RESEARCH ARTICLE

Predicting intercellular communication based on metabolite-related ligand-receptor interactions with MRCLinkdb

Yuncong Zhang[1†], Yu Yang[2,3†], Liping Ren[3†], Meixiao Zhan[1], Taoping Sun[1*], Quan Zou[4*] and Yang Zhang[2,4*]

Abstract

Background Metabolite-associated cell communications play critical roles in maintaining human biological function. However, most existing tools and resources focus only on ligand-receptor interaction pairs where both partners are proteinaceous, neglecting other non-protein molecules. To address this gap, we introduce the MRCLinkdb database and algorithm, which aggregates and organizes data related to non-protein L-R interactions in cell-cell communication, providing a valuable resource for predicting intercellular communication based on metabolite-related ligand-receptor interactions.

Results Here, we manually curated the metabolite-ligand-receptor (ML-R) interactions from the literature and known databases, ultimately collecting over 790 human and 670 mouse ML-R interactions. Additionally, we compiled information on over 1900 enzymes and 260 transporter entries associated with these metabolites. We developed Metabolite-Receptor based Cell Link Database (MRCLinkdb) to store these ML-R interactions data. Meanwhile, the platform also offers extensive information for presenting ML-R interactions, including fundamental metabolite information and the overall expression landscape of metabolite-associated gene sets (such as receptor, enzymes, and transporter proteins) based on single-cell transcriptomics sequencing (covering 35 human and 26 mouse tissues, 52 human and 44 mouse cell types) and bulk RNA-seq/microarray data (encompassing 62 human and 39 mouse tissues). Furthermore, MRCLinkdb introduces a web server dedicated to the analysis of intercellular communication based on ML-R interactions. MRCLinkdb is freely available at https://www.cellknowledge.com.cn/mrclinkdb/.

Conclusions In addition to supplementing ligand-receptor databases, MRCLinkdb may provide new perspectives for decoding the intercellular communication and advancing related prediction tools based on ML-R interactions.

Keywords Metabolite, Ligand-receptor interaction, Cell-cell communication, Single cell RNA sequencing, Spatial transcriptomics

---

†Yuncong Zhang, Yu Yang and Liping Ren contributed equally to this work.

*Correspondence:
Taoping Sun
d201578100@alumni.hust.edu.cn
Quan Zou
zouquan@nclab.net
Yang Zhang
zhy1001@alu.uestc.edu.cn; yangzhang@cdutcm.edu.cn
Full list of author information is available at the end of the article

Zhang *et al. BMC Biology*   (2024) 22:152

Page 2 of 12

Background

Cellular communication forms the foundation of intercellular connections within biological systems, transcending diverse biological forms and tissue structures [1, 2]. In multicellular organisms, intercellular communication plays a pivotal role in various physiological processes, such as immune responses, neural transmission, and developmental regulation [2–4]. Immune cells utilize intercellular communication to identify and combat invading pathogens, maintaining immune homeostasis within the organism [5]. Neuronal cells achieve rapid signal transmission through the release and reception of neurotransmitters [6]. During the organism's developmental journey, intercellular communication orchestrates cell differentiation, tissue formation, and organ development [7]. Furthermore, intervening in intercellular communication stands as a strategic approach, allowing for the modulation of aberrant signal transduction to restore normal physiological function [8, 9]. Therefore, a comprehensive understanding of intercellular communication mechanisms can offer a theoretical foundation for designing novel drugs and therapeutic approaches [10].
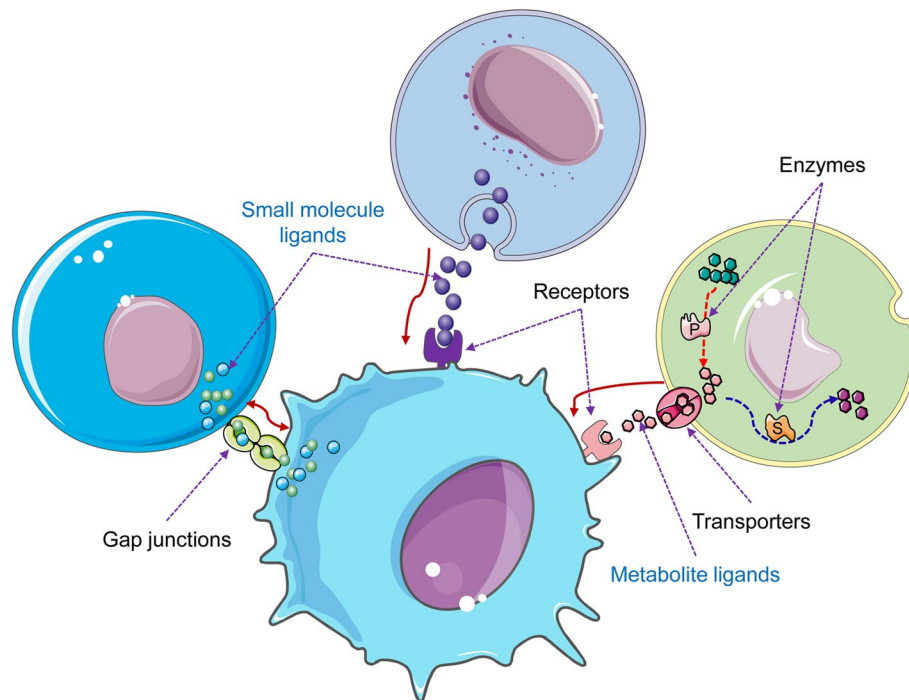
In recent years, propelled by the rapid advancement of single-cell sequencing and spatial transcriptomics technologies, the systematic exploration of intercellular communication within tissues and microenvironments has emerged as a focal point in current research [11–15]. Over the past 3 years, more than 80 algorithms and data resources pertaining to intercellular communication have been developed (https://www.cellknowledge.com.cn/mrclinkdb/LitetatureCCI.html). These include methods based on the expression of cell-to-cell ligand-receptor (L-R) interactions (such as CellChat [16], CellPhoneDB [17], SingleCellSignalR [18], NATMI [19], ICELLNET [20], etc.), methods based on downstream intracellular signaling network of L-R interactions (CellCall [21], CCCExplorer [22], SoptSC, NicheNet [23], CytoTalk [24], scMLnet [25], etc.), and methods incorporating spatial transcriptomics data (Cell2Cell [26], SpaOTsc [27], SVCA [28], stLearn [29], COMMOT [30], etc.).

The aforementioned cell-cell communication (CCC) prediction methods based on single-cell sequencing all rely on prior knowledge of L-R interactions [31]. Therefore, apart from the L-R interaction resources such as DLRP [32], IUPHAR/BPS [33], KEGG [34], and HPMR [35] existing before the single-cell sequencing era, specialized resources targeting CCC research have also emerged, such as Cellinker [36], connectomeDB2020 [17], CellTalkDB [37], OmniPath [38], and PlantPhoneDB [39]. Meanwhile, there are existing databases that provide cell-cell communication analysis results for individual datasets, which can be directly accessed and explored by researchers. These include CellCommuNet

[40], ABC portal [41], AgeAnno [42], HTCA [43], and SPEED [44]. However, most current methods and platforms primarily focus on those cell communications and L-R pairs in which both partners are proteinaceous. Yet, the material foundation of intercellular communication extends beyond proteins and includes numerous non-protein endogenous ligands such as carbohydrates, lipids, inorganic compounds, metal ions, and nucleic acid ligands (Fig. 1). Among these, metabolites are classical non-protein ligands that play significant roles in cell-to-cell interactions [45–48]. For instance, the interaction between tumor-derived histamine and macrophage histamine receptor H1 (HRH1) can lead to T cell dysfunction [49]. Neurotransmitters, typically non-peptide small molecules such as glutamate and gamma-aminobutyric acid (GABA), are crucial for communication between neurons, influencing various physiological processes and behaviors. Glutamate and GABA not only facilitate signal transmission between neurons but also play important roles in neurons-microglia interactions [50]. Additionally, cancer cells often export lactate, which induces a local inflammatory response in the tumor environment, ultimately promoting tumor cell growth and metastasis [51–53].

Presently, a limited number of studies have initiated the development of intercellular communication prediction methods based on non-protein L-R interactions [54]. Zheng et al. introduced the MEBOCOST algorithm, which leverages the expression of various enzymes related to metabolites to predict intercellular communication involving diverse metabolites [54]. Zhao et al. devised NeuronChat [55], a tool based on the expression of enzymes and transport proteins linked to neurotransmitters (including many non-protein ligands), enabling the inference of communication between neurons. In addition to extensively cataloging L-R protein-protein interactions (PPIs), our previous work, Cellinker [36], has also compiled over 400 small molecule-related L-R interaction data. These endeavors notably expand the application scope of intercellular communication prediction tools. Nonetheless, there remains a lack of a dedicated resource that aggregates and organizes data related to non-protein L-R interactions in CCC.

In this study, we developed MRCLinkdb, a resource of metabolite-ligand-receptor (ML-R) interactions for intercellular communication analysis (Fig. 2). The current version of MRCLinkdb documents more than 790 human and 670 mouse ML-R interactions. Additionally, the platform has compiled information on over 1900 enzyme entries and 260 transporter entries associated with these metabolites. The platform offers a user-friendly interface and a wealth of information, facilitating the browsing and presentation of ML-R

Fig. 1 Schematic diagram of ML-R interaction-based intercellular communication

Fig. 2 Construction and content of MRCLinkdb

interactions, and introduces a new webserver dedicated to the analysis of the ML-R interaction-based intercellular communication. MRCLinkdb is free available at https://www.cellknowledge.com.cn/mrclinkdb/.

Results

Statistics of ML-R interactions

The MRCLinkdb platform offers a comprehensive collection of ML-R interactions. For humans, the platform

Zhang *et al. BMC Biology*   (2024) 22:152

Page 4 of 12

hosts 794 ML-R interactions, referencing 239 metabolites and 418 receptor proteins or complexes (Fig. 3A). Furthermore, it documents 1962 enzyme entries (related to 935 unique enzymes) and 244 transporter entries (related to 71 unique transporters). For mouse, the platform

hosts 678 ML-R interactions, referencing 259 metabolites and 356 receptor proteins or complexes (Fig. 3B). It also documents 1913 enzyme entries (related to 900 unique enzymes) and 241 transporter entries (related to 68 unique transporters).
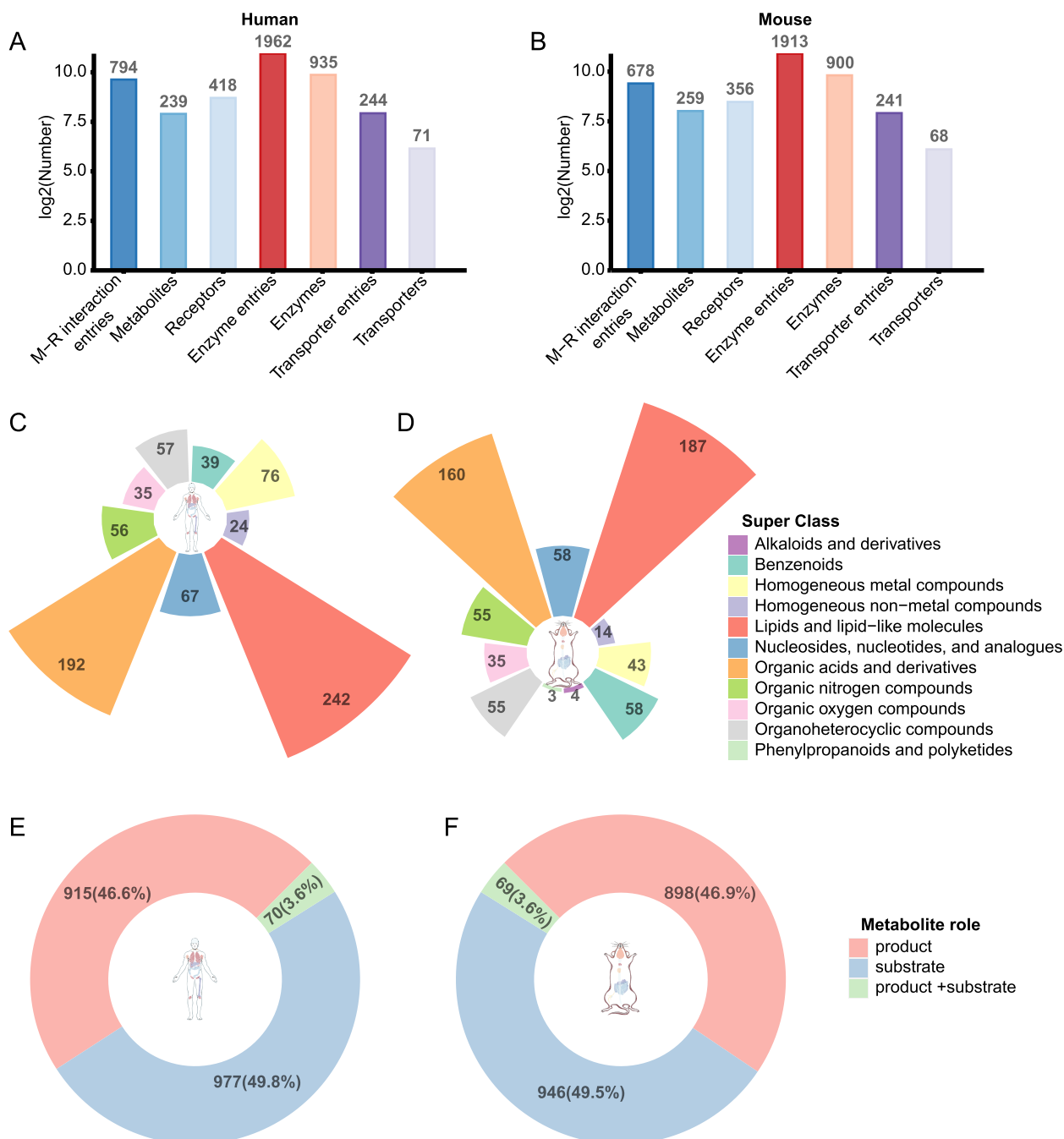
Fig. 3 Statistical data for MRCLinkdb. **A** Statistics for human ML-R interactions, enzymes, and transporter proteins. **B** Statistics for mouse ML-R interactions, enzymes, and transporter proteins. **C** Category distributions of metabolites related to human ML-R interaction. **D** Category distributions of metabolites related to mouse ML-R interaction. **E** Role of human enzymes in metabolite-related reactions. **F** Role of mouse enzymes in metabolite-related reactions

Categorically, human metabolites are distributed across nine types, including 39 benzenoids, 76 homogeneous metal compounds, 24 homogeneous non-metal compounds, 242 lipid-type interactions, 67 nucleosides, nucleotides, and analogues, 192 organic acids and derivatives, 56 organic nitrogen compounds, 35 organic oxygen compounds, and 57 organoheterocyclic compounds (Fig. 3C). Mouse metabolites are categorized into 11 classes, including 4 alkaloids and derivatives, 58 benzenoids, 43 homogeneous metal compounds, 14 homogeneous non-metal compounds, 187 lipids, 58 nucleosides, nucleotides, and analogues, 160 organic acids and derivatives, 55 organic nitrogen compounds, 35 organic oxygen compounds, 55 organoheterocyclic compounds, and 3 phenylpropanoids and polyketides (Fig. 3D).

Evaluating the role of enzymes, 915 human enzyme entries (46.6%) are identified as producers of their associated metabolites. Conversely, 977 entries (49.8%) participate in the consumption of the metabolite, with an additional 70 entries (3.6%) having the dual capacity to both generate and use the corresponding metabolite (Fig. 3E). A similar pattern is seen in mouse, with 898 enzyme entries (46.9%) for production, 946 entries (49.5%) for consumption, and 69 entries (3.6%) capable of both roles (Fig. 3F).

MRCLinkdb interface overview

The MRCLinkdb platform offers an intuitive interface composed of six functional modules designed to access and explore ML-R interaction data. These modules include "search", "browse", "webserver", "submit", "download & API", and a "list of CCI tools". The "browse" module allows easy navigation of all ML-R interactions, with a "more" option in each row granting detailed insights into selected entries. It offers extensive data annotations and strives to present the data as visually as possible. The "download & API" section ensures comprehensive data accessibility, while the "list of CCI tools" presents a curated collection of CCC prediction research from the past 3 years. Users are also encouraged to contribute missing ML-R interaction data through the "submit" page.

Data querying and result presentation

To facilitate user access and browsing of ML-R interaction data, MRCLinkdb has designed an elaborate search and browsing interface. The overall design and layout of the website are illustrated in Fig. 4. If users wish to search for specific metabolites and/or proteins, they can utilize the "search" interface. MRCLinkdb offers two distinct search methods: "Exact Search" and "Batch Search". For these searches, users are required to input or upload details such as the metabolite name, HMDB ID, PubChem CID, PubChem SID, gene symbol, Entrez ID, or UniProt ID of the metabolites/receptors (Fig. 4). The search outcomes are displayed in a tabular format accompanied by an embedded Sankey plot. For more in-depth information regarding specific ML-R interactions, users can navigate to the Detail page by clicking on "more". This detailed page provides extensive data, including (1) basic information about M-R interactions, such as species, metabolite name, HMDB ID, PubChem CID, PubChem SID, gene symbol, Entrez ID, UniProt ID, and source; (2) fundamental details of the metabolite, like its 2D structure, molecular formula, kingdom, class, canonical SMILES, and the associated enzymes and transporter proteins of metabolic reactions; (3) source and destination of metabolites, such as metabolite tissue, biospecimen, and cellular location information, and metabolite-cell interaction information; and (4) expression data for the receptor, enzymes, and transporter proteins based on scRNA-seq (covering tissues and cell types) and bulk RNA-seq/microarray data (encompassing tissues). Furthermore, by hovering over the bar chart of the scRNA-seq, users can highlight and view the distribution of tissues/cell types they are interested in. The right panel will then display the distribution of these tissues/cell types, helping users in understanding the expression distributions of receptors, enzymes, and transporter proteins more clearly.

Online tool for analyzing ML-R interaction-based intercellular communication

MRCLinkdb introduces a dedicated webserver to facilitate analysis of intercellular communication via metabolite and receptor interactions, leveraging scRNA-seq data (Fig. 5). Users are required to upload an expression matrix file (either counts or TPM). In this matrix, rows correspond to gene symbols while columns represent individual cells. An accompanying META file should be provided, detailing each cell's name and type. The platform offers a customizable experience. Users can designate a threshold "*N*", thereby filtering out enzymes, transporter proteins, or receptors expressed in fewer than "*N*%" of cells within a specific cell type. There is also flexibility to define the *p-value* threshold and stipulate the number of iterations for the permutation test. Upon analysis completion, outcomes, including MR_{score} and *p*-values, are displayed as a visually intuitive bubble plot on the result page. Additionally, users have the provision to download the data for offline perusal (Fig. 4).

Discussion

MRCLinkdb functions as a comprehensive platform for analyzing ML-R interactions in intercellular communication. However, it still presents certain limitations. Firstly,

Zhang *et al. BMC Biology*   (2024) 22:152

Page 6 of 12

Fig. 4 The Search page, Result page, and Detail page of MRCLinkdb

MRCLinkdb provides detailed annotations for ML-R interactions, yet it currently lacks the capacity to offer insights into the specific biological processes or pathways in which these interactions are involved. A more nuanced understanding of these processes is crucial for comprehending the broader implications of ML-R interactions within distinct biological contexts. Meanwhile, the resource predominantly focuses on enzymes directly associated with metabolites but does not encompass vital enzyme-related chemical reaction constants, such as K_m, K_s, and K_{cat}. These constants are indispensable for precisely estimating metabolite concentrations. Additionally,
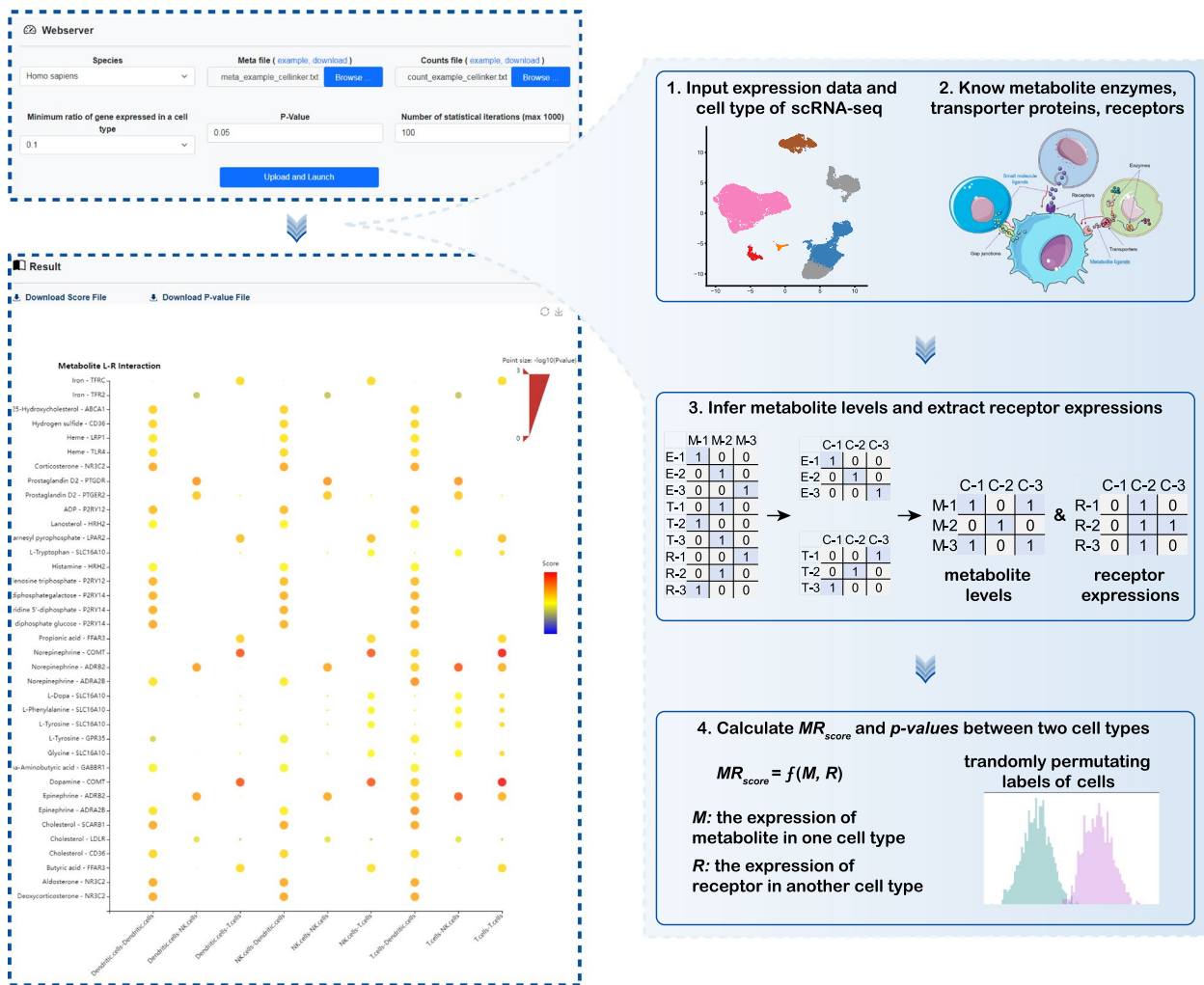
Fig. 5 Webserver and algorithm for ML-R interaction-based intercellular communication analysis

the scope of MRCLinkdb is primarily limited to data related to metabolite-linked transport proteins. However, the field of metabolism encompasses a wide array of transport mechanisms beyond transporter-mediated pathways. Mechanisms such as exosome and vesicle trafficking, along with their associated molecules, remain absent from the resource. Recognizing these diverse aspects of metabolism is essential for a more comprehensive understanding of intercellular communication.

MRCLinkdb's webserver tool employs methods from MEBOCOST and NeuronChat, using enzyme RNA expression levels to estimate the extracellular concentration of related metabolites and thus infer intercellular communication. However, this approach has limitations: (1) the relationship between enzyme mRNA expression levels and protein content is nonlinear, potentially introducing bias when inferring metabolite levels; (2) metabolite concentrations are also influenced by upstream substrates, downstream by-products, and environmental factors such as temperature and pH, which are not fully captured by transcription data alone; (3) high expression levels of metabolites and receptors do not necessarily indicate actual intercellular communication, adding uncertainty to predictions based solely on ligand-receptor levels. While single-cell metabolomics or proteomics data would provide more reliable insights, these techniques currently lag behind transcriptomics in molecular coverage and throughput. Therefore, single-cell and spatial transcriptomics data remain the primary resources for these inferences. Moreover, few algorithms are specifically designed to infer metabolite-receptor interactions. Despite potential high false positive rates, existing algorithms still offer valuable preliminary insights into intercellular communication mechanisms. In conclusion, although using enzyme and transporter protein expression levels

Zhang *et al. BMC Biology*   (2024) 22:152

Page 8 of 12

to infer metabolite-related intercellular communication has limitations, it remains a useful approach given current technological constraints. We look forward to advancements in biotechnology that will improve the accuracy and interpretability of these predictions.

In the evolving landscape of research concerning ML-R interactions, we are firmly committed to maintaining the integrity of our platform by conducting regular updates. Our primary objective is to ensure data accuracy and comprehensiveness as new insights continue to emerge. One of our foremost goals is to broaden the taxonomic scope of our platform, encompassing a wider array of model organisms and systems pertinent to various diseases. This expansion will facilitate a more comprehensive understanding of ML-R interactions in diverse biological contexts. Moreover, our future endeavors will focus on incorporating additional functional data. We intend to highlight the dynamic changes in ML-R interactions across different disease states and elucidate the intricate relationships between ML-R interactions and cellular fate determinations. This expansion will provide a richer set of annotations for ML-R data. To further enrich the dataset, we plan to integrate detailed metabolite-related information. This will include enzyme reaction constants and critical binding constants that serve as bridges between metabolites and receptors. This addition will significantly enhance the depth and breadth of our ML-R data annotations. At last, leveraging these enhanced datasets, we will proceed to refine ML-R prediction methodologies and further enhancing the functionalities of our online server.

Conclusions

MRCLinkdb represents an integrated platform that combines ML-R interaction data collection, curation, and intercellular communication analysis. It catalogs over 790 human and 670 mouse ML-R interactions and compiles information on more than 1900 enzyme entries and 260 transporter protein entries associated with these metabolites. The platform offers a user-friendly interface and comprehensive information, facilitating the exploration and visualization of ML-R interactions. Additionally, it introduces a dedicated web server for intercellular communication analysis based on ML-R interactions. In summary, we believe that this platform possesses invaluable potential to expedite the unraveling of intercellular communication intricacies and propel the progress of algorithms related to scRNA-seq and spatial transcriptomics data, especially those geared towards predicting CCC based on ML-R interactions.

Methods

Collection of ML-R interaction

The ML-R interaction data were manually curated from literatures and databases (Fig. 2). For curation, we search for literatures with the following keywords in Web of Science, PubMed and Google: "metabolite", "secreted metabolite", "metabolite cell–cell communication", "extracellular metabolite", "metal ions", "neurotransmitter", "amino acid", "carbohydrates", "ligand to receptor interaction", "metabolite to receptor interaction". The retrieved publications were then subjected to a preliminary check by expert curators to eliminate false-positive papers. Only experimentally supported ML-R interactions were included in the MRCLinkdb platform. Meanwhile, we also collected the ML-R interactions from three other known resources, including IUPHAR/BPS [33], KEGG [56], and MEBOCOST [54]. The location information of metabolites in tissues, biological specimens, and cellular was collected from HMDB. And the metabolite-cell interaction pairs were collected from MACC [57]. In addition, in order to better assist users in understanding the latest developments and dynamics in CCC prediction research, we have compiled a list of prediction tools and resources from the past 3 years. This compilation, available on our platform (https://www.cellknowledge.com.cn/mrclinkdb/LitetatureCCI.html), includes over 80 tools and databases and will be periodically updated to reflect the latest advancements.

Data annotation

To enhance the usability of the data within the platform, we have performed rigorous annotation of ML-R interaction data, referencing several authoritative resources (Fig. 2). For metabolite details, annotation such as the molecular name, HMDB ID, CID, SID, kingdom, super class, class, molecular formula, canonical SMILES, and 2D structure of metabolite were sourced from NCBI PubChem database [58]. The metabolic reactions, enzymes, and transporter proteins associated with the metabolite were sourced the Human Metabolome Database (HMDB) [59]. The relationship between metabolites and enzymes primarily arises from the chemical reactions involving small molecules catalyzed by enzymes. For gene details, we referenced the NCBI Gene database [60] to acquire the gene name and entrez ID. Protein-related details were obtained from UniProt database [61]. It is worth noting that while the HMDB exclusively offers information on human-related enzymes and transporter proteins, we bridged this gap for mouse enzymes and transporter proteins using gene homology mapping. Orthology data for enzymes and transporter proteins was gathered from HomoloGene [60].

Zhang *et al. BMC Biology*　　(2024) 22:152

Page 9 of 12

scRNA-seq and bulk RNA-seq data collection and processing

Human and mouse single cell atlas were collected from Mouse Cell Atlas (MCA) version 1.0 [62] and version 1.1 [63]. The human cell landscape contains 344,000 cells, consisting of 52 cell types across 35 tissues. The mouse cell landscape contains 174,455 cells, consisting of 44 cell types and 26 tissues. The bulk RNA-seq/microarray data for gene expression across different tissues in human and mouse were collected from the Human Protein Atlas (HPA) project (62 human tissues) [64] and the TISSUES 2.0 database (39 mouse tissues) [65], respectively.

For further processing, we procured scRNA-seq raw data and any supplementary data provided by the authors. The raw data of human was analyzed by the Seurat R package (version = 4.0.5) [66]. Besides otherwise stated, all parameters use the default values. Cells registering fewer than 100 detected genes were excluded to maintain data quality. The expression matrix, post-filtering, was normalized using Seurat's NormalizeData function (with a scale.factor set at 10,000). Leveraging the FindVariableFeatures function, 3000 highly variable genes were identified, forming the basis for subsequent principal component analysis. The tissue and cell type distributions were visualized through t-Distributed Stochastic Neighbor Embedding (tSNE) via the RunTSNE function (dims spanning 1:50). The average gene expression was determined for each tissue and cell type. A parallel analysis approach was adopted for the mouse scRNA-seq dataset.

Inferring ML-R interaction-based cell-cell communication

MRCLinkdb employed a webserver to infer cell communication based on metabolites and their receptor proteins. The methodological approach was inspired by the MEBOCOST [54] and NeuronChat [55] algorithms. By contrast, MRCLinkdb's adoption of the Euclidean norm in its algorithm for inferring intercellular communication scores improve sensitivity, interpretability, and robustness. The Euclidean norm ensures balanced consideration of metabolite concentration and receptor expression levels, enhancing sensitivity to their variations and allowing for additive effects or nonlinear relationships, while also improving robustness by penalizing outliers less severely than the product approach. Within the cell, the concentration of a specific metabolite is highly correlated with the enzymes responsible for its synthesis and utilization. Concurrently, for intercellular communication, these metabolites are relayed outside the cell through specific transport proteins. Given this, extracellular metabolite concentrations are intrinsically linked to their related enzymes and transport proteins. This prompted

us to estimate the extracellular concentration of particular metabolites using the expression levels of enzymes and transport proteins, providing insights into intercellular communication via ML-R interactions. The specific steps of the pipeline are as follows:

1) Estimating extracellular metabolite concentration

$$M = \sqrt[2]{E^2 + T^2} \tag{1}$$

where M is the extracellular metabolite concentration, and E is the expression of the enzymes, determined as:

$$E = \sqrt[m]{\prod_{i=1}^{m} P_i} - \sqrt[n]{\prod_{j=1}^{n} S_j} \tag{2}$$

where P is the average expression of the enzymes that produces the metabolite, and m represents the number of the enzymes that produces the metabolite. S is the average expression of the enzyme that consumes the metabolite. n represents the number of enzymes that consumes the metabolite. And T is the expression of the transporter proteins for the metabolite. The calculation of T is formulated as follows:

$$T = \sqrt[k]{\prod_{g=1}^{k} T_g} \tag{3}$$

where T_g is the average expression of the transporter proteins, and k represents the number of transporter proteins associated with the metabolite. g is the gth transporter proteins in the k transporter proteins.

2) Estimating receptor expression

Receptors, often a protein or complexes, bind or receive metabolites to trigger signaling. If a receptor is a complex containing q proteins, R is defined as the geometric mean of the expression value of all proteins. The formula is as follows:

$$R = \sqrt[q]{\prod_{h=1}^{q} R_h} \tag{4}$$

where R_h is the expression value of protein h in the receptor complex.

3) Inferring the cell-cell communication score

To infer the cell communication between different cell types based on metabolites and their receptor proteins. The MR_{score} is defined as the score of an ML-R interaction entry between different cell types, which is evaluated by the expression of metabolite and receptor. The formula is as follows:

Zhang *et al. BMC Biology*   (2024) 22:152

Page 10 of 12

$$MR_{score} = \sqrt[2]{M^2 + R^2} \qquad (5)$$

where M is the extracellular concentration of metabolite, and R is the expression of receptor, respectively.

4) Estimating statistical significance of CCC score

To examine the statistical significance of each ML-R interaction entry, the *p*-value was estimated using the permutation test (by randomly permutating group labels of cells and then recalculating the communication strength for each permutation).

Architecture

The MRCLinkdb runs on CentOS Linux. The architecture of the MRCLinkdb is composed of two parts. The first part is the frontend, which provides a visual interface of the webpage and an interactive experience with users by Hyper Text Markup Language 5 (HTML5), Cascading Style Sheets (CSS) and JavaScript. The last part is the backend. All data are stored in MySQL. PHP plays a key role in connecting the frontend and backend. PHP can execute different scripts in response to the user's various requests. To streamline the development process and enhance functionality, various plugins were incorporated: DataTables, an addition to the jQuery JavaScript library, this plugin augments HTML tables with advanced features; ECharts, a declarative framework designed for swift creation of web-based visualizations; Bootstrap, a robust frontend toolkit that provides numerous features; Smarty, facilitates the separation of frontend and backend, isolating logical programs from external content. This ensures ease in subsequent management and maintenance.

Abbreviations
MRCLinkdb	Metabolite-Receptor based Cell Link Database
ML-R	Metabolite-ligand-receptor
L-R	Ligand-receptor
CCC	Cell-cell communication
PPIs	Protein-protein interactions

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Guangdong Provincial Key Laboratory of Tumor Interventional Diagnosis and Treatment, Zhuhai Institute of Translational Medicine, Zhuhai People's Hospital (Zhuhai Clinical Medical College of Jinan University), Jinan University, Zhuhai, Guangdong, China. [2]Innovative Institute of Chinese Medicine and Pharmacy, Academy for Interdiscipline, Chengdu University of Traditional Chinese Medicine, Chengdu, China. [3]School of Healthcare Technology, Chengdu Neusoft University, Chengdu, China. [4]Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China.

References
1. Mittelbrunn M, Sánchez-Madrid F. Intercellular communication: diverse structures for exchange of genetic information. Nat Rev Mol Cell Biol. 2012;13(5):328–35.
2. Scadden DT. Nice neighborhood: emerging concepts of the stem cell niche. Cell. 2014;157(1):41–50.
3. Douam F, Gaska JM, Winer BY, Ding Q, Schaewen MV, Ploss A. Genetic dissection of the host tropism of human-tropic pathogens. Annu Rev Genet. 2015;49(1):21–45.
4. Li Y, Wang C, Miao Z, Bi X, Wu D, Jin N, Wang L, Wu H, Qian K, Li C, Zhang T, Zhang C, Yi Y, Lai H, Hu Y, Cheng L, Leung KS, Li X, Zhang F, Li K, Li X, Wang D. ViRBase: a resource for virus-host ncRNA-associated interactions. Nucleic Acids Res. 2015;43(Database issue):D578–82.
5. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Res. 2020;30(6):492–506.
6. Hantak MP, Einstein J, Kearns RB, Shepherd JD. Intercellular communication in the nervous system goes viral. Trends Neurosci. 2021;44(4):248–59.
7. Sonnen KF, Janda CY. Signalling dynamics in embryonic development. Biochem J. 2021;478(23):4045–70.
8. Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. Nature. 2016;529(7586):298–306.

Zhang *et al. BMC Biology*   (2024) 22:152

Page 11 of 12

9. Ning L, Cui T, Zheng B, Wang N, Luo J, Yang B, et al. MNDR v3.0: mammal ncRNA-disease repository with increased coverage and annotation. Nucleic Acids Res. 2021;49(D1):D160–4.

10. Polyakova N, Kalashnikova M, Belyavsky A. Non-classical intercellular communications: basic mechanisms and roles in biology and medicine. Int J Mol Sci. 2023;24(7):6455.

11. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, Park J-E, Stephenson E, Polański K, Goncalves A, Gardner L, Holmqvist S, Henriksson J, Zou A, Sharkey AM, Millar B, Innes B, Wood L, Wilbrey-Clark A, Payne RP, Ivarsson MA, Lisgo S, Filby A, Rowitch DH, Bulmer JN, Wright GJ, Stubbington MJT, Haniffa M, Moffett A, Teichmann SA. Single-cell reconstruction of the early maternal–fetal interface in humans. Nature. 2018;563(7731):347–53.

12. Zhang Q, He Y, Luo N, Patel SJ, Han Y, Gao R, Modak M, Carotta S, Haslinger C, Kind D, Peet GW, Zhong G, Lu S, Zhu W, Mao Y, Xiao M, Bergmann M, Hu X, Kerkar SP, Vogt AB, Pflanz S, Liu K, Peng J, Ren X, Zhang Z. Landscape and dynamics of single immune cells in hepatocellular carcinoma. Cell. 2019;179(4):829–45.e20.

13. Popescu D-M, Botting RA, Stephenson E, Green K, Webb S, Jardine L, Calderbank EF, Polanski K, Goh I, Efremova M, Acres M, Maunder D, Vegh P, Gitton Y, Park J-E, Vento-Tormo R, Miao Z, Dixon D, Rowell R, McDonald D, Fletcher J, Poyner E, Reynolds G, Mather M, Moldovan C, Mamanova L, Greig F, Young MD, Meyer KB, Lisgo S, Bacardit J, Fuller A, Millar B, Innes B, Lindsay S, Stubbington MJT, Kowalczyk MS, Li B, Ashenberg O, Tabaka M, Dionne D, Tickle TL, Slyper M, Rozenblatt-Rosen O, Filby A, Carey P, Villani A-C, Roy A, Regev A, Chédotal A, Roberts I, Göttgens B, Behjati S, Laurenti E, Teichmann SA, Haniffa M. Decoding human fetal liver haematopoiesis. Nature. 2019;574(7778):365–71.

14. Wang M, Liu X, Chang G, Chen Y, An G, Yan L, Gao S, Xu Y, Cui Y, Dong J, Chen Y, Fan X, Hu Y, Song K, Zhu X, Gao Y, Yao Z, Bian S, Hou Y, Lu J, Wang R, Fan Y, Lian Y, Tang W, Wang Y, Liu J, Zhao L, Wang L, Liu Z, Yuan R, Shi Y, Hu B, Ren X, Tang F, Zhao XY, Qiao J. Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. Cell Stem Cell. 2018;23(4):599–614.e4.

15. Sharma A, Seow JJW, Dutertre C-A, Pai R, Blériot C, Mishra A, Wong RMM, Singh GSN, Sudhagar S, Khalilnezhad S, Erdal S, Teo HM, Khalilnezhad A, Chakarov S, Lim TKH, Fui ACY, Chieh AKW, Chung CP, Bonney GK, Poh BGK, Chan JKY, Chow PKH, Ginhoux F, DasGupta R. Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. Cell. 2020;183(2):377–394.e21.

16. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using Cell Chat. Nat Commun. 2021;12(1):1088.

17. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. Cell PhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat Protoc. 2020;15(4):1484–506.

18. Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, Fau C, Lacroix M, Colinge J. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. Nucleic Acids Res. 2020;48(10):e55.

19. Hou R, Denisenko E, Ong HT, Ramilowski JA, Forrest ARR. Predicting cell-to-cell communication networks using NATMI. Nat Commun. 2020;11(1):5011.

20. Noël F, Massenet-Regad L, Carmi-Levy I, Cappuccio A, Grandclaudon M, Trichot C, Kieffer Y, Mechta-Grigoriou F, Soumelis V. Dissection of intercellular communication using the transcriptome-based framework ICELLNET. Nat Commun. 2021;12(1):1089.

21. Zhang Y, Liu T, Hu X, Wang M, Wang J, Zou B, Tan P, Cui T, Dou Y, Ning L, Huang Y, Rao S, Wang D, Zhao X. CellCall: integrating paired ligand–receptor and transcription factor activities for cell–cell communication. Nucleic Acids Res. 2021;49(15):8520–34.

22. Choi H, Sheng J, Gao D, Li F, Durrans A, Ryu S, Lee SB, Narula N, Rafii S, Elemento O, Altorki NK, Wong ST, Mittal V. Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. Cell Rep. 2015;10(7):1187–201.

23. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020;17(2):159–62.

24. Hu Y, Peng T, Gao L, Tan K. CytoTalk: De novo construction of signal transduction networks using single-cell transcriptomic data. Sci Adv. 2021;7(16):eabf1356.

25. Cheng J, Zhang J, Wu Z, Sun X. Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using scMLnet with an application to COVID-19. Brief Bioinform. 2021;22(2):988–1005.

26. Armingol E, Ghaddar A, Joshi CJ, Baghdassarian H, Shamie I, Chan J, Her HL, Berhanu S, Dar A, Rodriguez-Armstrong F, Yang O, O'Rourke EJ, Lewis NE. Inferring a spatial code of cell-cell interactions across a whole animal body. PLoS Comput Biol. 2022;18(11):e1010715.

27. Cang Z, Nie Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. Nat Commun. 2020;11(1):2084.

28. Arnol D, Schapiro D, Bodenmiller B, Saez-Rodriguez J, Stegle O. Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. Cell Rep. 2019;29(1):202–11.e6.

29. Pham D, Tan X, Xu J, Grice LF, Lam PY, Raghubar A, Vukovic J, Ruitenberg MJ, Nguyen Q. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. 2020:2020.05.31.125658.

30. Cang Z, Zhao Y, Almet AA, Stabell A, Ramos R, Plikus MV, Atwood SX, Nie Q. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. Nat Methods. 2023;20(2):218–28.

31. Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell–cell interactions and communication from gene expression. Nat Rev Genet. 2021;22(2):71–88.

32. Graeber TG, Eisenberg D. Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles. Nat Genet. 2001;29(3):295–300.

33. Armstrong JF, Faccenda E, Harding SD, Pawson AJ, Southan C, Sharman JL, Campo B, Cavanagh DR, Alexander SPH, Davenport AP, Spedding M, Davies JA. The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. Nucleic Acids Res. 2020;48(D1):D1006–21.

34. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1):D545–51.

35. Ben-Shlomo I, Yu Hsu S, Rauch R, Kowalski HW, Hsueh AJ. Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. Sci STKE. 2003;2003(187):Re9.

36. Zhang Y, Liu T, Wang J, Zou B, Li L, Yao L, Chen K, Ning L, Wu B, Zhao X, Wang D. Cellinker: a platform of ligand–receptor interactions for intercellular communication analysis. Bioinformatics. 2021;37(14):2025–32.

37. Shao X, Liao J, Li C, Lu X, Cheng J, Fan X. Cell TalkDB: a manually curated database of ligand–receptor interactions in humans and mice. Brief Bioinform. 2020;22(4):bbaa269.

38. Türei D, Valdeolivas A, Gul L, Palacio-Escat N, Klein M, Ivanova O, Ölbei M, Gábor A, Theis F, Módos D, Korcsmáros T, Saez-Rodriguez J. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. Mol Syst Biol. 2021;17(3):e9923.

39. Xu C, Ma D, Ding Q, Zhou Y, Zheng HL. PlantPhoneDB: a manually curated pan-plant database of ligand-receptor pairs infers cell-cell communication. Plant Biotechnol J. 2022;20(11):2123–34.

40. Ma Q, Li Q, Zheng X, Pan J. Cell CommuNet: an atlas of cell-cell communication networks from single-cell RNA sequencing of human and mouse tissues in normal and disease states. Nucleic Acids Res. 2024;52(D1):D597–606.

41. Gao X, Hong F, Hu Z, Zhang Z, Lei Y, Li X, Cheng T. ABC portal: a single-cell database and web server for blood cells. Nucleic Acids Res. 2023;51(D1):D792–804.

42. Huang K, Gong H, Guan J, Zhang L, Hu C, Zhao W, Huang L, Zhang W, Kim P, Zhou X. AgeAnno: a knowledgebase of single-cell annotation of aging in human. Nucleic Acids Res. 2023;51(D1):D805–15.

43. Pan L, Shan S, Tremmel R, Li W, Liao Z, Shi H, Chen Q, Zhang X, Li X. HTCA: a database with an in-depth characterization of the single-cell human transcriptome. Nucleic Acids Res. 2023;51(D1):D1019–28.

44. Chen Y, Zhang X, Peng X, Jin Y, Ding P, Xiao J, Li C, Wang F, Chang A, Yue Q, Pu M, Chen P, Shen J, Li M, Jia T, Wang H, Huang L, Guo G, Zhang W, Liu H, Wang X, Chen D. SPEED: single-cell pan-species atlas in the light of ecology and evolution for development and diseases. Nucleic Acids Res. 2023;51(D1):D1150–9.

45. Ochioni AC, ImbroisiFilho R, Esteves AM, Leandro JGB, Demaria TM, do Nascimento Junior JX, Pereira-Dutra FS, Bozza PT, Sola-Penna M, Zancan P. Clotrimazole presents anticancer properties against a mouse melanoma model acting as a PI3K inhibitor and inducing repolarization

Zhang *et al. BMC Biology*   (2024) 22:152

Page 12 of 12

of tumor-associated macrophages. Biochim Biophys Acta Mol Basis Dis. 2021;1867(12):166263.

46. Nakashima C, Kirita T, Yamamoto K, Mori S, Luo Y, Sasaki T, Fujii K, Ohmori H, Kawahara I, Mori T, Goto K, Kishi S, Fujiwara-Tani R, Kuniyasu H. Malic enzyme 1 is associated with tumor budding in oral squamous cell carcinomas. Int J Mol Sci. 2020;21(19):7149.

47. Talasila KM, Rosland GV, Hagland HR, Eskilsson E, Flones IH, Fritah S, Azuaje F, Atai N, Harter PN, Mittelbronn M, Andersen M, Joseph JV, Hossain JA, Vallar L, Noorden CJ, Niclou SP, Thorsen F, Tronstad KJ, Tzoulis C, Bjerkvig R, Miletic H. The angiogenic switch leads to a metabolic shift in human glioblastoma. Neuro Oncol. 2017;19(3):383–93.

48. Jena BC, Das CK, Banerjee I, Bharadwaj D, Majumder R, Das S, Biswas A, Kundu M, Roy PK, Kundu CN, Mandal M. TGF-beta1 induced autophagy in cancer associated fibroblasts during hypoxia contributes EMT and glycolysis via MCT4 upregulation. Exp Cell Res. 2022;417(1):113195.

49. Li H, Xiao Y, Li Q, Yao J, Yuan X, Zhang Y, Yin X, Saito Y, Fan H, Li P, Kuo WL, Halpin A, Gibbons DL, Yagita H, Zhao Z, Pang D, Ren G, Yee C, Lee JJ, Yu D. The allergy mediator histamine confers resistance to immunotherapy in cancer patients via activation of the macrophage histamine receptor H1. Cancer Cell. 2022;40(1):36–52.e9.

50. Czapski GA, Strosznajder JB. Glutamate and GABA in microglia-neuron cross-talk in Alzheimer's disease. Int J Mol Sci. 2021;22(21):11677.

51. Yabu M, Shime H, Hara H, Saito T, Matsumoto M, Seya T, Akazawa T, Inoue N. IL-23-dependent and -independent enhancement pathways of IL-17A production by lactic acid. Int Immunol. 2011;23(1):29–41.

52. Shime H, Yabu M, Akazawa T, Kodama K, Matsumoto M, Seya T, Inoue N. Tumor-secreted lactic acid promotes IL-23/IL-17 proinflammatory pathway. J Immunol. 2008;180(11):7175–83.

53. Wike-Hooley JL, Haveman J, Reinhold HS. The relevance of tumour pH to the treatment of malignant disease. Radiother Oncol. 1984;2(4):343–66.

54. Zheng R, Zhang Y, Tsuji T, Zhang L, Tseng Y-H, Chen K. MEBOCOST: metabolic cell-cell communication modeling by single cell transcriptome. 2022:2022.05.30.494067.

55. Zhao W, Johnston KG, Ren H, Xu X, Nie Q. Inferring neuron-neuron communications from single-cell transcriptomics through NeuronChat. Nat Commun. 2023;14(1):1128.

56. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):D353–61.

57. Gao J, Mo S, Wang J, Zhang M, Shi Y, Zhu C, Shang Y, Tang X, Zhang S, Wu X, Xu X, Wang Y, Li Z, Zheng G, Chen Z, Wang Q, Tang K, Cao Z. MACC: a visual interactive knowledgebase of metabolite-associated cell communications. Nucleic Acids Res. 2024;52(D1):D633–9.

58. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE. PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res. 2021;49(D1):D1388–95.

59. Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, Dizon R, Sayeeda Z, Tian S, Lee BL, Berjanskii M, Mah R, Yamamoto M, Jovel J, Torres-Calzada C, Hiebert-Giesbrecht M, Lui VW, Varshavi D, Varshavi D, Allen D, Arndt D, Khetarpal N, Sivakumaran A, Harford K, Sanford S, Yee K, Cao X, Budinski Z, Liigand J, Zhang L, Zheng J, Mandal R, Karu N, Dambrova M, Schioth HB, Greiner R, Gautam V. HMDB 5.0: the Human Metabolome Database for 2022. Nucleic Acids Res. 2022;50(D1):D622–31.

60. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, Comeau DC, Funk K, Kim S, Klimke W, Marchler-Bauer A, Landrum M, Lathrop S, Lu Z, Madden TL, O'Leary N, Phan L, Rangwala SH, Schneider VA, Skripchenko Y, Wang J, Ye J, Trawick BW, Pruitt KD, Sherry ST. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2021;49(D1):D10–7.

61. UniProt C. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51(D1):D523–31.

62. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, Huang D, Xu Y, Huang W, Jiang M, Jiang X, Mao J, Chen Y, Lu C, Xie J, Fang Q, Wang Y, Yue R, Li T, Huang H, Orkin SH, Yuan GC, Chen M, Guo G. Mapping the mouse cell atlas by Microwell-Seq. Cell. 2018;172(5):1091–107.e17.

63. Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y, Chen H, Wang J, Tang H, Ge W, Zhou Y, Ye F, Jiang M, Wu J, Xiao Y, Jia X, Zhang T, Ma X, Zhang Q, Bai X, Lai S, Yu C, Zhu L, Lin R, Gao Y, Wang M, Wu Y, Zhang J, Zhan R, Zhu S, Hu H, Wang C, Chen M, Huang H, Liang T, Chen J, Wang W, Zhang D, Guo

64. G. Construction of a human cell landscape at single-cell level. Nature. 2020;581(7808):303–9.

64. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. Proteomics. Tissue-based map of the human proteome. Science (New York, NY). 2015;347(6220):1260419.

65. Palasca O, Santos A, Stolte C, Gorodkin J, Jensen LJ. TISSUES 2.0: an integrative web resource on mammalian tissue expression. Database. 2018;2018:bay003.

66. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573–87.e29.

Publisher's Note