Deciphering cell–cell interactions and communication from gene expression

Erick Armingol, Adam Officer, Olivier Harismendy and Nathan E. Lewis

Abstract | 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. In this Review, we highlight discoveries enabled by analyses of cell–cell interactions from transcriptomic data and review the methods and tools used in this context.

Multicellular life relies on the coordination of cellular activities, which depend on cell–cell interactions (CCIs) across an organism’s diverse cell types and tissues. Thus, studies on cellular functions increasingly require consideration of the community context of each cell. CCIs leverage diverse molecules, including ions, metabolites, integrins, receptors, junction proteins, structural proteins, ligands and secreted proteins of the extracellular matrix. Some molecules support structural CCIs (for example, cell adhesion molecules), whereas ligands such as hormones, growth factors, chemokines, cytokines and neurotransmitters mediate cell–cell communication (CCC) (Fig. 1a). The signalling events behind CCC are often mediated by interactions of various types of protein, encompassing ligand–receptor, receptor–receptor and extracellular matrix–receptor interactions. Receiver cells trigger downstream signalling through cognate receptors, generally culminating in altered transcription factor activity and gene expression. These cells with altered expression further interact with their microenvironment. To understand the role of each cell within its local community, one must identify the protein messages passed between cells; measuring expressed messenger molecules and their associated pathways is fundamental to understanding the directionality, magnitude and biological relevance of CCC.

Direct measurement of proteins mediating CCC requires specialized biochemical assays and extensive domain knowledge; moreover, these proteins cannot always be studied in the native microenvironment. Traditional assays of the underlying protein–protein interactions (PPIs) include yeast two-hybrid screening, co-immunoprecipitation, proximity labelling proteomics, fluorescence resonance energy transfer imaging and X-ray crystallography. These techniques have identified many interactions between proteins that are secreted or displayed extracellularly to mediate intercellular communication. Proteomics and transcriptomics can further reinforce such studies as evidence of expression supports the presence of PPIs. This approach has been applied to, for instance, the analysis of communication between 144 human primary cell types, which provided insights into pairs of cells that are more likely to interact and the specific pathways they use to communicate. While proteomic technologies are preferable for these analyses owing to the direct measurement of protein abundances, RNA sequencing (RNA-seq) data sets are more numerous, easier to access and straightforward to analyse. They can also be generated from bulk samples, microdissected specimens or single-cell suspensions and enable studies of CCC at different resolutions, whereas proteomics at single-cell resolution is a technology still under development. Single-cell RNA-seq has benefits over bulk analysis, chiefly in quantifying expression in rare cell types and in identifying the cell type of origin of proteins mediating CCIs. Results from transcriptomics must be cautiously considered and validated to avoid misleading hypotheses; however, the ubiquity and ease of analysis have enabled many recent studies to infer CCC from gene expression, generating testable hypotheses across diverse disciplines. In particular, the coordinated gene expression of ligands and receptors can be used to infer intercellular communication.
Here, we start by providing an overview of the range of fields that RNA-based CCI analyses have been applied to, illustrating the types of insight that can be gleaned. We then discuss the computational strategies adopted in those studies, detailing the PPI databases and mathematical models commonly used to decipher CCC. Additionally, we introduce the computational tools that facilitate these analyses, describing their main features as well as their strengths and weaknesses. Finally, we review approaches to validate CCI-derived results and discuss remaining challenges and future directions in the field.

**Insights from RNA-based CCI analyses**

The study of intercellular interactions has greatly accelerated as transcriptomics, in particular bulk and single-cell RNA-seq, has become commonplace. These approaches use transcriptional profiling to decipher CCCs at any stage of development and in any multicellular community. Many studies focus on signals mediating cellular differentiation, interactions of cell types within tissues and organs, and immune responses (Fig. 1b; Table 1; see Supplementary Table 1 for more details). Here, we review these studies and illustrate the types of insights gained from analysing CCC.

**Interactions drive cellular differentiation and organ development.** Cellular differentiation depends on temporally and locally precise cell communication, so inspecting intercellular communication has increased our understanding of stem cell fates and revealed ligand–receptor interactions that initiate self-renewal and differentiation. For example, a CCC network of haematopoietic cells built using ligand–receptor pairs showed that fate decisions are regulated through precise coordination of an antagonistic feedback circuit involving megakaryocyte-derived stimulatory factors and monocyte-derived inhibitory factors. Another analysis of CCC networks interrogating how differentiated cells influence haematopoietic stem cell fate revealed that ligand production is cell type specific, whereby some cells can produce signals with the same function, whereas receptors are less specific. Given the promiscuity of receptors, physical compartmentalization of cells is key to limit ligand signalling and confer specificity to stem cell fates.

Tissue and organ development also depends on signals that progenitor cells send and receive. The analysis of brain CCC showed crosstalk involved in neurogenesis and identified novel mediators, such...
| Sample or organ | Key input | Scoring | CS value | CCC score | Validation | Study focus | Ref |
|----------------|-----------|---------|----------|-----------|------------|-------------|-----|
| **Cell development** | | | | | | | |
| Haematopoietic cells (human) | Microarray; LRIs | Expression thresholding | Binary | No score | Functional validation | Role of CCC between differentiated haematopoietic cells and HSCs in fate decisions | 15 |
| Brain (mouse embryonic cortex) | Microarray; LRIs | Expression thresholding | Binary | No score | Functional validation | Role of microenvironment in self-renewal versus differentiation decision of neural precursor cells during neurogenesis | 16 |
| Liver and iPS cells (human) | scRNA-seq; LRIs | Expression thresholding | Binary | Normalized sum of CS | Functional validation | 3D liver bud organoid from iPS cells to characterize CCC shaping hepatogenesis | 24 |
| Placenta (human) | scRNA-seq; LRIs | Expression thresholding | Binary | No score | Functional validation | CCI in the fetus–placenta interface before and after decidualization | 29 |
| Brain (mouse) | Bulk RNA-seq; LRIs | Expression thresholding | Binary | Sum of CS | Colocalization | Ligand–receptor pathways active during neural development; CCC between neural, vascular and microglial cells | 22 |
| iPS cells (mouse) | scRNA-seq; LRIs | Expression product | Continuous | Sum of CS | Functional validation | CCC at the beginning of differentiation | 19 |
| Bone marrow (mouse) | scRNA-seq; LRIs | RNA-Magnet | Continuous | No score | Colocalization | CCC and interactions between bone marrow cells | 53 |
| **Tissue interactions** | | | | | | | |
| Multiple lineages (human) | scRNA-seq; LRIs | Expression thresholding | Binary | Sum of CS | None | CCC between multiple cell lineages | 7 |
| Lungs (human) | popRNA-seq; LRIs | Expression thresholding | Binary | Sum of CS | Colocalization functional validation | Signals sent by mesenchymal cells in lungs that are key for self-renewal of epithelial progenitors after tissue injury | 40 |
| Heart (mouse) | scRNA-seq; LRIs | Expression thresholding | Binary | Sum of CS | Functional validation | Transcriptional profiles of non-myocyte cells in heart and their CCC | 35 |
| Lungs (mouse) | scRNA-seq; LRIs | Expression correlation (Spearman) | Continuous | No score | Expression; colocalization functional validation | CCC between and within immune and non-immune cells during development | 41 |
| Immune system and structural cells (mouse) | Low-input RNA-seq; LRIs | Differential combinations | Binary | Odds ratios | Functional validation | Role of structural cells in immune responses | 58 |
| Heart (mouse) | scRNA-seq; LRIs | Differential combinations | Binary | Sum of CS | Expression; colocalization functional validation | CCC of cardiomyocytes and non-cardiomyocytes in human heart in health and under failure | 54 |
| Placenta (human) | scRNA-seq; LRIs | CellPhoneDB | Continuous | No score | Colocalization | Key ligand–receptor pairs based on subunit architecture; CCC at maternal–fetal interface | 30 |
| **Tumour microenvironment** | | | | | | | |
| Melanoma (human) | scRNA-seq; LRIs | Expression thresholding | Binary | Sum of CS | None | CCI network of isolated cells | 70 |
| HNSCC (human) | scRNA-seq; LRIs | Expression thresholding | Binary | No score | Colocalization | CCC in patients with HNSCC generated by HPV or environmental carcinogens (HPV negative) | 61 |
| Five cancer types (mouse) | scRNA-seq; LRIs | Expression product | Continuous | No score | None | CCC within a tumoural microenvironment | 60 |
| Nine cancer types (human) | Microarray; LRIs | Expression correlation (Pearson) | Continuous | No score | None | Correlation between autocrine signalling pathways and mRNA levels of ligands and receptors | 69 |
with each other. Interaction network would example, a protein–protein interaction would involve a set of nodes with defined metabolic processes. Signal transduction or communication involves interactions between two proteins, often mediated by proteins (PPis). Physical interaction between proteins, often mediated by protein–protein interactions (PPis), is a fundamental mechanism for intracellular signal transduction and communication. Signals are sent between or within cells and generate biochemical signals that support to neighbouring cells. Extracellular matrix (ECM) provides structural and functional support to the cells and ECM proteins, such as collagen, fibronectin, and laminin, play crucial roles in cell adhesion, migration, and communication. In extracellular space, it provides a physical and chemical microenvironment that influences cell behaviour and communication. Proteins that bind to other molecules to receive or amplify a signal. They are a common membrane-bound but can also be found in the cytoplasm. Ligands are biomolecules that bind to receptors and change the activity, conformation or other biological properties of the receptor, triggering a signalling event.

Extracellular matrix is a three-dimensional organization of molecules located in the extracellular space. It provides structural and functional support to neighbouring cells.

Cell–cell communication (CCC) is critical for the development and function of the immune system. CCC involves the interaction of immune cells, such as T cells and natural killer cells, with other immune cells and non-immune cells. These interactions are mediated by ligands and receptors, and the specificity of these interactions is crucial for the regulation of immune responses. NicheNet is a computational approach that can be used to predict ligand–target interactions based on single-cell RNA sequencing data.

Cell-type communication in tissue and organ homeostasis is essential for maintaining tissue integrity and function. Expression profiling of different cell types in adult tissues and organs has shown how intercellular communication contributes to organ function. To date, this approach has been applied to brain, heart, kidney, liver, lungs, placenta, retina, and visual cortex. Remarkably, this approach revealed new roles of cells within a tissue and helped explain how ageing, diseases, infections, and injuries shape multicellular organization. Intercellular cross-talk has been investigated in healthy and diseased liver, kidneys, during homeostasis and rejection kidney transplants, heart during failure and recovery conditions, and healthy and asthmatic lungs. RNA-based strategies for studying CCCs and CC have helped elucidate, for instance, how cells communicate during ageing of the mouse brain. This revealed that CST3 and CXCL12 are mediators that differentiate intercellular interactions in young and old brains and may modulate ageing-related processes.

Immune interactions in disease. The immune system receives signals from multiple tissues, but only specific signals allow it to coordinate healthy immune responses. For instance, CCL2- and CXCL1-mediated communication coordinate the recruitment and positioning of immune cells, as determined from single-cell transcriptomes. Specifically, these CCCs were associated with the recruitment of monocytes that later became resident macrophages and the positioning of mononuclear phagocytes in kidney, which are crucial processes to combat ascending uropathogenic Escherichia coli infection. CCC is also involved in the response to viral infections. Studies of respiratory diseases investigated the cross-talk between lung and T cells in Sendai virus-infected mice or CCC associated with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, where interactions between immune and epithelial cells correlated with COVID-19 severity. CCC-based studies have also provided more general insights. For example, they helped build a social network of immune cells by identifying communication pathways between immune cells. Furthermore, they revealed the roles of structural cells in immune responses by elucidating how fibroblasts, endothelial cells and epithelial cells are primed for organ-specific immune gene activation through upregulation of ligands and receptors, including β₂M, CD74, CXCL10, VCAM1 and TNFRSF1A.

Table 1 (cont.) | Illustrative studies and their strategies for deciphering cell–cell interactions and communication

| Sample or organ | Key input | Scoring | CS value | CCC score | Validation | Study focus | Ref |
|-----------------|-----------|---------|----------|-----------|------------|-------------|-----|
| Lungs (human)   | scRNA-seq; LRs | Differential combination; expression thresholding | Binary | No score | Functional validation | Tumour–stroma CCC in lung cancer; introduced CCCExplorer | 67 |
| Ovary (human)   | Microarray; LRs; downstream target genes | Differential combination; expression thresholding | Binary | No score | Expression; functional validation | CCC between stromal and ovarian cancer cells | 66 |
| Head and neck and immune system (human) | scRNA-seq; LRs; downstream target genes | NicheNet | Continuous | No score | None | Prediction of ligand–target links between interacting cells; tested on HNSCC data set | 54 |

CCC, cell–cell communication; CCI, cell–cell interaction; CS, communication score; HNSCC, head and neck squamous cell carcinoma; HPV, human papillomavirus; HSC, haematopoietic stem cell; iPSC cell, induced pluripotent stem cell; LRI, ligand–receptor interaction; popRNA-seq, population RNA sequencing; RNA-seq; RNA sequencing; scRNA-seq, single-cell RNA sequencing. For additional studies and details, see Supplementary Table 1.
Tumours and their surrounding microenvironments are complex communities of cells that modify local immune cell functions. Studying CCC within these communities can reveal how cells communicate in these ecosystems and help guide the development of effective cancer immunotherapies just as the inhibition of CCC through PD1 and PDL1 has revolutionized the field. CCC analyses have also elucidated cross-talk between tumour and stroma and communication pathways used by tumours. Several studies have developed statistical models to connect inferred CCC mechanisms to cancer phenotypes. One case observed clear correlations between the activity of specific ligand–receptor interactions and the degree of regulatory T cell infiltration and tumour growth. In another study, active ligand–receptor pairs were associated with invasiveness and proliferation of malignant cells under a partial epithelial–mesenchymal transition programme in patients with head and neck squamous cell carcinoma. Moreover, the expression levels of key mediators of CCC were used as inputs for training a decision tree to predict prognosis for patients with glioma; this model classified patients into high-risk and low-risk groups, defined on the basis of the difference in patient survival time. Thus, studying CCC within the tumour microenvironment provides opportunities to identify druggable pathways and develop new cancer therapeutics.

Deciphering CCC

The aforementioned studies provide a glimpse of the insights attainable when studying CCC. While the methods and tools that these studies used have in common that they infer CCC from gene expression, a diverse range of strategies can be applied. For simplicity, we refer to them as methods for studying CCC, but these strategies can decipher any type of gene product-based CCIs, including proteins that participate in structural interactions between cells. Furthermore, although we focus on mammalian CCIs, the approaches apply to any prokaryotic or eukaryotic cells with a characterized interactome (for example, Drosophila melanogaster and Caenorhabditis elegans).

Building from PPIs. Inferring CCC from transcriptomics relies on gene co-expression, whereby one gene in a given pair comes from one interacting cell and the other gene comes from the second interacting cell. Several studies focused on intercellular signalling using co-expression of all genes or specific cell markers, the similarity between expression profiles or the properties of regulatory networks. However, most studies rely on literature-curated lists of interacting proteins, which facilitates the biological interpretation of results. Although several studies have used interactions between any class of cell-surface protein and secreted protein, the predominant class of interactions used for studying CCC are known ligands and their cognate receptors. By focusing on ligands and receptors, an early study investigated autocrine signalling loops in cancer. The authors analysed the correlation between ligand expression and cognate receptor expression. They first created a

literature-curated list of ligand–receptor pairs (Database of Ligand-Receptor Partners), which was subsequently integrated with microarray-based expression data to analyse autocrine signalling in cancer cells. At the time, the Database of Ligand-Receptor Partners consisted of 451 interactions in humans; since then, many other databases have catalogued substantially more interacting ligand–receptor pairs, enabling more comprehensive testing of communication processes between cells.

Many studies have enumerated known ligand–receptor pairs using different approaches. To facilitate further use and comparison, we collated publicly available lists into a single ligand–receptor pair repository. One extensively used database contains ~2,400 human ligand–receptor pairs, obtained from multiple databases and literature curation. Similar approaches allowed researchers to increase the number of known ligand–receptor pairs and to build databases for other organisms. Integrating multiple sources of data is challenging and requires reconciliation of the different ways ligand–receptor pair confidence was assessed or how orthologues were determined. Furthermore, few predicted PPIs in databases have been validated, raising the concern of false positives.

While early efforts increased the number of reported ligand–receptor pairs, many lacked information about protein complexes involving multiple subunits. This scenario greatly increases false-positive CCC predictions. Some proteins, such as transforming growth factor-β receptors and cytokine receptors, require multisubunit assembly for function. A lack of expression of any subunit blocks ligand–receptor interactions and the resulting communication. Thus, more recent computational tools such as CellPhoneDB, CellChat and ICELLNET include multimeric proteins and interactions between complexes of both ligands and receptors. Accounting for subunit co-expression better represents functional ligand–receptor interactions. For example, CellChat includes ~2,000 ligand–receptor interactions, ~48% of which represent interactions of heterodimers, a substantial increase from the ~900 ligand–receptor pairs in CellPhoneDB.

Additional efforts have incorporated information beyond ligand–receptor pairs to reveal other aspects of CCC, such as the interchange of metabolites or activation of intracellular signalling. For example, metabolite secretion data have been included in communication studies by considering the expression of the producing enzymes. In this regard, the production of specific metabolites can be inferred from transcriptomic data; however, metabolite concentrations cannot be directly predicted, so concentration-dependent interactions cannot be assessed. Among other efforts, information about downstream signalling gene products and gene regulatory networks can be included by weighing the potential of using a communication pathway given the intracellular effect on a receiver cell. For this, genes in a downstream pathway to each receptor are obtained from signalling and regulatory networks; then their expression provides additional inputs for scoring the respective ligand–receptor pairs. However, more extensive
1. Collection of samples or cells for transcriptomic analysis

2. Data preprocessing and generation of expression matrix

3. Interacting proteins or ligand–receptor pairs

4. Filtering by interacting proteins or ligand–receptor pairs

5. Cell–cell interactions and communication analysis

6. Interpretation and visualization of results

**b**

- **Expression thresholding**
  - Ligand expression
  - Receptor expression
  - \((L > x)\) and \((R > x)\)
  - Bulk, single cell
  - Binary

- **Expression product**
  - Ligand expression
  - Receptor expression
  - \(L \times R\)
  - Single cell
  - Continuous

- **Expression correlation**
  - Receptor expression
  - Ligand expression
  - Bulk, single cell
  - Continuous

- **Differential combinations**
  - Fold change between ligand and receptor samples
  - Bulk, single cell
  - Binary
Fig. 2 | Analysis workflow for inferring cell–cell interactions and communication from gene expression. a | 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). b | 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.

---

**False negatives**
A false negative occurs when a positive example is assigned a negative label. For the purpose of this Review, this means a true interacting ligand–receptor pair is labelled as not interacting.

**Post-translational modifications**
Covalent modification of amino acid residues on a protein, commonly altering function, structure or localization. Phosphorylation, acetylation and glycosylation are among the most common.

**Signalling pathways**
The network of biomolecules that serve to transmit signals and induce cellular responses. Post-translational modification of proteins is the most common way signals are propagated.

---

Information on ligand–receptor pairs is required to include downstream genes. A limitation of this information is that rules of gene regulation are not considered, leading to potential false positives and false negatives.

Adding information can be laborious, requiring careful curation of mediator molecules and development of algorithms that incorporate these data. Furthermore, inference methods based on PPI databases are sensitive to the quality of the underlying databases. Nevertheless, PPIs, especially ligand–receptor pairs, have been central to deciphering CCC in all strategies.

### Which communication pathways are cells using?
To elucidate which biological processes are used by interacting cells to communicate, a score is usually estimated for each pair of interacting proteins (BOX 1) using the expression of their cognate genes as input to a scoring function (Fig. 2a). Importantly, the main assumption of these methods is that (1) gene expression reflects protein abundance and (2) protein abundance is sufficient to infer the PPI strength, ignoring essential factors for their binding, such as post-translational modifications or multisubunit complex assembly. Here, we focus on ligand–receptor pairs as PPIs (Fig. 2a, step 3); however, these strategies can be extended to any intercellular PPIs.

Communication scores may be binary or continuous (Fig. 2b), each providing different insights into the signalling pathways that cells use. Binary scores are simpler, whereas continuous scores enable more precise quantification of intercellular signalling. For both types of communication scores, we identified two core scoring functions in the literature. The ‘expression thresholding’ and ‘differential combinations’ methods are defined for the binary category, whereas the ‘expression product’ and ‘expression correlation’ approaches quantify continuous scores.

In binary scoring, expression thresholding is widely used because of its easy implementation and interpretation (TABLE 1). By thresholding expression values of both interacting partners in each ligand–receptor pair, we can measure all communication processes used between cells. If both genes are expressed above a threshold, the ligand–receptor pair is considered ‘active’; otherwise it is ‘inactive’ (assigning ones and zeros, respectively; FIG. 3). Different thresholds can be used for the ligand and receptor\(^{[47]}\). By contrast, the differential combinations method uses any approach to quantify differential gene expression\(^{[48]}\) to identify ‘active’ interacting partners. Thus, this strategy measures communication pathways in a sample- or cell type-specific manner.

Regarding input data sets, both binary functions are suitable for either bulk or single-cell data (FIG. 2b); however, bulk-derived samples may miss certain interacting ligand–receptor pairs, as averaging gene expression across all cells can mask signals from low-abundance cells. These binary methods both assume that higher expression is needed for interaction and require the choosing of a gene expression threshold (either significance or gene expression magnitude), which can lead to false positives and/or false negatives. General thresholds may fail as (1) some proteins have different bioactivities in a concentration-dependent manner\(^{[49]}\) and (2) mRNA–protein level correlations vary across genes\(^{[50]}\). Hence, gene-specific thresholds should be developed, as a general threshold may not properly represent the presence and activity of proteins.

Modelling interactions as non-binary processes provides biologically relevant information\(^{[51]}\) and helps infer active communication pathways. The expression product method computes a continuous value by multiplying the expression of both interacting proteins, and it has successfully found differences in ligand–receptor pair use between cells\(^{[52]}\). For example, communication pathways that were previously described to enhance the expansion of regulatory T cells were also identified to have communication scores linearly correlated with the infiltration of those cells in human metastatic melanoma\(^{[53]}\), which would not have been possible with binary scores. However, this approach may become problematic if interacting proteins have vastly different transcript levels wherein one protein dominates the interaction signal. Additional data preprocessing, such as cell-type normalization of gene expression using housekeeping genes\(^{[54]}\) or accounting for correlation between transcripts and proteins\(^{[55]}\), may mitigate this effect. Although this method can be applied to bulk data, microdissected or single-cell data sets are preferred as they capture the full heterogeneity of expression across cells\(^{[56]}\), resulting in clearer differences in expression\(^{[57]}\).

The second continuous scoring function is the expression correlation method. This score is the correlation between expression values of the interacting proteins across many samples. Thus, the score represents the general behaviour of each ligand–receptor pair in the evaluated groups rather than the individual importance for each pair of samples or cells. Whereas the previous methods can use any gene expression data type, the expression correlation method requires data sets containing many measurements of two populations to compute correlation for each ligand–receptor pair. A potential obstacle for this method is the sparsity of single-cell data sets, which can increase or decrease correlation coefficients in undesirable ways\(^{[58]}\), leading to correlation values that measure sparsity, rather than biology. In addition, the formation of subgroups of data...
points (representing pairs of cells) due to low expression variance of the ligand or the receptor in an interaction is known to bias correlation and can make this score inappropriate\textsuperscript{95}.

In addition to these core scoring functions, more advanced approaches can generate interaction scores, such as RNA-Magnet\textsuperscript{53}, a method using fuzzy logic to identify candidate ligands and receptors active in cell communication. As mRNAs encoding surface receptors frequently show low abundance, this strategy better accounts for the variance in receptor expression. Other advanced approaches have been wrapped into tools (see the section entitled A growing toolbox to facilitate CCC analysis). Each strategy can decipher CCC but relies on different assumptions that may influence the results. Therefore, knowing the limitations and following best practices in data set preprocessing\textsuperscript{93,96,97} are crucial to minimize false discoveries.

**Distinguishing which pair of cells are more likely to interact.** Measurement of individual communication scores facilitates the study of CCC, exposing the roles of specific signalling mechanisms; however, it does not reveal the entire interaction state between cells. Thus, it may be desirable to use an aggregate score to define the interactions between pairs of cells\textsuperscript{84}.

From the few studies that aggregated communication scores into an overall score (FIG. 2a, step 5), the most common approach quantifies the number of active ligand–receptor pairs between cells (that is, the sum of binary communication scores). This score suggests which cells interact more strongly and enables

---

**Box 1 | Computational strategies for building lists of interacting proteins**

To reduce the dependency on laborious experiments, lists of interacting proteins can be compiled computationally using the resources mentioned in the table below. These strategies include a heuristic search of gene products as candidates for interaction and manual curation of interacting protein pairs by using available databases or previously curated lists of interacting partners from the literature.

Several high-quality lists of ligand–receptor pairs exist for model organisms\textsuperscript{7,14–16,22,28,31,33,35,38,53–55,67,84}. However, studying cell–cell communication in organisms with fewer protein–protein interaction (PPI) data requires additional effort, especially in imputing ligand–receptor pairs through heuristic approaches and subsequent manual curation to increase the confidence in the list. For instance, some studies use orthologues of human ligand–receptor pair proteins from databases as indicated in the table below to adapt those pairs to the organism of interest, thus delineating core pathways of communication\textsuperscript{33,34,38}. When reference PPI networks or orthologue data are not available, other heuristic procedures to impute protein interactions could be used. A preliminary selection of candidate genes can be done according to their functional annotations. Some genes may lack annotation, so their potential interactions can be imputed with strategies such as protein co-evolution\textsuperscript{106}, modelling of protein structures\textsuperscript{105} and other in silico techniques for inferring PPIs. Additionally, machine learning\textsuperscript{94,100} and data mining\textsuperscript{96} can also be used, facilitating the search of candidate ligand–receptor pairs.

| Resource                                           | URL                                           | Ref. |
|----------------------------------------------------|-----------------------------------------------|------|
| **Functional annotations**                         |                                               |      |
| Gene Ontology                                      | http://geneontology.org/                     | 167  |
| UniProt                                            | https://www.uniprot.org/                      | 165  |
| KEGG                                               | https://www.genome.jp/kegg/                   | 169  |
| **Orthologues**                                    |                                               |      |
| OrthoDB                                            | https://www.orthodb.org/                      | 170  |
| gProfiler                                          | https://biit.cs.ut.ee/gprofiler/orth          | 171  |
| **Databases of protein–protein interactions**      |                                               |      |
| The Human Protein Atlas                            | https://www.proteinatlas.org/                 | 172  |
| STRING                                             | https://string-db.org/                        | 173  |
| BioGRID                                            | https://thebiogrid.org/                       | 174  |
| PICKLE                                             | http://www.pickle.gr                          | 175  |
| APID                                               | http://apid.dep.usal.es                       | 176  |
| IntAct                                             | https://www.ebi.ac.uk/intact                  | 177  |
| Pathway Commons                                    | http://www.pathwaycommons.org/               | 178  |
| **Ligand–receptor interactions**                   |                                               |      |
| IUPHAR/BPS Guide to Pharmacology                   | https://www.guidetopharmacology.org/          | 179  |
| Resources for cell–cell interactions from the Bader laboratory | http://baderlab.org/CellCellInteractions | 32   |
| Compendium of ligand–receptor pairs in the literature | https://github.com/LewisLabUCSD/Ligand-Receptor-Pairs | https://doi.org/10.1038/s41576-020-00292-x |

BPS, British Pharmacological Society; IUPHAR, International Union of Basic and Clinical Pharmacology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

---

**Fuzzy logic**

Extension of standard Boolean logic to define truth values of variables, encompassing real values between 0 and 1 both inclusive, instead of binary values.
Permutation
Random reassignment of sample labels, frequently used to compute null models in biological systems.

Differentially expressed genes
Genes identified as more highly (or lowly) expressed in one condition versus the other after comparison of their expression values between two conditions.

PageRank algorithm
Algorithm that takes as input a network and quantifies the importance of each node on the basis of centrality and connectedness to other central nodes.

Null model
Statistical model under which there is no interaction or difference between the groups being tested.

Multisubunit protein complexes
Quaternary structures of proteins involving the non-covalent interaction of two or more proteins to generate a functional unit.

The building of CCC networks to perform graph-based analyses. However, different cell pairs could have similar counts but completely different circuits of communication, causing inaccuracies. Newer methods attempted to deal with those inaccuracies by weighting CCIIs and CCC on the basis of additional data. Hence, these approaches enable the comparison of interactions weighted by their importance, among those that are considered active. One approach computes a probability of intracellular communication using a given ligand–receptor pair between a sender and a receiver cell and then aggregates this into a global CCC score\(^2\). Another study proposed a statistical model to evaluate expression variation of individual genes given a cell's CCIIs in its neighbourhood\(^3\). This method yields coefficients interpretable as an overall CCI score. Here, there is no aggregation of communication scores as the coefficient is computed from the covariance of different factors, such as the distance between cells, the cellular composition of the neighbourhood and transcriptionomes including all genes instead of only ligands and receptors. Although these approximations seem more biologically relevant than just counting the expressed ligand–receptor pairs, they require more detailed information that might not be available in all experimental settings.

Alternative strategies to quantify CCC include Euclidian-like or Jaccard-like distance metrics that represent the number of active ligand–receptor pairs in relation to all ligand–receptors in interacting cells. Furthermore, spatial transcriptomics technologies\(^4\)\(^\text{–}\)\(^8\) may help train machine learning models to predict a physical distance-dependent potential of CCIIs\(^9\). Expression of ligands and receptors can serve as inputs and distances between cells as outputs.

Although there are many approaches, it remains unclear which metric best captures the underlying biology. Moreover, any strategy relying on gene expression will remain limited to signals captured in the ligand–receptor list, which may miss ways that cells can communicate. Thus, there remains a need for strategies that determine an aggregate CCC potential and functional–spatial relationships of cells.

A growing toolbox to facilitate CCC analysis. In addition to the core scoring functions (Fig. 2b), many tools use advanced statistical methods to identify intercellular communication (TABLE 2). Existing computational tools can be grouped into one of four categories on the basis of the mathematical models used for identifying ligand–receptor interactions, classified here as (1) differential combination-based, (2) network-based, (3) expression permutation-based and (4) array-based tools.

In differential combination-based tools, significantly differentially expressed genes between cell clusters in single-cell RNA-seq data are identified, and these lists are analysed for ligand–receptor pairs. Two such tools, iTALK\(^10\) and CellTalker\(^11\), use slightly different downstream analysis methods to curate the final list of significantly interacting ligand–receptor pairs. Similarly, PyMINER\(^12\) annotates interacting cell types but labels interactions as ‘activating’ or ‘inhibitory’ according to public interaction databases. Another tool, CCCExplorer\(^13\), uses differential gene expression and PPI database-guided network analysis of downstream targets and transcription factors to determine activated or deactivated signalling events between two groups of samples or cells. These methods are powerful for discerning ligand–receptor interactions in the background of the rest of the data set but are blind to interactions that are common between all groups.

Network methods are used by several tools and exploit properties of connections between genes. For example, CCCExplorer uses them to identify concerted movements in the expression levels of genes involved in ligand–receptor signalling\(^13\). The tools SoptSC\(^14\) and NicheNet\(^15\) use known interactions between ligands, receptors and downstream targets to build a network of ligand–receptor relationships. The former computes a probability, while the latter uses a personalized PageRank algorithm, in both cases to evaluate the effect of ligand–receptor co-expression on downstream signalling genes in the receiver cell and, thus, obtain a continuous score for ranking ligands and receptors based on this effect. Most recently, a method called ‘SpaOTsc’\(^16\) formulates intercellular communication as an optimal transport problem\(^17\) using RNA-seq and spatial information. All these tools use not only the expression levels of ligands and receptors to compute interaction scores but also expression levels of downstream signalling targets, which is intended to be a strength of these techniques. However, they do not account for the rules of gene regulation, so a limitation of these methods is that they are blind to signalling crosstalk, where the signals triggered by one receptor could interfere with the signals triggered by another receptor. Pitfall scenarios could be when an intracellular pathway is highly scored because of the expression of its downstream genes, but its activity may be strongly diminished because of inactivation due to another activated pathway that regulates it with post-translational modifications rather than expression control, as may happen in some cytokine signalling pathways\(^18\), leading to false positives or false negatives not seen with other strategies.

Expression permutation-based tools are the most widely used among those listed in this Review. These methods compute a communication score for each ligand–receptor pair and evaluate its significance either through cluster label permutation, non-parametric tests to assess differences with the null model, or through empirical methods. To increase confidence, CellPhoneDB\(^19\)\(^,\)\(^20\), CellChat\(^21\) and ICELLNET\(^22\) address one of the limitations common to most CCC methods: not considering multisubunit protein complexes. Lists with multimeric proteins are used to assess whether all subunits are simultaneously expressed to identify likely functional ligand–receptor interactions. CellChat also allows multisubunit complexes and incorporates positive and negative effectors into its Hill function-inspired framework. Also, other important features are present in different tools in this category: Giotto\(^23\) includes spatial expression information, accounting for the potential of cells to interact given their physical proximity, whereas ICELLNET is the only tool that returns a global...
**a**

Cell type A

| Ligand   | 1   | 2   | 3   | 4   | 5   |
|----------|-----|-----|-----|-----|-----|
| 1        | 3.81| 3.46| 4.32| 2.32| 3.00|
| 2        | 3.17| 1.58| 2.32| 2.32| 0.00|
| 3        | 3.00| 3.46| 4.91| 2.32| 6.64|
| 4        | 4.00| 5.32| 5.64| 7.71| 8.02|
| 5        | 5.32| 3.32| 4.32| 3.91| 2.32|

Cell type B

| Receptor | 1   | 2   | 3   | 4   | 5   |
|----------|-----|-----|-----|-----|-----|
| 1        | 4.70| 4.58| 4.17| 3.46| 3.61|
| 2        | 3.00| 4.09| 4.46| 6.91| 7.13|
| 3        | 4.32| 2.00| 3.00| 3.17| 1.00|

**b**

| Ligand   | 1   | 2   | 3   | 4   | 5   |
|----------|-----|-----|-----|-----|-----|
| 1        | 3.81| 3.46| 4.32| 2.32| 3.00|
| 2        | 3.17| 1.58| 2.32| 2.32| 0.00|
| 3        | 3.00| 3.46| 4.91| 2.32| 6.64|
| 4        | 4.00| 5.32| 5.64| 7.71| 8.02|
| 5        | 5.32| 3.32| 4.32| 3.91| 2.32|

**c**

Ligand expression

\( (L > 3.3) \)

and

Receptor expression

\( (R > 3.3) \)

**d**

Ligand expression

\( (L) \)

and

Receptor expression

\( (R) \)

**e**

Ligand-receptor pairs

**f**

Expression correlation

\[ \text{Expression correlation} = \text{log FC} \]

Diffenential combinations
Two primary inputs are used for quantifying communication scores: a preprocessed gene expression matrix (part a) and a list of interacting proteins to supervise the analysis (for example, ligand–receptor pairs) (part b). Then a communication score (CS) can be computed for every ligand–receptor pair in a given pair of cells. Here, we show how to perform these calculations for four core functions (parts c–f). These are applied to elucidate paracrine (parts c,d) and autocrine (parts e,f) communication. To assess cell–cell communication, a CS can be computed for each ligand–receptor pair by accounting for the presence of both partners if their expression is greater than a given threshold, which for demonstrative purposes was set arbitrarily to a value of 3.3 (part c), or by multiplying their expression values (part d). Similarly, the CS for each ligand–receptor pair can be the correlation score obtained from their expression across all cell types for autocrine communication (part e). To reveal non-autocrine interactions, the correlation can be computed across pairs of different cells. Particular signatures of each cell type can be extracted through analysing differentially expressed ligands and receptors. Using the cell type-specific differentially expressed genes, we can assign a binary CS and study the ligand–receptors used for autocrine communication (part f). In this example, autocrine communication is evaluated for cell type A by using its differentially expressed genes with respect to cell type B (cell type A-specific genes are located in the coloured quadrant). Analogously to the correlation score, for non-autocrine communication we would need to consider differentially expressed genes in each of the cell types or samples. For a given pair of cells, we can say that a communication pathway is active when the ligand is differentially expressed in one cell and its cognate receptor is differentially expressed in the other. FC, fold change.

Assessing predicted CCC mechanisms

CCC inference should be considered a data-driven process for generating hypotheses, which can lead to different results depending on the strategy adopted (FIG. 3). Thus, validating inferred mechanisms is crucial to confirm associations with phenotypes and behaviours of cell communities. In this section, we discuss current approaches used for this purpose, emphasizing both computational methods to minimize false discoveries and experiments used to validate results, and illustrate hallmark studies that successfully implemented this important step.

Computational minimization of false discoveries. Robust inferences are essential to minimize false positives and false negatives and help reduce the number of validations to perform, which is especially useful when experiments are expensive. From the total pool of inferred ligand–receptor interactions between cells, statistical modeling can assess whether they may result from the null hypothesis and help discard artefactual or non-specific CCC inferences.

Permutation-based analyses can help discard results arising from random noise by prioritizing cell type-enriched ligand–receptor pairs. The cluster labels, representing the cell type, are permuted for each single cell, and the mean gene expression within each permuted cluster is computed, followed by a recalculation of the communication scores. With all communication scores generated for each ligand–receptor pair in a given pair of clusters, a null distribution is built and a P value of the measured communication score can be computed. A full list of tools that use this method is included in the discussion of permutation-based tools (see the section entitled A growing toolbox to facilitate CCC analysis and TABLE 2). As a representative example, the CellPhoneDB tool was applied on single-cell RNA-seq of human first-trimester placentae to understand the regulation of the immune response and how it prevents harmful maternal responses.

Subsampling analysis has been applied to evaluate the robustness of the CCC results. Random subsamples of the original data set are used to rerun the CCC analysis. Using the subsampled results, one can compute true-positive and false-positive rates with respect to the original data set results and quantify how variable the inferences are, given subtle changes in the composition of cell clusters. CellChat is one tool that has applied this method to compare its performance with that of other tools and measure robustness when identifying the role of a specific population of mouse myeloid cells, termed ‘MYL-A’, in wound healing through transforming growth factor-β signalling.

Enrichment analyses have also helped reduce false discoveries. A recent study applied this strategy to study the role of structural cells (that is, fibroblasts, endothelial cells and epithelial cells) in priming immune responses and how they interact with immune cells. It identified ligand–receptor pairs that were enriched between interacting cell types using a Fisher’s exact test over all possible pairs of differentially expressed genes. Supported by other experiments, these interactions revealed the

**Fig. 3** | Toy examples of using core functions to compute communication scores. The tools described also include powerful visualization features that facilitate the interpretation of results. Several of the more common visualization methods are outlined in FIG. 4 and display data by directly plotting ligand–receptor co-expression patterns and communication scores (FIG. 4a–c) and provide higher-level intuition concerning overall CCI levels and the directionality of these effects between cell types (FIG. 4d–f). Thus, several tools not only quantify CCIs and CCC but also facilitate their analysis and interpretation.
| Tool               | Method overview                                                                 | Output                                                                 | Visualization                                      | Available in          | URL                                                      | Refs  |
|--------------------|---------------------------------------------------------------------------------|------------------------------------------------------------------------|----------------------------------------------------|-----------------------|----------------------------------------------------------|-------|
| **Differential combinations** |                                                                                  |                                                                        |                                                    |                       |                                                          |       |
| CellTalker         | Uses differentially expressed ligands and receptors in each cluster to identify unique interactions between clusters | Upregulated and downregulated interactions between all clusters         | Circos plot of differential interactions between clusters | R                     | https://github.com/arc85/celltalker                        | 61    |
| iTALK              | Enumerates differentially expressed ligand and receptor values to identify LRIIs between different clusters | Upregulated and downregulated interactions between all clusters         | CCI networks, Circos plots and boxplots              | R                     | https://github.com/Coolgenome/iTALK                        | 106   |
| PyMINEr            | Uses differentially expressed ligand and receptor pairs to identify altered signalling pathways. Detects both activation and inhibition | Upregulated and downregulated interactions between all clusters         | Network visualization and Circos plots              | Python and standalone application | https://www.sciencescott.com/pyminer                     | 107   |
| **Expression permutation** |                                                                                  |                                                                        |                                                    |                       |                                                          |       |
| CellChat           | Modelling of LRI is generalized from the Hill equation, including expression of both agonists and antagonists. Significance is computed by permutation | Likelihood of CCC between all clusters for all interactions | Alluvial and Circos plots of communication pathways, dot plots of interactions between clusters | R and Web interface | https://github.com/sqjin/CellChat                          | 79    |
| CellPhoneDB        | Randomly permutes cluster labels to generate a null distribution of LRI scores using protein complex subunit architecture to identify significant interactions | Upregulated and downregulated interactions between all clusters         | Heatmap of significant interaction counts, dot plot of LRIIs and cluster combinations | Python and Web interface | https://github.com/Teichlab/cellphonedb                    | 30,83 |
| Giotto             | Randomly permutes cluster labels to generate a null distribution of LRI scores using spatial information to identify significant interactions | Upregulated and downregulated interactions between all clusters         | Heatmap of significant interaction counts, dot plot of LRIIs and cluster combinations | R                     | https://github.com/RubD/Giotto                           | 111   |
| ICCELLNET          | Sums the product of all LRI scores between two clusters to compute an overall CCI score | Intergroup communication scores, P values for these scores and LRI scores | Stacked bar plot of LRIIs, network visualization of interacting groups and pathway-level analysis | R                     | https://github.com/soumelis-lab/ICCELLNET                 | 84    |
| SingleCellSignalR  | Uses a regularized ligand–receptor expression product to measure extent of CCC   | Interaction scores for each LRI between all clusters in the dataset   | Circos plot, tables and graph visualizations of interactions between clusters | R                     | https://github.com/SCA-IRCMM                              | 112   |
| **Graph or network** |                                                                                  |                                                                        |                                                    |                       |                                                          |       |
| CCCExplorer        | A graph of all signalling pathways is built, then a Fisher-like statistic is computed using ligand, receptor and downstream TF expression to identify significant interactions | Graph visualizations of all interactions                               | Interactive directed graphs                        | Standalone application | https://github.com/methodistsmab/CCCEXPLORER              | 67    |
| NicheNet           | A network of ligand–receptor pathway interactions is used to measure the predictive power of the ligand for its downstream pathway targets as an interaction score, based on a personalized PageRank algorithm | Ligand interaction scores and expressing cell types for provided target pathway | Circos plot of interactions between cells or clusters | R                     | https://github.com/saeyslab/nichenetr                      | 54    |
Table 2 (cont.) | Existing tools for measuring cell–cell communication

| Tool     | Method overview                                                                 | Output                                      | Visualization                                      | Available in                  | URL                                                                 | Refs |
|----------|---------------------------------------------------------------------------------|---------------------------------------------|---------------------------------------------------|-----------------------------|----------------------------------------------------------------------|------|
| SoptSC   | Integrates downstream signalling measurements into an LRI scoring function      | Upregulated and downregulated interactions between all clusters | Circos plot of interactions between cells          | MATLAB and R               | https://github.com/WangShuxiong/SoptSC                                | 25   |
| SpaOTsc  | An optimal transport model is used to infer CCC from ligand, receptor and downstream component expression | Likelihood of CCC between all clusters for all LRs | Not implemented                                   | Python                      | https://github.com/zcang/SpaOTsc                                      | 108  |

Tensor based

| scTensor | Tucker decomposition on a tensor of order three to identify key LRIs present in certain cell types | HTML file with summaries of clustering, decomposition and interaction components | Many options for interaction, expression and pattern visualization | R                           | https://github.com/rikenbit/scTensor                                  | 113  |

CCC, cell–cell communication; CCI, cell–cell interaction; LRI, ligand–receptor interaction; TF, transcription factor.

Although providing only tangential information, rather than validation, studies can confirm the presence of proteins to test the potential of signals contributing to CCC events. Immunohistochemistry or western blots have measured the cell type-specific presence of involved signalling complex members. A recent analysis of human liver development used immunohistochemistry to verify VCAM1 production in macrophages and ITGA4 production in erythroid cells of the human fetal liver to suggest a possible interaction between these two cell types and their relevance to lineage differentiation. However, these approaches only support and do not validate a predicted interaction.

Immunohistochemistry and other tagging techniques, such as single-molecule fluorescence in situ hybridization, are frequently used to gain further information about cellular localization for predicted CCIs. This set-up verifies the spatial colocalization of CCC mediators and therefore supports the potential occurrence of a CCC event. For example, in addition to immunostaining to confirm protein expression, statistically significant co-occurrence of IL-33 and its receptor was validated on adjacent AT-2 cells and basophils, respectively, whereas another study used imaging flow cytometry to show interaction between macrophages and liver progenitor cells. In addition, immunofluorescence was used to colocalize the expression of four ligand–receptor interactions (APOE–LRP1, APOE–LDLR, VTN–KDR and LAMA4–ITGB1) to adjacent cells in mouse brain on embryonic day 14.5, supporting the relevance of these interactions to the developing brain. Thus, these approaches confirm spatial colocalization of predicted interactions.

Functional assessment is the most informative validation technique. This approach evaluates whether the phenotypes observed in the interacting cells are the result of specific CCC events. For example, the interaction between VEGFA and KDR (also known as VEGFR2) in liver bud development was confirmed by crucial role of structural cells in priming immune response in an organ-specific fashion.

Data generated with other technologies, such as proteomics, have shaped CCC methods and their potential performance. Specifically, they have helped tune the parameters of methods before making predictions on the gene expression data set of interest, addressing issues originating from imperfect mRNA–protein correlations. For example, Browaeys et al. used several gene and protein expression data sets to optimize NicheNet. Similarly, another study used gene and protein expression data sets to evaluate false-positive rates and tune the parameters of CCC measurement methods. Thus, integrating other types of data is also important to benchmark and optimize tools.

As discussed here, computational strategies can help reduce false discoveries and facilitate the selection of important results for further research or experimental validation. Although we highlighted cases for which these methods helped drive new hypotheses, many other strategies can be used, such as Wilcoxon’s test, random generation of data sets, cross-validation and Welch’s test.

Experimental validation. Many studies use experimental methods to validate CCC mechanisms inferred computationally (Table 1; Supplementary Table 1). These cover three levels of experimental tests: (1) confirmation of the expression of proteins thought to be involved in CCC (for example, through proteomics, enzyme-linked immunosorbent assay, western blot or immunohistochemistry); (2) visualization of interactions between gene products expressed in neighbouring cells (for example, through microscopy coupled with immunostaining, single-molecule fluorescence in situ hybridization or measurement of co-occurrence through flow cytometry); and (3) assessment of the functional role of CCC mediators by performing in vivo or in vitro experiments using activators or inhibitors of the interacting molecules or genetic manipulation of cells to observe effects consistent with the predicted communication.
dosing KDR inhibitors in vitro into micro-liver buds, significantly reducing hepatoblast abundance and impairing differentiation. Elsewhere, 33 candidate ligands affecting haematopoietic differentiation were dosed in vitro in haematopoietic stem cells, and 27 significantly affected the differentiation rate and trajectory of these cells. Similarly, organ structural cells thought to contribute to the immune response were exposed to candidate cytokines, which induced expression changes similar to those seen upon lymphocytic choriomeningitis virus infection, as determined with a mouse model of lymphocytic choriomeningitis virus infection as a reference.

Challenges and future directions

As the methods measuring CCIs improve and the associated results are experimentally validated, new research opportunities will arise that may improve the reliability of inferred CCIs. Furthermore, novel insights will emerge through the study of interactions between cells from different species and the engineering of phenotypes by manipulating communication events.

Multi-omics integration. Although mRNA and protein levels are qualitatively correlated, transcriptomics may not represent a fully accurate view of intercellular communication, as transcript and protein abundances can be uncoupled by post-transcriptional and post-translational processes. For example, the inferred presence of ligand–receptor pairs from transcriptomics may not coincide with their actual presence in proteomic data. Borrowing information from other omics technologies can improve confidence in results, especially for emerging technologies such as Nativeomics, which detects intact ligand–receptor assemblies using mass spectrometry, and INs-seq, which couples single-cell RNA-seq with intracellular protein measurements to simultaneously profile transcription factors, signaling activity and metabolism. Moreover, emerging technologies such as single-cell proteomics will become important inputs for CCI methods and complement single-cell RNA-seq to improve CCI predictions.

Omics integration can extend beyond gene expression. Mammalian cells are covered by a thick glycoalyx, and most hormones and receptors are glycoproteins.
Hence, glycomic data can add contextual information to CCI analyses. Glycosylation has an impact on protein interactions, especially in ligand–receptor binding as many ligands are glycosylated120, and glycosylation can change receptor affinity121–123. This phenomenon may rewire CCIs and affect, for example, development or T cell activation124. Moreover, glycans are involved in interspecies crosstalk, contributing to the specificity of communication125–127. Therefore, integration of additional omics technologies, such as proteomics and glycomics, will help identify additional CCIs that are missed with use of RNA-based methods.

**Adding a spatial dimension.** Ligand mRNA expression and ligand abundance are not the only factors required for communication: the ligand must also be localized in the correct cellular compartment, which RNA-based methods are blind to, and interacting cells are usually close to each other, which routine single-cell RNA-seq experiments cannot measure. To improve reliability of computed CCIs, it is crucial to account for spatial positions of mediators and interacting cells. Single-cell analyses have considered cell location in mouse bone marrow53,128, demonstrating that cell proximity is key to the study of intercellular communication. The study of spatial proximity in interacting cells is an emerging approach129. One technology for profiling these physical interactions, PIC-seq, uses cell sorting to acquire and transcriptionally profile physically interacting cells (PICs) through massively parallel single-cell RNA-seq130. The study authors presented an algorithm for deconvolving the data to capture signals from intercellular physical interactions. One advantage of this method is that it captures more details in the crosstalk of PICs than single-cell RNA-seq. Although this technology is promising, it is currently limited to studying PICs from only two populations of cells. Nevertheless, similarly to other PIC-based methods130,131,132, PIC-seq can be readily integrated into the pipelines described here and leverage the analysis of CCIs by including the intrinsic spatial information that PICs encode. As more approaches emerge for spatial-based transcriptomics133–134 and proteomics135,136, future studies and algorithms should include this information. Accounting for the physical distance between cells will lead to the generation of new scoring functions that may better capture the potential of cells to communicate and interact137,138,139. As an example, ligand–specific diffusion constants can be considered to reflect how effectively gene products can mediate long-distance communication140,141.

**Adding a temporal dimension.** Time is another important factor for studying CCC. It can help elucidate how communication evolves and detect important changes in dynamic processes, such as cellular differentiation and the immune response. One can use samples taken at different time points or infer pseudotimes from the RNA-seq data set as happens when one is studying cellular differentiation through lineage tracing142,143. Then CCC analyses on each (pseudo) time point can identify active ligand–receptor pairs and test the overall interaction potential between cells144. However, time-dependent behaviours might be uncoupled from mRNA levels, hindering the detection of changes through transcriptomics-based CCC predictions. mRNA expression can be temporarily disconnected from the activity of the products owing to long half-lives or storage of the products145. It can also take longer to reach the appropriate abundance of a ligand or a receptor than the production of the cognate mRNA146. For example, endothelial cells store cell adhesion molecules, such as P-selectins, in granules, which are quickly mobilized to the cell surface to start the recruitment of immune cells, instead of expressing those proteins de novo147. Thus, time-dependent analyses can improve CCI discoveries, but other data types should be integrated to distinguish whether phenotypes stem from CCC or other dynamics of gene expression.

**Shedding light on interspecies interactions.** Comprehensive lists of ligand–receptor pairs or interacting surface-secreted proteins are crucial for algorithms that study interspecies CCIs. However, study of interspecies interactions — for example, humans and pathogens148— requires lists of molecular interactions between species. An opportunity awaits to define interspecies PPI lists with omics methods. Considering the avalanche of data generated in the microbiome field149,150 and recent approaches for modelling microbial communities150, lists of interacting proteins or cross-species ligand–receptor pairs to enable CCI analyses would evidently yield novel discoveries. For example, a study mapped the interaction between inclusion membrane proteins secreted by *Chlamydia trachomatis* and cognate human proteins151, providing insights into the host machinery this pathogen uses to establish the intracellular niche needed for infection. Thus, even small-scale lists of interacting partners between the host and a single microorganism, instead of the whole microbiota, will open great opportunities to use CCC methods to understand infections and diseases on the basis of host–pathogen interactions152–154. Although this endeavour will require validation of putative interactions, as done for human–virus protein interactions155–158, results from such studies will have a considerable impact on biomedical and microbiome fields.

**Predicting and manipulating phenotypes.** Models for predicting phenotypes were previously trained with the active communication pathways underlying intercellular interactions159. As more models are developed and their predictive power is increased, they will enable the identification of drug targets and manipulation of phenotypes through cell engineering. Particularly, removing or adding communication pathways with genome editing and cell engineering technologies160 will modify cellular phenotypes to control how pairs of cells interact. This approach will have a great impact on many fields, such as developmental biology161, wherein CCIs are fundamental for cell differentiation into specific lineages. Biomedicine and biotherapeutics will benefit further from controlling CCIs, particularly in modifying disease courses, as is currently done with immune checkpoint inhibitors162,163. As a proof of concept, a tool that induces gene activation when specific cells interact or are directly in contact was recently built by combining a synthetic Notch receptor.
Adaptations applied have elucidated fundamental roles of cells within their communities and how they shape cellular functions, with great potential for future applications, especially in biomedicine and biotherapeutics. Each approach for inferring CCIs and CCC has its own assumptions and limitations to consider; when one is using such strategies, it is important to be aware of these strengths and weaknesses and to choose appropriate parameters for analyses. Methodological and technological challenges remain, but many opportunities exist to increase our understanding of intercellular interactions.

Published online: 09 November 2020
56. Chua, R. L. et al. COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. Nat. Biotechnol. https://doi.org/10.1038/s41598-020-0662-z (2020).

57. Reckermann, T. et al. Spatial network architecture of human immune cells unveiled by quantitative proteomics. Nat. Immunol. 18, 583–593 (2017).

58. Krausgruber, V. et al. Gene expression signatures of transcriptional analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell 171, 1611–1624.e24 (2017).

59. Kumar, M. P. et al. Analysis of single-cell RNA-seq identifies cell-type communication associated with tumor characteristics. Cell Rep. 25, 1658–1668.e4 (2018).

This study is an example of using expression products for measuring intercellular communication and for finding relationships between ligand–receptor pairs and cancer phenotypes.

60. Gillo, A. R. et al. Immune landscape of viral- and carcinoma-driven head and neck cancer. Immunity https://doi.org/10.1016/j.immuni.2019.11.014 (2019).

61. Zhang, Q. et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. Cell 179, 8245–8258.e22 (2019).

62. Song, Q. et al. Dissecting intratumoral myeloid cell heterogeneity. Immunity https://doi.org/10.1016/j.immuni.2020.05.039 (2020).

63. Finotello, F., Rieder, D., Hackl, H. & Trajanoski, Z. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing. BMC Genomics 20, 40 (2019).

This study is an example of using differential gene expression analysis of ligand–receptor pairs reveals important cell-to-cell interactions inside glioma.

Cell Commun. Signal. 17, 48 (2019).

This is pioneering work that trains a machine learning model to predict the diagnosis of patients with glioma from the ligand–receptor pairs used by glioma cells.

64. Murali, T. K. et al. Datasets 2011: a comprehensive integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila.

65. Thurmond, J. et al. FlyBase 2.0: the next generation.

66. Girard, L. R. et al. WormBook: the online review of genetics and genomics. Mol. Biol. Cell 30, 1–10 (2019).

67. Zhang, Q. et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. Cell 179, 8245–8258.e22 (2019).

68. Chang, Y. W. et al. ITALK: an R package to characterize and illustrate intercellular communication. Preprint at bioRxiv https://doi.org/10.1101/2020.08.01.201751 (2020).

69. Wang, Y. et al. Architecture and single cell RNA-seq data analysis of paraneoplastic neuroendocrine tumors. Cell Rep. 26, 1951–1964.e8 (2019).

70. Yang, Y. et al. Functional and structural analysis of SWI/SNF complexes in human skeletal myelofibrosis. Cell Stem Cell 25, 1714–1726.e9 (2019).

71. Conesa, A. et al. A survey of best practices for differential expression analysis: an extended review on quality metrics and data visualization. BioRxiv e1007185 (2019).

In this pioneering work, the authors evaluated gene expression variability due to CCIs by integrating spatial transcriptomics.

Cell Commun. Signal. 17, 113–124 (2019).

This work introduces Proximold, a computational approach for building a cell–cell network from physical and virtual associations between different spatial transcriptomic data. Nature Biotechnol. 2019, 537–543 (2018).

72. Song, Q. et al. Dissecting intratumoral myeloid cell heterogeneity. Immunity https://doi.org/10.1016/j.immuni.2020.05.039 (2020).

73. Finotello, F., Rieder, D., Hackl, H. & Trajanoski, Z. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing. BMC Genomics 20, 40 (2019).

This study is an example of using differential gene expression analysis of ligand–receptor pairs reveals important cell-to-cell interactions inside glioma.

Cell Commun. Signal. 17, 48 (2019).

This is pioneering work that trains a machine learning model to predict the diagnosis of patients with glioma from the ligand–receptor pairs used by glioma cells.

64. Murali, T. K. et al. Datasets 2011: a comprehensive integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila.

65. Thurmond, J. et al. FlyBase 2.0: the next generation.

66. Girard, L. R. et al. WormBook: the online review of genetics and genomics. Mol. Biol. Cell 30, 1–10 (2019).

67. Zhang, Q. et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. Cell 179, 8245–8258.e22 (2019).

68. Chang, Y. W. et al. ITALK: an R package to characterize and illustrate intercellular communication. Preprint at bioRxiv https://doi.org/10.1101/2020.08.01.201751 (2020).

69. Wang, Y. et al. Architecture and single cell RNA-seq data analysis of paraneoplastic neuroendocrine tumors. Cell Rep. 26, 1951–1964.e8 (2019).

70. Yang, Y. et al. Functional and structural analysis of SWI/SNF complexes in human skeletal myelofibrosis. Cell Stem Cell 25, 1714–1726.e9 (2019).

In this pioneering work, the authors evaluated gene expression variability due to CCIs by integrating spatial transcriptomics.

Cell Commun. Signal. 17, 113–124 (2019).

This work introduces Proximold, a computational approach for building a cell–cell network from physical and virtual associations between different spatial transcriptomic data. Nature Biotechnol. 2019, 537–543 (2018).

72. Song, Q. et al. Dissecting intratumoral myeloid cell heterogeneity. Immunity https://doi.org/10.1016/j.immuni.2020.05.039 (2020).

73. Finotello, F., Rieder, D., Hackl, H. & Trajanoski, Z. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing. BMC Genomics 20, 40 (2019).

This study is an example of using differential gene expression analysis of ligand–receptor pairs reveals important cell-to-cell interactions inside glioma.

Cell Commun. Signal. 17, 48 (2019).

This is pioneering work that trains a machine learning model to predict the diagnosis of patients with glioma from the ligand–receptor pairs used by glioma cells.

64. Murali, T. K. et al. Datasets 2011: a comprehensive integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila.

65. Thurmond, J. et al. FlyBase 2.0: the next generation.

66. Girard, L. R. et al. WormBook: the online review of genetics and genomics. Mol. Biol. Cell 30, 1–10 (2019).

67. Zhang, Q. et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. Cell 179, 8245–8258.e22 (2019).

68. Chang, Y. W. et al. ITALK: an R package to characterize and illustrate intercellular communication. Preprint at bioRxiv https://doi.org/10.1101/2020.08.01.201751 (2020).

69. Wang, Y. et al. Architecture and single cell RNA-seq data analysis of paraneoplastic neuroendocrine tumors. Cell Rep. 26, 1951–1964.e8 (2019).

70. Yang, Y. et al. Functional and structural analysis of SWI/SNF complexes in human skeletal myelofibrosis. Cell Stem Cell 25, 1714–1726.e9 (2019).

In this pioneering work, the authors evaluated gene expression variability due to CCIs by integrating spatial transcriptomics.

Cell Commun. Signal. 17, 113–124 (2019).

This work introduces Proximold, a computational approach for building a cell–cell network from physical and virtual associations between different spatial transcriptomic data. Nature Biotechnol. 2019, 537–543 (2018).

72. Song, Q. et al. Dissecting intratumoral myeloid cell heterogeneity. Immunity https://doi.org/10.1016/j.immuni.2020.05.039 (2020).

73. Finotello, F., Rieder, D., Hackl, H. & Trajanoski, Z. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing. BMC Genomics 20, 40 (2019).

This study is an example of using differential gene expression analysis of ligand–receptor pairs reveals important cell-to-cell interactions inside glioma.

Cell Commun. Signal. 17, 48 (2019).

This is pioneering work that trains a machine learning model to predict the diagnosis of patients with glioma from the ligand–receptor pairs used by glioma cells.
data. Protein Cell https://doi.org/10.1007/s13238-020-00727-5 (2020).

130. Giladi, A. et al. Dissecting cellular crosstalk by sequencing physically interacting cells. Nat. Biotechnol. https://doi.org/10.1038/s41587-020-0442-2 (2020).

This work describes PI-Seq, a transcriptomics technology to study physically interacting cells.

131. Halpern, K. B. et al. Paired-cell sequencing spatially resolves gene expression in liver and lung tissue. Nat. Biotechnol. 38, 131–139 (2020).

132. Szczepan, B. M. et al. Neutrophilic cell therapy for inflammatory diseases. Nature 569, 111–121 (2019).

133. Schapero, D. et al. elsCAT analysis of cell phenotypes and interactions in multiplex image cytometry data. Nat. Methods 14, 875–876 (2017).

134. Lundberg, E. & Borner, G. H. Spatial proteomics: a powerful discovery tool for cell biology. Nat. Rev. Mol. Cell Biol. 20, 285–302 (2019).

135. Ren, X. et al. Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. Cell Res. https://doi.org/10.1038/s41422-020-0333-2 (2020).

136. Francis, K. & Palsson, B. O. Effective intercellular communication distances are determined by the relative time constants for cytokine secretion and diffusion. Proc. Natl Acad. Sci. USA 94, 12258–12262 (1997).

137. Atakan, B. in Molecular Communications and Nanonetworks: From Nature to Practical Systems (ed. Atakan, B.) 105–143 (Springer, 2014).

138. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. Nat. Rev. Genet. 21, 410–427 (2020).

139. Zhang, J., Nie, Q., Legacy, and interactions in multiplex image cytometry data. ACS Infect. Dis. 5, 336–344 (2019).

140. Penaranda, C. & Hung, D. T. Single-cell RNA sequencing to understand host:pathogen interactions. Nat. Rev. Genet. 20, 355–357 (2018).

141. Jager, S. et al. Global landscape of HIV-human protein complexes. Nature 481, 365–370 (2011).

142. Bhat, J. et al. Protein interaction mapping identifies HBBP6 as a negative regulator of Ebola virus replication. Cell 175, 1917–1930.e13 (2018).

143. Shah, P. S. et al. Comparative flavivirus-host protein interaction mapping reveals mechanisms of Dengue and Zika virus pathogenesis. Cell 175, 1931–1945.e18 (2018).

144. Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature https://doi.org/10.1038/s41586-020-2286-9 (2020).

145. Adli, M. The CRISPR tool kit for genome editing and beyond. Nat. Commun. 9, 1911 (2018).

146. Hsu, M.-N. et al. CRISPR technologies for stem cell engineering and regenerative medicine. Biotechnol. Adv. 37, 1074–1107 (2019).

147. Kwon, E. D. et al. Manipulation of T-cell costimulatory and inhibitory signals for immunotherapy of prostate cancer. Proc. Natl Acad. Sci. USA 94, 8099–8105 (1997).

148. Xu, W., Atkins, M. B. & McDermott, D. F. Checkpoint inhibitor immunotherapy in kidney cancer. Nat. Rev. Urol. 13, 137–150 (2016).

149. Huang, H. et al. Cell-cell contact-induced gene editing/activation in mammalian cells using a synthetic CRISPR-Cas9 system. Protein Cell 11, 299–305 (2020).

150. de Juan, D., Pazo, F. & Valencia, A. Emerging methods in protein co-evolution. Nat. Rev. Genet. 14, 249–261 (2013).

151. Kuhlman, B. & Bradley, P. Advances in protein structure prediction and design. Nat. Rev. Mol. Cell Biol. 20, 681–697 (2019).

152. Chupakhin, V., Marcou, G., Baskin, I., Varnek, A. & Rognan, D. Predicting ligand binding modes from protein fingerprints. J. Chem. Inf. Model. 53, 765–772 (2013).

153. Fenyo, E., Izzazaga, J. M. C., Jurtz, V., Brunak, S. & Nielsen, M. A generic deep convolutional neural network framework for prediction of receptor-ligand interactions-NetPhosPan application to kinase dual phosphorylation prediction. Bioinformatics 35, 1098–1107 (2019).

154. Kvet, K. et al. Immune-centric network of cytokines and cells in disease context identified by computational mining of Pubmed. Nat. Biotechnol. 36, 651–659 (2018).

155. Ashburner, M. et al. Gene ontology: tool for the general functional annotation of genes. Nucleic Acids Res. 25, 25–29 (2000).

156. UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 43, D204–D212 (2015).

157. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27–30 (2000).

158. Waterhouse, R. M., Tegenfeldt, F., Li, J., Zdobnov, E. M. & Koonvateua, E. V. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res. 41, D538–D565 (2015).