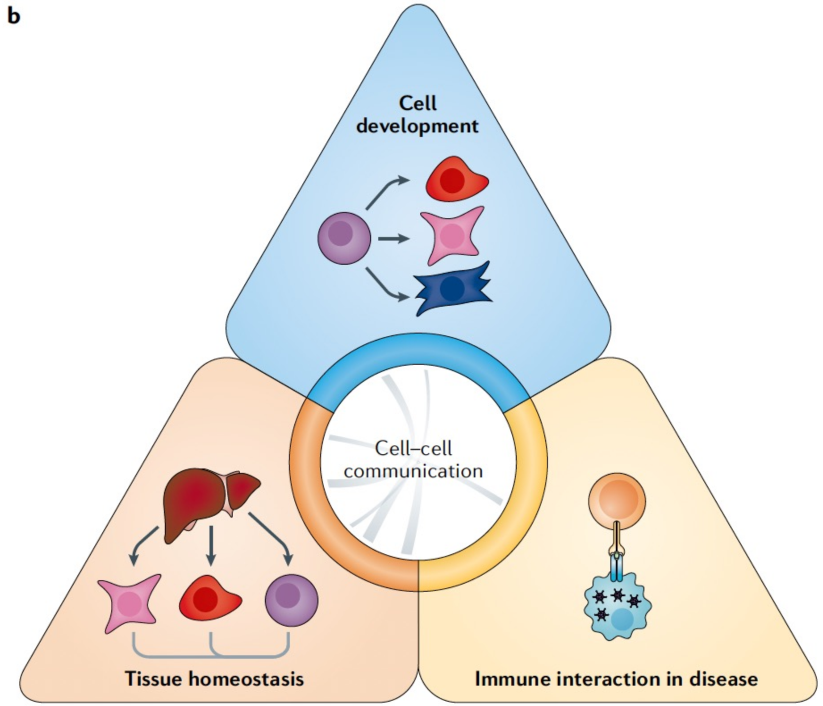
Cell cell communications in spatial transcriptomics

Dr. Juexin Wang
Department of BioHealth Informatics
Luddy School of Informatics, Computing, and Engineering

07/12/2024

Cell–cell interactions

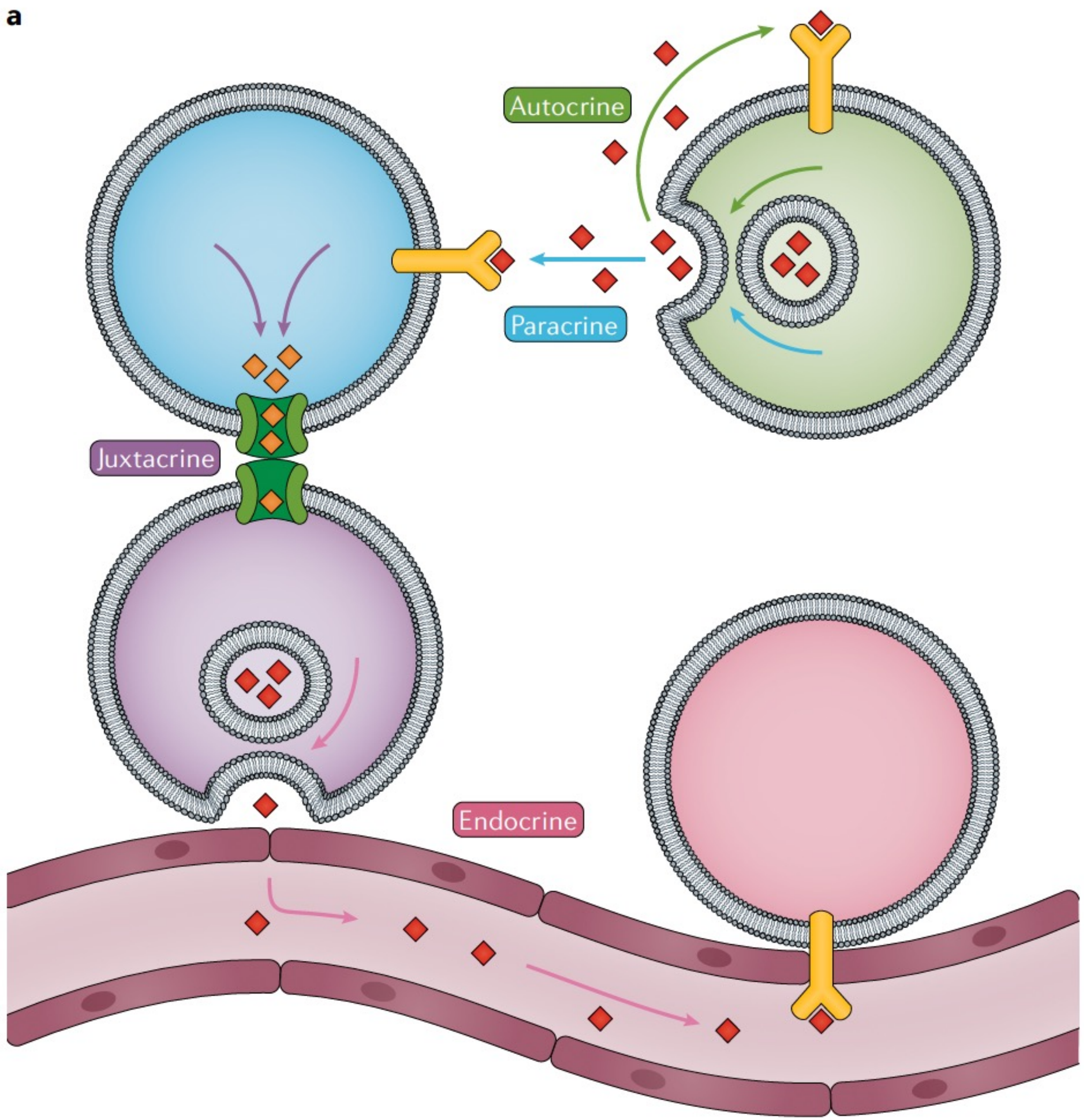
- **Cell–cell interactions** orchestrate organismal development, homeostasis and single-cell functions.
- When cells do not properly interact or improperly decode molecular messages, **disease** ensues.
- Thus, the identification and quantification of **intercellular signalling pathways** has become a common analysis performed across diverse disciplines.
- The expansion of protein–protein interaction databases and recent advances in RNA sequencing technologies have enabled routine analyses of intercellular signalling from gene expression measurements of bulk and single-cell data sets.
- In particular, **ligand–receptor pairs** can be used to infer intercellular communication from the coordinated expression of their cognate genes.



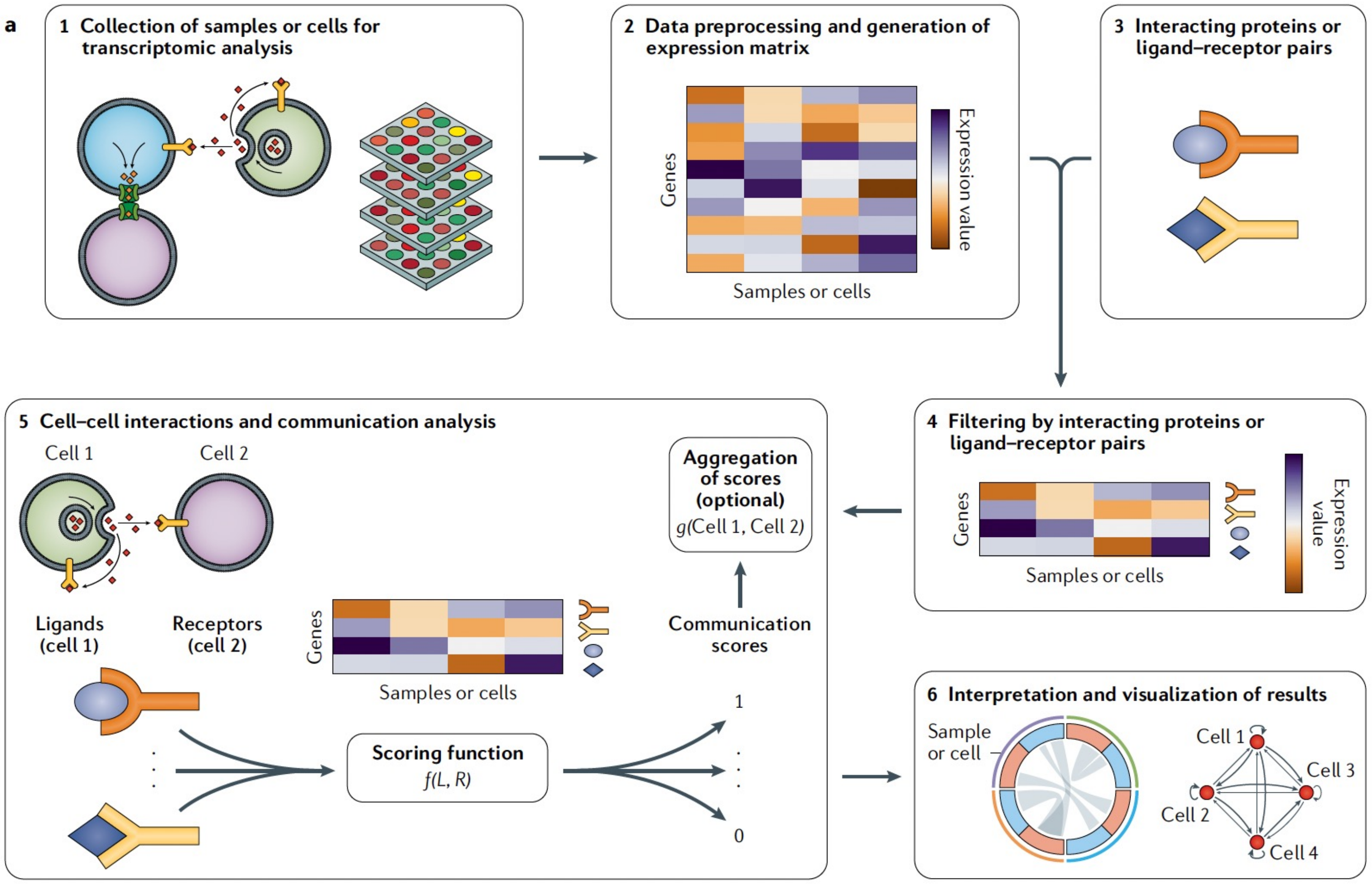
Overview of the main applications of cell–cell interaction methods: cell development, tissue and organ homeostasis, and immune interactions in disease

Interpretable

Armingol, E., Officer, A., Harismendy, O. *et al.* Deciphering cell–cell interactions and communication from gene expression. *Nat Rev Genet* **22**, 71–88 (2021).

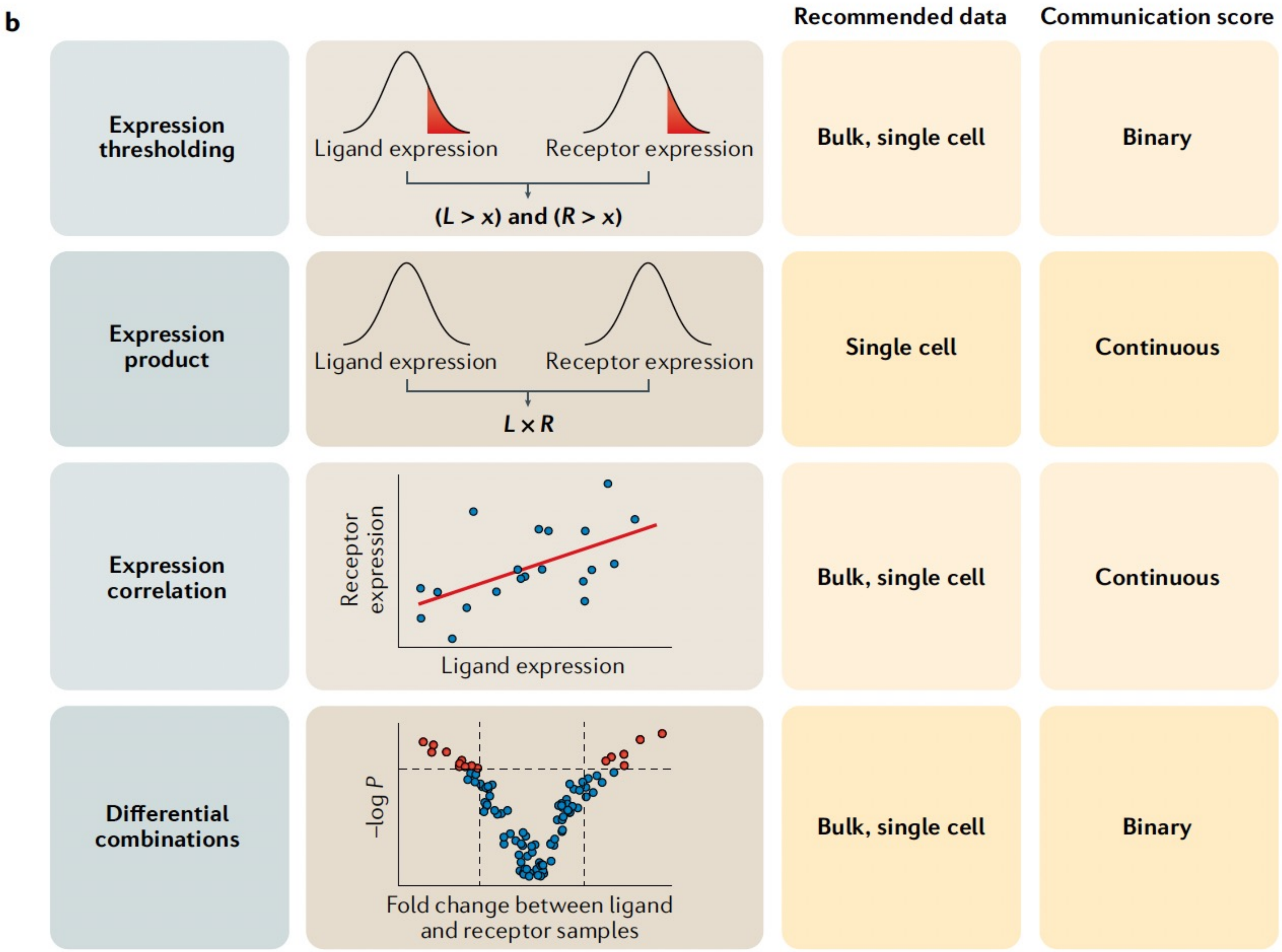


- **a | ‘Autocrine signalling’** refers to intracellular communication whereby cells secrete ligands that are used to induce a cellular response through cognate receptors for those molecules expressed on the <u>same cell</u>. **Paracrine** cell–cell communication does not require cell–cell contact, rather depending on the <u>diffusion of signalling molecules</u> from one cell to another after secretion. **Juxtacrine**, that is, contact-dependent, cell–cell communication relies on gap junctions or other structures such as membrane nanotubes to pass signalling molecules <u>directly between cells</u>, without secretion into the extracellular space. **Endocrine** cell–cell communication represents intercellular communication whereby signalling molecules are secreted and travel long distances through <u>extracellular fluids</u> such as the blood plasma; typical mediators of this communication are hormones.

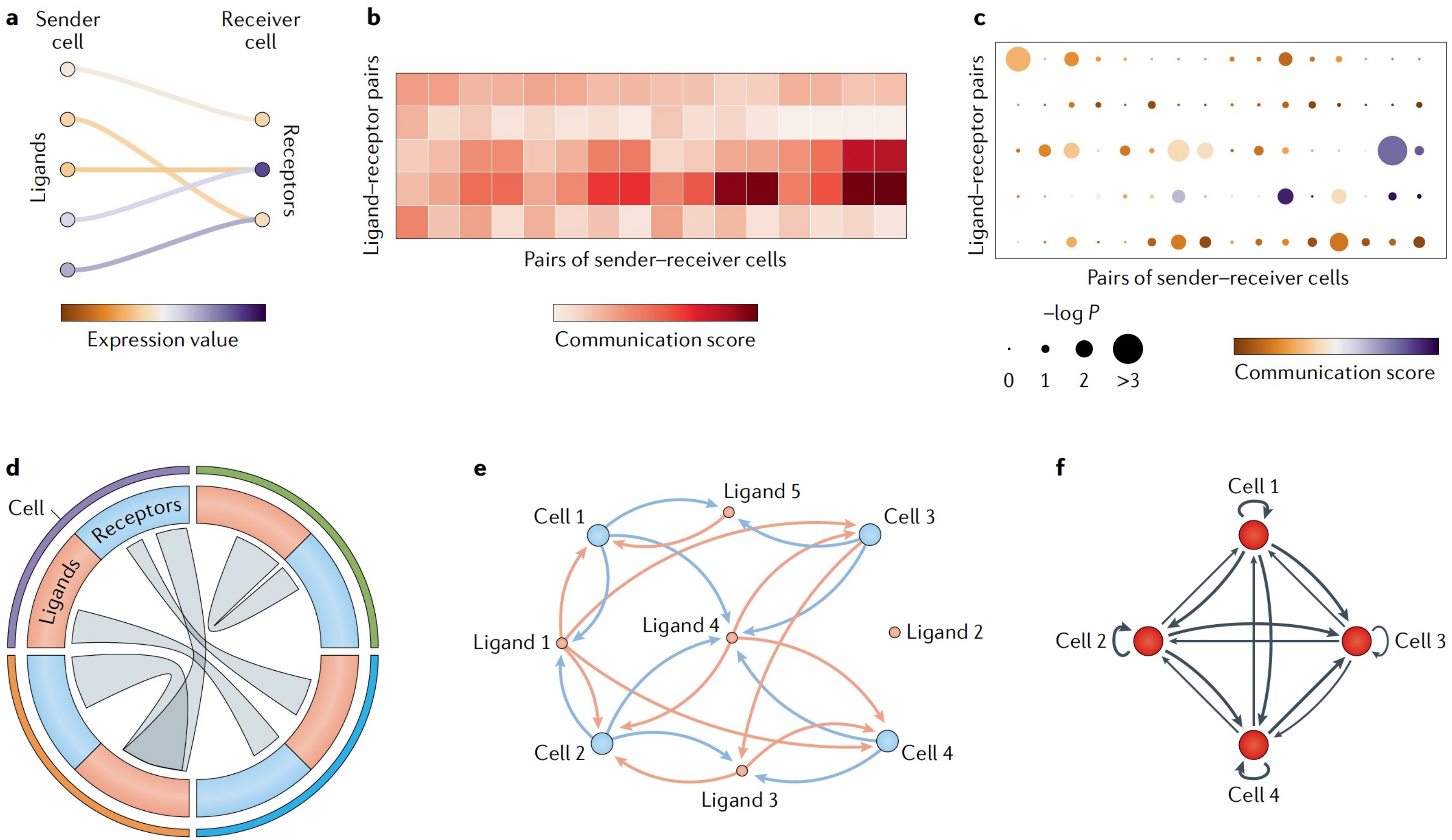


Routine SRT analysis protocol

- Samples or cells are analysed by transcriptomics to measure the expression of genes (**step 1**).
- Then the data generated are preprocessed to build a gene expression matrix, which contains the transcript levels of each gene across different samples or cells (**step 2**).
- A list of interacting proteins that are involved in intercellular communication is generated or obtained from other sources (**step 3**), often including interactions between secreted and membrane-bound proteins (commonly ligands and receptors, respectively).
- Only the genes associated with the interacting proteins are held in the gene expression matrix (**step 4**).
- Their expression levels are used as inputs to compute a communication score for each ligand–receptor pair using a scoring function (function $f(L, R)$, where L and R are the expression values of the ligand and the receptor, respectively). These communication scores may be aggregated to compute an overall state of interaction between the respective samples or cells using an aggregation function (function g(Cell 1, Cell 2), where Cell 1 and Cell 2 are all communication scores of those cells or corresponding samples) (**step 5**).
- Finally, communication and aggregated scores can be represented by, for instance, Circos plots and network visualizations to facilitate the interpretation of the results (**step 6**).



- Main **scoring functions** of communication pathways based on the expression of their components. Recommended data to use with these functions and the type of their resulting communication score are indicated.



- **a** | A **Sankey diagram** for connecting key ligands from a sender cell to cognate receptors in the receiver cell. Node colour (ligand or receptor) indicates the expression level. **b** | **Heatmap** to represent the communication scores for each ligand–receptor interaction in each cell pair. **c** | **Dot plot** to show the communication score (colour of dots) and at the same time its significance (size), often obtained from a statistical model or permutation analysis. **d** | **Circos plot or chord diagram** to show key communication pathways used by different cell types to communicate. The links start from a ligand (red) and end in a receptor (blue), which are grouped for each cell type (coloured outer arcs). **e** | **Bipartite network** where nodes can be either cells or ligands. Edges can be directed only from a cell to a ligand it produces or from a ligand to a cell that expresses its cognate receptor. **f** | **Cell–cell interaction network** to represent the potential of cells to interact. Nodes correspond to cells and edges correspond to their interactions. These are directed from a sender cell to a receiver cell, and their thicknesses are proportional to the respective global cell–cell communication scores (for example, number of active ligand–receptor pairs).

Parts of the CCC methods

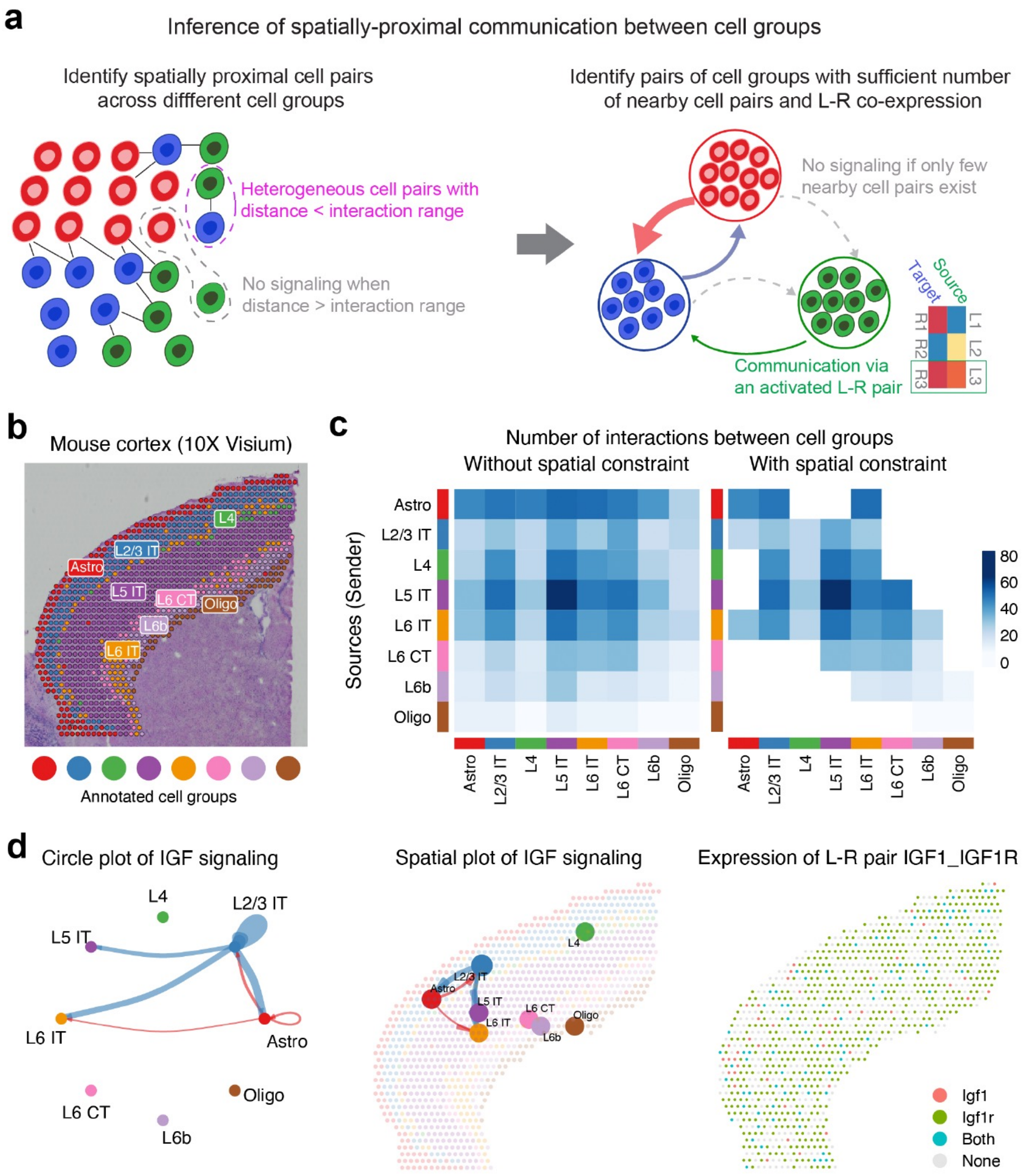
Tools	Method	Subunit	Prior knowledge	Language	Ref.
Statistical-based tools					
CellCall	Embedded pathway activity analysis for activity score; hypergeometric testing for significance of pathway activity	Single subunit	Ligand-receptor pairs; downstream TF regulation	R	[15]
CellChat	Law of mass action for communication probability; permutation test for significant interactions	Multi-subunit	Ligand-receptor pairs; signaling cofactors and pathways	R	[11]
CellPhoneDB	The mean of average ligand and receptor expression values for interaction enrichment; permutation test for significant interactions	Multi-subunit	Ligand-receptor pairs	Python	[10]
ICELLNET	Product of ligand and receptor expression values for communication score; geometric mean for multi-subunit complexes; Wilcoxon statistical test for highly potential interactions	Multi-subunit	Ligand-receptor pairs	R	[16]
iTALK	Finding differentially expressed ligand and receptor genes between cell types	Single subunit	Ligand-receptor pairs	R	[17]
SingleCellSignalR	Regularized product of ligand and receptor for lr-score; estimate lr-score cutoff for filtering interactions	Single subunit	Ligand-receptor pairs	R	[18]
Network-based tools					
Connectome	Cell types as nodes, interactions as edges; gene-wise *z*-score of ligand and receptor expression values as edge weights; system-wide Wilcoxon rank sum test for significant edges filtering	Single subunit	Ligand-receptor pairs	R	[19]
CytoTalk	Integrate two de novo intracellular signaling networks by known ligand-receptor interactions; optimal subnetwork searching for significant interactions	Single subunit	Ligand-receptor pairs	R	[20]
Domino	Construction global signaling network; cluster specific signaling subnetwork for prediction	Multi-subunit	Ligand-receptor pairs; TF regulation	R	[21]
NATMI	Cell types as nodes, interactions as edges; mean expression or specificity for edge weights; edge weight ranks for confident interactions	Single subunit	Ligand-receptor pairs	Python	[22]
NicheNet	Weighted network prior knowledge model; compute ligand activity and regulatory potential score using network propagation; select interactions by potential score	Single subunit	Ligand-receptor pairs; ligand-target pairs; receptor-target pairs	R	[12]
scMLnet	Construct primary ligand-receptor, TF-target, receptor-TF subnetworks using highly expressed genes; merge three subnetworks as final output	Single subunit	Ligand-receptor pairs; receptor-TF pairs; TF-target pairs	R	[23]
ST-based tools					
CellPhoneDB v3	L-R expression for enrichment; permutation test for significance; filter interactions based on spatial microenvironment	Multi-subunit	Ligand-receptor pairs; spatial microenvironment	Python	[13]
Giotto	Spatial proximity for interacting cell types; spatial co-expression for interactions	Single subunit	Ligand-receptor pairs; cell type colocalization; L-R co-expression	R	[24]
stLearn	Identify interactions by L-R co-expression and cell type density	Single subunit	Ligand-receptor pairs; cell type colocalization; L-R co-expression	Python	[25]

However, these non-spatial studies often contain significant false positives given that CCC takes place **only within limited spatial distances** that are not measured in scRNA-seq datasets.

Liu, Z., Sun, D. & Wang, C. Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information. *Genome Biol* **23**, 218 (2022).

Spatial

- Inference of cell-cell communication can be naturally extended to spatial context by **first identifying pairs of cells that are physically close to one another to have biologically realistic interactions based on maximal possible molecular interaction/diffusion ranges**, and then identifying combinations of cell groups that have enough nearby cell-cell pairs.
- The diffusive spatial distance of molecules depends on **many factors**, including molecule size, its covalent modifications, touristy of the spatial tissues, and the molecule's regulators on the cell membrane and in the extracellular environments. All these factors usually reduce diffusion. For example, large molecules have shorter diffusion distance, leading to more restricted spatial range in diffusion.
- CellChat v2 uses the <u>ideal diffusion range in a free medium</u>, that is the maximally allowable transport distance for small diffusive molecules (by default 250 μm). In this way, CellChat v2 will not remove any interactions that are spatially plausible. For the <u>contact-dependent signaling</u>, the interaction range is restricted to the nearest neighbors of each cell, such that signaling and target cells are in direct contact.



CellChat, CellChatDB, CellChat V2

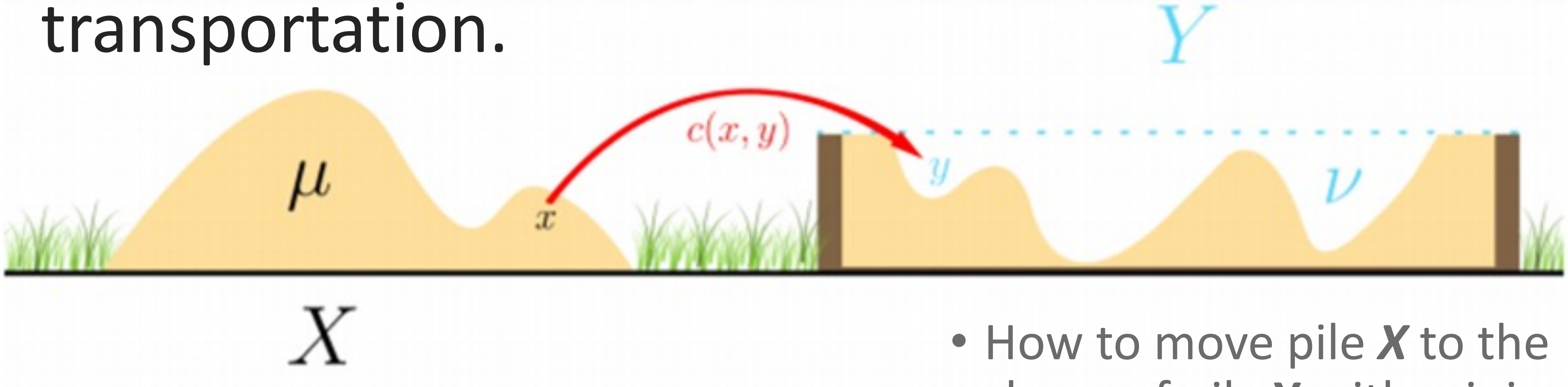
- CellChat is an R package designed for inference, analysis, and visualization of cell-cell communication from single-cell and spatially resolved transcriptomics. CellChat aims to enable users to **identify and interpret cell-cell communication** within an easily interpretable framework, with the emphasis of clear, attractive, and interpretable visualizations.
- CellChatDB is a **manually curated database** of literature-supported ligand-receptor interactions in mutiple species, leading to a comprehensive recapitulation of known molecular interaction mechanisms including multi-subunit structure of ligand-receptor complexes and co-factors.
- CellChat V2:
  - inference of spatially proximal cell-cell communication between interacting cell groups from spatially resolved transcriptomics
  - expanded database CellChatDB v2 by including more than 1000 protein and non-protein interactions (e.g. metabolic and synaptic signaling) with rich annotations. A function named updateCellChatDB is also provided for easily updating CellChatDB.
  - new functionalities enabling easily interface with other computational tools for single-cell data analysis and cell-cell communication analysis
  - interactive web browser function to allow exploration of CellChat outputs of spatially proximal cell-cell communication

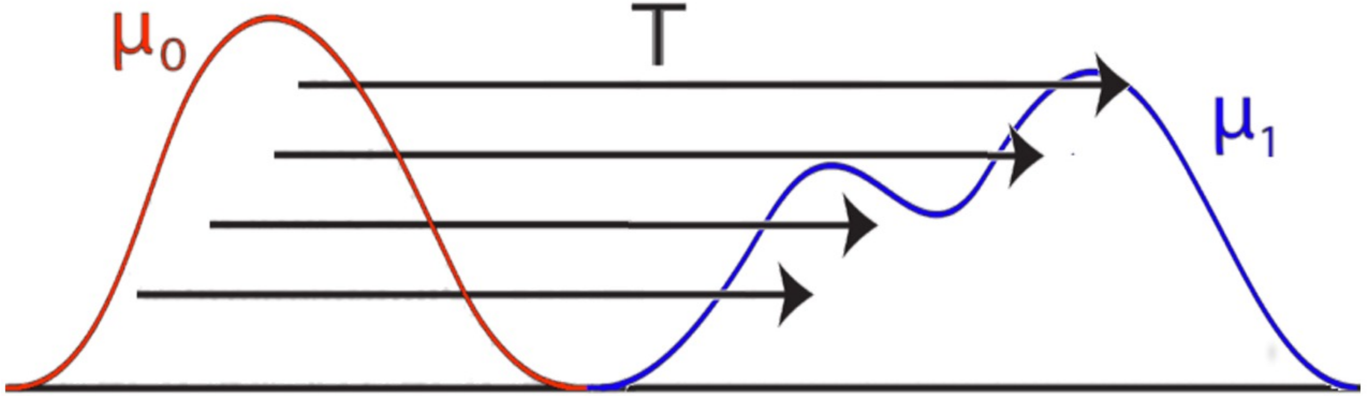
- Suoqin Jin et al., CellChat for systematic analysis of cell-cell communication from single-cell and spatially resolved transcriptomics, bioRxiv 2023 [CellChat v2]
- Suoqin Jin et al., Inference and analysis of cell-cell communication using CellChat, Nature Communications 2021 [CellChat v1] **Citation:2896**

CCC on spatial transcriptomics

- Giotto builds a spatial proximity graph to identify interactions through membrane-bound ligand–receptor pairs[23];
- CellPhoneDB v3 restricts interactions to cell clusters in the same microenvironment defined based on spatial information[25];
- stLearn relates the co-expression of ligand and receptor genes to the spatial diversity of cell types[24];
- SVCA[26] and MISTy[27] use probabilistic and machine learning models, respectively, to identify the spatially constrained intercellular gene–gene interactions;
- NCEM fits a function to relate cell type and spatial context to gene expression[28].

Optimal Transport: is an optimization problem which goal is to minimize the cost of transportation.

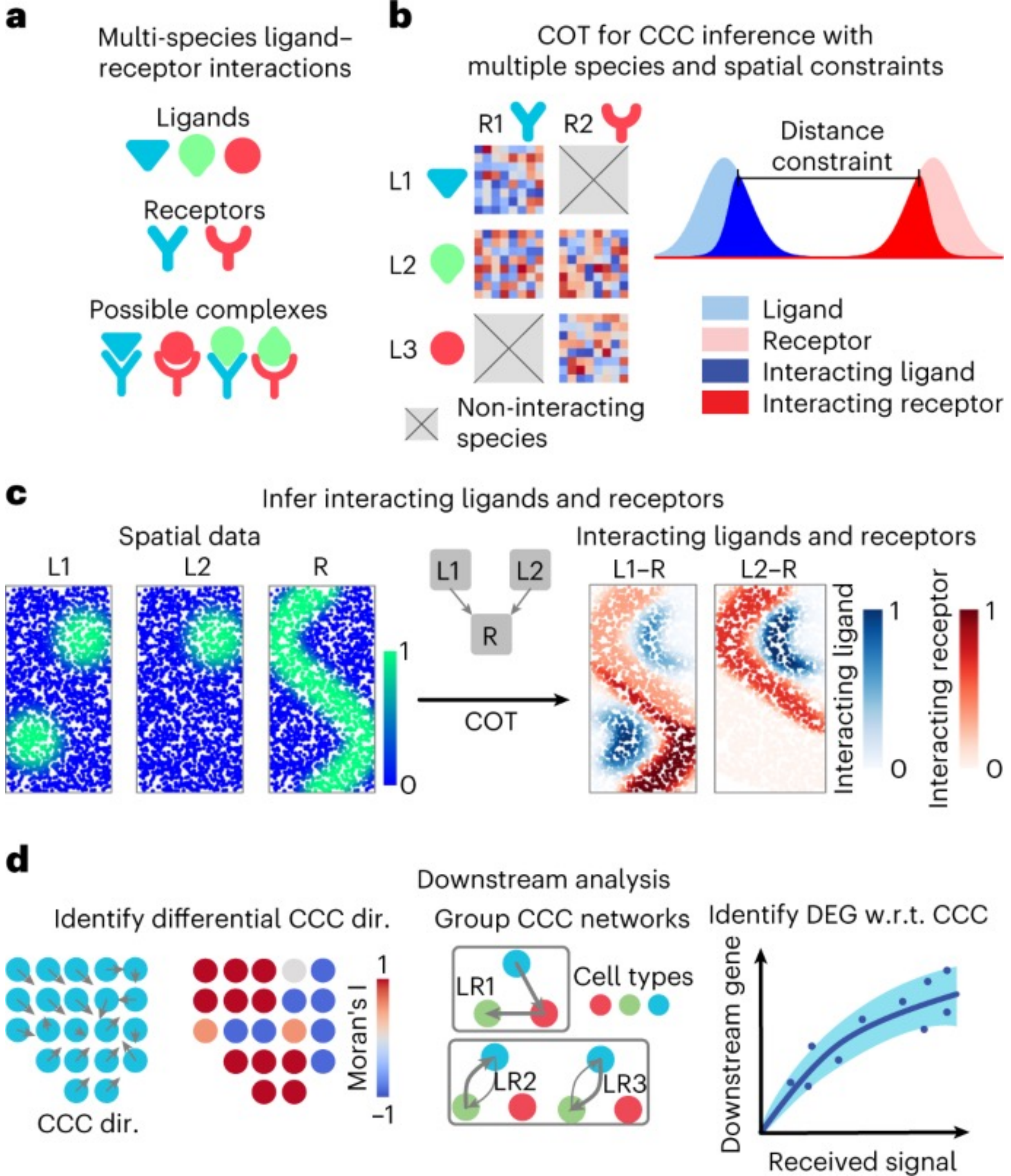




- How to move pile ***X*** to the shape of pile **Y** with minimal effort? I.e. given a cost function **c(x,y)** of moving a grain **x∈X** to the position **y∈Y**, what is the optimal displacement of all **X** to **Y**?

French Mathematician: Gaspard Monge

N. Papadakis, Optimal Transport for Image Processing, habilitation à diriger des recherches, Université de Bordeaux, Dec. 2015



COMMOT
(COMMunication analysis by Optimal Transport)

- Ligands and receptors often interact with multiple species and within limited spatial ranges (Fig. 1a).
- Considering this, we present **collective optimal transport** (Fig. 1b) with three important features:
- first, the use of non-probability mass distributions to control the marginals of the transport plan to maintain comparability between species;
- second, enforcement of spatial distance constraints on CCC to avoid connecting cells that are spatially far apart;
- last, the transport of multi-species distributions (ligands) to multi-species distributions (receptors) to account for multi-species interactions (Fig. 1c).

Summary

- The complex structures and functions of multicellularity are achieved through the coordinated activities of various cells.
- Cells make decisions and accomplish their goals by interacting with an environment consisting of external stimuli and other cells.
- There are huge needs and spaces in CCC

Expert opinion

"With COMMOT, Cang et al. present an elegant mathematical solution to the problem of inferring cell–cell communication from spatial transcriptomics data based on a variant of optimal transport. The method is applied to spatial datasets of different sizes and technologies, and robustness of results is shown. Further, the authors show how their method can be used in different biological contexts, including human breast cancer and mouse brain samples." **Fabian Theis and Marius Lange, Helmholtz Munich, Germany**