Interactive analysis and exploration of cell-cell communication in single-cell transcriptomics with InterCellar

Marta Interlandi (✉ Marta.Interlandi@uni-muenster.de)

University of Muenster

Kornelius Kerl

Department of Pediatric Hematology and Oncology, University Children’

<https://orcid.org/0000-0002-5676-8102>

Martin Dugas

University of Münster <https://orcid.org/0000-0001-9740-0788>

Article

Keywords: cell-cell communication, single-cell transcriptomics, InterCellar

Posted Date: June 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-525466/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Communications Biology on January 11th, 2022. See the published version at <https://doi.org/10.1038/s42003-021-02986-2>.

Abstract

Deciphering cell-cell communication is a key step in understanding the physiology and pathology of multicellular systems. Recent advances in single-cell transcriptomics have contributed to unraveling the cellular composition of tissues and enabled the development of computational algorithms to predict cellular communication mediated by ligand-receptor interactions. Despite the existence of various tools capable of inferring cell-cell interactions from single-cell RNA sequencing data, the analysis and interpretation of the biological signals often require deep computational expertise. Here we present InterCellar, an interactive platform empowering lab-scientists to analyze and explore predicted cell-cell communication without requiring programming skills. InterCellar guides the biological interpretation through customized analysis steps, several visualization options and the possibility to link biological pathways to ligand-receptor interactions. By analyzing COVID-19 and melanoma cell-cell interactions, we show that InterCellar resolves data-driven patterns of communication and highlights molecular signals through the integration of biological functions and pathways.

Introduction

Cell-cell communication plays a major role in multicellular organisms by driving cellular differentiation, contributing to tissue and organ homeostasis as well as coordinating both favorable and noxious immune responses in disease¹⁻⁴. The advent of single-cell transcriptomics has offered the unprecedented opportunity to decipher the cellular heterogeneity of tissues and organs and, at the same time, investigate how different cell populations communicate at the molecular level. For example, a recent study examined the cellular composition of the human heart and revealed, through the analysis of cell-cell communication, that specific non-cardiomyocyte cells represent major communication hubs and might participate in maintaining heart contraction⁵. Moreover, due to the ongoing pandemic caused by the new coronavirus (SARS-CoV-2), multiple studies used single-cell transcriptomics to investigate which cell types might be more vulnerable to infection and how cellular communication might impact a positive or negative course of the disease⁶⁻⁸. In particular, these studies highlighted a central role of the immune system in the communication with epithelial cells of the lungs and illustrated how communication patterns might be distinguishing patients with mild symptoms from those who developed severe disease.

In the last few years, a number of methods have been developed to predict the occurrence of cell-cell communication by ligand-receptor interaction in single-cell RNA sequencing (scRNA-seq) data. In the seminal paper by Ramilowski *et al.*⁹ the authors provided for the first time a manually curated collection of 2,422 ligand-receptor interactions that were used to predict cell-cell communication. Thereafter, extensive efforts have been made by the scientific community to improve such databases and provide the necessary statistical framework to infer cellular communication between different cell types¹⁰. For example, more comprehensive ligand-receptor collections have been assembled through the inclusion of other molecular entities such as multi-subunit complexes and cofactors^{11,12}. Several computational

strategies rely on the co-expression of a given interaction pair (int-pair), evaluated on a cell cluster level. Specifically, for autocrine interactions, the expression levels of both genes (or gene complexes) comprising the int-pair are evaluated on the same cell cluster. On the contrary, for paracrine interactions, two different cell clusters are tested for gene expression of the first and the second gene (or gene complex). Therefore, this type of inference of cell-cell communication relies on data pre-processing followed by clustering, with the aim of identifying relevant cell types that might take part in the communication¹³. Although many computational tools exist for the inference of cell-cell communication^{10,13}, the majority only provides limited downstream analysis functionalities, hindering the biological interpretation of the predicted interactions. Moreover, computational expertise is often required to visualize and produce interpretable results, thus representing a limiting factor for scientists with a deep biological background but basic programming skills.

Here we present InterCellar, an interactive platform implemented in R/Shiny that enables the downstream analysis and exploration of cell-cell communication based on scRNA-seq data. InterCellar's primary goal is to assist analysts in the biological interpretation of cellular interactions, by combining data-driven results with the possibility to customize the analysis. Requiring no programming skills, InterCellar is conceived as a final step in the analysis pipeline, fostering the collaboration between computational and non-computational scientists.

The workflow of InterCellar builds upon pre-computed, inferred ligand-receptor interactions, which can be uploaded as the first step in the form of a cell-cell interactions dataset (CCI-data) (**Fig. 1**). The analyst has the option to provide either CCI-data generated by InterCellar-supported tools (e.g., CellPhoneDB¹¹ or SingleCellSignalR¹⁴), or CCI-data result of custom algorithms (e.g., Kumar *et al.*¹⁵) that predict cell-cell interactions. The second step allows the user to focus the analysis on three biological domains, named *universes*: cell clusters, genes and biological functions. In each *universe* (*cluster-verse*, *gene-verse* and *function-verse*), different filtering and visualization options are provided, giving the opportunity to interactively subset the data and get qualitative and quantitative insights. Particularly, in the so-called *function-verse*, InterCellar provides multiple resources to perform functional annotation of enriched interactions. Lastly, InterCellar enables a focused inspection of cell-cell communication through the analysis of the so-called int-pair modules, which are defined as groups of interactions with similar functional patterns. The user may select one cell cluster as *viewpoint* and one *flow* of communication among: (i) *directed-outgoing*, i.e., the viewpoint cluster sends ligands to other clusters; (ii) *directed-incoming*, i.e., the viewpoint cluster expresses receptors; and (iii) *undirected*, as in the case of receptor-receptor pairs (see **Methods**). A subset of interactions is then used to define int-pair modules that can be visualized and analyzed together with their significant functions. This last step can be iterated for all cell clusters and flows of interest. Importantly, InterCellar ensures reproducibility of the analysis and facilitates collection and finalization of results by providing multiple download options for tables and figures. In the following, we present InterCellar's main features and demonstrate its general applicability on two datasets, concerning coronavirus disease 2019 (COVID-19) and melanoma.

Results

InterCellar highlights data-driven patterns of cell-cell communication in scRNA-seq data

In order to demonstrate InterCellar's functionalities implemented in the three biological domains (so-called, *universes*), we considered a publicly available COVID-19 dataset from Chua *et al.*⁶, containing nasopharyngeal and bronchial samples from 19 patients and from five healthy controls. In particular, we adopted the clinical classification of patients provided by the study, consisting of 8 moderate cases and 11 critical cases. Moreover, we retained the original cell type labeling performed by the authors and ran CellPhoneDB (v2.0)¹¹ to obtain predicted cell-cell interactions as input to InterCellar. Thus, three separate input datasets were generated, corresponding to control, moderate and critical cases. Cell types can be grouped in (a) epithelial cells, namely basal, ciliated, ciliated-differentiating (ciliated-diff), FOXN4+, ionocyte, *IFNG* responsive cell (IRC), secretory, secretory-differentiating (secretory-diff) and squamous and (b) immune cells, namely B cell, cytotoxic T cell (CTL), regulatory T cell (Treg), natural killer (NK), natural killer T cell (NKT), natural killer T cell-proliferating (NKT-p), neutrophil (Neu), plasmacytoid dendritic cell (pDC), monocyte-derived dendritic cell (moDC), mast cell (MC), resident macrophage (rMa), non-resident macrophage (nrMa) and monocyte-derived macrophage (MoMa).

As the first step, we focused on the analysis of the *cluster-verse* and considered the total number of interactions per cell type while comparing critical to moderate cases (**Fig. 2a**). We observed rather small differences in numbers of interactions for cell clusters belonging to epithelial cells (with the exception of ionocytes), while immune cell types showed a higher variability. In particular, we could confirm the findings of Lin *et al.*⁷, who described a higher number of interactions for macrophages (MoMa and nrMa) as well as a lower number of interactions for T cells (CTL and Treg), when comparing critical to moderate cases. Moreover, in the same comparison, comprehensive consideration of all cell types highlights a striking gain of interactions for mast cells (MC) as well as proliferating natural killer T cells (NKT-p). This novel finding holds true when examining the number of interactions between a certain cell type of interest and all other cell types (**Fig. 2b** and **Supplementary Fig. 1a,b**). Specifically, both immune cells (e.g., MoMa, nrMa, CTL and Treg) and epithelial cells (e.g., ciliated, ciliated-diff, secretory and secretory-diff) show a consistent pattern of communication in critical cases, characterized by a higher number of interactions occurring between all cell types and MC or NKT-p cells. Interestingly, MCs have recently been hypothesized to have a major role in driving hyperinflammation in severe cases of COVID-19, due to their dysfunctional phenotype related to mast cell activation syndrome^{16,17}.

InterCellar unravels the composition of cellular communication by examining ligand-receptor pairs

Another biological domain relevant for cell-cell communication is represented by the proteins (and thus genes, in the case of scRNA-seq) that compose ligand-receptor pairs. InterCellar's *gene-verse* offers the possibility to investigate precisely which cluster-pairs communicate through which genes. We call "int-pair/cluster-pair couplet" the occurrence of an int-pair in a certain cluster-pair. These couplets are visually represented by InterCellar in dot plots, where the analyst interactively chooses int-pairs and cluster-pairs

to examine. Thus, we proceeded in the analysis of the COVID-19 datasets, by plotting the interaction score (calculated by CellPhoneDB, see **Methods**) of int-pairs occurring in the communication between secretory cells, secretory-diff and all other cell types. In particular, by looking at int-pairs composed of selected chemokine ligands mentioned in Chua *et al.*⁶(*CXCL1*, *CXCL3*, *CXCL6*, *CXCL16*, *CXCL17*), we could observe a clear enrichment of ligand-receptor pairs in the communication between secretory as well as secretory-diff cells and neutrophils (**Fig. 3a**). Specifically, these int-pairs were detected at varying levels of expression in the two disease conditions, while they were completely absent in control samples. This finding confirms the hypothesis of the authors concerning neutrophil recruitment induced by secretory cells⁶. However, no relevant difference could be recognized between moderate and critical cases.

To further investigate the latter phenomenon, we made use of InterCellar's *gene-verse* functionalities by highlighting only int-pair/cluster-pair couplets that are unique to a certain phenotype (i.e., control, moderate or critical). We considered two groups of chemokine ligands, namely the *CC*- and *CXC*-subfamilies and selected all possible int-pairs found in each phenotype. *CXCL*-pairs showed a predominant enrichment of interactions unique to critical cases, in both epithelial (47%) and immune cells (57%) (**Fig. 3b** and **Supplementary Fig. 2a**). Interestingly, due to this unbiased view of all int-pairs and cell clusters, we could notice, in moderate cases, neutrophil recruitment carried out by epithelial cells such as basal, ciliated and *FOXN4*⁺ cells, mainly through the pairs *CXCL1* & *CXCR1* and *CXCL1* & *CXCR2*. On the contrary, critical cases showed enrichment of *CXCL3* & *CXCR1* and *CXCL3* & *CXCR2* in the communication from IRC or squamous cells to neutrophils (**Fig. 3b**). Regarding immune cells, critical cases were characterized by an outgoing cellular communication from moDC, MoMa and NKT-p cells towards multiple other immune cell types. Moderate cases showed a unique pattern of communication promoted by NK cells and directed towards immune as well as epithelial (secretory and ionocyte) cells (**Supplementary Fig. 2a**). For *CCL*-pairs, we focused on selected immune cell types, separated into three groups: macrophages (MoMa, nrMa and rMa), NK cells (NK, NKT and NKT-p) and T cells (Treg and CTL) (respectively **Fig. 3c** and **Supplementary Fig. 2b,c**). While for macrophages the proportion of unique interactions favors the critical phenotype (56%), an inverse tendency could be noticed for NK (41% critical) and T cells (20% critical). In particular, among macrophages, MoMa and nrMa displayed many chemokine interactions that were unique to critical cases. These interactions involved recipient cell types such as MC, moDC, neutrophils and NKT-p. Specifically, MoMa communicates with these other cell types via *CCL4L2* while nrMa via *CCL7*. At the same time, interactions of rMa via *CCL3* in controls are absent in COVID-19 cases (**Fig. 3c**). On the contrary, NK and NKT as well as Treg and CTL showed an enrichment of interactions unique to moderate cases, directed towards secretory cells (among others) (**Supplementary Fig. 2b,c**). Altogether, these results are in line with two findings by the authors: on the one hand, critical cases are characterized by a highly inflammatory profile for MoMa and nrMa; on the other hand, a well-balanced immune response distinguishes moderate cases, in which the communication between immune cells and epithelial cells underlie an effective response to the viral infection⁶.

InterCellar links biological functions and signaling pathways to cellular communication

InterCellar's *function-verse* investigates the third biological domain of interest and offers the opportunity to annotate cell-cell communication with biological functions and pathways. We performed functional annotation for each COVID-19 dataset, using all functional databases provided by InterCellar (see **Methods**). Then, we examined functional terms of interest through a sunburst visualization, which combines all int-pair/cluster-pair couplets that have been annotated to that particular function. To investigate signals of a "cytokine storm"¹⁸ in critical COVID-19 cases, as previously suggested by the analysis of chemokines, we selected the functional term "viral protection interaction with cytokine and cytokine receptor" (annotated from KEGG¹⁹) and compared the sunburst plots of moderate and critical cases (**Fig. 4**). As a first observation, we noticed that moderate cases had a higher total number of int-pairs annotated to this function (681 pairs, compared to 553 in critical cases). However, when considering only unique int-pairs, critical cases had the highest number (51 vs. 46). While CTL and Treg are predominant in moderate cases (with the highest total int-pairs, respectively 104 and 101), these two cell types are replaced by MoMa and nrMa in critical cases, occupying third and fourth place in the ranking with almost half of their int-pairs (55 and 54, respectively). Interestingly, control cases are comparable to moderate ones with respect to the ranking of CTL, B cells and Tregs (**Supplementary Fig. 3**). Furthermore, when examining the specific int-pairs, CTL communicated in moderate cases with nrMa (as well as other cell types) through a cytokine-encoding ligand called *LTA* (*lymphotoxin-alpha*), which is completely absent in critical cases. On the other hand, MoMa and nrMa profiles seem to be fairly conserved in the two disease phenotypes, with only a few int-pairs unique to critical cases (e.g., *IL 18* & *IL 18 receptor* for MoMa::CTL; *CCL 7* & *CCR1*, *CXCL 5* & *CXCR1*, *CXCL 5* & *CXCR2* for nrMa::Neu). Overall, these results suggest once again a dysfunctional immune response in COVID-19 critical cases, where a balanced lymphocyte-mediated regulation seems to be overthrown by excessive activation of macrophages.

InterCellar guides a focused analysis of cellular communication through the definition of functionally distinct modules

The workflow of InterCellar concludes with the analysis of int-pair modules, defined as groups of functionally similar interactions. In order to demonstrate this last step, as well as the general usability of InterCellar, we considered a second scRNA-seq dataset, composed of metastatic melanoma samples from Tirosh *et al.*²⁰. This dataset includes both malignant cells and cells from the tumor microenvironment (TME) of 19 patients. We retained the cell type labels assigned by the authors, namely melanoma-malignant cells, T cells, B cells, macrophages (Macro), endothelial cells (Endo), cancer-associated fibroblasts (CAF) and natural killer (NK) cells. Once again, we ran CellPhoneDB (v2.0)¹¹ to obtain ligand-receptor interactions as input to InterCellar.

Since cancer cells have been described as actively recruiting and reprogramming normal cells of the tumor ecosystem²¹, we used InterCellar to perform an in-depth analysis of the interactions that characterize malignant cells in their communication with the TME. Thus, we chose the malignant cell cluster as *viewpoint* in the analysis and focused on *directed-outgoing* interactions as those int-pairs where the ligand is expressed (and sent) by malignant cells to receptors expressed on all other cell types.

Briefly, InterCellar's definition of int-pair modules is based on the functional annotation performed in the *function-verse*: each module is composed of a subset of int-pairs that share common functional patterns. These modules can be visualized in a two-dimensional plot as a UMAP²², which represents int-pairs clustered by functional similarity (**Fig. 5a**, see **Methods**). Moreover, for each module, InterCellar displays a circle plot to identify precisely which cluster-pairs and genes are involved (**Fig. 5b-e**). Hence, a total number of nine int-pair modules was defined for the selected interactions and, for each module, we could associate functional terms found to be statistically significant (using a one-sided permutation test, see **Methods**). Interestingly, by examining the circle plots, the relevance of many int-pairs could be validated by further literature research. For example, we investigated the underlying interactions for modules #1, #3, #4 and #7. Int-pair module #1, characterized by the functional terms “extracellular matrix (ECM) organization” and “integrin signaling pathways”, is composed of collagen-encoding genes expressed by malignant cells, which interact with integrin-complexes expressed by endothelial cells, CAF, B cells and T cells, as well as with malignant cells themselves. Notably, the importance of collagen-integrin interactions in promoting cell invasiveness has been already described in the literature by Zhou *et al.*²³. EPH-Ephrin signaling, which distinguishes, among others, int-pair module #3, has been associated with tumor neovascularization²⁴ and is mainly promoted by interactions between malignant cells and endothelial cells or CAFs. Int-pair module #4 shows greater heterogeneity and comprises interactions associated with vascular endothelial growth factors (*VEGFA* and *VEGFB*), platelet-derived growth factors (*PDGFA*) as well as fibroblast growth factor (*FGF5*) that are enriched to a higher extent between malignant cells and CAFs, compared to the other cell types. Notably, the role of malignant cells in promoting CAFs proliferation has been previously investigated by Izar *et al.*²⁵. However, the authors focused on the interaction between transforming growth factor-alpha (*TGF-alpha*) and epidermal growth factor receptor (*EGFR*), which was not found in this data. Finally, mechanisms of metastatic invasiveness in melanoma have been associated with the overexpression of annexin A1 in malignant cells²⁶. Interactions involving this gene can be seen in the circle plot of int-pair module #7: *ANXA1*, sent by malignant cells, interacts with formyl peptide receptor (*FPR*) exclusively expressed by macrophages, suggesting a deleterious cross-talk between these two cell types.

Discussion

In the context of single-cell transcriptomics, the investigation of cellular cross-talk has been regarded as a valuable method for understanding biological mechanisms in health and disease¹⁰. Despite the existence of many algorithms for predicting the occurrence of cell-cell communication based on ligand-receptor interactions, further analysis and interpretation of the associated biological signals are not trivial due to the complexity of results and the lack of a standardized approach to perform such analyses¹³.

We developed InterCellar as an interactive analysis platform designed to guide scientists in the interpretation of complex cell-cell communication results obtained from scRNA-seq data.

By demonstrating InterCellar's functionalities on two different scRNA-seq datasets, we were able to obtain comprehensive insights on the modalities of cell-cell communication and the underlying ligand-receptor interactions, which could be further corroborated by literature review. Moreover, we have shown that InterCellar is a powerful tool for identifying and highlighting previously unknown molecular interactions inferred from single-cell transcriptomic data. For example, an in-depth analysis of cellular interactions in COVID-19 patients confirmed the primary role of immune response in discriminating between different courses of the disease⁶. In addition, the analysis suggested a novel, consistent pattern of communication for patients with critical COVID-19, involving cell types such as mast cells. These results may provide new exciting hypotheses to deepen the understanding of severe SARS-CoV-2 cases.

Implemented in R/Shiny, InterCellar is distributed as an open-source Bioconductor package, ensuring software robustness, maintenance and compatibility with multiple operating systems. With an existing installation of R and Bioconductor on the user's personal computer, installing the InterCellar package and launching the Shiny app can be accomplished by running two lines of code, which, together with further information, can be found in the online user guide (<https://bioconductor.org/packages/InterCellar/>). Notably, the local installation of InterCellar avoids issues related to data privacy or sharing of unpublished patient data, which is a typical problem for web-based platforms running on external servers. Moreover, each step of InterCellar's workflow is generally completed in a very short time frame (from seconds to few minutes), ensuring a truly interactive and performant analysis.

InterCellar directly builds upon the results of other existing tools developed to infer cell-cell communication, which can be freely chosen by the user. Alongside the automatic import from supported tools, InterCellar accepts custom data with specific format and necessary information (see **Methods**). Since prediction methods often rely on different reference databases, and build their results on diverse statistical and mathematical assumptions¹⁰, evaluating advantages and disadvantages of each method is critical for choosing the method that best fits to the data of the user. Lastly, we want to stress the importance of further experimental validation of the predicted interactions, which could be achieved, for example, by single-molecule fluorescence in situ hybridization or measurement of co-occurrence through flow cytometry¹⁰.

To benchmark InterCellar's features, we identified 18 published tools from two recent review papers^{10,13} as well as from manual search (as of April 2021, **Supplementary Table 1**). We collected methods allowing the analysis of cell-cell communication and excluded 13 methods that required programming skills (e.g., extensive coding in R/Python to perform the analysis) from a systematic comparison, thus focusing on the remaining web-based or standalone systems (**Fig. 6**). We found that one tool, CCCExplorer²⁷, had comparable functionalities to InterCellar, providing a local, interactive analysis that focuses on enriched biological pathways. However, this software was developed for bulk RNA sequencing, therefore lacking a straightforward application in scRNA-seq. Three platforms, namely pyMINER²⁸, CellChat Explorer¹² and Cellinker²⁹ lack a proper interactive analysis, the possibility of customizing the data and download options; therefore, they could be regarded mostly as explorative tools providing static results. Moreover,

CellChat Explorer and Cellinker run remotely on a server, thus requiring data sharing and possibly leading to privacy issues. The authors of CellChat recently provided a Shiny app with the same set of functionalities as CellChat Explorer and the advantage of local installation through a docker container.

In the future, we plan to further extend the input options to other published tools. This might, on the one hand, facilitate internal comparison and validation of biological conclusions drawn from interactions predicted by different methods and, on the other hand, give more flexibility to the end-user.

In conclusion, InterCellar empowers lab-scientists to interactively analyze cell-cell interactions without programming skills. Moreover, it highlights data-driven patterns of cell-cell communication, by providing several graphical outputs and the possibility to link cellular interactions to biological functions. We believe InterCellar will contribute to accelerate the generation of hypotheses and deepen the understanding of the cellular cross-talk in various pathological and physiological systems.

Methods

Implementation

InterCellar is implemented as an R/Shiny application and structured as an R package, which is available on Bioconductor at <https://bioconductor.org/packages/InterCellar/>. We used the R package *golem* to ease the development of a robust Shiny application, as well as multiple R packages to build the user interface (*shiny*, *shinydashboard*, *shinyFiles*, *shinycssloaders*, *shinyFeedback*, *shinyalert*, *htmltools* and *htmlwidgets*). Data handling is achieved through custom R functions that build upon public R/Bioconductor packages, such as *readxl*, *utils*, *DT*, *plyr*, *dplyr*, *tidyr* and *tibble*. The functional annotation of InterCellar's *function-verse* is performed through two R packages that query functional databases, namely *biomaRt*³⁰ and *graphite*³¹. Lastly, graphical outputs are generated by custom R functions based upon R/Bioconductor packages such as *ggplot2*, *plotly*, *circlize*³², *fmsb*, *umap*, *visNetwork*, *igraph*, *dendextend*³³, *factoextra*, *colourpicker*, *scales* and *grDevices*.

Input tools and preprocessing

InterCellar's input data consists of a pre-computed dataset of predicted cell-cell interactions (CCI-data). The application accepts either the output of supported tools (CellPhoneDB v2 and SingleCellSignalR) or of custom analyses performed by the user (for example, as in Kumar *et al.*¹⁵). In the first case, InterCellar automatically parses the output generated by the supported tool, requiring no further data manipulation by the user. In the second case, the custom input data must contain relevant information structured as a table with the following columns: (i) interaction pairs, containing the names of two molecular components that participate in the interaction (e.g., *compA_compB*); (ii) communicating cell clusters, containing names or numbers of the cell populations (divided into two columns, e.g., *clustA* and *clustB*); (iii) molecular types, either ligand or receptor (in two columns, corresponding to *compA* and *compB*); and (iv) a numeric value representing a score for each interaction (e.g., average expression of *compA_compB* over *clustA* and *clustB*).

For both input options, InterCellar will preprocess the data to generate a standardized dataset. These preprocessing steps involve: (1) mapping int-pairs to the associated gene symbols; (2) annotation of the molecular type of each int-pair by combining the provided information with a manually curated set of int-pairs (available at <https://github.com/martaint/InterCellar-reproducibility>); and (3) reordering int-pairs listed as receptor-ligand to ligand-receptor, to ensure consistency in the definition of the communication flow. To this date, InterCellar supports human genes; the analysis of cell-cell communication obtained from other species can be performed by converting gene names to human orthologs.

Filtering and customization

We implemented multiple filtering options to enable a flexible interactive analysis of cell-cell communication. In particular, we distinguish between two types of filters: those applied to the input data or filters applied to visualization options. In the first case, the user can (i) remove entire clusters from the analysis, for example in the event of unknown or poorly defined cell populations (in *cluster-verse*); (ii) refine int-pairs selecting either a minimum interaction score or a maximum p-value (in *cluster-verse*); (iii) select int-pairs by annotation source (as in the case of CellPhoneDB) or by specificity (as computed by SingleCellSignalR) (in *gene-verse*). These three filtering options are applied to the input dataset, therefore the following analyses will be performed on the data subset. The second filtering type concerns graphical outputs and includes features to customize visualizations. These options are only applied to the selected plot and will not affect the underlying dataset. For example, a dot plot visualization available in the *gene-verse* displays int-pair/cluster-pair couplets based on the user's selection of int-pairs of interest. The user can further select a subset of clusters to consider, as well as change the color scale of the dot plot.

Functional annotation of interaction pairs

InterCellar provides interactive annotation of int-pairs with the aim to link biological functions and pathways to cell-cell communication. Specifically, we use two Bioconductor packages, *biomaRt*³⁰ and *graphite*³¹, to automatically query existing databases of functional terms: Gene Ontology^{34,35}, KEGG¹⁹, Reactome³⁶, Biocarta³⁷, PID:NCI-Nature³⁸, Panther³⁹ and PharmGKB⁴⁰. InterCellar's functional annotation (available in the *function-verse*) is required before proceeding to the definition of int-pair modules. Importantly, the annotation of a certain functional term to an int-pair is fulfilled only when all components (e.g., ligand and receptor genes) of the int-pair are enriched by that functional term. Although the user can freely choose which functional sources to consider when performing the annotation, we suggest including as many databases as possible to maximize the number of int-pairs which will be annotated by at least one term. Only these annotated pairs are further considered in the definition of int-pair modules.

Definition of interaction-pair modules

Defining int-pair modules requires three key, user-driven decisions: (1) choice of the *viewpoint* cell cluster, representing the cell type of interest; (2) selection of a communication *flow*, among *directed-outgoing* (in which the viewpoint cluster sends ligands to other clusters), *directed-incoming* (in which the viewpoint cluster expresses receptors) and *undirected* (as in the case of receptor-receptor pairs); and (3) definition

of the number of int-pair modules to consider. In particular, InterCellar subsets the result of the functional annotation (i.e., a binary matrix of int-pairs by functional terms) according to steps (1) and (2). This filtered matrix is given as input to a dimensionality reduction algorithm called UMAP (Uniform Manifold Approximation and Projection²²), which calculates a 2D-embedding of the interaction pairs considered (using cosine similarity as metric to compute distances between data points). Thus, the low-dimensional embedding reflects the similarity of int-pairs based on their functional profiles. Finally, using these UMAP coordinates, a hierarchical clustering defines modules of int-pairs that share similar functional profiles (with euclidean distance and *ward.D2* clustering algorithm). The UMAP as well as a further dendrogram result of hierarchical clustering can be visualized and downloaded. In addition, two plots provide guidance to the choice of the optimal number of int-pair modules: (i) total within-cluster sum of squares (WSS) (i.e., elbow method) and (ii) average silhouette width. Both metrics are computed and visualized using functions provided in the R package *factoextra*. The first metric indicates the compactness of clusters, by looking at intra-cluster variation. Specifically, one seeks to minimize the total WSS, represented as a function of the number of clusters. The elbow method, in particular, defines the optimal number of clusters by looking at a bend (i.e., the elbow) in the plot. We automated this process by using the R package *akmedoids*, which provides a function to find the maximum curvature. The average silhouette width evaluates how well each data point lies within its cluster and, on the contrary, should be maximized. By default, InterCellar uses the optimal number of modules computed by the elbow method to generate UMAPs and dendrograms; this value can however be freely changed by the user. When the total number of unique int-pairs considered (by choosing *viewpoint* and *flow*) is less or equal to 10, only one module is defined to prevent noisy results due to insufficient data points.

Significant functional terms

To facilitate the biological interpretation of the defined int-pair modules, InterCellar calculates an empirical p-value for each annotated functional term indicating statistical significance of a term to a certain int-pair module. To this end, we implemented a one-sided permutation test which considers as test statistic the proportion of int-pairs annotated to each function, per module (called module ratio). By randomly shuffling the module assignment of each int-pair (1,000 times, without replacement) and calculating the test statistic, we generate the distribution under the null hypothesis. Finally, we compute for each function the proportion of the module ratios which are higher or equal to the actual module ratio, thus obtaining an empirical p-value for the specificity of a function to a certain int-pair module. Functional terms are then ranked by p-values, and the threshold for significance can be chosen by the user (0.05 by default).

Datasets

COVID-19

This dataset comprises single-cell RNA sequencing data of 24 total patients, divided into 5 healthy controls, 8 COVID-19 moderate cases and 11 COVID-19 critical cases. A total number of 32 samples were

derived from nasopharyngeal protected specimen brush and bronchial lavage. Classification in moderate and critical cases was performed by the original authors following WHO guidelines. Preprocessed, normalized data were retrieved from Chua et al.⁶, as well as cell type assignment. We removed two cell clusters whose label assignment was poorly defined, namely “unknown epithelial” and “outlier epithelial”, corresponding to ~1.5% of the total number of cells. Moreover, moderate and critical datasets were randomly subsampled to 10,000 cells each (without losing any cell label), while for the control dataset we retained all cells, corresponding to a total of 2,966 cells. Finally, for each of the three datasets, we run CellPhoneDB v2 statistical analysis using default parameters.

Melanoma

For the melanoma dataset, a preprocessed gene expression matrix was downloaded from NCBI Gene Expression Omnibus (GEO) ([GSE72056](#)), containing transcript-per-million gene expression values for a total of 4,645 cells, comprising both melanoma malignant cells and cells from the tumor microenvironment. We removed ~11% of these cells, due to a missing or unknown cell type label, retaining 4,097 final cells. Once again, we run CellPhoneDB v2 statistical analysis with default parameters.

Data Availability

All data analyzed in this study are publicly available. In particular, for the COVID-19 datasets⁶, we retrieved preprocessed data for the three conditions (control, moderate and critical) from FigShare at <https://doi.org/10.6084/m9.figshare.12436517.v2> (data object named “covid_nbt_main.rds”).

For the melanoma dataset²⁰, preprocessed data were downloaded from GEO with Accession Number [GSE72056](#). InterCellar input datasets used in this study are available at <https://github.com/martaint/InterCellar-reproducibility> (CellPhoneDB output). All other relevant data supporting the key findings of this study are available within the article and its [Supplementary Information](#) files or from the corresponding authors upon reasonable request.

Code Availability

The InterCellar package is available on Bioconductor v3.13 (requiring R4.1). A package version compatible with R4.0 can be installed following the instructions on the GitHub repository (<https://github.com/martaint/InterCellar>), which contains source code and user guidelines. The code used to generate InterCellar input data for the two datasets analyzed in this paper (COVID-19 and melanoma) is available at <https://github.com/martaint/InterCellar-reproducibility>.

Declarations

Acknowledgments

We would like to acknowledge and thank members of the K.K. lab, as well as members of IMI, for their insightful comments and suggestions. A special thanks to our colleagues Carolin Walter and Thomas K. Albert for their critical reading of the manuscript; to Flavia Watusi de Faria, Viktoria Melcher, Monika Graf and Teresa Gerling for testing InterCellar on different datasets and giving feedback for improvements. M.D. is supported by funds by European Union grant MDS-RIGHT [634789]. K.K. is supported by funds by the Deutsche Krebshilfe (70113653), the Deutsche Forschungsgemeinschaft (KE 2004/4-1) and the Wilhelm Sander Stiftung (2019.158.1).

Author Contributions

M.I., K.K. and M.D. conceived the project. M.D. supervised the bioinformatics part of the research while K.K. supervised the biological aspects. M.I. developed and implemented the software, analyzed the data and prepared the figures. M.I., K.K. and M.D. wrote the paper. All authors read and approved the final manuscript.

Competing Interests

The authors declare no competing interests.

References

1. Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 1517 (2019).
2. Skelly, D. A. *et al.* Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. *Cell Rep.* **22**, 600–610 (2018).
3. Rajbhandari, P. *et al.* Single cell analysis reveals immune cell–adipocyte crosstalk regulating the transcription of thermogenic adipocytes. *Elife* **8**, e49501 (2019).
4. Melcher, V. *et al.* Macrophage-tumor cell interaction promotes ATRT progression and chemoresistance. *Acta Neuropathol.* **139**, 913–936 (2020).
5. Wang, L. *et al.* Single-cell reconstruction of the adult human heart during heart failure and recovery reveals the cellular landscape underlying cardiac function. *Nat. Cell Biol.* **22**, 108–119 (2020).
6. Chua, R. L. *et al.* COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* **38**, 970–979 (2020).
7. Lin, Y., Loo, L., Tran, A., Moreno, C. & Hesselson, D. Characterization of cell-cell communication in COVID-19 patients. *bioRxiv* (2020).
8. Qi, F., Qian, S., Zhang, S. & Zhang, Z. Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. *Biochem. Biophys. Res. Commun.* **526**, 135–140 (2020).

9. Ramilowski, J. A. *et al.* A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 1–12 (2015).
10. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2020).
11. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
12. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
13. Shao, X., Lu, X., Liao, J., Chen, H. & Fan, X. New avenues for systematically inferring cell-cell communication: through single-cell transcriptomics data. *Protein Cell* **11**, 866–880 (2020).
14. Cabello-Aguilar, S. *et al.* SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2020).
15. Kumar, M. P. *et al.* Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. *Cell Rep.* **25**, 1458–1468.e4 (2018).
16. Theoharides, T. C. Potential association of mast cells with coronavirus disease 2019. *Ann. Allergy Asthma Immunol.* **126**, 217–218 (2021).
17. Afrin, L. B., Weinstock, L. B. & Molderings, G. J. Covid-19 hyperinflammation and post-Covid-19 illness may be rooted in mast cell activation syndrome. *Int. J. Infect. Dis.* **100**, 327–332 (2020).
18. Coperchini, F., Chiovato, L., Croce, L., Magri, F. & Rotondi, M. The cytokine storm in COVID-19: An overview of the involvement of the chemokine/chemokine-receptor system. *Cytokine Growth Factor Rev.* **53**, 25–32 (2020).
19. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
20. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
21. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
22. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).
23. Zhou, J. X., Taramelli, R., Pedrini, E., Knijnenburg, T. & Huang, S. Author Correction: Extracting Intercellular Signaling Network of Cancer Tissues using Ligand-Receptor Expression Patterns from Whole-

tumor and Single-cell Transcriptomes. *Sci. Rep.* **8**, 17903 (2018).

24. Hess, A. R., Margaryan, N. V., Seftor, E. A. & Hendrix, M. J. C. Deciphering the signaling events that promote melanoma tumor cell vasculogenic mimicry and their link to embryonic vasculogenesis: role of the Eph receptors. *Dev. Dyn.* **236**, 3283–3296 (2007).
25. Izar, B. *et al.* Bidirectional cross talk between patient-derived melanoma and cancer-associated fibroblasts promotes invasion and proliferation. *Pigment Cell Melanoma Res.* **29**, 656–668 (2016).
26. Boudhraa, Z. *et al.* Characterization of pro-invasive mechanisms and N-terminal cleavage of ANXA1 in melanoma. *Arch. Dermatol. Res.* **306**, 903–914 (2014).
27. Choi, H. *et al.* Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep.* **10**, 1187–1201 (2015).
28. Tyler, S. R. *et al.* PyMINER Finds Gene and Autocrine-Paracrine Networks from Human Islet scRNA-Seq. *Cell Rep.* **26**, 1951–1964.e8 (2019).
29. Zhang, Y. *et al.* Cellinker: a platform of ligand–receptor interactions for intercellular communication analysis. (2021).
30. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
31. Sales, G., Calura, E., Cavalieri, D. & Romualdi, C. graphite - a Bioconductor package to convert pathway topology to gene network. *BMC Bioinformatics* **13**, 1–12 (2012).
32. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
33. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2015).
34. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
35. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).
36. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).
37. Nishimura, D. BioCarta. *Biotech Software & Internet Report* **2**, 117–120 (2001).
38. Schaefer, C. F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res.* **37**, D674–9 (2009).

39. Mi, H. & Thomas, P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol. Biol.* **563**, 123–140 (2009).
40. Thorn, C. F., Klein, T. E. & Altman, R. B. PharmGKB: The Pharmacogenomics Knowledge Base. in *Pharmacogenomics: Methods and Protocols* (eds. Innocenti, F. & van Schaik, R. H. N.) 311–320 (Humana Press, 2013). doi:10.1007/978-1-62703-435-7_20.

Figures

Figure 1

Representation of InterCellar's workflow. The analysis workflow is composed of three main steps: (1) upload of user-provided data, containing predicted, pre-computed cell-cell interactions (CCI-data); (2) inspection of three biological domains (so-called, universes), whereby the analysis focuses on cell clusters, genes and biological functions; and (3) definition and inspection of interaction-pair (int-pair) modules, after selecting a viewpoint cell cluster and a flow of communication (among directed-outgoing, directed-incoming and undirected). For each step, filtering options (e.g., on cell clusters, genes and functional databases) and multiple visualizations (e.g., networks, barplots, dotplots and circle plots) are provided to conduct a customized analysis. Moreover, all tables and figures can be downloaded in multiple formats, helping the interpretation of complex cell-cell interaction datasets and finalization of results.

Figure 2

InterCellular generates comprehensive insights on the number of interactions per cell cluster. (a) Differences in the total number of interactions per cell type are shown in back-to-back bar plots generated by InterCellular's cluster-verse. All interaction flows are considered (directed-outgoing, directed-incoming and undirected). Dots on the black curve represent differences between the number of interactions in critical compared to moderate cases. Both autocrine and paracrine interactions are considered and plotted separately in each bar. Cell types are automatically color-coded by InterCellular for consistency throughout the analysis. (b) Relative differences in the number of interactions are displayed using radar plots, comparing moderate to critical cases. Each plot displays the number of interactions between a specific cell type and all other cell types participating in the communication, ordered alphabetically, counter-clockwise. Cell clusters belonging to the immune cell group are represented in the first row, while epithelial cell types are shown in the second row. Cell types are abbreviated as: -diff - differentiating; CTL - cytotoxic T cell; FOXN4 - FOXN4+ epithelial cells; IRC - IFNG responsive cell; MC - mast cell; moDC - monocyte-derived dendritic cell; MoMa - monocyte-derived macrophage; Neu - neutrophil; NK - natural

killer cell; NKT - NK T cell; NKT-p - proliferating NKT; nrMa - non-resident macrophage; pDC - plasmacytoid dendritic cell; rMa - resident macrophage; Treg - regulatory T cell.

Figure 3

InterCellar allows in-depth analysis of interaction pairs and their enriched cell clusters. (a) Selected chemokine pairs enriched in secretory and secretory-diff cells are shown in a dot plot, split by control, moderate and critical cases. Cluster-pairs are represented on the x-axis (e.g., cluster1::cluster2), while

selected int-pairs are shown on the y-axis (e.g., geneA & geneB). Only statistically significant interactions are considered (one-sided permutation test by CellPhoneDB, $p < 0.05$). For control cases, int-pairs involving CXCL1, CXCL3 and CXCL6 were not significant and thus are not shown. The score represents the mean interaction-score calculated by CellPhoneDB as the average of the mean expression of the interacting genes, evaluated in the respective clusters. (b) Int-pairs belonging to the CXC-chemokine subfamily are evaluated in epithelial cell clusters. The dot plot represents only unique occurrences of int-pair/cluster-pair couplets, for each phenotype (control, moderate, critical). The overall contribution of each phenotype is summarized in a pie chart. (c) Int-pairs belonging to the CC-chemokine subfamily are evaluated in macrophage cell clusters. As in (b), only unique couplets are represented and overall contributions are summarized in a pie chart.

Figure 4

InterCellar associates cell-cell communication with biological pathways. Sunburst plots for the selected functional term “viral protein interaction with cytokine and cytokine receptor”, in COVID-19 moderate (left) and critical cases (right). Only int-pairs annotated to the selected functional term are considered in the generation of the plot. For a generic int-pair ‘geneA & geneB’, enriched in ‘cluster1::cluster2’, the first cell cluster (cluster1, expressing geneA) is represented in the inner circle, while the second (cluster2, expressing geneB) is shown in the outer circle. The width of each section indicates the relative fraction of int-pairs enriched in that cell type. Sections are arranged in descending order, counterclockwise, starting from the horizontal black line. The total numbers of interactions per cell type are reported in brackets. Gray boxes show the list of int-pairs enriched for CTL::nrMa, MoMa::CTL and nrMa::Neu; int-pairs that are marked by an asterisk are unique to a certain phenotype.

Figure 5

InterCellar drives the definition of interaction-pair modules uncovering distinct functional profiles of communication. (a) UMAP of int-pairs enriched in melanoma malignant cells, when subsetting to directed-outgoing interactions. Each dot represents one int-pair. For each module, manually selected functional terms are shown. (b) Circle plot for int-pair module #1. Cell types are shown in the outer circle, while int-pairs are displayed in the inner one. Directional links originate from ligands and are pointing

towards receptors; they are colored by a scaled (-2,2) interaction score to highlight relative differences in expression. (c), (d) and (e) show circle plots for int-pair module #3, #4 and #7, respectively. s.p. - signaling pathway.

	InterCellar	CCCExplorer	pyMINEr	CellChat Explorer	Cellinker
Programming skills					
Interactive analysis					
Local or remote	local	local	local	remote	remote
Data type	scRNA-seq	bulk RNA-seq	scRNA-seq	scRNA-seq	scRNA-seq
Customization/filtering					
Visualization					
Cluster-based					
Gene-based					
Function-based					
Download					
Figures					
Tables					

Figure 6

Feature comparison between InterCellar and other related tools. Seven main categories evaluate InterCellar’s functionalities against other open-source software allowing the analysis of cell-cell communication. For this comparison, only tools that do not require programming skills were considered. Visualization options are divided into three sub-categories to account for specific features implemented with regards to cell clusters, genes and functional pathways. Automatic download of figures and tables is considered separately.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [InterCellarsupplementary.docx](#)
- [nrreportingsummary011.pdf](#)