Comparison of Resources and Methods to infer Cell-Cell Communication from Single-cell RNA Data

Daniel Dimitrov[1], Dénes Türei[1], Charlotte Boys[1], James S. Nagai[3], Ricardo O. Ramirez Flores[1], Hyojin Kim[1], Bence Szalai[2], Ivan G. Costa[3], Aurélien Dugourd[1], Alberto Valdeolivas[4], Julio Saez-Rodriguez[1,*]

[1] Heidelberg University, Faculty of Medicine, and Heidelberg University Hospital, Institute for Computational Biomedicine, Bioquant, Heidelberg, Germany
[2] Faculty of Medicine, Department of Physiology, Semmelweis University, Budapest, Hungary
[3] Institute for Computational Genomics, Faculty of Medicine, RWTH Aachen University, Aachen 52074 Germany
[4] Roche Pharma Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, Basel, Switzerland

* Corresponding author: julio.saez@uni-heidelberg.de

ORCIDs:
AD: 0000-0002-0714-028X
AV: 0000-0001-5482-9023
BS: 0000-0002-9320-5704
CB: 0000-0003-3955-5110
DD: 0000-0002-5197-2112
DT: 0000-0002-7249-9379
JSN: 0000-0002-7655-7206
JSR: 0000-0002-8552-8976
IC: 0000-0003-2890-8697
RORF: 0000-0003-0087-371X
HK: 0000-0003-1553-4702

Abstract

The growing availability of single-cell data has sparked an increased interest in the inference of cell-cell communication from this data. Many tools have been developed for this purpose. Each of them consists of a resource of intercellular interactions prior knowledge and a method to predict potential cell-cell communication events. Yet the impact of the choice of resource and method on the resulting predictions is largely unknown. To shed light on this, we created a framework, available at https://github.com/saezlab/ligrec_decoupler, to facilitate a comparative assessment of methods for inferring cell-cell communication from single cell transcriptomics data and then compared 15 resources and 6 methods. We found few unique interactions and a varying degree of overlap among the resources, and observed uneven coverage in terms of pathways and biological categories. We analysed a colorectal cancer single cell RNA-Seq dataset using all possible combinations of methods and resources. We found major differences among the highest ranked intercellular interactions inferred by each method even when using the same resources. The varying predictions lead to fundamentally different biological interpretations, highlighting the need to benchmark resources and methods.

Findings

- Built a framework to systematically combine 15 resources and 6 methods to estimate cell-cell communication from single-cell RNA data
- Cell-cell communication resources are often built from the same original databases and very few interactions are unique to a single resource. Yet overlap varies among resources and certain biological terms are unevenly represented
- Different methods and resources provided notably different results
- The observed disagreement among the methods could have a considerable impact on the interpretation of results

Keywords

cell-cell communication; ligand-receptor; single-cell; transcriptomics;

1. Introduction

The growing availability of single-cell RNA sequencing (scRNA-Seq) data is helping us improve our understanding of the cellular heterogeneity of tissues. Furthermore, Spatial Transcriptomics has recently emerged as a technology to measure gene expression while preserving the spatial distribution of cells in a sample, thus providing an unprecedented opportunity to decipher tissue architecture and organization [1]. These advancements have in turn led to an increased interest in the development of tools for cell-cell communication (CCC) inference. CCC commonly refers to interactions between secreted ligands and plasma membrane receptors. This picture can be broadened to include secreted enzymes, extracellular matrix proteins, transporters, and interactions that require the physical contact between cells, such as cell-cell adhesion proteins and gap junctions [2]. For simplicity, we refer to all of these events involving protein-protein interactions as CCC. CCC events are essential for homeostasis, development, and disease, and their estimation is becoming a routine approach in scRNA-seq data analysis [3].

A number of computational tools and resources have emerged that can be further classified as those that predict CCC interactions alone [4–13], and those that additionally estimate intracellular pathway activities related to CCC [14–18]. Here, we focus on the former (Table 1). These CCC tools typically use gene expression information obtained by scRNA-Seq. In general, single cells are clustered by their gene expression profile and cell type identities are assigned to the clusters based on known gene markers. Then, CCC tools can predict intercellular crosstalk between any pair of clusters, one cluster being the source and the other the target of a CCC event. CCC events are thus typically represented as a one-to-one interaction between a 'transmitter' and 'receiver' protein, expressed by the source and target cell clusters, respectively. The information about which transmitter binds to which receiver is extracted from diverse sources of prior knowledge. Roughly, CCC tools then estimate the likelihood of crosstalk based on the expression level of the transmitter and the receiver in the source and target clusters, respectively. Every

tool has two major components: a resource of prior knowledge on CCC (interactions), and a method to estimate CCC from the known interactions and the dataset at hand. Most tools have been published as the combination of one resource and one method, but in principle any resource could be combined with any method.

Despite the aforementioned common premises to explore CCC events, each tool uses a different method, such as permutation of cluster labels, regularizations, and scaling, to prioritize interactions according to the input datasets (Table 1). In turn, these different approaches result in diverse scoring systems that are difficult to compare and evaluate. The difficulties are further exacerbated by the lack of an appropriate gold standard to benchmark the performance of CCC methods [3,19]. Nevertheless, different strategies have been used to indirectly evaluate the methods' performance, including a presumed correlation between CCC activity and spatial adjacency [13,18], recovering the effect of receptor gene knockouts [18], robustness to subsampling [13], agreement with proteomics [11], simulated scRNA-Seq data [8], and the agreement among methods [9,11,13,18].

The available prior knowledge resources, largely composed of ligand-receptor, extracellular matrix, and adhesion interactions, are typically distinct but often show partial overlap [2,20]. Some of these resources also provide additional details for the interactions such as information about protein complexes [2,7,13,21,22], subcellular localisation [2,13], and classification into signalling pathways and categories [13,21] (Supp. Table 1). CCC resources are often manually curated and/or built from other resources, with varying proportions of expert curation and literature support [2,20]. Some databases gather and harmonize the information contained in the individual resources [2]. Despite the fact that CCC inference is constrained by the prior knowledge used, the impact of resource choice is largely unexplored, with the only exception, to our knowledge, of a descriptive comparison of 4 resources with one method [20]. It remains thus unclear how the choice of resource and method affects the results and thereby the biological interpretation of the scRNA-seq data.

Here, we systematically compared all combinations of 15 resources and 6 CCC methods (Figure 1). First, we explored the degree of overlap among resources and whether certain resources are biased toward specific biological terms, such as pathways and functional cancer states. Then, we analysed how different combinations of resources and methods influence CCC inference, by decoupling the methods from their corresponding resources. In particular, we explored their impact on the predicted CCC using a publicly available colorectal cancer scRNA-Seq dataset [23]. Our framework, available at https://github.com/saezlab/ligrec_decoupler, establishes a uniform interface to all the resources and methods in any combination. We see this work as a platform for further analyses, benchmarks, and method development, and we invite all interested parties to join us in this endeavour.
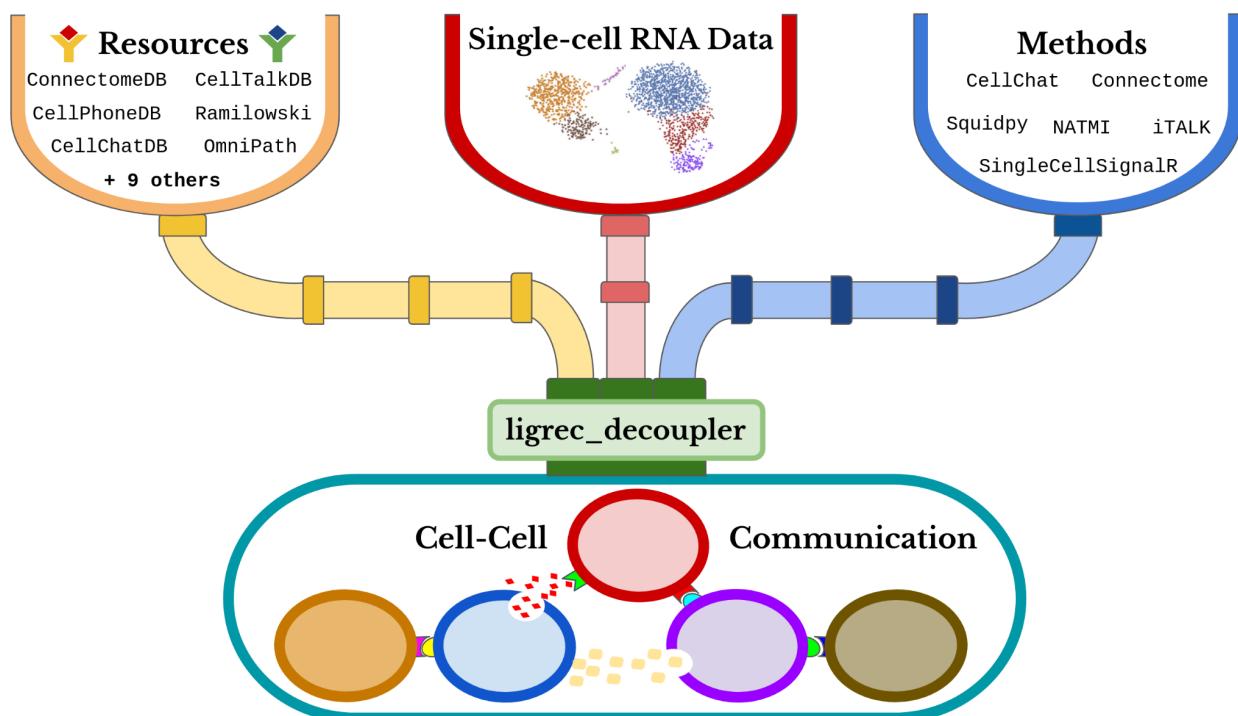

Figure 1. The Cell-Cell Communication Framework.

Table 1. *The tools included in the framework.*

Each tool uses a resource and a method with a specific scoring system. Each method considers expression at the cell cluster level, and all of the scoring systems presented here are based on the expression of transmitters and receiver genes in the source and target cells, respectively. CellPhoneDBv2 algorithm was included via its implementation in Squidpy [24]. For further details, check the original references.

| Tool	Resource	Methods' scoring systems
CellChat[13]	CellChatDB	1) **Probability** - based on the expression of differentially expressed transmitter and receiver genes and their mediators, calculated with the law of mass action 2) **P-values†** - significance identified via permutation of cell cluster labels and recalculating the probabilities for each cell pair and each transmitter-receiver interaction
Squidpy#[24] (CellPhoneDB [7,25,26])	OmniPath or CellPhoneDB	1) **Truncated Mean** - average expression of transmitter and receivers, the minimum expression (by default) of heteromeric complex of subunits 2) **P-values†** - significance identified via permutation of cell cluster labels to determine a null distribution of means for each receiver-transmitter interaction
Connectome[9]	Ramilowski	1) **weight_norm** - derived via the product (by default) of the normalized expression of transmitter and receiver genes 2) **weight_scale†** - derived from the function (mean, by default) of the z-scores of the transmitter and the receiver, scaled according to cell cluster specificity
iTALK[5]	iTALK	1) **Expression Z-score products** - based on the differentially expressed transmitter and receiver genes between clusters
NATMI[10]	ConnectomeDB 2020	1) **Mean-expression edge weight** - transmitter and receiver gene expression product 2) **Specificity-based edge weight†** - the mean expression of the transmitter and receiver are divided by the sum of the means of the same transmitters/receivers across all cell clusters
SingleCellSignalR[11] (SCA)	LRdb	1) **LRScore** - a regularized score calculated using the squared expression of the transmitter and receiver (sqTRE) divided by sum of the mean of the count matrix and sqTRE
† Explicitly measures cell-cluster specific communication (referred as "cluster-specific measures") # Here we refer to the re-implementation of the CellPhoneDB method as 'Squidpy', even though Squidpy is a spatial transcriptomics framework with a much broader range of functionalities.		

6

2. Results

2.1 Resource Uniqueness and Overlap

To investigate the lineages of CCC resources, we manually gathered information about the origins of each resource. Many of these resources share the same original data sources, including general biological databases such as KEGG [27,28], Reactome [29], and STRING [30] (Figure 2A). Moreover, interactions from Guide to Pharmacology [31], CellPhoneDB [7], and in particular Ramilowski [32], were incorporated into subsequently published resources. All these resources are integrated into OmniPath's CCC resource [2], along with additional CCC interactions from other sources (e.g. SIGNOR [33], Adhesome [34], SignaLink [35], and others). We filtered the OmniPath CCC interactions by quality (4.1 Methods).

As a consequence of their common origins, we noted limited uniqueness across the resources, with mean percentages of 4.6 unique receivers, 5.3 unique transmitters, and 16.8% unique interactions, for all resources (Figure 2B; Supp. Table 1). OmniPath and CellChatDB [13] had the largest degree of uniqueness, with 4, 16, and 46% for OmniPath and 17, 12, and 50% for CellChatDB in terms of receivers, transmitters, and interactions, respectively. Despite the sparse uniqueness among the resources, the pairwise overlap between them varied (Figure 2C; Supp. Figure 1). Particularly high similarity was observed between CellTalkDB [20], ConnectomeDB [10], talklr [36], iTALK [5], LRdb [11], and Ramilowski (Figure 2C; Supp. Figure 2). The aforementioned resources, together with OmniPath, contained on average more than 65% the interactions present in the other resources (Supp. Figure 3), largely explained by each including a large proportion (>80%) of the interactions present in Ramilowski. CellChatDB, CellPhoneDB and Baccin [22] showed limited similarity with other resources, as each included ~45% of the interactions present in any other resource, on average. These latter resources include protein complexes, which were dissociated and treated as distinct protein subunits in our resource analyses. The smaller resources ICELLNET [12], Guide to Pharmacology, HMPR [37] and Kirouac2010 [38] were most dissimilar with the

remainder of the resources and included on average only 21, 28, 17, and 7% of the interactions present in the other resources, respectively. The similarity among the resources was generally higher when considering transmitters, and receivers in particular (Supp. Figure 1-3).

In summary, our results indicate that many of the transmitters, receivers, and interactions are not unique to any single resource, due to their common origins. However, different resources include varying proportions of the collective CCC prior knowledge.

C)

	Baccin2019	CellChatDB	CellPhoneDB	CellTalkDB	ConnDB2020	EMBRACE	GuidePharm	HPMR	ICELLNET	iTALK	Kirouac2010	LRdb	OmniPath	Ramilowski	talklr
Baccin2019	1	0.324	0.42	0.393	0.501	0.539	0.172	0.119	0.2	0.459	0.027	0.414	0.265	0.56	0.464
CellChatDB	0.324	1	0.503	0.447	0.398	0.401	0.132	0.087	0.153	0.449	0.021	0.456	0.384	0.364	0.451
CellPhoneDB	0.42	0.503	1	0.423	0.443	0.394	0.202	0.107	0.225	0.421	0.029	0.424	0.293	0.403	0.424
CellTalkDB	0.393	0.447	0.423	1	0.581	0.565	0.114	0.089	0.124	0.802	0.018	0.904	0.543	0.615	0.795
ConnDB2020	0.501	0.398	0.443	0.581	1	0.495	0.137	0.112	0.161	0.631	0.022	0.607	0.446	0.739	0.635
EMBRACE	0.539	0.401	0.394	0.565	0.495	1	0.125	0.104	0.152	0.696	0.018	0.593	0.327	0.595	0.707
GuidePharm	0.172	0.132	0.202	0.114	0.137	0.125	1	0.137	0.415	0.137	0.081	0.117	0.068	0.166	0.137
HPMR	0.119	0.087	0.107	0.089	0.112	0.104	0.137	1	0.112	0.107	0.062	0.094	0.058	0.135	0.107
ICELLNET	0.2	0.153	0.225	0.124	0.161	0.152	0.415	0.112	1	0.148	0.093	0.13	0.081	0.183	0.147
iTALK	0.459	0.449	0.421	0.802	0.631	0.696	0.137	0.107	0.148	1	0.021	0.85	0.455	0.747	0.985
Kirouac2010	0.027	0.021	0.029	0.018	0.022	0.018	0.081	0.062	0.093	0.021	1	0.018	0.01	0.027	0.021
LRdb	0.414	0.456	0.424	0.904	0.607	0.593	0.117	0.094	0.13	0.85	0.018	1	0.522	0.647	0.839
OmniPath	0.265	0.384	0.293	0.543	0.446	0.327	0.068	0.058	0.081	0.455	0.01	0.522	1	0.353	0.456
Ramilowski	0.56	0.364	0.403	0.615	0.739	0.595	0.166	0.135	0.183	0.747	0.027	0.647	0.353	1	0.758
talklr	0.464	0.451	0.424	0.795	0.635	0.707	0.137	0.107	0.147	0.985	0.021	0.839	0.456	0.758	1

Jaccard Index
1.00
0.75
0.50
0.25

Figure 2. Dependencies and overlap between CCC resources. A) The lineages of CCC interaction database knowledge. General biological knowledge databases, CCC-dedicated resources used in this work, Literature curation, resources included in iTALK, and OmniPath are in blue, magenta, yellow, cyan, and green respectively. Arrows show the data transfers between resources. B) Shared and Unique Interactions, Receivers and Transmitters. C) Similarity between the interactions from different resources (Jaccard Index).

2.2 Resource Prior Knowledge Bias

Since CCC inference methods rely on prior knowledge to estimate intercellular communication events, the choice of resource and any potential bias in it is expected to impact the results. We therefore explored whether the coverage of interactions in the resources is biased toward specific subcellular locations or functional categories when compared to the collection of all resources.

2.2.1 Subcellular Localisation

We obtained protein subcellular localisation annotations from OmniPath [2], which combines this information from 20 resources. We then matched the localisations to receivers and transmitters from each resource with the aim to assess the

9

localisation profile of different resources. On average 90% of transmitters and 79% of receivers were annotated as secreted and transmembrane proteins, respectively (Supp. Figure 4). We further used the localisations of transmitters and receivers to categorize the interactions as secreted or direct-contact signaling. We reasoned that, interactions between transmitters annotated as secreted and receivers annotated as membrane-bound represent solute mediated (secreted) signalling events. On the contrary, an interaction between two membrane-bound proteins requires direct contact between cells. Building on this, we observed that all resources were predominantly (74% on average) composed of interactions associated with secreted signalling, while direct-contact signalling constituted a substantially smaller (16% on average) proportion of interactions (Figure 3A; Supp. Figure 5). Interactions categorized as neither secreted nor direct-contact were labeled as 'Other' and made up the remainder of the interactions (Supp. Note 1). The proportions of secreted and direct-contact signalling varied between resources, as some of them, such as Baccin, ConnectomeDB, CellPhoneDB, HPMR, and OmniPath had an over-representation of direct-contact signalling when compared to the collective, while the opposite was noted for the case of secreted signalling (Figure 3B). Direct contact interactions were particularly under-represented in Guide to Pharmacology (4%), which was more focused on secreted signalling (87%). CellChatDB showed an overrepresentation of interactions matched to the category Other.

Our results suggest that localisations of transmitters and receivers were largely uniformly distributed and that secreted signalling was predominant across all resources. Yet, differences were noted between the relative abundance of secreted and direct-contact signalling interactions.

2.2.2 Functional Term Enrichment

To examine whether specific pathways and biological functions are unevenly represented in specific resources, we matched the interactions, receivers and transmitters from each resource to well-known pathways and functional categories from SignaLink [35], NetPath [39], CancerSEA [40], and HGNC [41].

We observed that the Receptor tyrosine kinase (RTK), JAK/STAT, TGF, and WNT pathways covered the largest proportions of interactions matched to SignaLink, with analogous results observed for receivers and transmitters (Supp. Figure 6). The interactions from Ramilowski, ConnectomeDB, LRdb, iTALK and talklr showed a similar pattern, which can be explained by the high overlap of these resources. On the contrary, interactions associated with innate immune pathways and T-cell receptor categories were under-represented in Guide to Pharmacology, Baccin2019, EMBRACE, Kirouac2010, ICELLNET, CellPhoneDB, and HMPR (Figure 3C). The innate immune pathway category was also diminished in OmniPath. In contrast, when we used NetPath instead of SignaLink to define the T-cell receptor pathway, the under-representation in Baccin2019 and OmniPath was not observed, and an over-representation was instead noted for ICELLNET and CellPhoneDB (Figure 3D; Supp. Figure 7). Moreover, we observed a considerable over-representation for the RTK pathway in OmniPath. The Signalink WNT pathway was under-represented in ICELLNET and Guide to Pharmacology, while for the NetPath WNT pathway this was only true for Guide to Pharmacology. In contrast, CellChat showed a relative abundance for both the SignaLink and NetPath WNT pathways. These observations for the WNT pathway were further supported by the relative abundance of HGNC (Supp. Figure 8). Functional cancer cell states from CancerSEA were also unevenly represented in sets of receivers and transmitters across the resources (Supp. Figure 9). For example, Cell Cycle, DNA repair, and DNA damage states were over-represented in LRdb. Hence, our results indicated heterogenous biases towards certain pathways and categories across the different CCC resources.
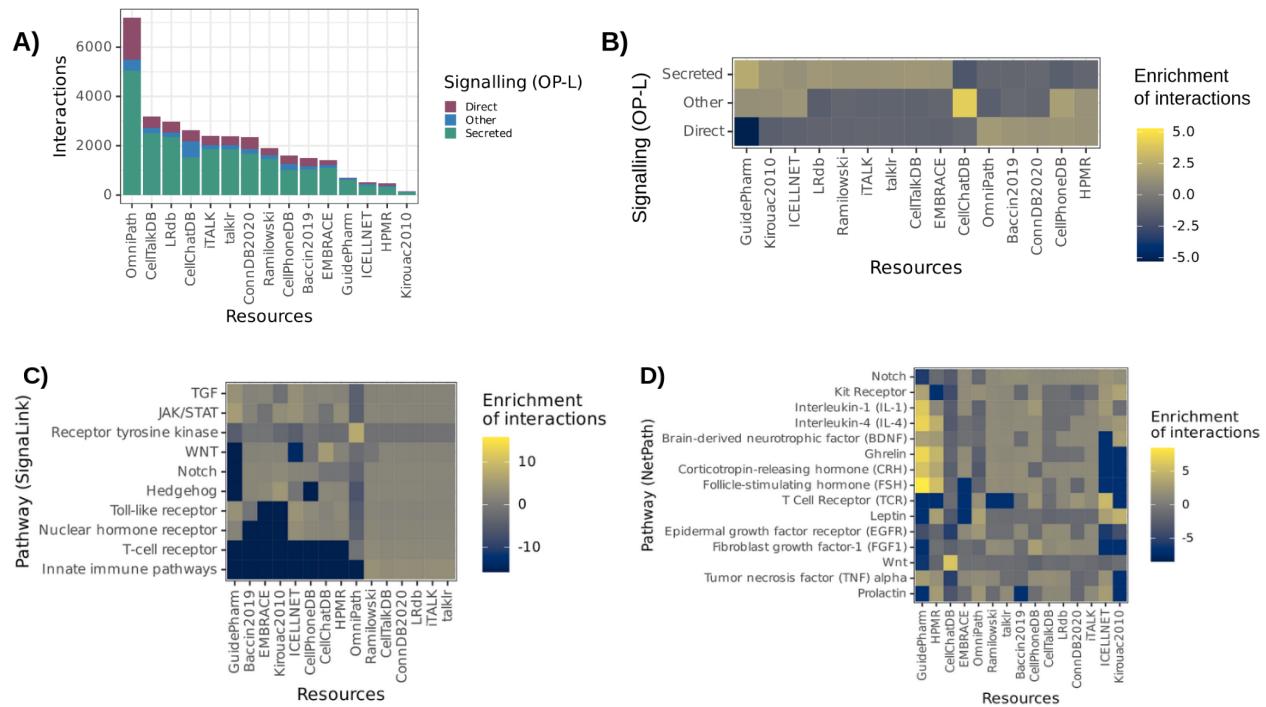
Figure 3. Localisation and Functional term abundance in CCC resources. A) *Number and* **B)** *Relative abundance of signalling categories based on OmniPath-derived protein locations (OP-L) by resource. Relative abundance of* **C)** *SignaLink and* **D)** *NetPath annotations matched to interactions from each resource.*

2.3 Agreement in CCC predictions in a Colorectal Cancer data set

To estimate the relative agreement between CCC methods and the importance of resources, we developed a framework to decouple tools from their inbuilt resources. We chose a well-annotated colorectal cancer (CRC) scRNA-Seq dataset [23] with 65,362 cells from a heterogeneous cohort of 23 Korean CRC patients. The 38 cell annotations in the dataset included stromal, immune, tumour and healthy epithelial cell types/states, as well as 3 unknown subtypes of Myeloid, B and T cells. We focused on the interactions between tumour cells subclassified by their resemblance of CRC consensus molecular subtypes (CMS) and immune cells from tumour samples (Supp. Table 3), reasoning that this subset of cell types represents a complex example where CCC events are known to have an important role. In addition to the 15 CCC resources reported in the descriptive resource analysis

(Supp. Table 2), we also included the default or inbuilt resource for each of the tools, except Squidpy (Table 1), as well as a reshuffled control resource (4.3 Framework).

2.3.1 Interaction overlap

We then used each method-resource combination to infer CCC interactions, assuming that different methods should generally agree on the most relevant CCC events for the same resource and expression data. To measure the agreement between method-resource combinations, we looked at the overlap between the 500 highest ranked interactions as predicted by each method. Whenever available, author recommendations were used to filter out the false-positive interactions (4.4 Method-Resource Specifics). Our analysis showed considerable differences in the interactions predicted by each of the methods regardless of the resource used, as the mean Jaccard index per resource ranged from 0.01 to 0.06 (mean = 0.024) when using different methods. These large discrepancies in the results were further supported by the pairwise comparisons between methods using the same resource, with mean Jaccard indices ranging from 0.063 (CellChat-SingleCellSignalR) to 0.110 (Connectome-NATMI). The overlap among the top predicted interactions was slightly higher when using the same method but with different resources, as Jaccard indices ranged from 0.113 to 0.203 per method (mean = 0.167) (Supp. Figure 11). Consequently, the highest ranked interactions for each method-resource combination largely showed stronger clustering by method than resource (Figure 4), with similar results observed when considering the highest ranked 100, 250, and 1000 interactions (Supp. Figure 10). In particular, method-resource combinations involving Squidpy, SingleCellSignalR, and Connectome clustered exclusively by method, suggesting that the overlap between these combinations occurs predominantly when using the same method regardless of the resource (Supp. Figure 11). The combinations involving NATMI also clustered by method, with the only exceptions being the Kirouac2010 [38] and ICELLNET [21] resources, which were the smallest resources (Supp. Table 2).

13

Moreover, CellChat and the CellPhoneDB (Squidpy) methods [7,24], account for heteromeric transmitter-receiver complexes, as such we examined the proportion of complex-containing interactions for these methods using the complex-containing resources (Supp. Note 2). This analysis showed that the proportion of complexes among the highest ranked hits was 2-23% for CellChat and 10-38% for Squidpy, largely reflecting the relative complex content in each resource.

Our results suggest that the overlap between methods when using the same resource was low (Supp. Figure 12). This was largely supported by the analysis of two additional scRNA-Seq data of cord blood mononucleated cells and pancreatic islet cells (Supp. Figure 13), even though we observed slightly higher agreement between methods in the former dataset. The overlap when using the same method with different resources, albeit higher than that between different methods, was also modest (Supp. Figure 14). Hence, our results indicate that both the method and the resource had a considerable impact on the predicted interactions.
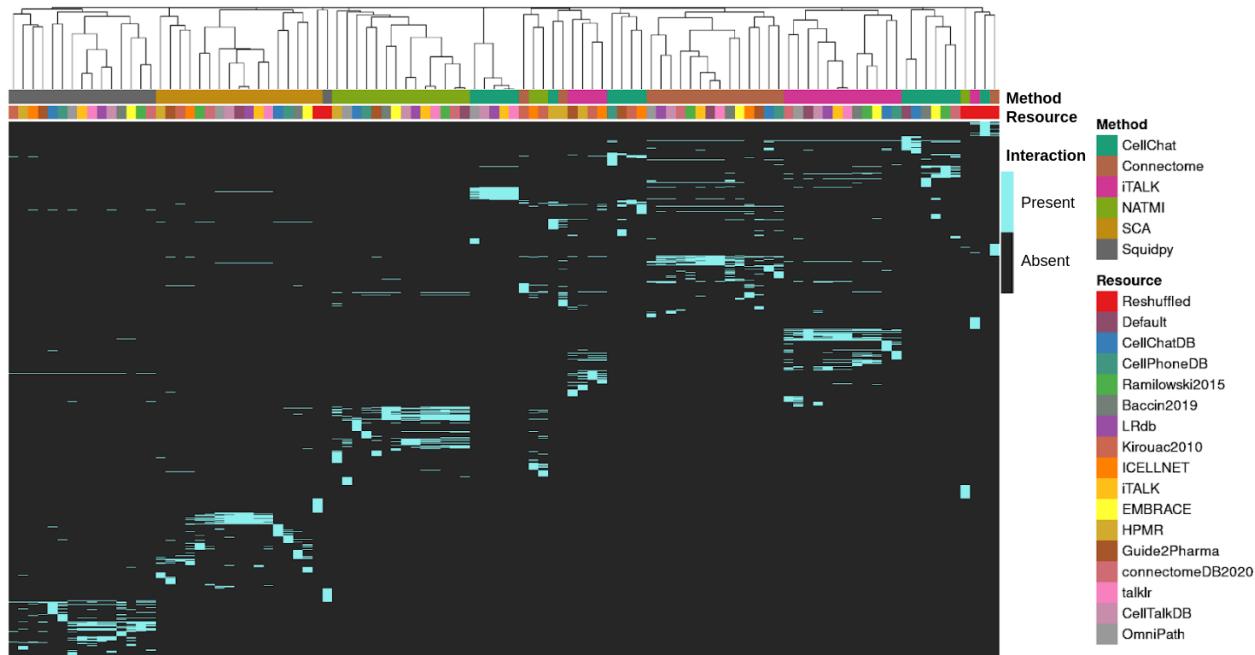
Figure 4. *Overlap in the 500 highest ranked CCC interactions between different combinations of methods and resources. Method-resource combinations were clustered according to binary (Jaccard index) distances. SCA refers to the SingleCellSignalR method.*

2.3.2 Communicating cell types

Next, we asked whether the discrepancies observed between the methods stem from the differences in the cell types inferred as most active in terms of CCC interactions. To this end, we used the 500 highest ranked interactions to examine the cell type activities, defined as the proportion of interactions per cell type, separately as a source and a target of CCC events (Figure 5). The results largely reiterated our observations from the CCC interaction overlap analysis above, as each method largely clustered by itself, regardless of the resource used, including the reshuffled resource. These results were further supported by the average interaction ranks per communicating pairs of clusters, as again the method-resource combinations largely grouped by method (Supp. Figure 15). We reasoned that the observed disagreement in regards to the most actively communicating cell types was likely caused by the methods' distinct approaches to handle cell cluster specificity. We thus performed a complementary analysis using the alternative, non cell-type specific scoring systems of the methods. The higher

agreement in this case suggested that these different approaches are indeed in part responsible for the disagreement (Supp. Note 3).

We further argued that the choice of method can have a major impact on the interpretation of CCC results. For example, regardless of the resource, SingleCellSignalR predicted CMS4-like cells to be a major source of signalling within the system, which was in disagreement with the majority of method-resource combinations (Figure *5*). Nevertheless, given that CMS4 is characterized by the exertion of immunosuppressive pressure on immune cells via stromal cells [42], it can be argued that SingleCellSignalR appropriately recognized CMS4 as the most active source of signalling. Regardless of the resource, NATMI highlighted CMS1-, CMS2-, and CMS3-like tumour cells as both major sources and targets of signalling, while the inferred activity of immune cells was overall sparse. NATMI's predictions were supported by CMS2 and CMS3 tumor subtypes being associated with having low immune and inflammatory molecular signatures [43]. Yet, this is not expected to be the case for CMS1-like cells, as CMS1 tumors are characterized by the infiltration of cytotoxic lymphocytes [42,43]. Moreover, CD4+ T cells and CD8+ T cells were the two most abundant immune cell types within the dataset (Supp. Table 3), but they were estimated to be among the least active cell types in the system across most method-resource combinations. Secreted phosphoprotein 1 (*SPP1*)$^+$ macrophages were observed to be important sources and targets of cell-cell communication in the system across most method-resource combinations. This observation is largely supported by the enrichment of SPP1+ macrophages in tumour tissues, and their potential key role in tumor progression and immune suppression [23].

The analysis of activities per cell type largely reiterated the results from the interaction overlap analysis, particularly as each method largely clustered by itself, regardless of the resource. As a consequence, the disagreement between the methods in which cell types are the most active is expected to have a major impact on the biological interpretation of CCC communication predictions.

Figure 5. *Activity per Cell type, inferred as the proportion of interaction edges that stem from Source Cell clusters or lead to Target Cell clusters in the highest ranked interactions.*

3. Discussion

The growing interest in CCC inference has led to the recent emergence of multiple methods and prior knowledge resources dedicated to studying intercellular crosstalk. To shed light on the impact of the choice of method and resource on the inference of CCC events, we built a framework to systematically combine 15 resources and 6 methods. We used this framework to describe in detail the content of the different resources and to estimate cell-cell communication from scRNA-Seq in a colon cancer case study. Our results suggest that both the method and resource can considerably impact CCC inference.

3.1. Resource Overlap and Bias

Despite their common origins, different resources cover varying proportions of the collective prior knowledge. Particularly, a large share of the observed overlap among resources stemmed from the inclusion of Ramilowski [32] into other

17

resources. Moreover, across the resources, the WNT, RTK, T-cell receptor and Innate immune pathways, among others, were present in varying proportions. The high abundance of interactions associated with the RTK pathway in OmniPath could be due to the ~1,600 expert curated RTK ligand-receptor interactions from SIGNOR [33] and the large size of RTK pathway in SignaLink [35]. The results presented here highlight an inherent limitation of knowledge-based inference, and hence of CCC methods, as any prior knowledge resource has its own biases and only represents a limited proportion of biological actuality. Consequently, these inherent limitations should be kept in mind for the interpretation of CCC predictions.

3.2. Impact of Methods and Resources

As a further step, we carried out a systematic analysis of the impact of resources and methods on CCC inference results using a public colorectal cancer dataset [23]. Although possibly over-simplistic, our binary overlap assessment enabled the direct comparison of the diverse scoring systems of the methods. We found that both resources and methods had a considerable effect on the predicted interactions, and the impact of methods outweighed that of the resource.

A potential explanation for the disagreement among the methods could be the distinct approaches they use to identify the most relevant interactions (Table 1). A common assumption among the methods is that cluster-specific interactions are more informative than those related to multiple clusters [7,9,10,13]. An experimental proof of this assumption and an evaluation of the distinct approaches is yet to be carried out. By focusing on the cluster-specific interactions in the dataset, these methods report the most specifically-interacting cell types [11], rather than the most actively communicating ones. Hence, the predicted CCC events typically do not capture processes that are common between multiple cell types. As an example, CD4+ T cells and CD8+ T cells, the two most abundant immune cell types found within tumours, were assigned a low CCC communication activity. In terms of the agreement in CCC event predictions we found that only a few biological patterns are robust across many methods, namely the SPP1+ macrophages have been

predicted to be main players of CCC signalling, supported by their enrichment in CRC tumour tissues [23] and frequent association with pro-metastatic role [44,45]. Nevertheless, we acknowledge that the pooled analysis of the CRC dataset largely limits the interpretation of results and warrants the future analysis of the same dataset on a per-patient/phenotype basis. Collectively, these results suggest that the common practice to highlight the most actively communicating cell clusters based on the CCC inference [23,46] should be considered with caution.

3.3. Previous Comparisons and limitations of our integration

Interestingly, our results did not recover some of the previously reported agreement between tools [9–11,13]. Contrary to previous comparisons[11], we saw little overlap between the SingleCellSignalR, CellPhoneDB and iTALK methods. Furthermore, despite their relatively similar approaches, the limited agreement between CellChat and CellPhoneDB was not observed here [13]. It is to note that we used CellChat's probabilities instead of p-values to obtain the highest ranked interactions. These probabilities do not deliberately reflect cell cluster specificity in regards to the inferred interactions[13]. However, our results support the previously observed low agreement between CellChat and both iTALK and SingleCellSignalR[13].

Some methods, namely CellChat [13] and the CellPhoneDB algorithm [7], as well as resources, such as Baccin, CellChatDB, CellPhoneDB, and ICELLNET, take protein complexes into account. This largely complicates the conversion of the resources and hence the comparison with methods and resources which do not consider complexes. Furthermore, CellChat, and hence CellChatDB, goes a step further than other methods and resources, as it considers interaction mediator molecules, which are absent in the remainder of the resources [13]. Thus, even though any resource can be used with any method, we acknowledge that some combinations put certain methods at a disadvantage.

3.4. CCC Inference Assumptions and Benchmarking

Our results further point to certain limitations of the CCC inference methods. In particular, CCC events are mainly predicted based on the average gene expression at the cluster or cell type/state level. Such an assumption inherently suggests that gene expression is informative of the activity of transmitters and receivers. However, gene expression provided by scRNA-Seq is typically limited to protein coding genes and the cells within the dataset, and hence does not capture secreted signalling events driven by non-protein molecules or long-distance endocrine signalling events. Further, CCC inference from scRNA-Seq data assumes that the product of the gene expression of a transmitter and a receiver is a good proxy for their joint activity, and thus does not consider any of the processes preceding transmitter-receiver interactions, including protein translation and processing, secretion, and diffusion.

We therefore believe that it is essential to establish a benchmark to comprehensively assess the predictive power of CCC methods. However, a gold standard for benchmarking is currently not available and the biological ground truth is largely unknown [3,19]. The field needs to identify experimental settings capable of establishing the biological ground truth. So far, intercellular interactions were mainly supported by the spatial colocalization of proteins and the functional deregulation of intracellular signalling [47], and the physical-interaction of cell types [48]. Yet these approaches are only applicable for the post-hoc and indirect validation of CCC interactions. Thus, until an experimental gold standard becomes available, simulated datasets might be used instead. However, any in silico benchmark is by definition only a simplified approximation of reality, with its own biases [49]. To our knowledge, appropriate benchmarks for resources and methods used in CCC inference are yet to be defined, although some proposals exist [19], that we elaborate on in Supp. Note 4.

3.5. Conclusion

Considerable efforts have been made to develop CCC inference, and we believe that further advancements will be key for the systems-level analysis of single-cell data. This will likely further increase by the rapidly emerging spatial transcriptomics [1] and single-cell proteomics [50], and the future applications of CCC inference approaches to interspecies communication [51,52]. Acknowledging the limitations of our work, we believe that it points at the interpretation inconsistencies that could arise as a consequence of the method and resource of choice. We thus regard the results and comparative framework presented here as steps towards an understanding of the strengths and weaknesses of CCC methods, and thereby towards their improvement. Future developments of this work will include extending the number of datasets in the comparative analysis as well as the benchmark of methods and resources. As such, we here extend an open invitation to all interested parties willing to join us in this endeavour.

4. Methods

4.1. Descriptive analysis of resources

The connections between resources shown in the dependency plot were manually gathered from the publications and the web pages of each CCC resource.

The CCC resources used in the analyses were queried from the OmniPath database [2]. The contents of the resources are identical to their original formats, apart from minor processing differences (Supp. Table 2).

The OmniPath CCC resource is a composite resource which contains interactions from all of the CCC dedicated resources compared here, along with some additional resources[2]. OmniPath's interactions were filtered according to the following criteria: i) we only retained interactions with literature references, ii) we

21

kept interactions only where the receiver protein was plasma membrane transmembrane or peripheral according to at least 30% of the localisation annotations, and iii) we only considered interactions between single proteins (interactions between complexes are also available in OmniPath). OmniPath's intra- and intercellular components are both available via the OmnipathR package (https://github.com/saezlab/OmnipathR).

We defined unique and shared interactions, receivers and transmitters between the CCC resources if they could be found in only one or at least two of the resources, respectively. We used pheatmap [53] and UpSetR [54] to generate the heatmaps and upset plots, respectively.

To identify uneven distributions of transmitters, receivers, and interactions toward biological terms or protein localisations, we used Fisher's exact test to compare each individual resource to the collection of all the resources. We obtained protein localisations from OmniPath which collects this information from 20 databases[2]. Then we kept consensus protein localisations where at least 50% of the annotations agreed. We classified CCC interactions using the localisation combinations of proteins involved in the interactions, which included secreted, plasma membrane peripheral and transmembrane proteins. Interactions, receivers and transmitters were independently matched to the 10 pathways from SignaLink [35], and the 15 largest categories from CancerSEA [40], HGNC [41], and NetPath [39]. Each of the aforementioned general functional annotation databases was also obtained via OmniPath. In case of signalling pathway databases (SignaLink and NetPath), we focused on the enrichment of annotations matched to interactions, while for the functional state databases (CancerSEA and HGNC), we presented the merged sets of transmitters and receivers matched to functional categories. Annotation matches for transmitters and receivers were examined independently using the aforementioned functional annotation databases, but they were not the focus of discussion presented here. Our approach allowed the same protein or interaction to be matched to multiple pathways or functional categories from the same database.

To enable a comparison of annotations across resources, we expanded protein complexes from Baccin2019 [22], CellChatDB [13], CellPhoneDB [7], and ICELLNET [10].

4.2. Single-cell Transcriptomics data

The processed single cell RNA-Seq data [23] for 23 Korean colorectal cancer patients is available at GSE132465. The analysis presented here focused on the CCC interactions between colorectal cancer subtypes and immune cells, and the remainder of cell types, including unknown immune cell subtypes, were filtered out. This resulted in a subset of 18 cell types and 42,544 cells. We kept the original subtype labels, reformatted the names to work with each CCC method (Supp. Table 3), and sparsified the counts into a Seurat[55] object.

The labelled scRNA-Seq data for pancreatic islet [56] and cord blood mononuclear cells [57] were obtained via SeuratData, normalized with Seurat [55], and used for CCC inference without any further formatting and filtering.

4.3. Framework

For the method-resource comparison, we used Seurat [55,56] objects which were converted into the appropriate data format when calling each method. We used the recommended conversion method or wrapper whenever available.

The resources were obtained from OmniPath and then converted to the appropriate format for each method. A reshuffled version of ConnectomeDB2020 was generated with BiRewire [58] and referred to as the reshuffled control resource. Each tool was run with its default or inbuilt resource, except Squidpy. The Default resource of Squidpy's **ligrec** function is OmniPath, which is already part of our benchmark set. The framework enabling the use of any resource and method combination, as well as the results, are available at https://github.com/saezlab/ligrec_decoupler.

4.4. Overlap Analysis

To compare the overlap between the interactions predicted by each method-resource combination, as a default we kept the 500 highest ranked interactions. We also considered the highest ranked 100, 250, and 1000 interactions for the CRC scRNA-Seq dataset. In case of ties, we considered the higher number of predicted interactions. We then generated a presence-absence matrix of predicted interactions with method-resource combinations. These matrices were subsequently used to calculate Jaccard indices and to cluster the results.

Activity per cell type was calculated using the highest ranked 500 hits for each method-resource combination. Cell type activity represents the proportion of interactions (or edges) that stem from or lead to a Source or Target cell type, respectively. In other words, a Source cell with a high cell type activity, in the broadest terms, can be inferred as an active 'secretor of ligands'. We used the z-normalized average interaction rank for each possible combination of communicating cell types to estimate the cell pair ranks for each method-resource combinations. These patterns of pairwise communication activities we presented in a PCA plot. We created the heatmaps with pheatmap [53] (v1.0.12), using binary distances for the overlap heatmaps and euclidean distances for the other heatmaps.

Connectome, NATMI, and iTALK do not provide an explicit threshold to control for false positives and the highest ranking 500 hits were kept for each without any preceding filtering. For methods where a threshold was proposed by the authors, as in the case of CellChat, Squidpy, and SingleCellSignalR, we first filtered their results accordingly and the highest ranked interactions were obtained afterwards. Further, we used cluster-specific interaction measures for each method whenever available (Table 1).

The same analysis was also carried out using the cluster-unspecific measures from each method. The scaling done in Connectome (weight_scale) and NATMI (Specificity-based edge weight), and in particular the cluster label permutation of

CellChat (p-values) and CellPhoneDB (p-values), explicitly reflect cell-cluster specific communication, thus we used their alternative measures. SingleCellSignalR and iTALK provide a single measure each and were hence excluded from this analysis.

4.5. Method Specifics

4.5.1 CellChat

CellChat was run using default settings with 1000 permutations and the gene expression diffusion-based smoothing process was omitted. CellChat returned a number of significant interactions with p-values of 0 ranging from 221 to 12,208 depending on the resource (2,988 with its inbuilt resource), these made a considerable proportion of the significant hits (p-value <= 0.05), as they ranged from 237 to 12,971 (3,041 with its inbuilt resource). As such, because obtaining the highest ranked interactions based on p-values was infeasible, CellChat results were filtered according to p-values (p-value <= 0.05) and the highest probability scores were instead used in the method-resource analysis.

4.5.2 Connectome

Connectome was run with its default settings using a Seurat object with processed gene expression counts. Results were filtered for differentially expressed genes (p-value <= 0.05), as identified via a Wilcoxon test, and Connectome's scaled weights were used in the method-resource analysis.

4.5.3 iTALK

iTALK was run with its default settings using the 'DEG' option which returns corrected p-values and logFold changes for each gene. Then transmitters and receivers with q-value <= 0.05 were kept. A differential expression product was calculated using z-scores of transmitters and receivers and subsequently used in the method-resource analysis.

4.5.4 SingleCellSignalR

SingleCellSignalR was run with the processed gene counts, considering differentially expressed genes with a log2 fold change threshold of 1.5 or above. The highest LRscores which passed the recommended threshold of 0.5 were used in the method-resource comparison. The number of interactions predicted by SingleCellSignalR ranged between 159 to 7,240 (LRscore >= 0.5). The source code of SingleCellSignalR was modified to work with external resources (available at https://github.com/CostaLab/SingleCellSignalR_v1).

4.5.5 NATMI

NATMI's implementation is command-line based, thus a system command is invoked via R that calls the NATMI python module and passes the appropriate command line arguments. NATMI was run with its default settings using the processed gene expression matrix, converted from Seurat, and the specificity-based edge weights were used in the method-resource comparison. NATMI's lrc2p resource was used as the default.

4.5.6 Squidpy

Squidpy is called via reticulate [59] (https://rstudio.github.io/reticulate/) and the Seurat object is converted to anndata [60] (https://anndata.readthedocs.io/) format in Python. The CellPhoneDB algorithm implementation was run via the Squidpy framework with 10,000 permutations, threshold of cells expressing transmitters and receivers of 0.1, and the minimum component expression was considered for complexes. For the method-resource comparison, we used the rank of p-values (p-value <= 0.05). Squidpy's number of significant hits ranged between 60 to 2,927 depending on the resource.

5. Acknowledgements

6. Conflict of interests

JSR has received funding from GSK and Sanofi and consultant fees from Travere Therapeutics. AV is currently employed by F. Hoffmann-La Roche Ltd. The authors declare that they have no other competing interests.

7. Authors contributions

JSR conceived the project. DD set up the framework used in the comparison of method-resource combinations, with the help of JSN and DT. JSN was supervised by IC. DT and CB created the resource analysis plots and pipeline, later extended by DD. JSR supervised the project with the help of AV and AD. DD wrote the manuscript together with AV, CB, DT, and JSR. HK, RORF, and BS performed preliminary and supplementary analyses that helped shape the work presented here and the future directions of this project. All authors contributed to the ideas presented in this work and all revised the final version of the manuscript.

8. References

1. Chen, X., Teichmann, S.A., and Meyer, K.B. (2018). From Tissues to Cell Types and Back: Single-Cell Gene Expression Analysis of Tissue Architecture. Annu. Rev. Biomed. Data Sci. *1*, 29–51.

2. Türei, D., Valdeolivas, A., Gul, L., Palacio-Escat, N., Klein, M., Ivanova, O., Ölbei, M., Gábor, A., Theis, F., Módos, D., et al. (2021). Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. Mol. Syst. Biol. *17*, e9923.

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

4. Cillo, A.R., Kürten, C.H.L., Tabib, T., Qi, Z., Onkar, S., Wang, T., Liu, A., Duvvuri, U., Kim, S., Soose, R.J., et al. (2020). Immune Landscape of Viral- and Carcinogen-Driven Head and Neck Cancer. Immunity *52*, 183-199.e9.

5. Wang, Y., Wang, R., Zhang, S., Song, S., Jiang, C., Han, G., Wang, M., Ajani, J., Futreal, A., and Wang, L. (2019). iTALK: an R Package to Characterize and Illustrate Intercellular Communication. BioRxiv.

6. Tyler, S.R., Rotti, P.G., Sun, X., Yi, Y., Xie, W., Winter, M.C., Flamme-Wiese, M.J., Tucker, B.A., Mullins, R.F., Norris, A.W., et al. (2019). PyMINEr Finds Gene and Autocrine-Paracrine Networks from Human Islet scRNA-Seq. Cell Rep. *26*, 1951-1964.e8.

7. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat. Protoc. *15*, 1484–1506.

8. Tsuyuzaki, K., Ishii, M., and Nikaido, I. (2019). Uncovering hypergraphs of cell-cell interaction from single cell RNA-sequencing data. BioRxiv.

9. Raredon, M.S.B., Yang, J., Garritano, J., Wang, M., Kushnir, D., Schupp, J.C., Adams, T.S., Greaney, A.M., Leiby, K.L., Kaminski, N., et al. (2021). *Connectome* : computation and visualization of cell-cell signaling topologies in single-cell systems data. BioRxiv.

10. Hou, R., Denisenko, E., Ong, H.T., Ramilowski, J.A., and Forrest, A.R.R. (2020). Predicting cell-to-cell communication networks using NATMI. Nat. Commun. *11*, 5011.

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

12. Noël, F., Massenet-Regad, L., Carmi-Levy, I., Cappuccio, A., Grandclaudon, M.,

Trichot, C., Kieffer, Y., Mechta-Grigoriou, F., and Soumelis, V. (2020). ICELLNET: a transcriptome-based framework to dissect intercellular communication. BioRxiv.

13. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. *12*, 1088.

14. Choi, H., Sheng, J., Gao, D., Li, F., Durrans, A., Ryu, S., Lee, S.B., Narula, N., Rafii, S., Elemento, O., et al. (2015). Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. Cell Rep. *10*, 1187–1201.

15. Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat. Methods *17*, 159–162.

16. Wang, S., Karikomi, M., MacLean, A.L., and Nie, Q. (2019). Cell lineage and communication network inference via optimization for single-cell transcriptomics. Nucleic Acids Res. *47*, e66.

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

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

19. Almet, A.A., Cang, Z., Jin, S., and Nie, Q. (2021). The landscape of cell-cell communication through single-cell transcriptomics. Current Opinion in Systems Biology.

20. Shao, X., Liao, J., Li, C., Lu, X., Cheng, J., and Fan, X. (2020). CellTalkDB: a manually curated database of ligand-receptor interactions in humans and mice. Brief. Bioinformatics.

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

22. Baccin, C., Al-Sabah, J., Velten, L., Helbling, P.M., Grünschläger, F., Hernández-Malmierca, P., Nombela-Arrieta, C., Steinmetz, L.M., Trumpp, A., and Haas, S. (2020). Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. Nat. Cell Biol. *22*, 38–48.

23. Lee, H.-O., Hong, Y., Etlioglu, H.E., Cho, Y.B., Pomella, V., Van den Bosch, B., Vanhecke, J., Verbandt, S., Hong, H., Min, J.-W., et al. (2020). Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer.

Nat. Genet. *52*, 594–603.

24. Palla, G., Spitzer, H., Klein, M., Fischer, D.S., Schaar, A.C., Kuemmerle, L.B., Rybakov, S., Ibarra, I.L., Holmberg, O., Virshup, I., et al. (2021). Squidpy: a scalable framework for spatial single cell analysis. BioRxiv.

25. Garcia-Alonso, L., Handfield, L.-F., Roberts, K., Nikolakopoulou, K., Fernando, R.C., Gardner, L., Woodhams, B., Arutyunyan, A., Polanski, K., Hoo, R., et al. (2021). Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. BioRxiv.

26. Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.-E., Stephenson, E., Polański, K., Goncalves, A., et al. (2018). Single-cell reconstruction of the early maternal-fetal interface in humans. Nature *563*, 347–353.

27. Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

28. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. *49*, D545–D551.

29. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. Nucleic Acids Res. *46*, D649–D655.

30. Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45*, D362–D368.

31. Harding, S.D., Sharman, J.L., Faccenda, E., Southan, C., Pawson, A.J., Ireland, S., Gray, A.J.G., Bruce, L., Alexander, S.P.H., Anderton, S., et al. (2018). The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY. Nucleic Acids Res. *46*, D1091–D1106.

32. Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V.P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., et al. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. Nat. Commun. *6*, 7866.

33. Licata, L., Lo Surdo, P., Iannuccelli, M., Palma, A., Micarelli, E., Perfetto, L., Peluso, D., Calderone, A., Castagnoli, L., and Cesareni, G. (2020). SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update. Nucleic Acids Res. *48*, D504–D510.

34. Winograd-Katz, S.E., Fässler, R., Geiger, B., and Legate, K.R. (2014). The integrin adhesome: from genes and proteins to human disease. Nat. Rev. Mol. Cell Biol. *15*, 273–288.

35. Fazekas, D., Koltai, M., Türei, D., Módos, D., Pálfy, M., Dúl, Z., Zsákai, L., Szalay-Bekő, M., Lenti, K., Farkas, I.J., et al. (2013). SignaLink 2 - a signaling pathway resource with multi-layered regulatory networks. BMC Syst. Biol. *7*, 7.

36. Wang, Y. (2020). talklr uncovers ligand-receptor mediated intercellular crosstalk. BioRxiv.

37. Ben-Shlomo, I., Yu Hsu, S., Rauch, R., Kowalski, H.W., and Hsueh, A.J.W. (2003). Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. Sci. STKE *2003*, RE9.

38. Kirouac, D.C., Ito, C., Csaszar, E., Roch, A., Yu, M., Sykes, E.A., Bader, G.D., and Zandstra, P.W. (2010). Dynamic interaction networks in a hierarchically organized tissue. Mol. Syst. Biol. *6*, 417.

39. Kandasamy, K., Mohan, S.S., Raju, R., Keerthikumar, S., Kumar, G.S.S., Venugopal, A.K., Telikicherla, D., Navarro, J.D., Mathivanan, S., Pecquet, C., et al. (2010). NetPath: a public resource of curated signal transduction pathways. Genome Biol. *11*, R3.

40. Yuan, H., Yan, M., Zhang, G., Liu, W., Deng, C., Liao, G., Xu, L., Luo, T., Yan, H., Long, Z., et al. (2018). CancerSEA: a cancer single-cell state atlas. Nucleic Acids Res. *47*,.

41. Povey, S., Lovering, R., Bruford, E., Wright, M., Lush, M., and Wain, H. (2001). The HUGO gene nomenclature committee (HGNC). Hum. Genet. *109*, 678–680.

42. Dienstmann, R., Vermeulen, L., Guinney, J., Kopetz, S., Tejpar, S., and Tabernero, J. (2017). Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. Nat. Rev. Cancer *17*, 79–92.

43. Becht, E., de Reyniès, A., Giraldo, N.A., Pilati, C., Buttard, B., Lacroix, L., Selves, J., Sautès-Fridman, C., Laurent-Puig, P., and Fridman, W.H. (2016). Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy. Clin. Cancer Res. *22*, 4057–4066.

44. Cheng, S., Li, Z., Gao, R., Xing, B., Gao, Y., Yang, Y., Qin, S., Zhang, L., Ouyang, H., Du, P., et al. (2021). A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell *184*, 792-809.e23.

45. Zhang, L., Li, Z., Skrzypczynska, K.M., Fang, Q., Zhang, W., O'Brien, S.A., He, Y., Wang, L., Zhang, Q., Kim, A., et al. (2020). Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. Cell *181*, 442-459.e29.

46. Zheng, Y., Lu, P., Deng, Y., Wen, L., Wang, Y., Ma, X., Wang, Z., Wu, L., Hong, Q., Duan, S., et al. (2020). Single-Cell Transcriptomics Reveal Immune Mechanisms of the Onset and Progression of IgA Nephropathy. Cell Rep. *33*, 108525.

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

48. Deczkowska, A., David, E., Ramadori, P., Pfister, D., Safran, M., At the, B., Giladi, A., Jaitin, D.A., Barboy, O., Cohen, M., et al. (2021). XCR1+ type 1 conventional dendritic cells drive liver pathology in non-alcoholic steatohepatitis. Nat. Med.

49. Weber, L.M., Saelens, W., Cannoodt, R., Soneson, C., Hapfelmeier, A., Gardner, P.P., Boulesteix, A.-L., Saeys, Y., and Robinson, M.D. (2019). Essential guidelines for computational method benchmarking. Genome Biol. *20*, 125.

50. Mahdessian, D., Cesnik, A.J., Gnann, C., Danielsson, F., Stenström, L., Arif, M., Zhang, C., Le, T., Johansson, F., Shutten, R., et al. Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. Nature *590*, 649–654.

51. Gul, L., Modos, D., Fonseca, S., Madgwick, M., Thomas, J.P., Sudhakar, P., Stentz, R., Carding, S.R., and Korcsmaros, T. (2021). Extracellular vesicles produced by the human commensal gut bacterium Bacteroides thetaiotaomicron affect host immune pathways in a cell-type specific manner that are altered in inflammatory bowel disease. BioRxiv.

52. Westermann, A.J., and Vogel, J. (2021). Cross-species RNA-seq for deciphering host-microbe interactions. Nat. Rev. Genet.

53. Kolde, R. (2019). pheatmap: Pretty Heatmaps.

54. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938–2940.

55. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

56. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

57. Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods *14*, 865–868.

58. Iorio, F., Bernardo-Faura, M., Gobbi, A., Cokelaer, T., Jurman, G., and Saez-Rodriguez, J. (2016). Efficient randomization of biological networks while preserving functional characterization of individual nodes. BMC Bioinformatics *17*,

542.

59. Ushey, K., Allaire, J.J., and Tang, Y. (2020). Interface to "Python" [R package reticulate version 1.20].

60. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.

61. Sheikh, B.N., Bondareva, O., Guhathakurta, S., Tsang, T.H., Sikora, K., Aizarani, N., Sagar, Holz, H., Grün, D., Hein, L., et al. (2019). Systematic Identification of Cell-Cell Communication Networks in the Developing Brain. IScience *21*, 273–287.

62. Bausch-Fluck, D., Hofmann, A., Bock, T., Frei, A.P., Cerciello, F., Jacobs, A., Moest, H., Omasits, U., Gundry, R.L., Yoon, C., et al. (2015). A mass spectrometric-derived cell surface protein atlas. PLoS ONE *10*, e0121314.

63. Lomize, A.L., Lomize, M.A., Krolicki, S.R., and Pogozheva, I.D. (2017). Membranome: a database for proteome-wide analysis of single-pass membrane proteins. Nucleic Acids Res. *45*, D250–D255.

64. Wykosky, J., Palma, E., Gibo, D.M., Ringler, S., Turner, C.P., and Debinski, W. (2008). Soluble monomeric EphrinA1 is released from tumor cells and is a functional ligand for the EphA2 receptor. Oncogene *27*, 7260–7273.

65. Armingol, E., Joshi, C.J., Baghdassarian, H., Shamie, I., Ghaddar, A., Chan, J., Her, H.-L., O'Rourke, E.J., and Lewis, N.E. (2020). Inferring the spatial code of cell-cell interactions and communication across a whole animal body. BioRxiv.

66. Dries, R., Zhu, Q., Dong, R., Eng, C.-H.L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., Bao, F., et al. (2021). Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome Biol. *22*, 78.

67. Tanevski, J., Ramirez Flores, R.O., Gabor, A., Schapiro, D., and Saez-Rodriguez, J. (2020). Explainable multi-view framework for dissecting inter-cellular signaling from highly multiplexed spatial data. BioRxiv.

Supplementary Materials

Supplementary Figure S1. *Jaccard Indices of A) Receivers and B) Transmitters from different resources.*

## A) Receivers
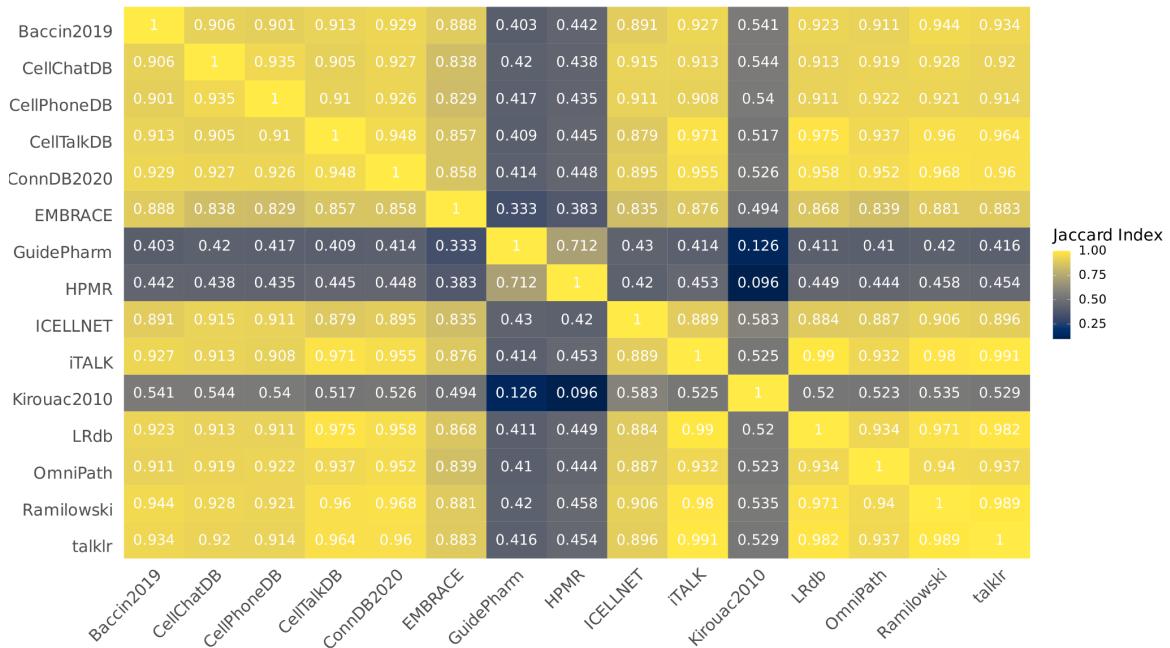
B) Transmitters
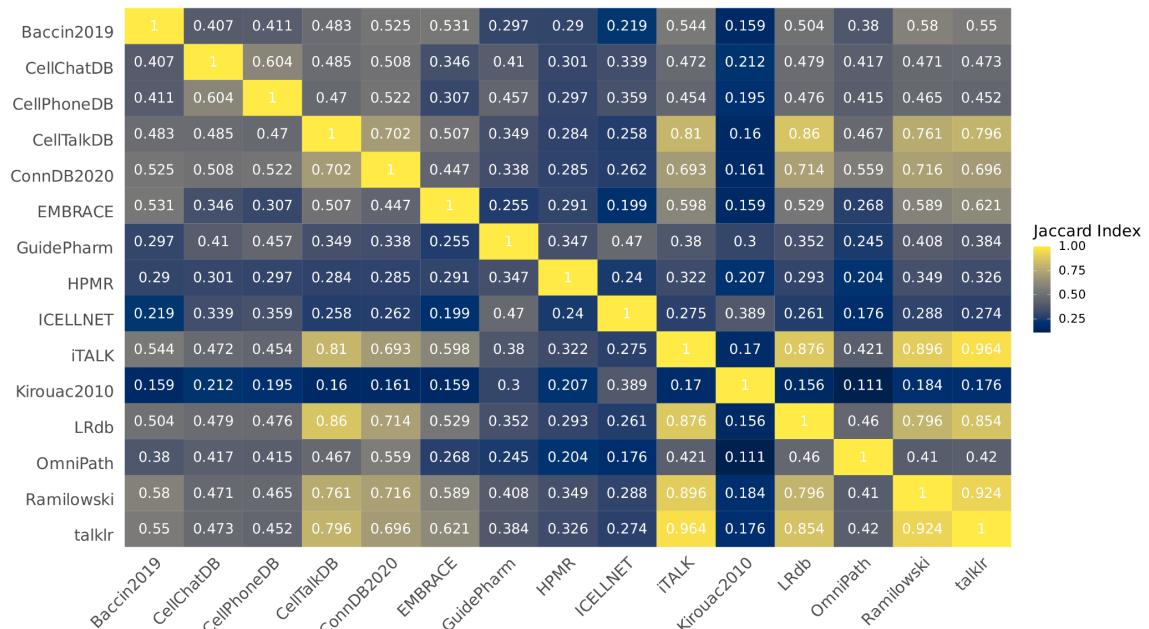
Supplementary Figure S2. *Upset plots representing the shared Interactions, Receivers, and Transmitters between all resources (A-C) and all resources except OmniPath (D-F).*

A) Interactions
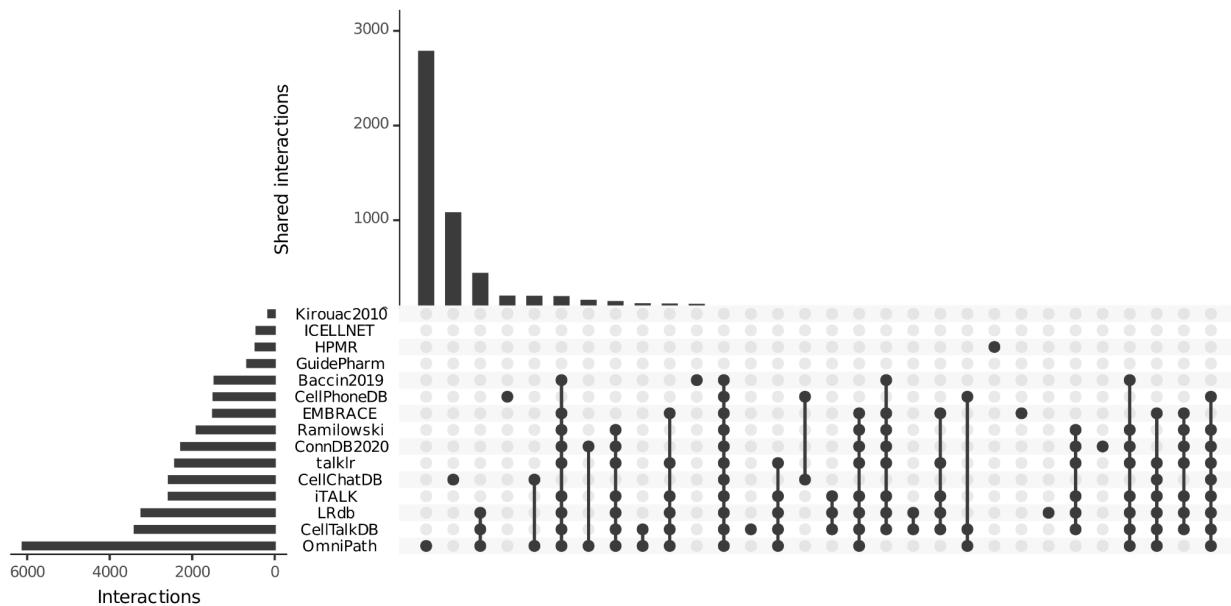
B) Receivers
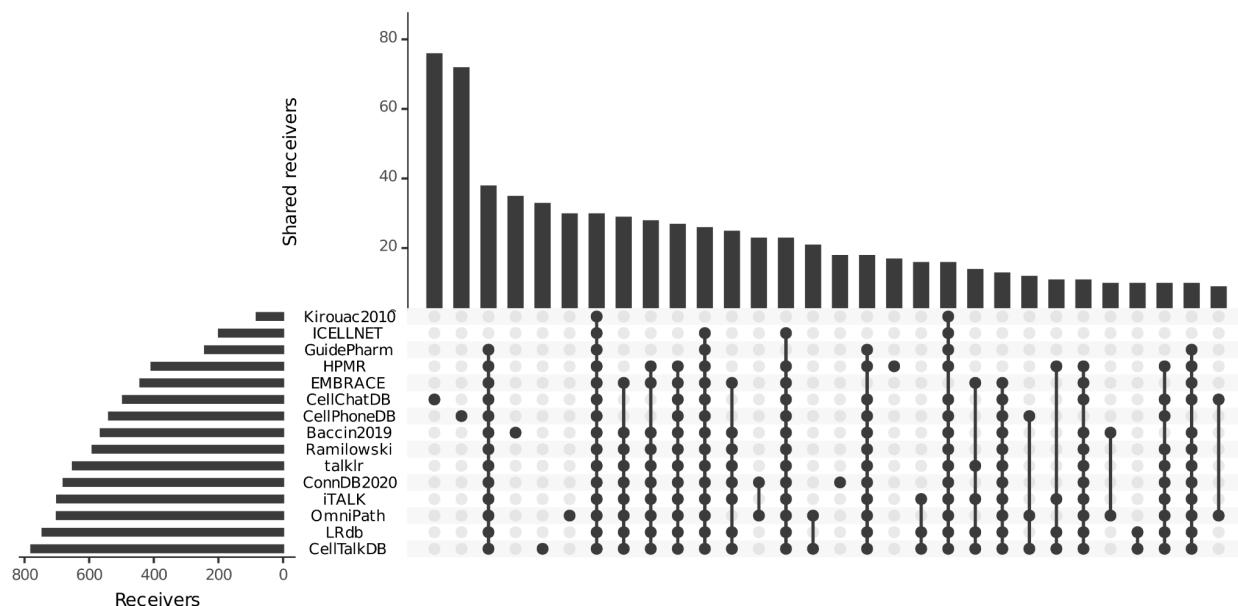
C) Transmitters
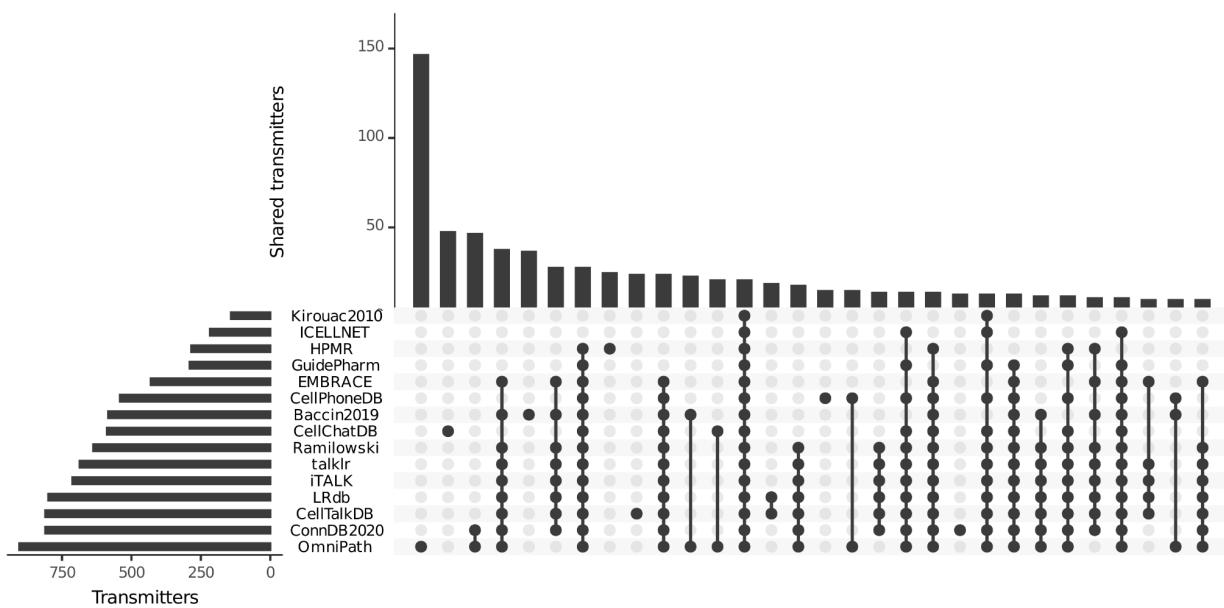
D) Interactions (except OmniPath)
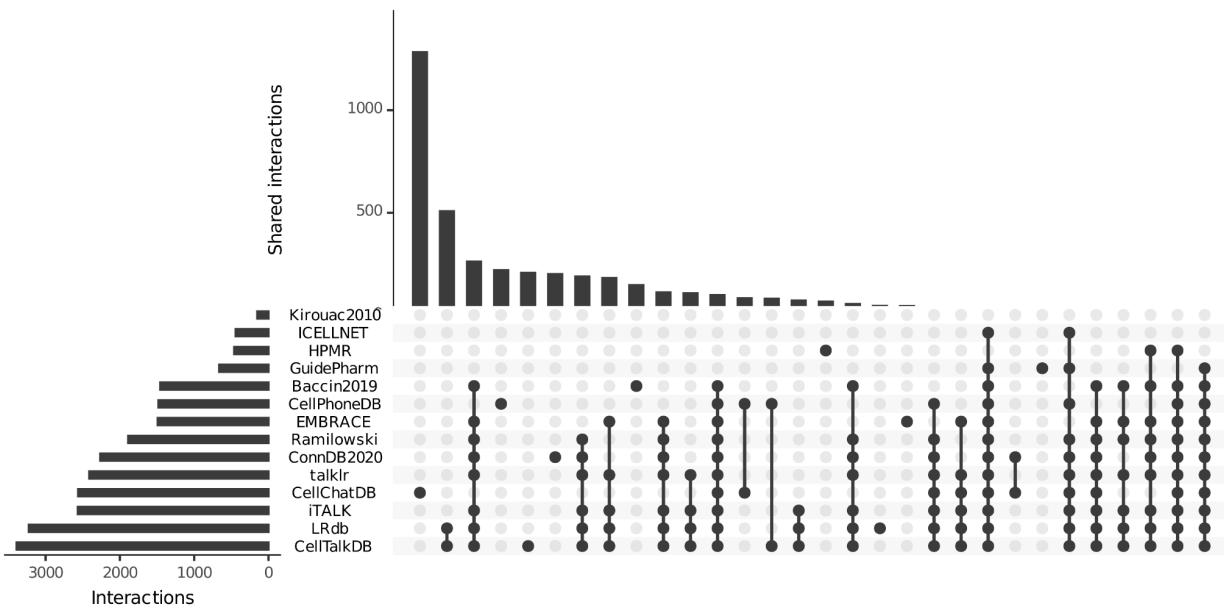
E) Receivers (except OmniPath)
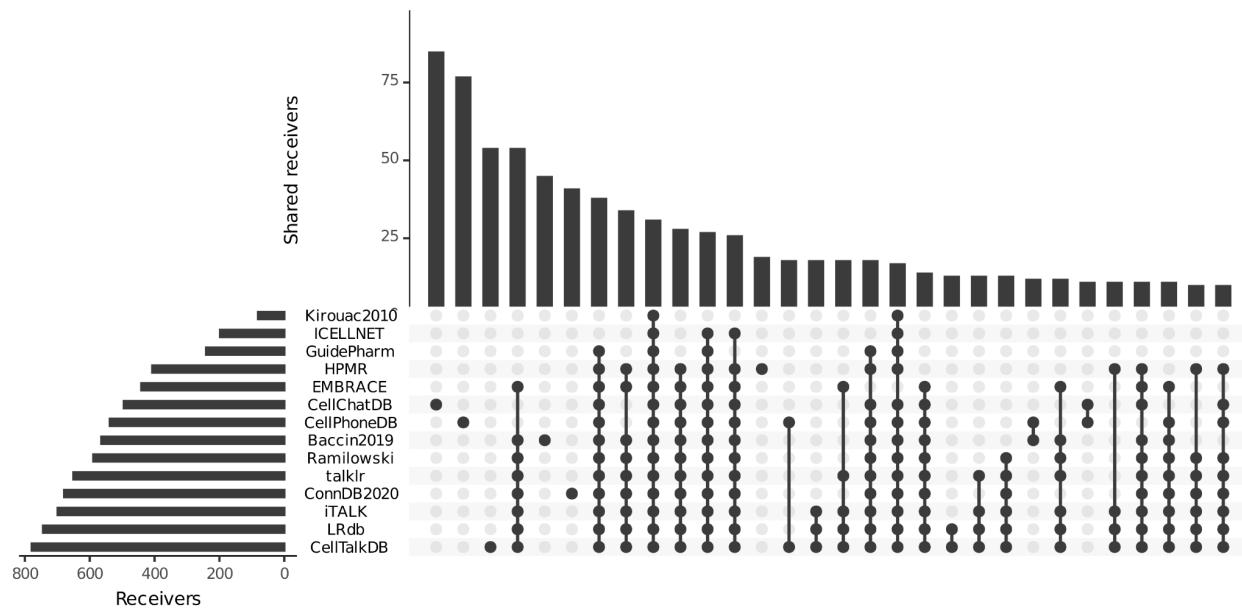

F) Transmitters (except OmniPath)
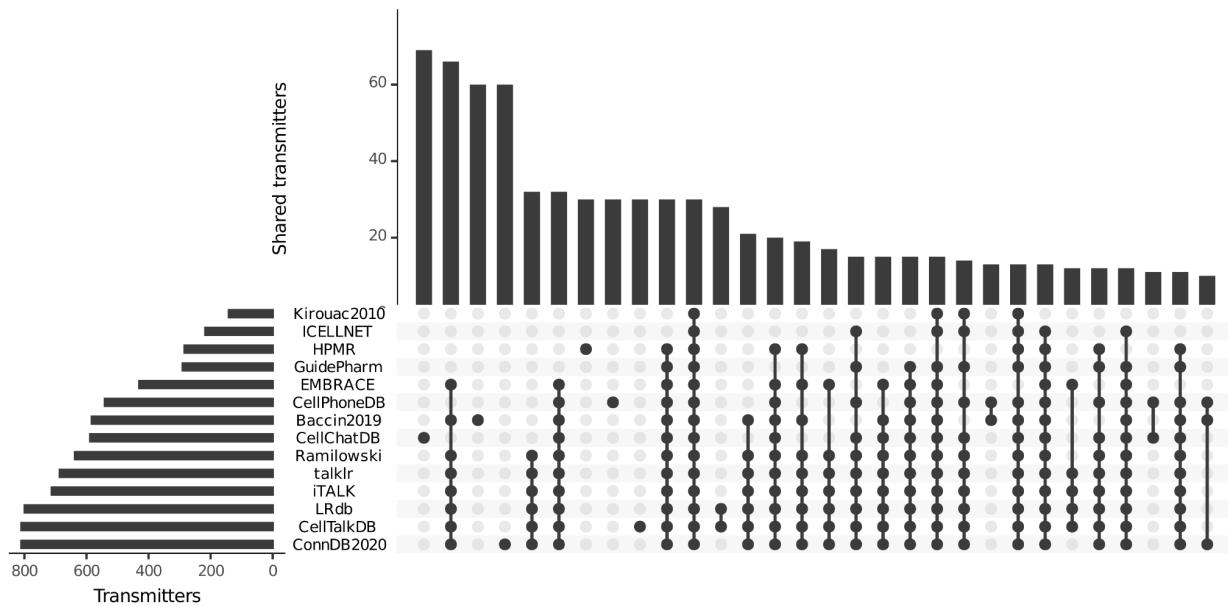
Supplementary Figure S3. *A) Interactions B) Receivers and C) Transmitters present in each resource when taken from the rest of the resources. Note these plots are asymmetric and represent the % of interactions from the resources on the X axis found in each resource on the Y axis.*

A)

	Baccin2019	CellChatDB	CellPhoneDB	CellTalkDB	ConnDB2020	EMBRACE	GuidePharm	HPMR	ICELLNET	iTALK	Kirouac2010	LRdb	OmniPath	Ramilowski	talklr	Mean Shared
Baccin2019	100	24.4	42.1	34.5	50.8	61.2	55.9	56	54.3	46.1	59.3	37	18	60.7	48.9	46.4
CellChatDB	43	100	60.3	30.2	42.4	38.3	68.3	55.1	81.6	36.8	80	31.4	20.9	42.6	38.8	47.8
CellPhoneDB	42.8	34.9	100	30	41.5	32.3	71.5	49.5	80.9	33.8	70	28.8	16.9	42.3	35.7	43.6
CellTalkDB	80.5	39.9	68.6	100	82.5	91.6	86.1	79.4	86.4	93.8	79.3	93.5	43.3	97	94.8	79.8
ConnDB2020	79	37.4	63.5	55.1	100	70	81.3	76.4	87.5	70.1	77.3	58.2	31.3	92.5	74.2	68.1
EMBRACE	62.7	22.3	32.6	40.3	46.1	100	38.1	45.1	39.3	56	39.3	43.7	18.2	58.1	59.6	42.9
GuidePharm	25.4	17.6	32	16.8	23.8	16.9	100	39.3	65	22.1	59.3	17.7	8.8	28	22.8	28.2
HPMR	17.7	9.9	15.4	10.8	15.5	14	27.3	100	25	14.7	36.7	11.6	5.7	18.6	14.7	17
ICELLNET	16.4	14	24.1	11.2	17	11.6	43.2	23.9	100	14.3	62.7	11.8	6.5	18.4	14.8	20.7
iTALK	81.3	36.9	58.6	70.9	79.5	96.3	85.6	81.6	83.4	100	76.7	78	31.3	100	100	75.7
Kirouac2010	6.1	4.7	7.1	3.5	5.1	4	13.4	11.9	21.4	4.5	100	3.6	1.8	5.9	4.8	7
LRdb	82.1	39.5	62.8	88.9	83	94.5	86.1	81.3	86.6	98.1	77.3	100	40.4	99.6	98.4	79.9
OmniPath	75.6	49.8	69.9	78	84.4	74.4	81.4	75.9	90.7	74.4	74.7	76.4	100	81.5	78.7	76.1
Ramilowski	78.8	31.4	53.9	54	77.1	73.6	79.8	76.4	78.9	73.6	74.7	58.3	25.2	100	78.2	65.3
talklr	81	36.5	58.1	67.4	79	96.3	83.1	77	81.4	94	76.7	73.5	31.1	99.9	100	73.9

% Present: 100 / 75 / 50 / 25

B)

	Baccin2019	CellChatDB	CellPhoneDB	CellTalkDB	ConnDB2020	EMBRACE	GuidePharm	HPMR	ICELLNET	iTALK	Kirouac2010	LRdb	OmniPath	Ramilowski	talklr	Mean Shared
Baccin2019	100	64	59.3	59.7	68.6	85.5	82.2	74.7	78.8	67	93.8	64.2	60.7	78.4	71.4	72
CellChatDB	56.2	100	64.1	47.8	57	54.6	85.5	66.8	85.4	51.4	96.3	50.5	54.1	59.4	54.5	63.1
CellPhoneDB	56.6	69.7	100	52	59.9	52.4	85.5	66.3	87.4	51.9	92.6	52.4	57.9	59.3	54.6	64.2
CellTalkDB	82.4	75.2	75.3	100	86	95.2	99.2	93.9	94.9	93.8	97.5	93	80.1	97.1	94.3	89.9
ConnDB2020	82.6	78.2	75.7	75	100	85.9	97.1	89.4	97	80.7	98.8	80	79.7	93	85.5	85.6
EMBRACE	66.8	48.7	42.9	53.9	55.8	100	58.9	61.9	61.1	62.8	69.1	59	47.7	68.8	67.5	58.9
GuidePharm	35.1	41.6	38.3	30.7	34.5	32.2	100	46.9	63.6	33.8	79	32	31.3	38.7	35.2	40.9
HPMR	53.9	54.9	50.2	49	53.6	57.1	79.3	100	73.2	55.2	82.7	51.7	50.3	63	57.4	59.4
ICELLNET	27.7	34.1	32.2	24.1	28.3	27.4	52.3	35.6	100	26.2	86.4	25.4	25.9	30.2	27.5	34.5
iTALK	83	72.5	67.5	84.2	83.1	99.5	97.9	94.8	92.4	100	97.5	93.3	72.9	100	100	88.5
Kirouac2010	13.5	15.8	13.9	10.1	11.8	12.7	26.6	16.5	35.4	11.3	100	10.6	10.9	13.4	12.2	15.3
LRdb	84.8	76	72.5	88.8	87.6	99.5	98.8	94.6	95.5	99.3	97.5	100	76.6	99.8	99.8	90.8
OmniPath	75.4	76.6	75.3	72	82.2	75.7	90.9	86.5	91.4	73	93.8	72	100	81.5	76.9	80.2
Ramilowski	81.9	70.7	64.9	73.4	80.7	91.8	94.6	91.2	89.9	84.3	97.5	79	68.6	100	90.5	82.8
talklr	82.3	71.5	66	78.7	81.9	99.5	95	91.6	90.4	93	97.5	87.2	71.4	99.8	100	86.1

% Present: 100 / 75 / 50 / 25

38

C)

	Baccin2019	CellChatDB	CellPhoneDB	CellTalkDB	ConnDB2020	EMBRACE	GuidePharm	HPMR	ICELLNET	iTALK	Kirouac2010	LRdb	OmniPath	Ramilowski	talklr	Mean Shared
Baccin2019	100	56.6	59.5	55.9	59.3	81.6	68.3	69	65	64	70.2	57.9	44.7	70.3	65.6	63.4
CellChatDB	57.1	100	77.8	56.3	57.9	60.2	87.9	71.5	93.5	58.4	90.8	55.9	48	61.4	59.5	66.9
CellPhoneDB	55.2	71.6	100	53.2	57	52.3	90	66.9	92.2	54.8	79.4	53.8	46.2	58.6	55.5	63.3
CellTalkDB	77.7	77.6	79.7	100	82.3	96.3	96.9	84.9	95.4	95.5	92.2	93.1	60.5	98	96.5	87.6
ConnDB2020	82.3	79.8	85.4	82.3	100	88.4	94.8	85.2	96.8	87.4	92.9	83.6	68.1	94.7	89.4	86.5
EMBRACE	60.2	44	41.6	51.1	46.9	100	49	56.3	47.5	59.7	54.6	52.9	31	61.7	62	51.3
GuidePharm	34	43.4	48.2	34.7	34	33	100	51.8	74.2	38.3	70.2	35.2	25.8	41.8	39.1	43.1
HPMR	33.6	34.5	35.1	29.8	29.9	37.2	50.7	100	44.2	34	51.1	30.7	22.4	37.4	34.7	36.1
ICELLNET	24.2	34.5	37	25.6	25.9	24	55.5	33.8	100	27.7	71.6	25.9	18.4	29.5	27.8	33
iTALK	78.2	70.7	72.1	84	76.8	96.8	94.1	85.2	90.8	100	87.2	88.4	53	100	100	84.2
Kirouac2010	17	21.8	20.7	16	16.2	17.9	34.1	25.4	46.5	17.3	100	15.8	11.4	18.8	17.9	21.2
LRdb	79.4	76	79.5	91.9	82.5	98.4	96.9	86.3	95.4	99.2	89.4	100	59.4	99.8	99.7	88.1
OmniPath	69.3	73.6	77.1	67.4	75.9	65.1	80.3	71.1	76.5	67.3	73	67.1	100	70.3	68.5	71.6
Ramilowski	76.8	66.5	68.9	77	74.4	91.4	91.7	83.8	86.6	89.5	85.1	79.6	49.6	100	92.6	79.5
talklr	77.2	69.4	70.4	81.7	75.7	98.8	92.4	83.8	88	96.3	87.2	85.6	52	99.7	100	82.7

% Present
100
75
50
25

Supplementary Figure S4. *Numbers and Percentages of Subcellular locations annotations of Receivers (A-B) and Transmitters (C-D) for each CCC resource. S, P and T stand for Secreted, Peripheral plasma membrane, and Transmembrane plasma membrane proteins, respectively.*

A)

Location (OP-L)
P
S
T

B)

Location (OP-L)
P
S
T

39

C)

D)

Supplementary Figure S5. *Percentages per Signalling category according to OmniPath locations (OP-L) for each CCC resource.*
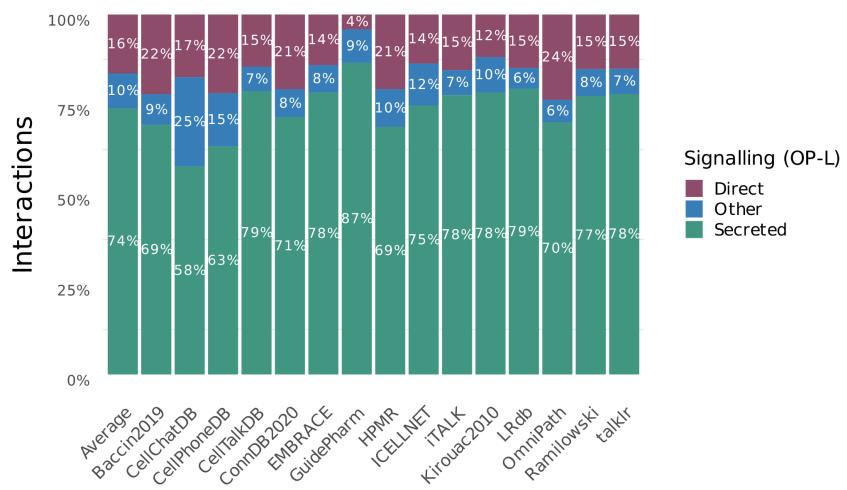
Supplementary Figure S6. *Number of matches to A) Interactions, B) Receivers and C) Transmitters, Enrichment Scores for their Receivers and Transmitters (D-E), and the Percentages of Interactions, Receivers and Transmitters (F-H) matched to the SignaLink database per resource.*
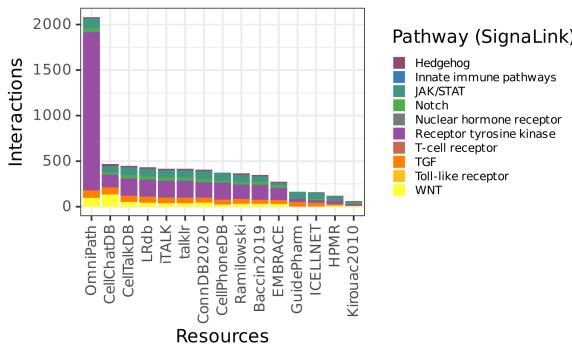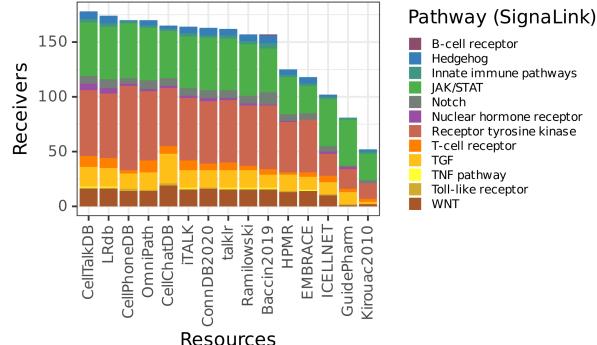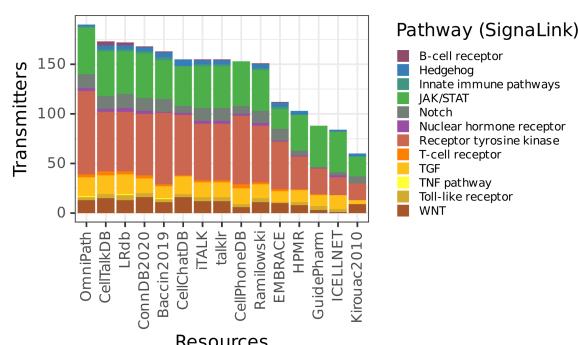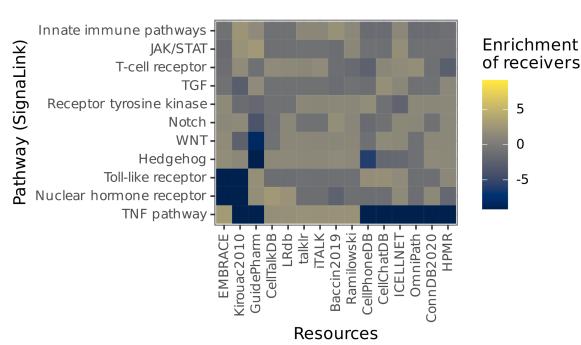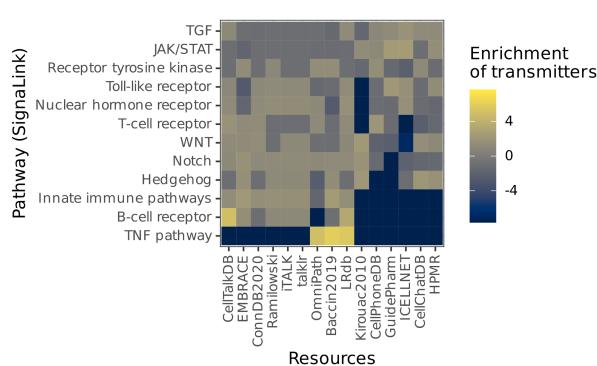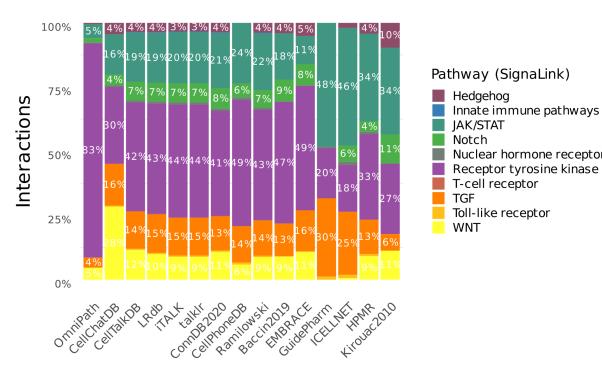
G)

H)
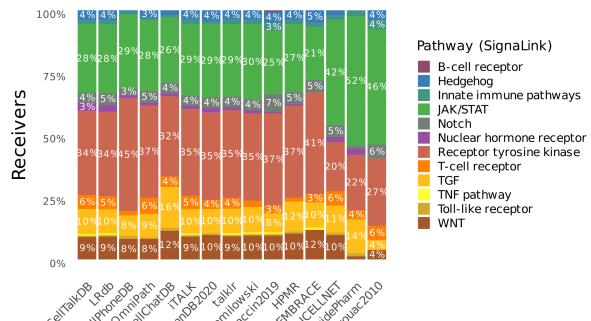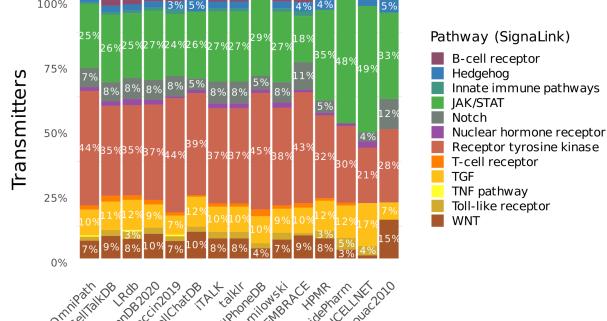
Supplementary Figure S7. *Number of matches to A) Interactions, B) Receivers and C) Transmitters, Enrichment Scores for their Receivers and Transmitters (D-E), and the Percentages of Interactions, Receivers and Transmitters (F-H) matched to the NetPath database per resource.*

A)

B)

C)

D)
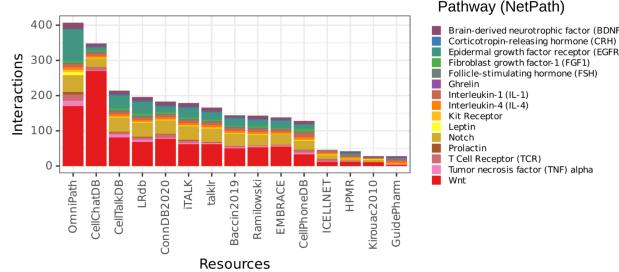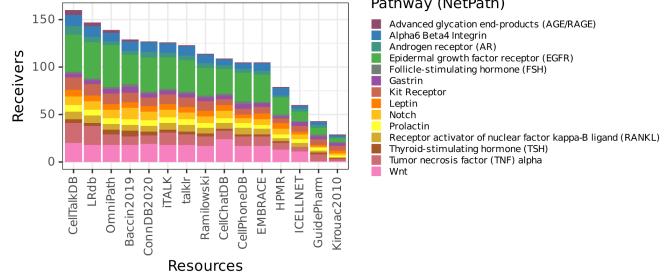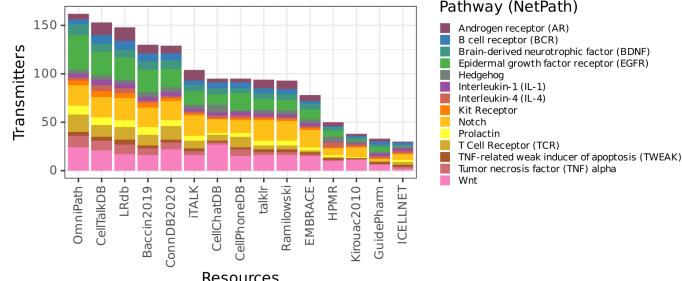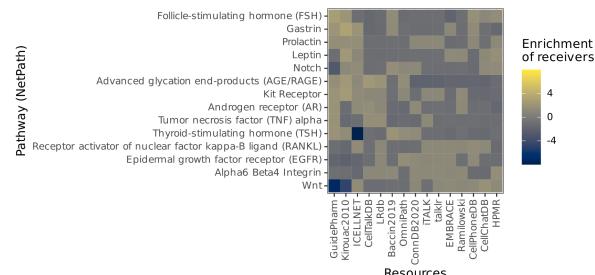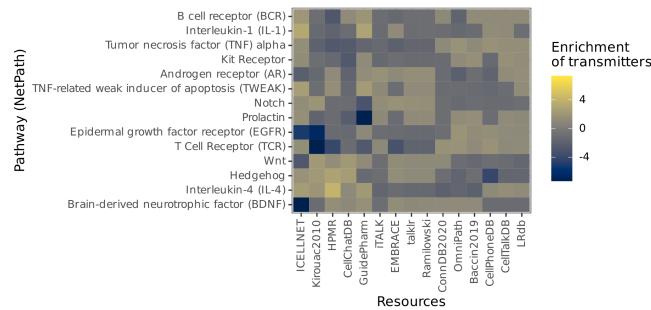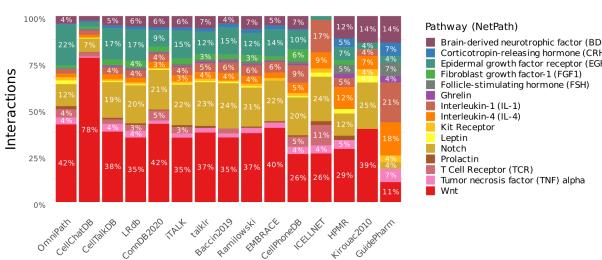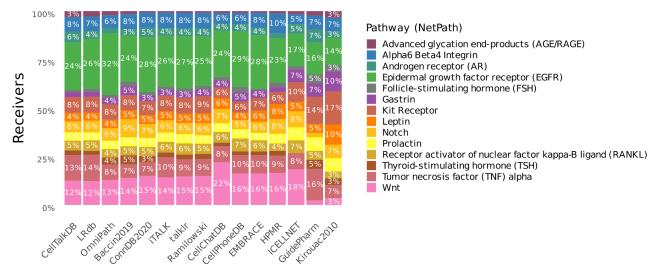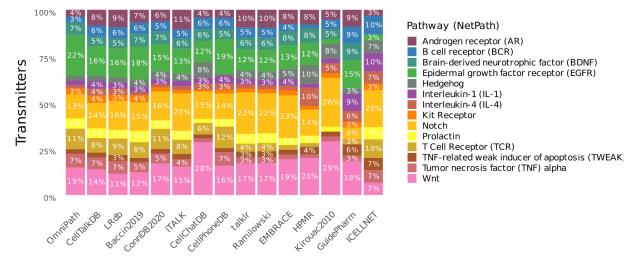
Supplementary Figure S8. *Number of matches to A) Merged Sets of Receivers and Transmitters, B) Receivers and C) Transmitters, their corresponding Enrichment Scores (D-F), and Percentages (G-I) per resource matched to the HGNC database.*
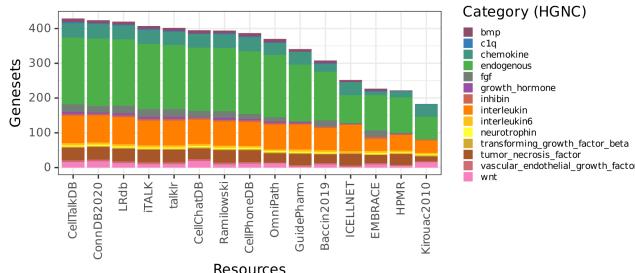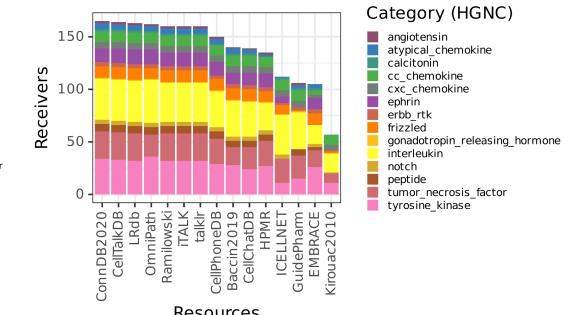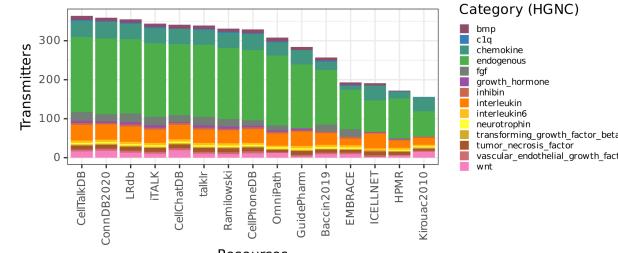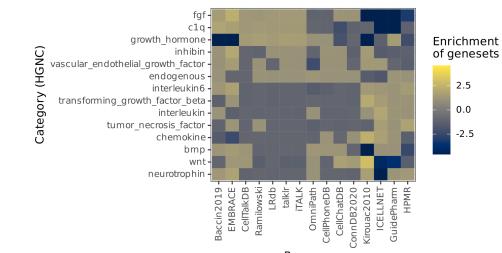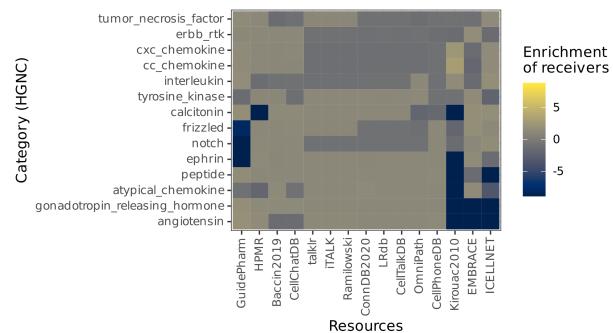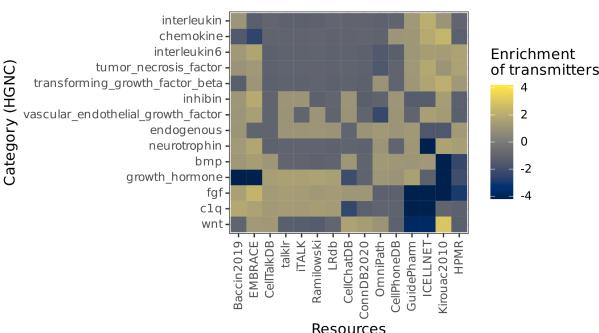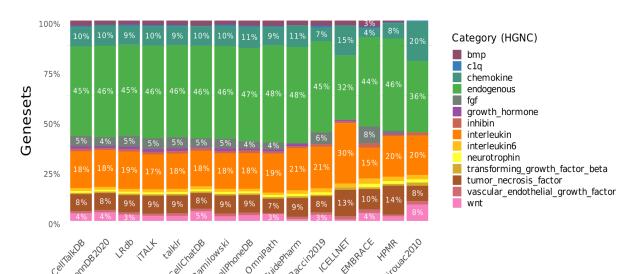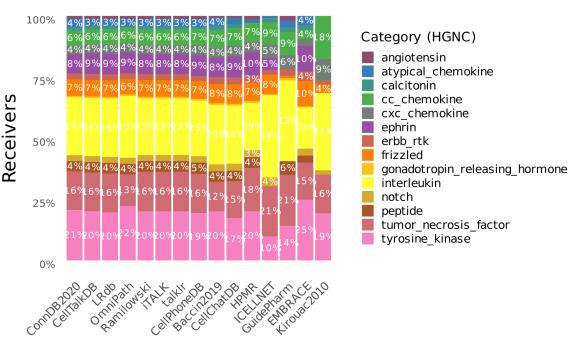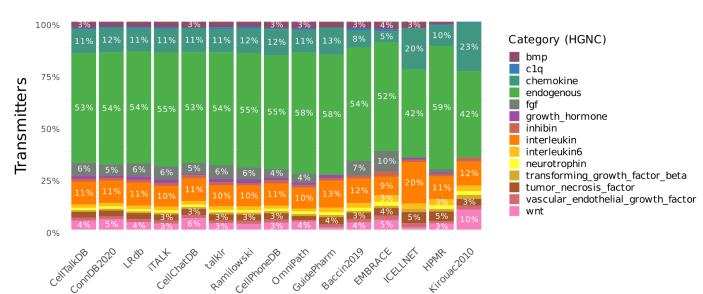
Supplementary Figure S9. *Number of matches to A) Merged Sets of Receivers and Transmitters, B) Receivers and C) Transmitters and their corresponding (D-F) Enrichment Scores, and Percentages (G-I) per resource matched to the CancerSEA database.*
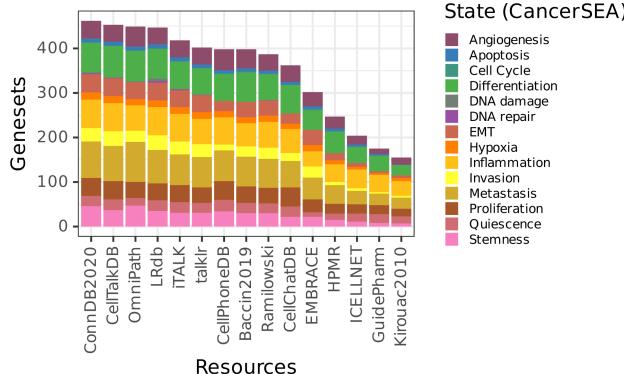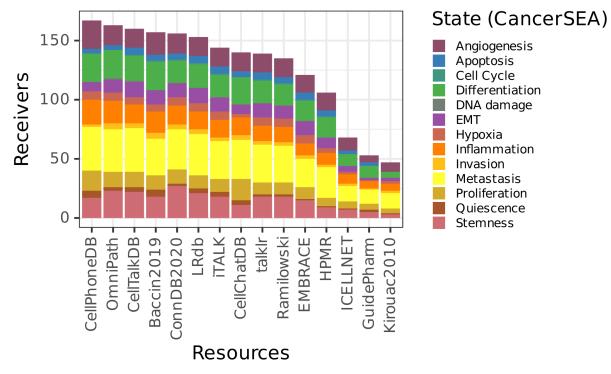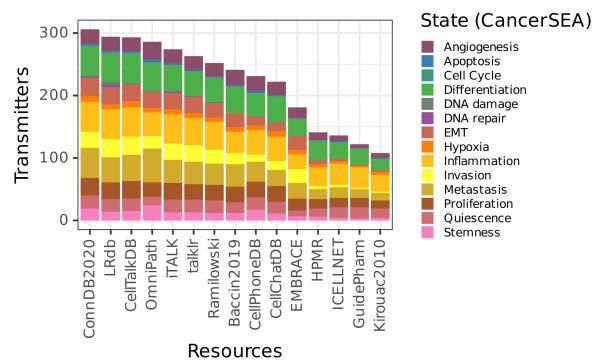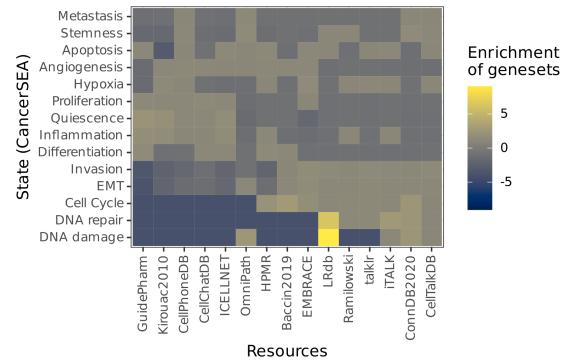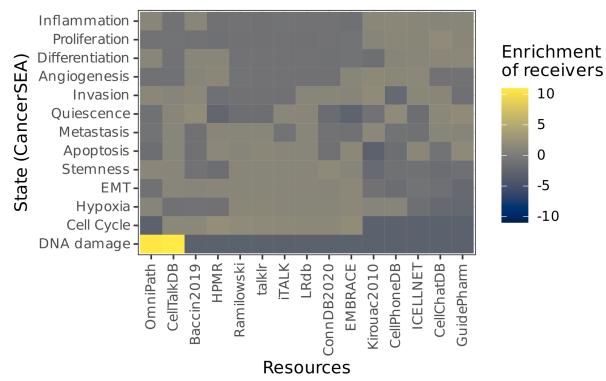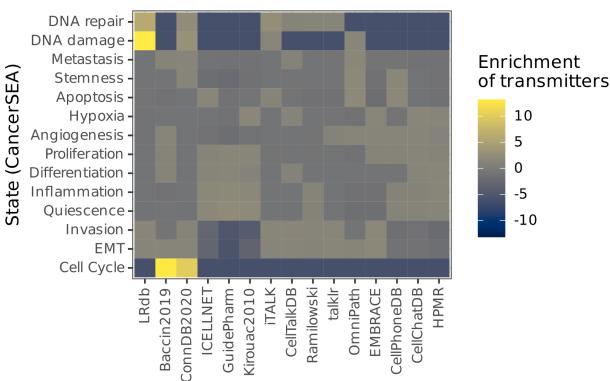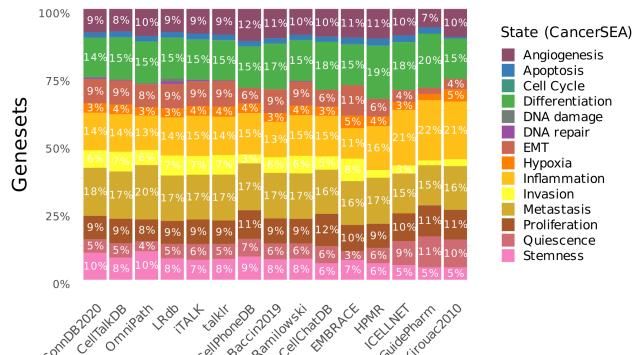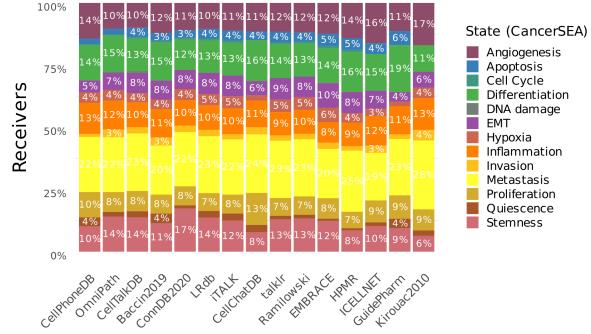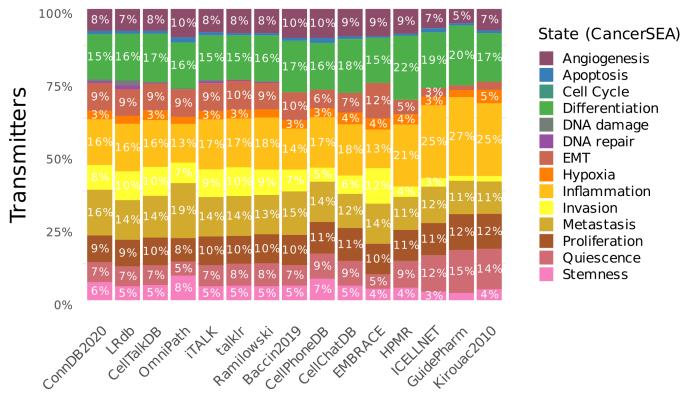
Supplementary Figure S10. *Overlap in the A) 100, B) 250, and C) 1000 highest ranked CCC interactions between different combinations of methods and resources. Method-resource combinations were clustered according to binary (Jaccard index) distances. SCA refers to the SingleCellSignalR method.*

A)

B)

C)

48
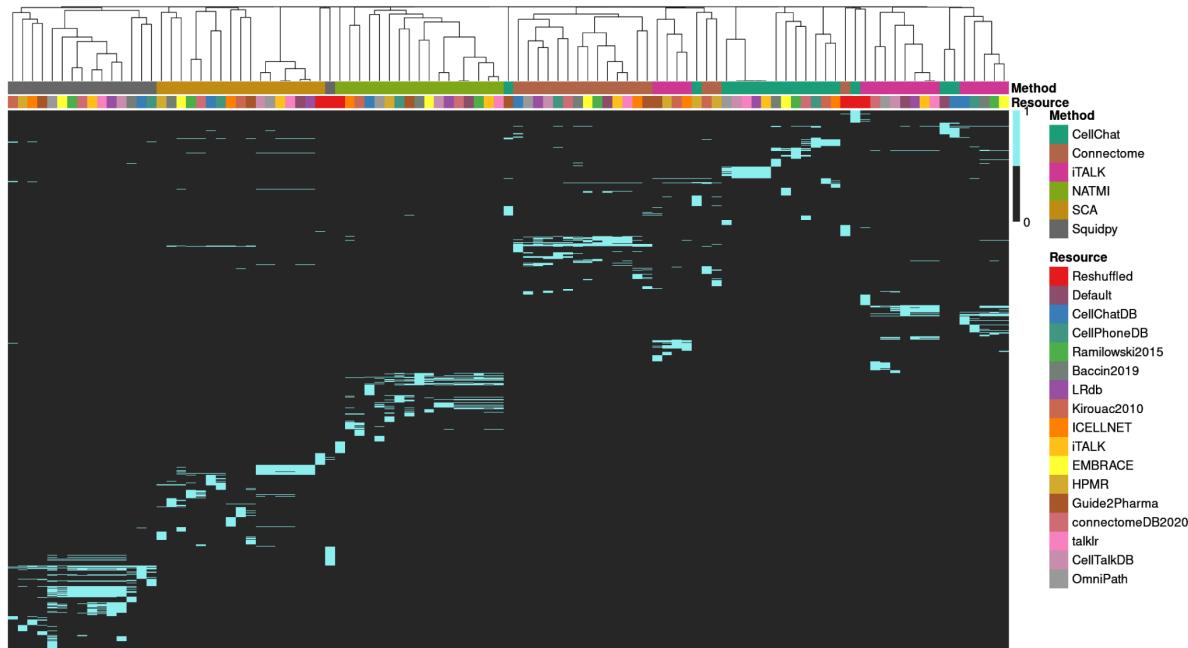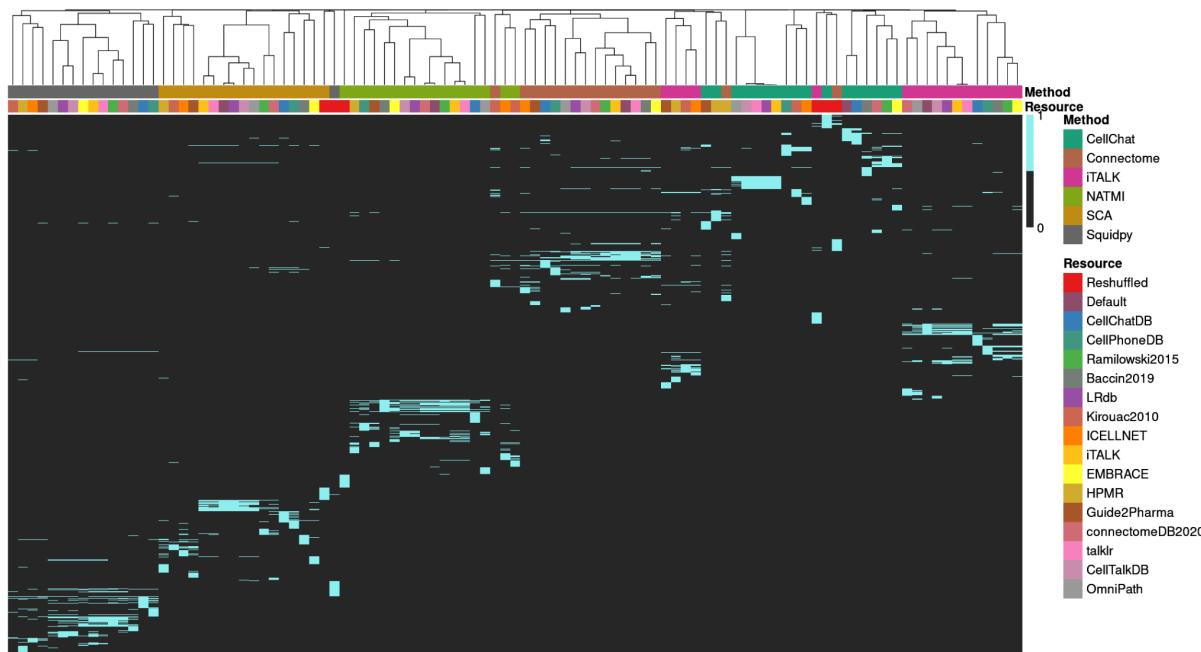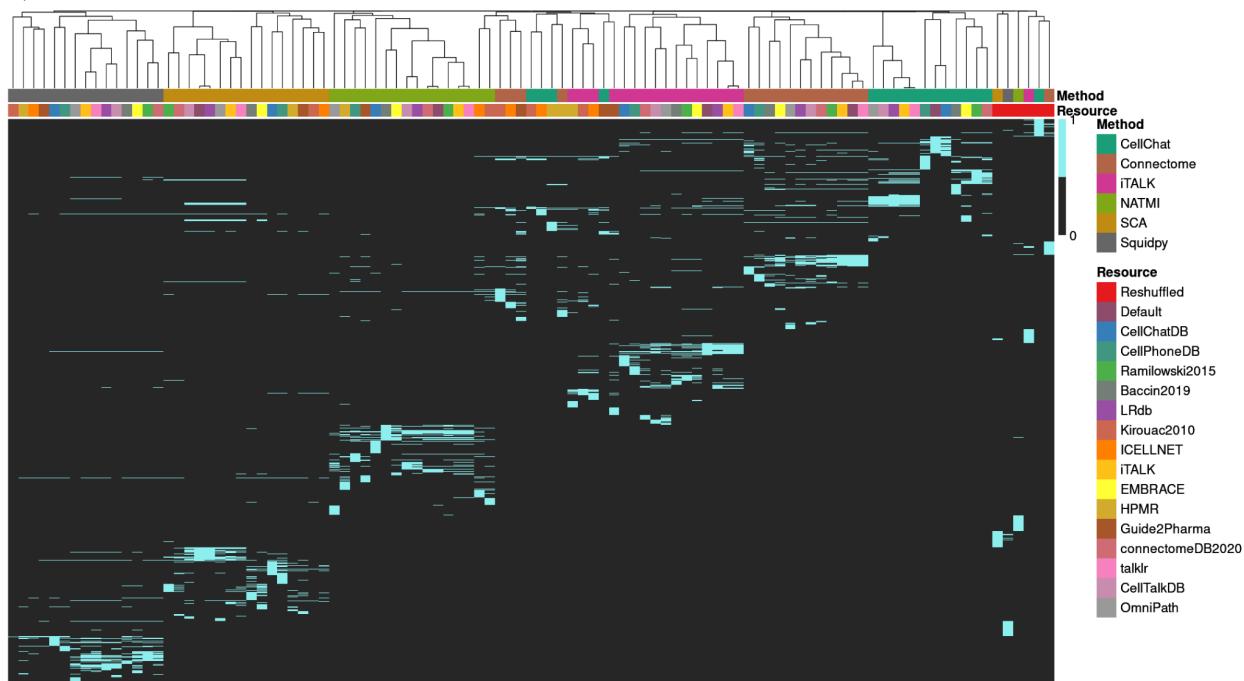
Supplementary Figure S11. *Jaccard indices for the 500 highest ranked interactions obtained from each method-resource combination.*
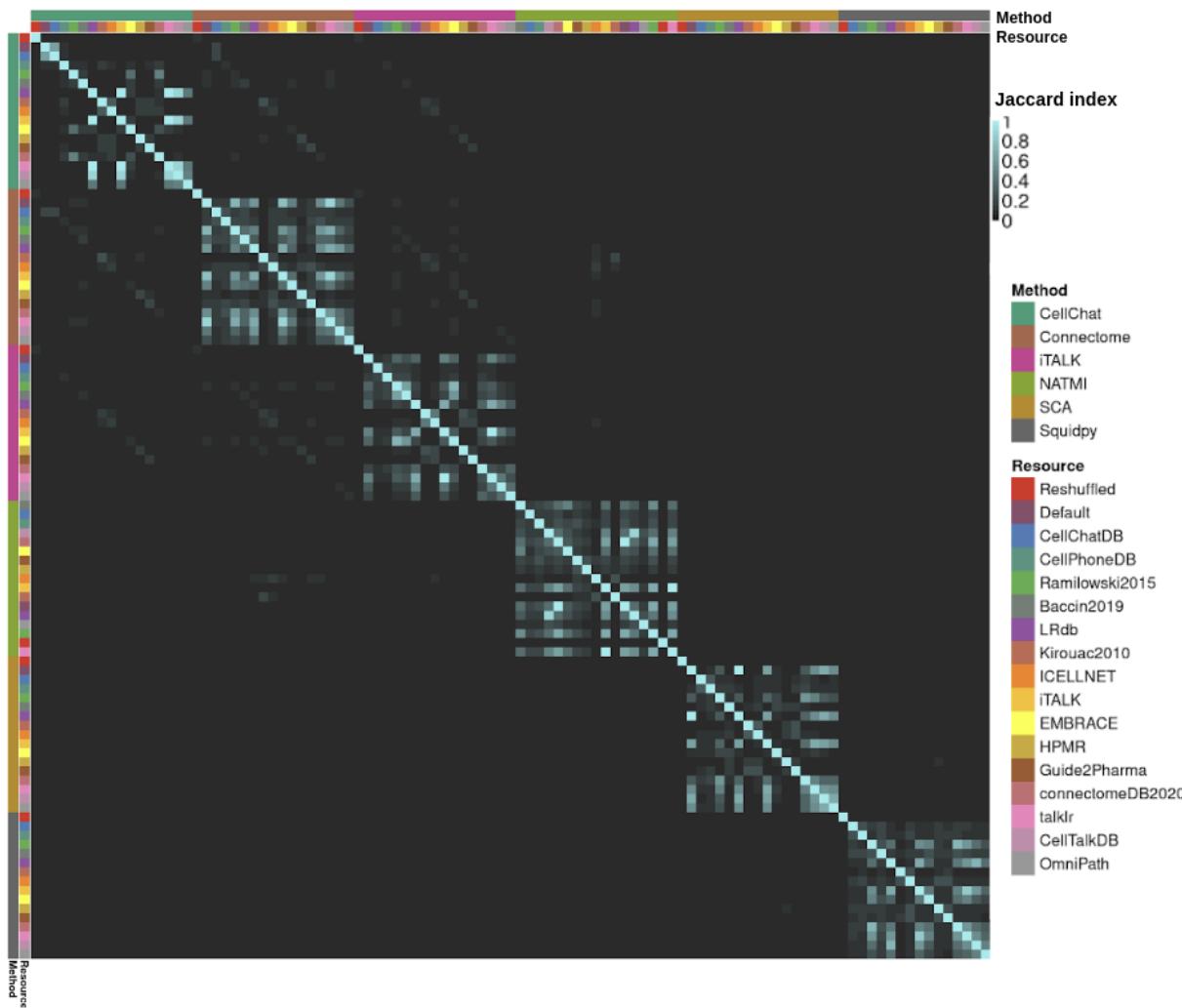
Supplementary Figure S12. - *Upset plot showing the overlap between the 500 highest ranked interactions using the same method with all resources.*

A) *Using CellChat (probabilities) with all resources.*
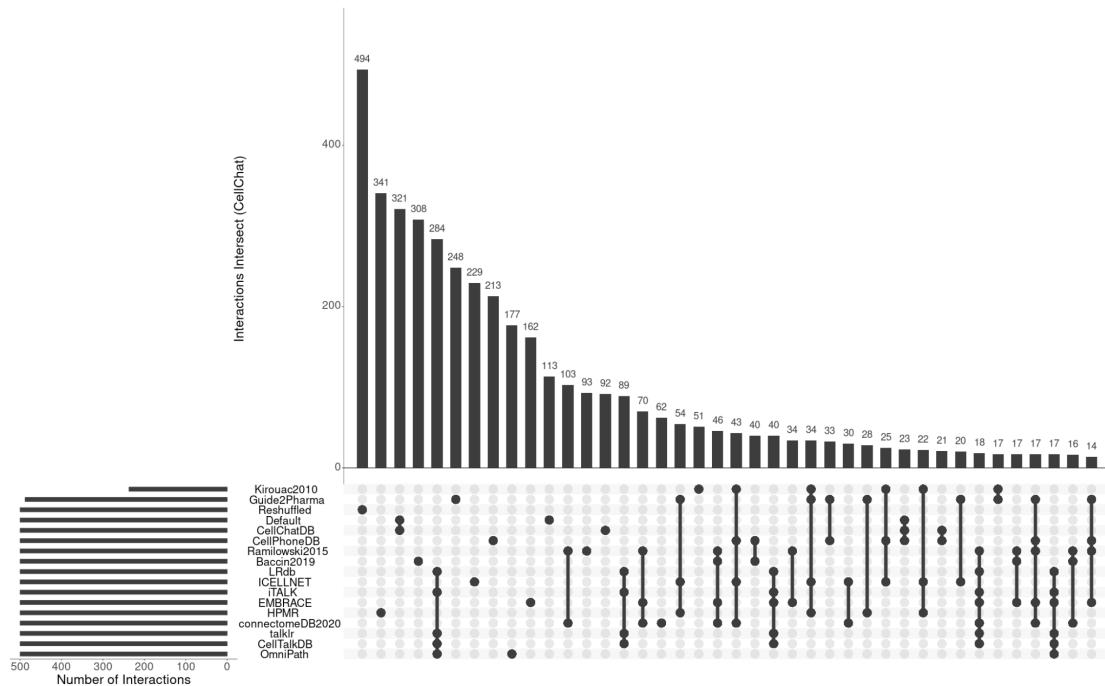
B) *Using Connectome with all resources.*

50
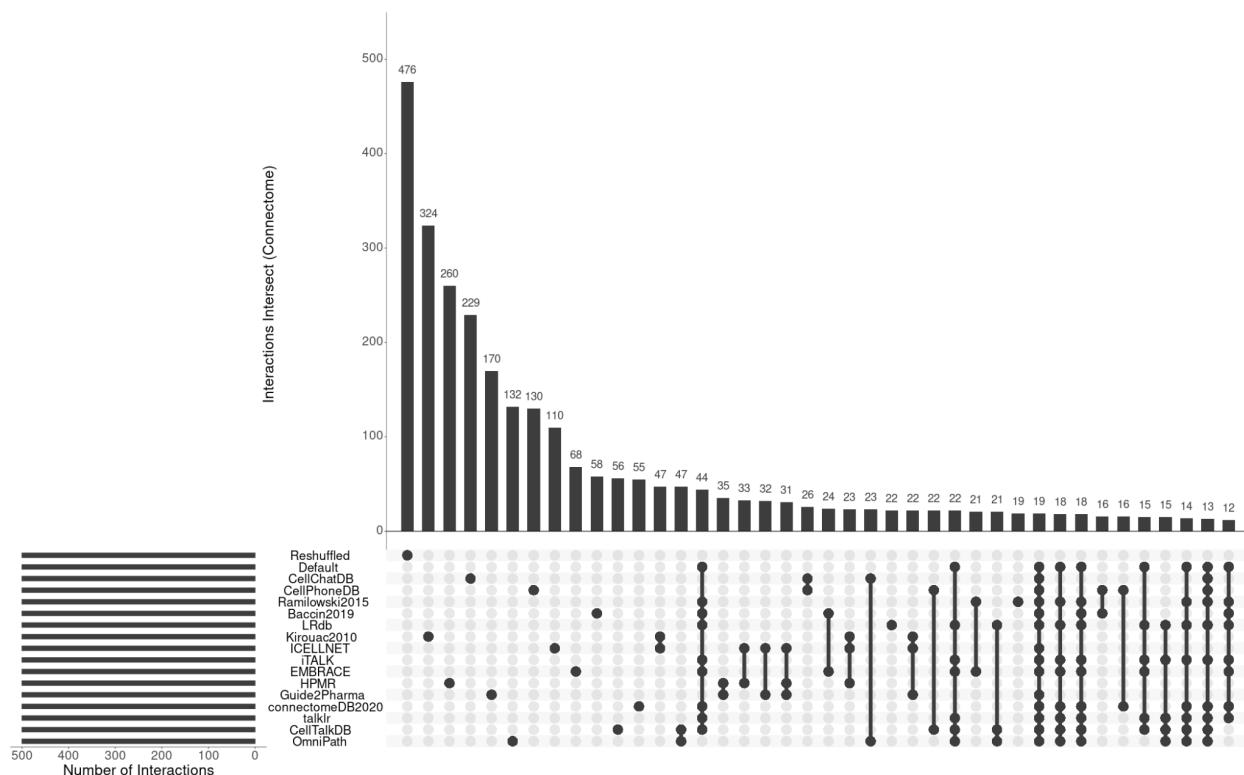
C) Using iTALK with all resources.
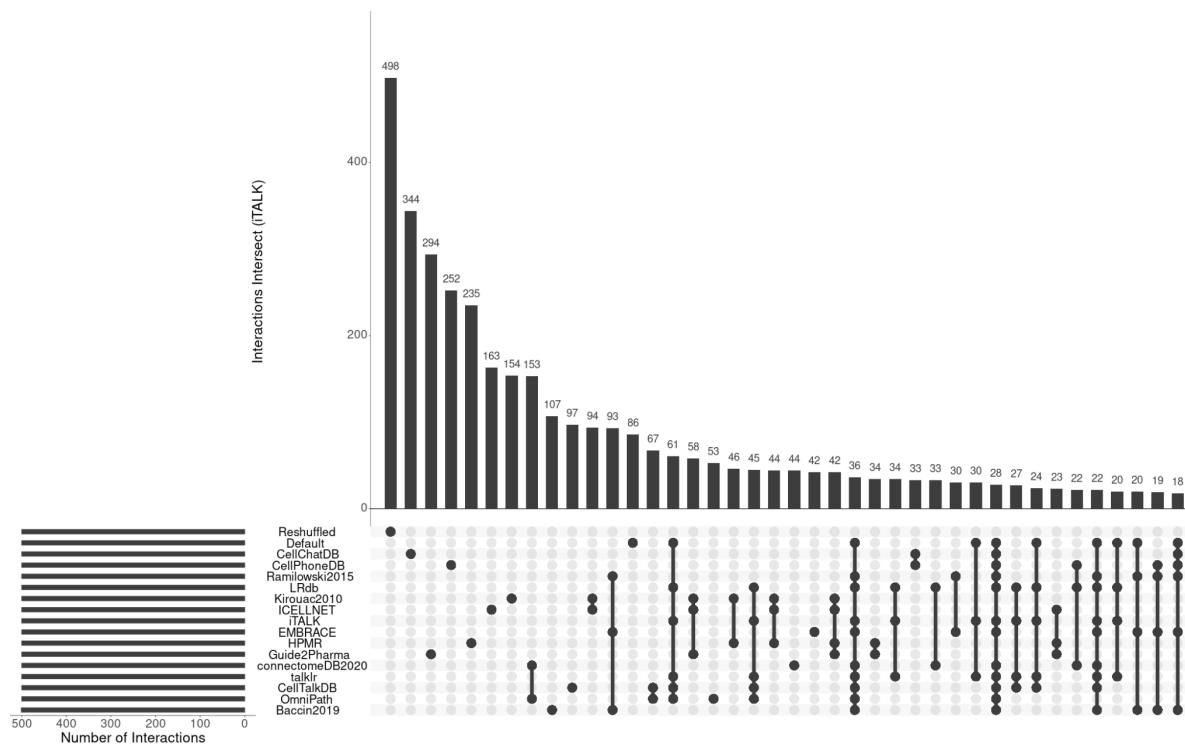
D) Using NATMI (edge specificity) using all resources.
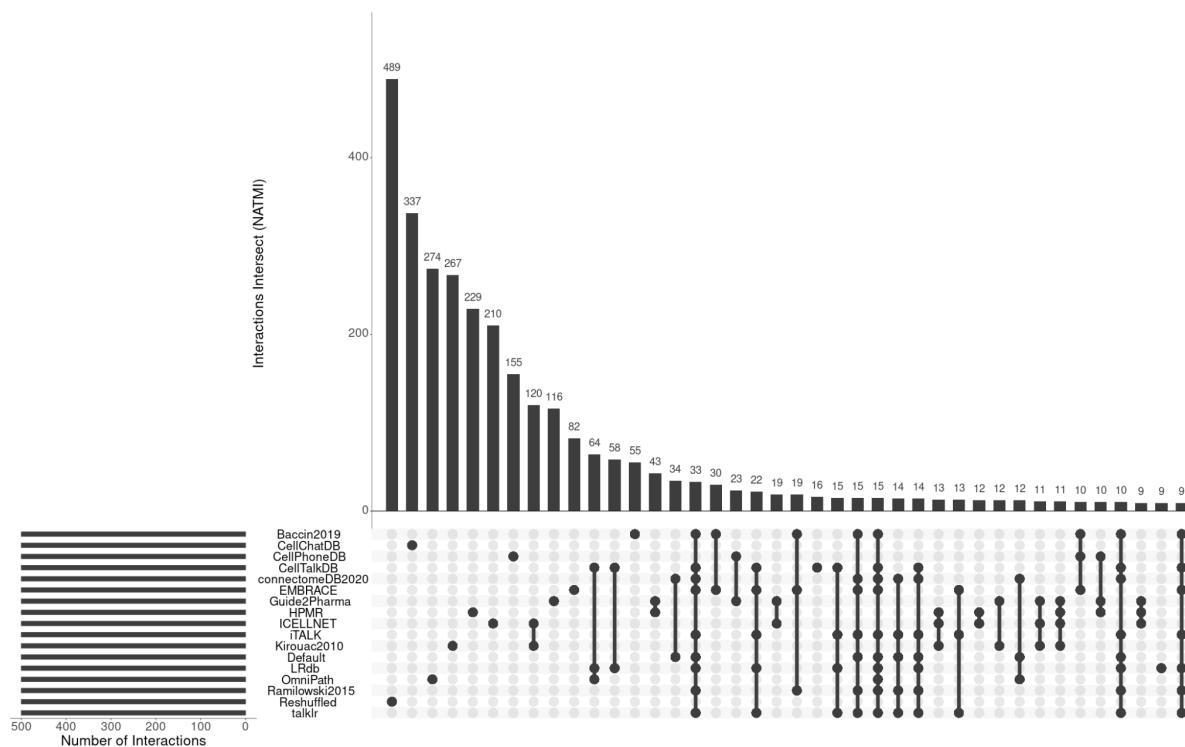
E) Using SingleCellSignalR with all resources.
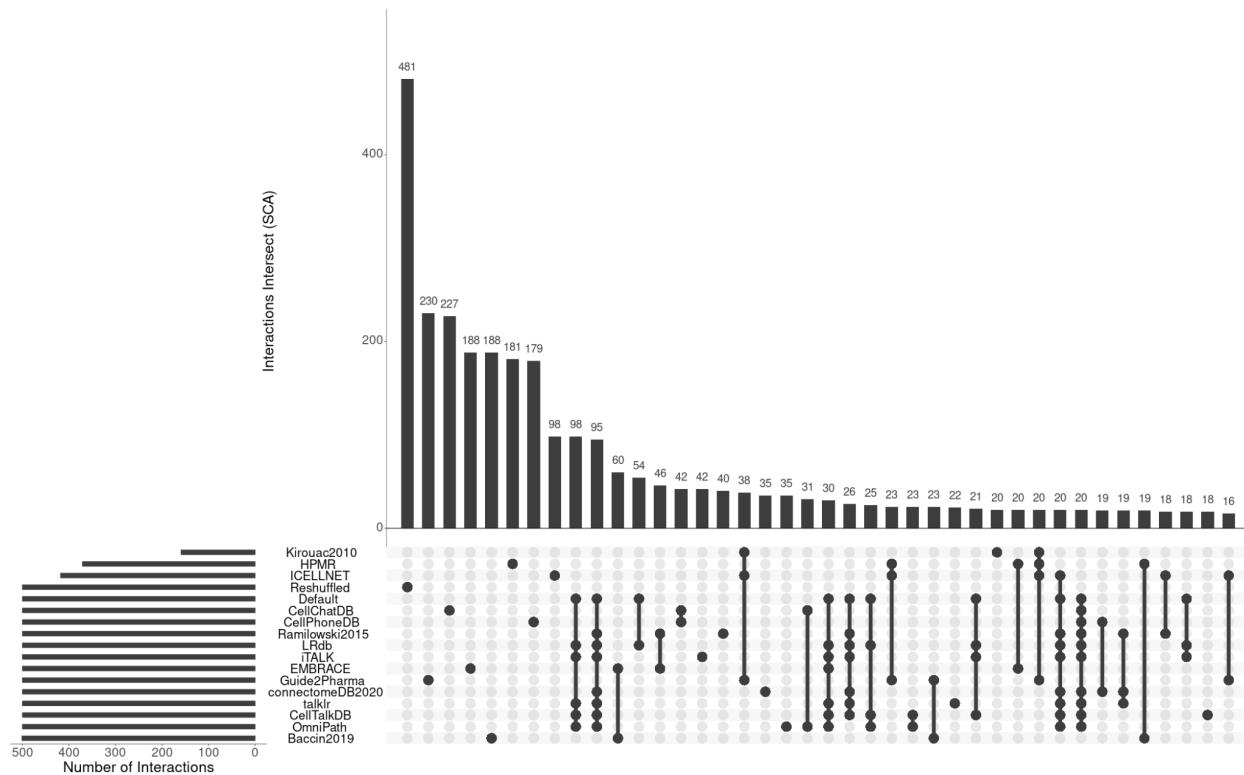
F) Using Squidpy with all resources.
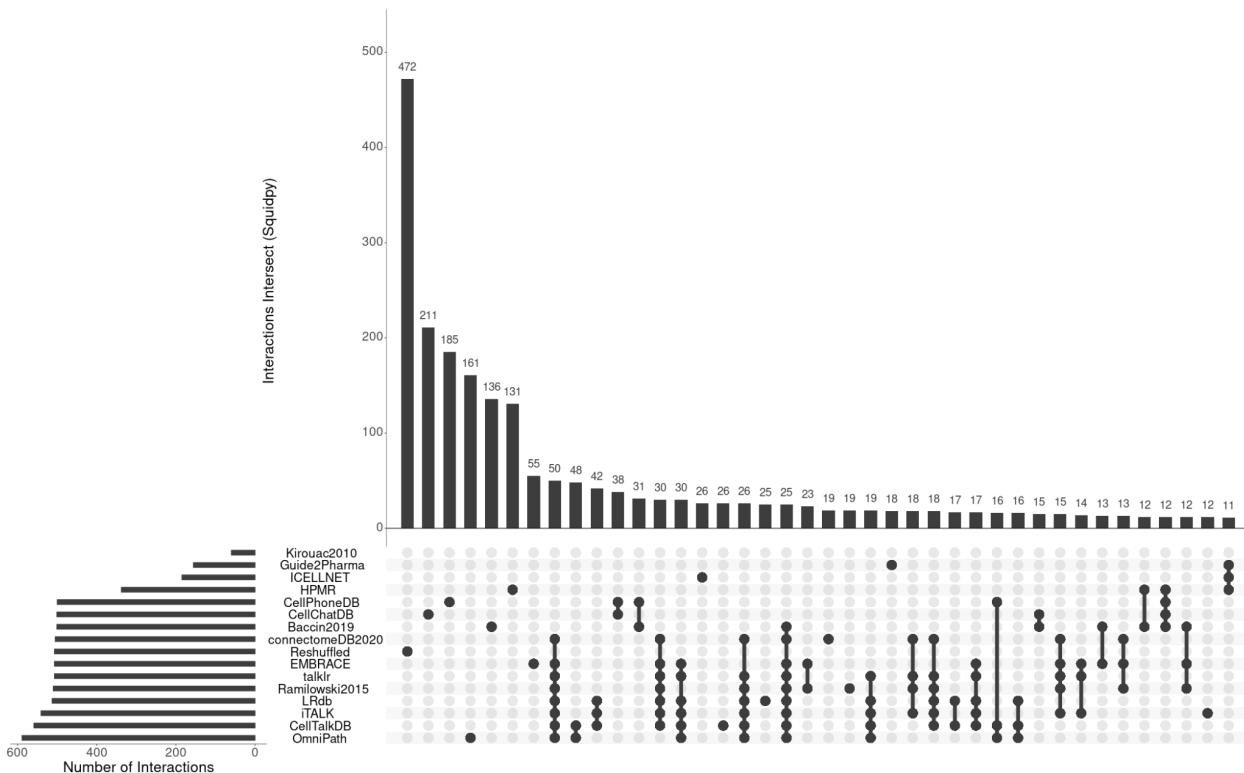
Supplementary Figure S13. *Overlap, Jaccard indices, and Activity per Cell type in the 500 highest ranked interactions between different combinations of methods and resources for Pancreatic islet (A-C) and Cord Blood Mononuclear Cells (D-F) scRNA-Seq datasets, respectively.*

A)
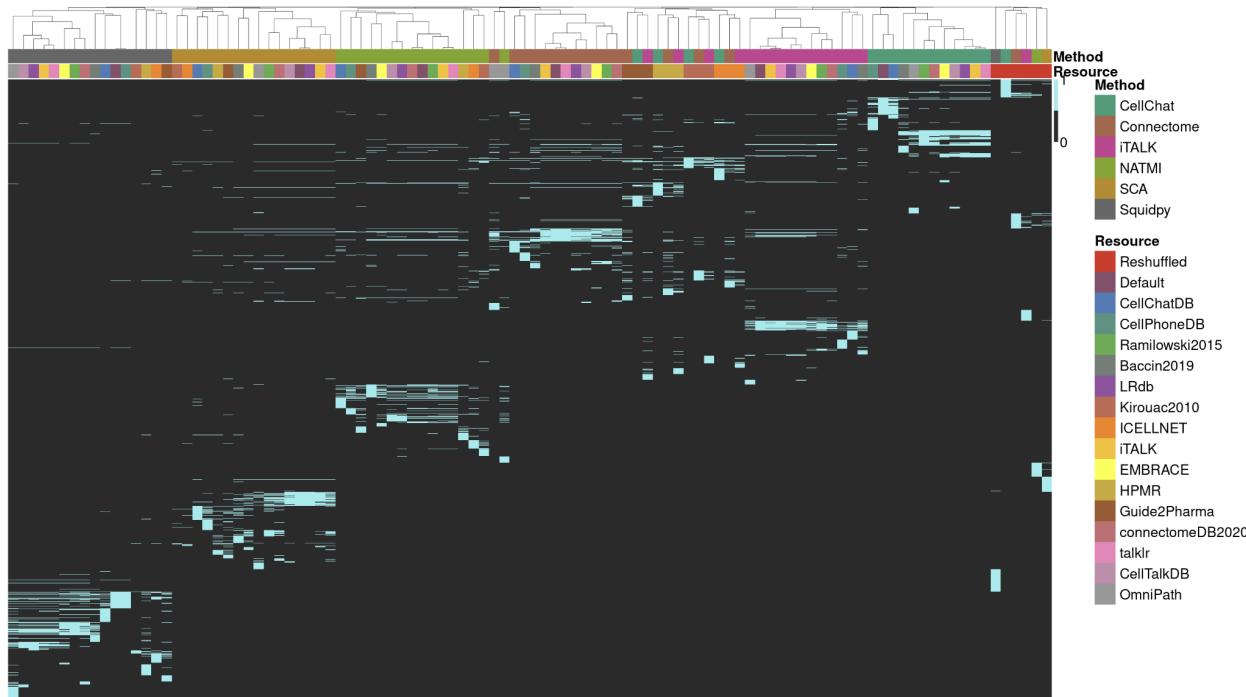
B)

C)
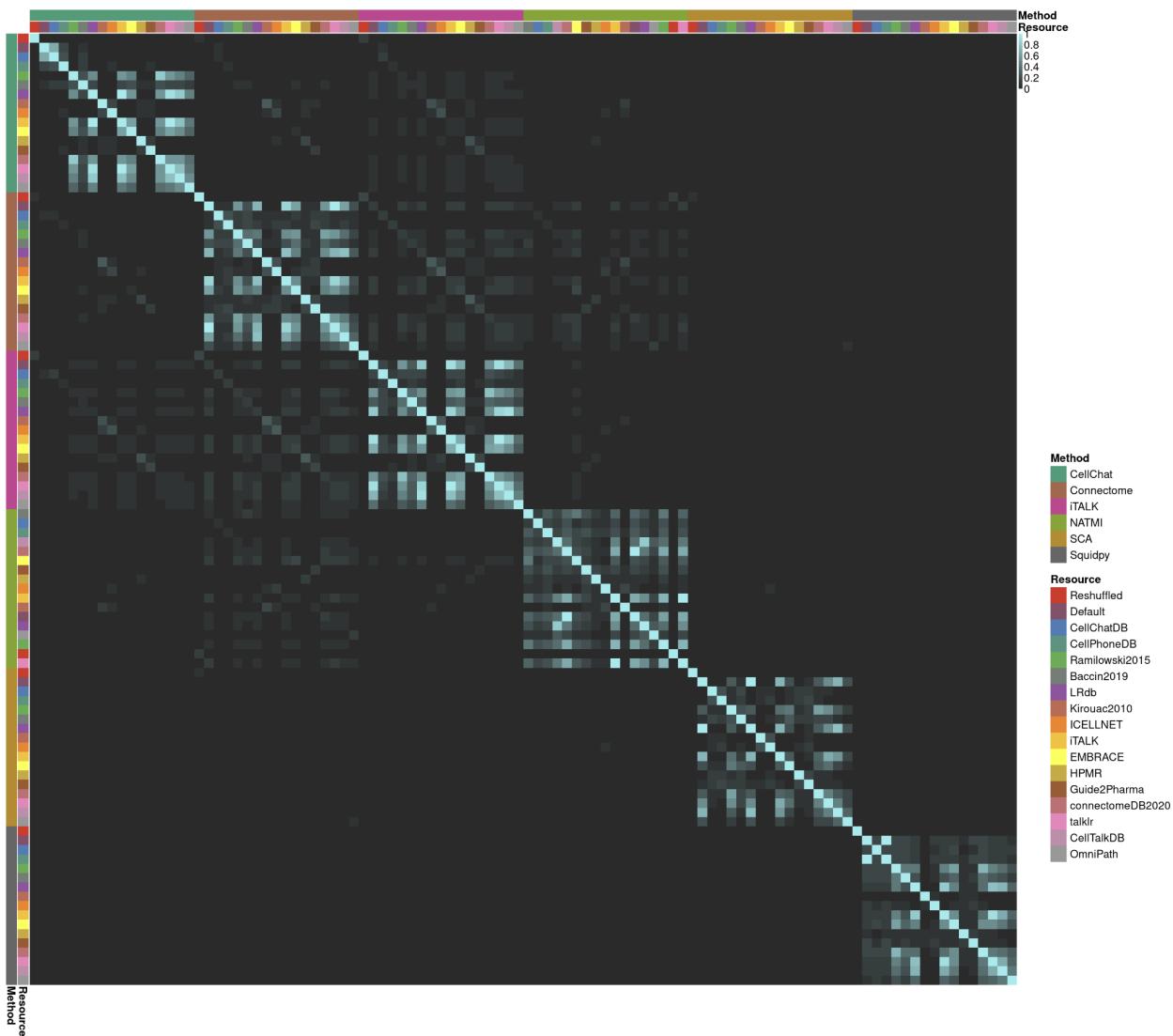
D)
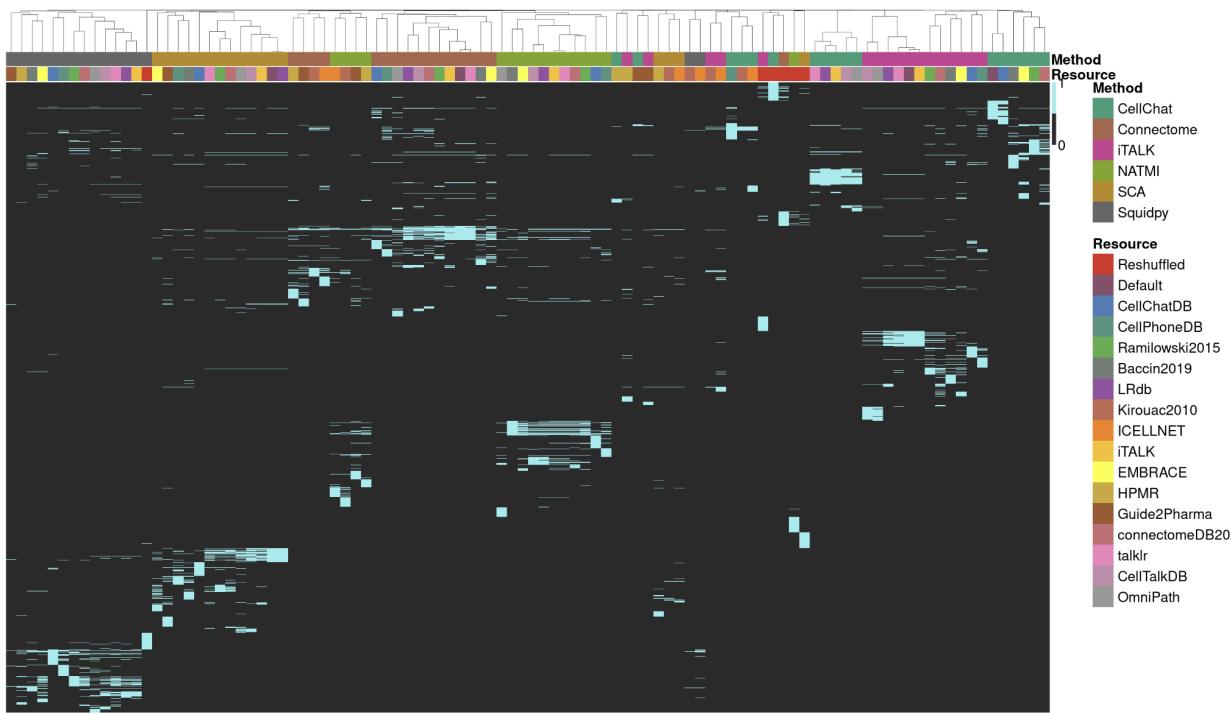
E)

F)
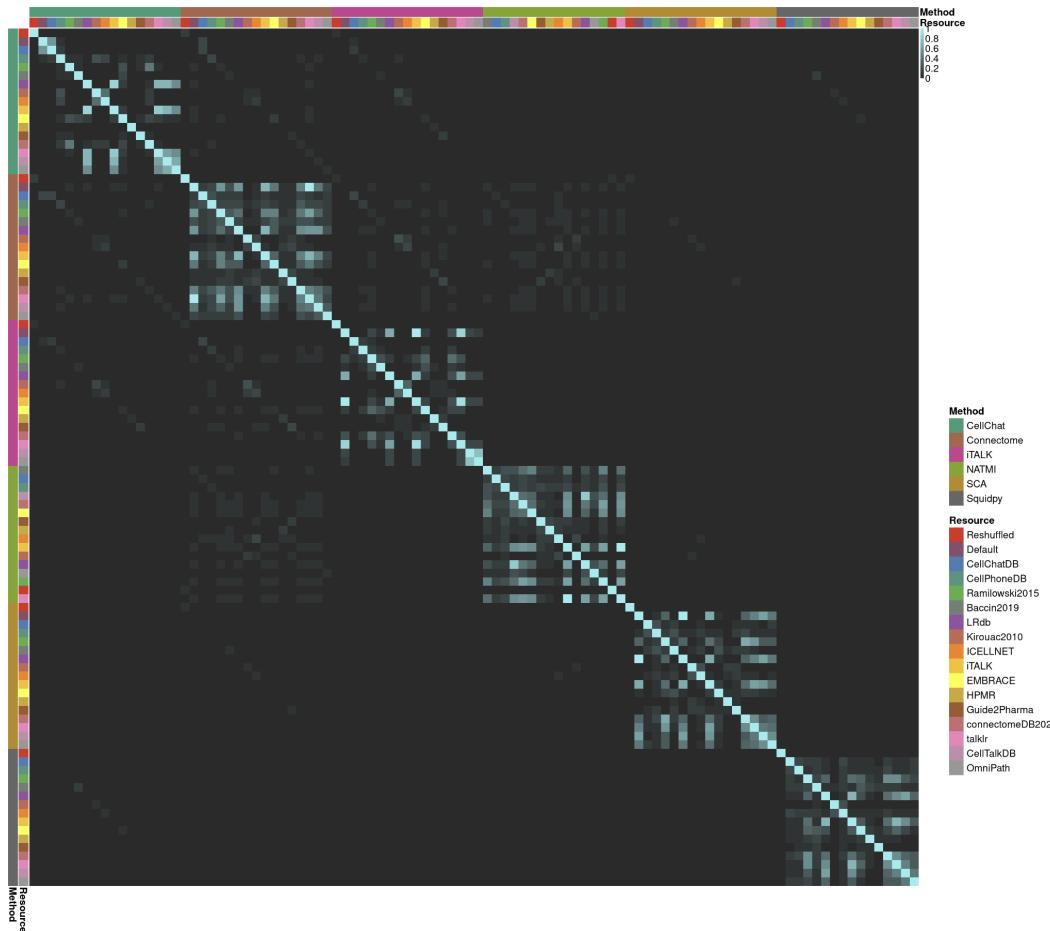
Supplementary Figure S14. *Upset plot showing overlap of most relevant interactions for each method with the same resource*

A) *Each method with the OmniPath Resource*
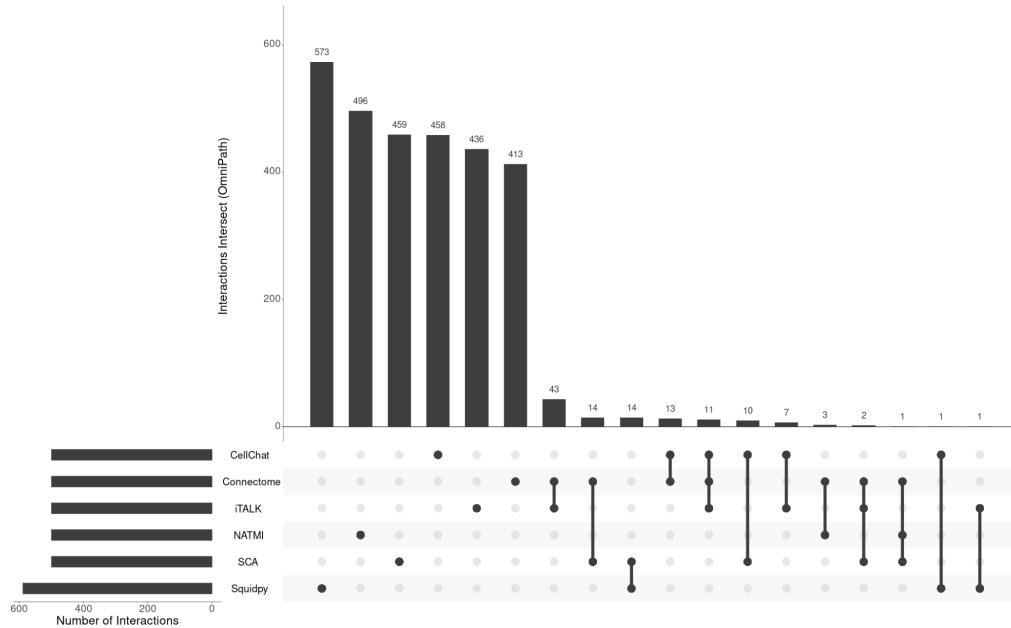
B) *Each method with its inbuilt Default resource*
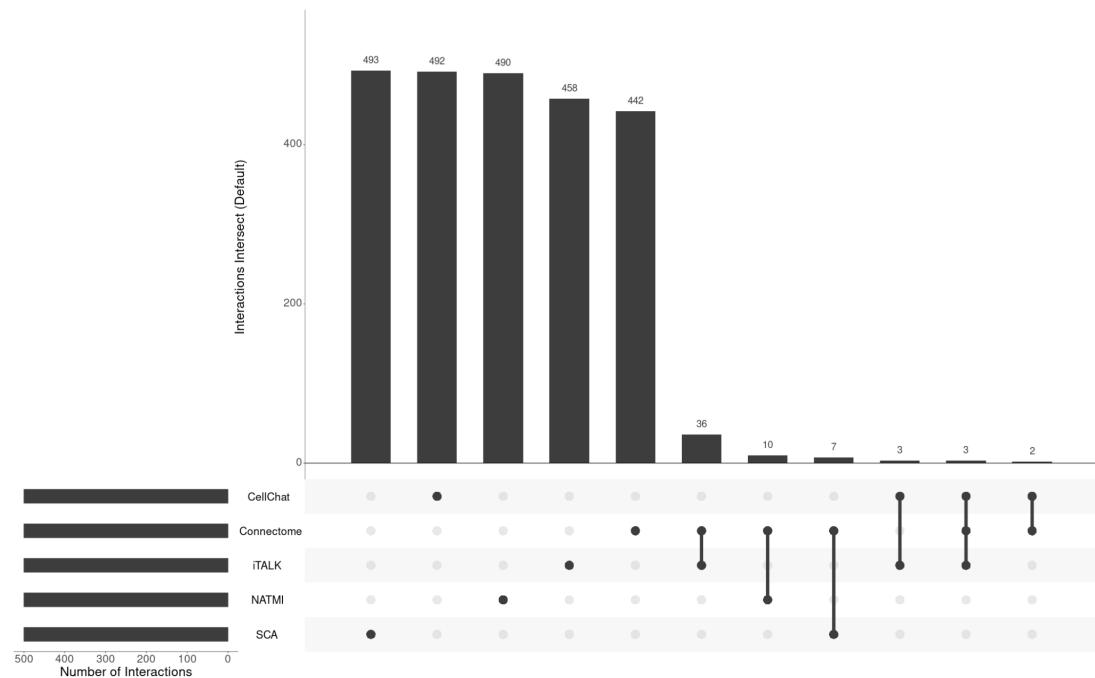
C) *Each method with the Reshuffled Control resource*
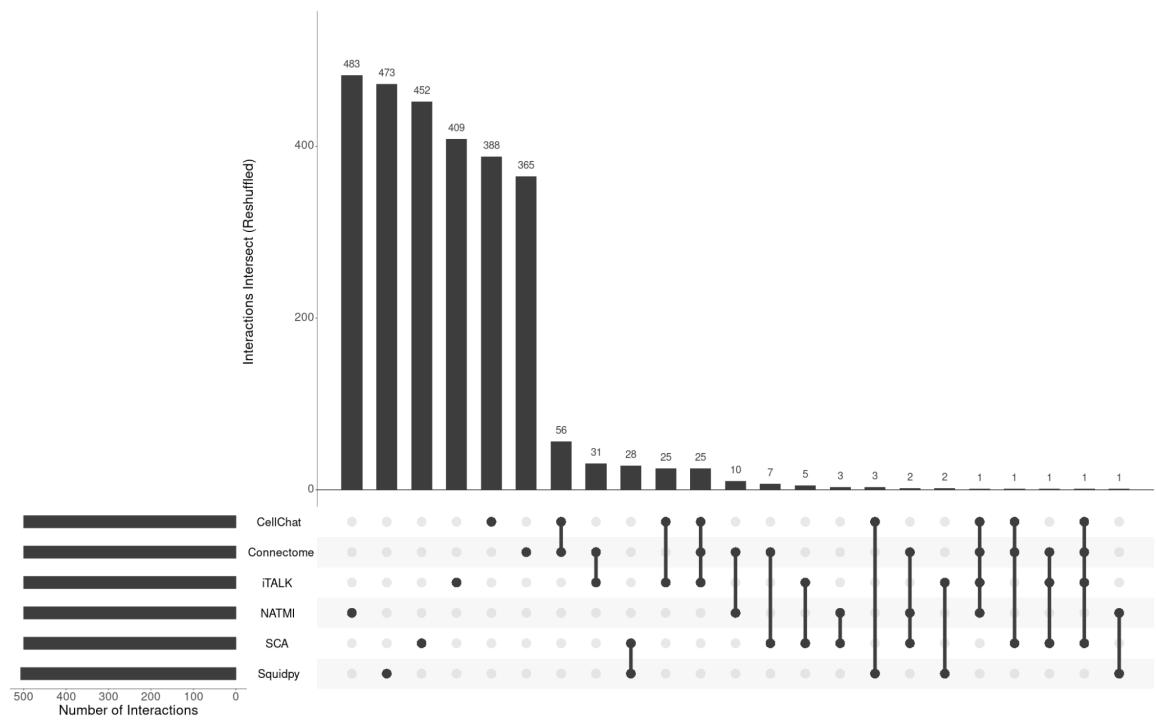
D) *Each method with the Baccin2019 resource*
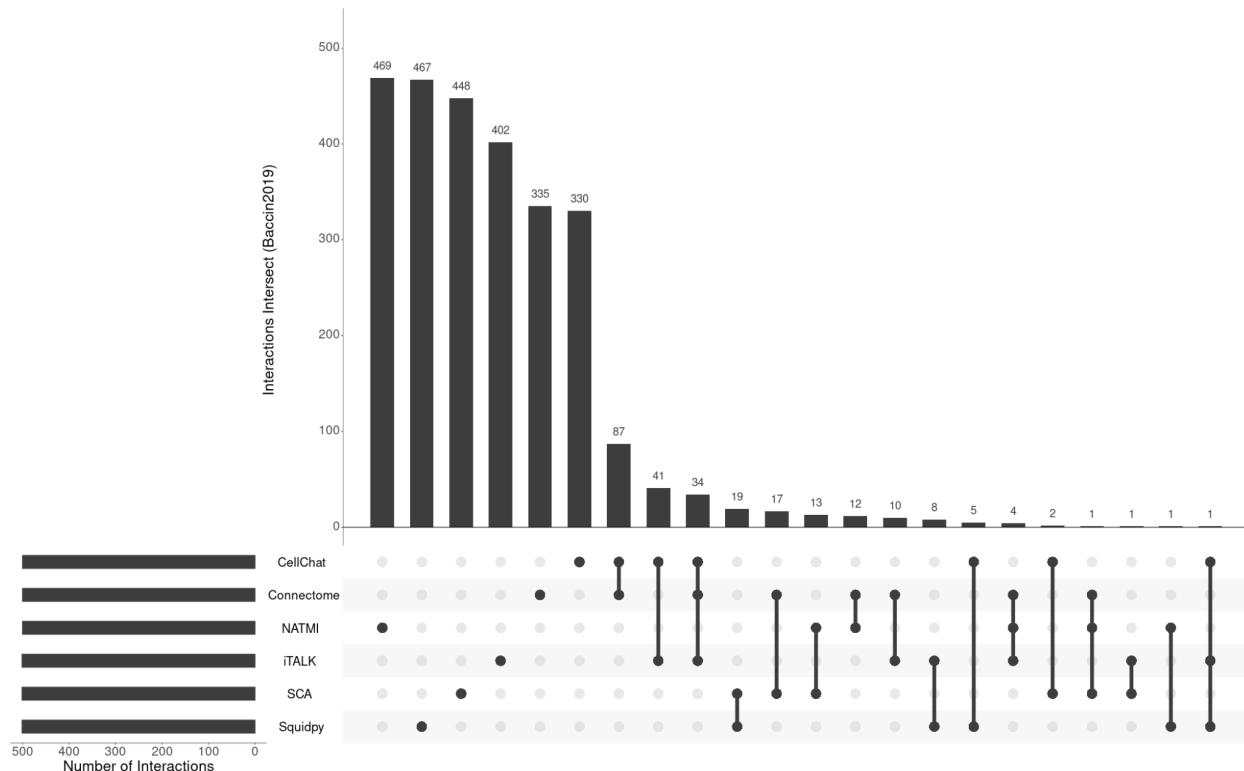
E) *Each method with the CellChatDB resource*
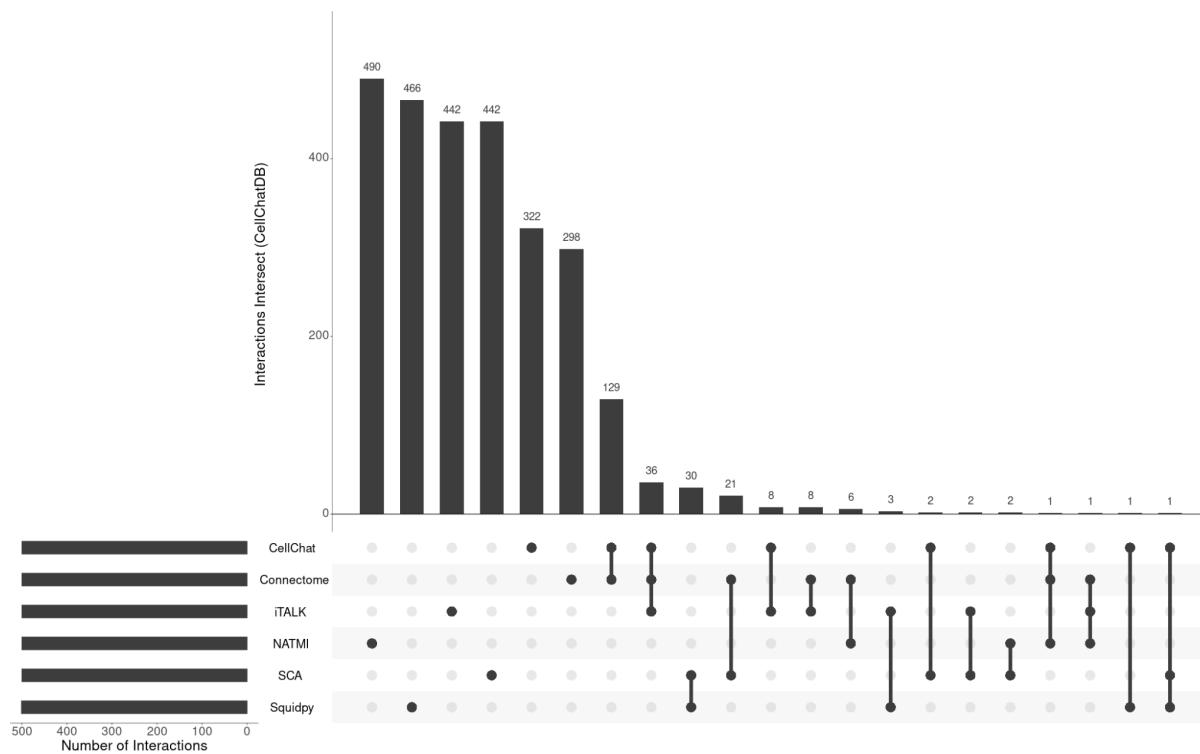
F) *Each method with the CellPhoneDB resource*

59
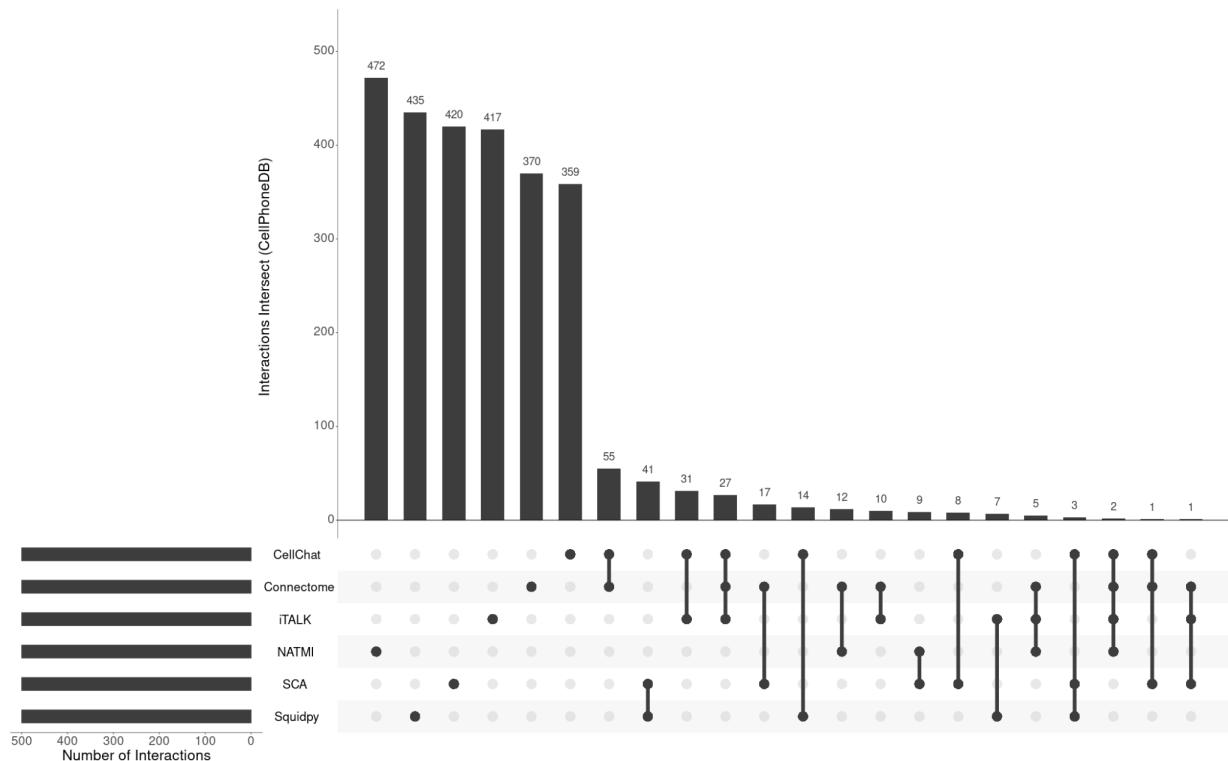
G) *Each method with the CellTalkDB resource*
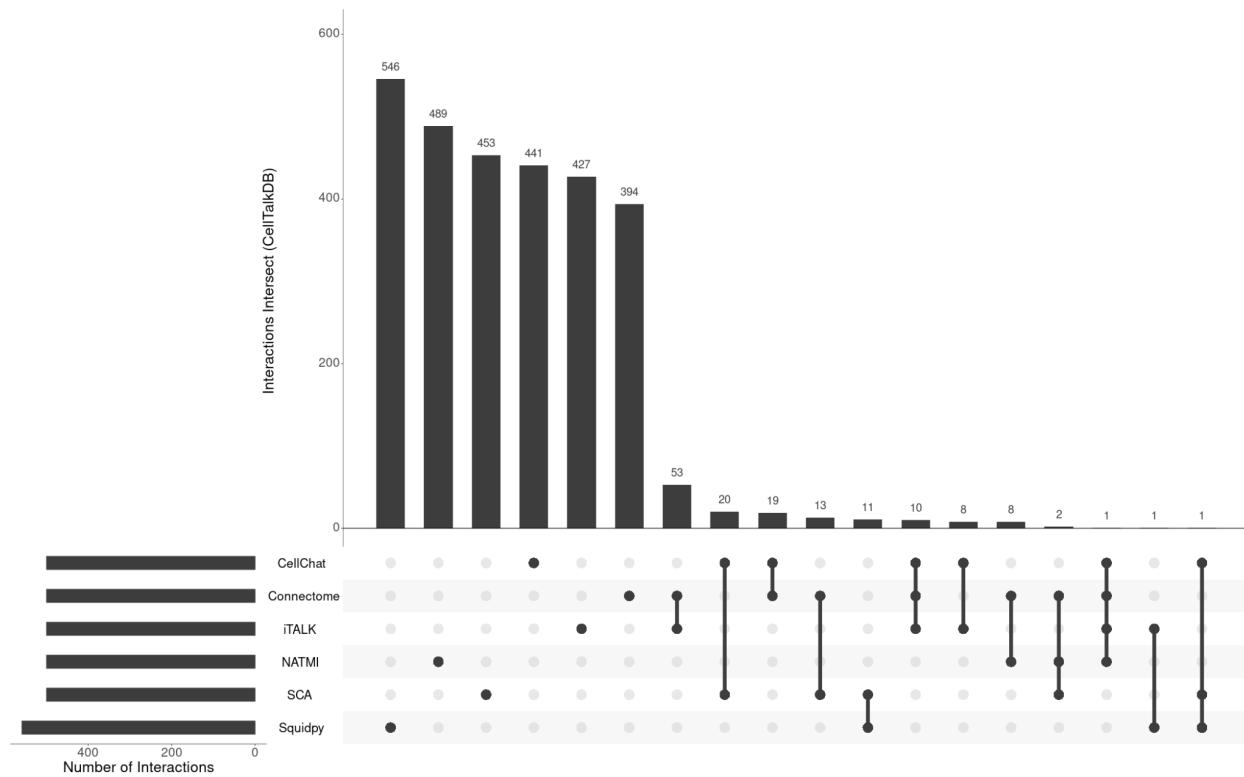
H) *Each method with the ConnectomeDB2020 resource*


60
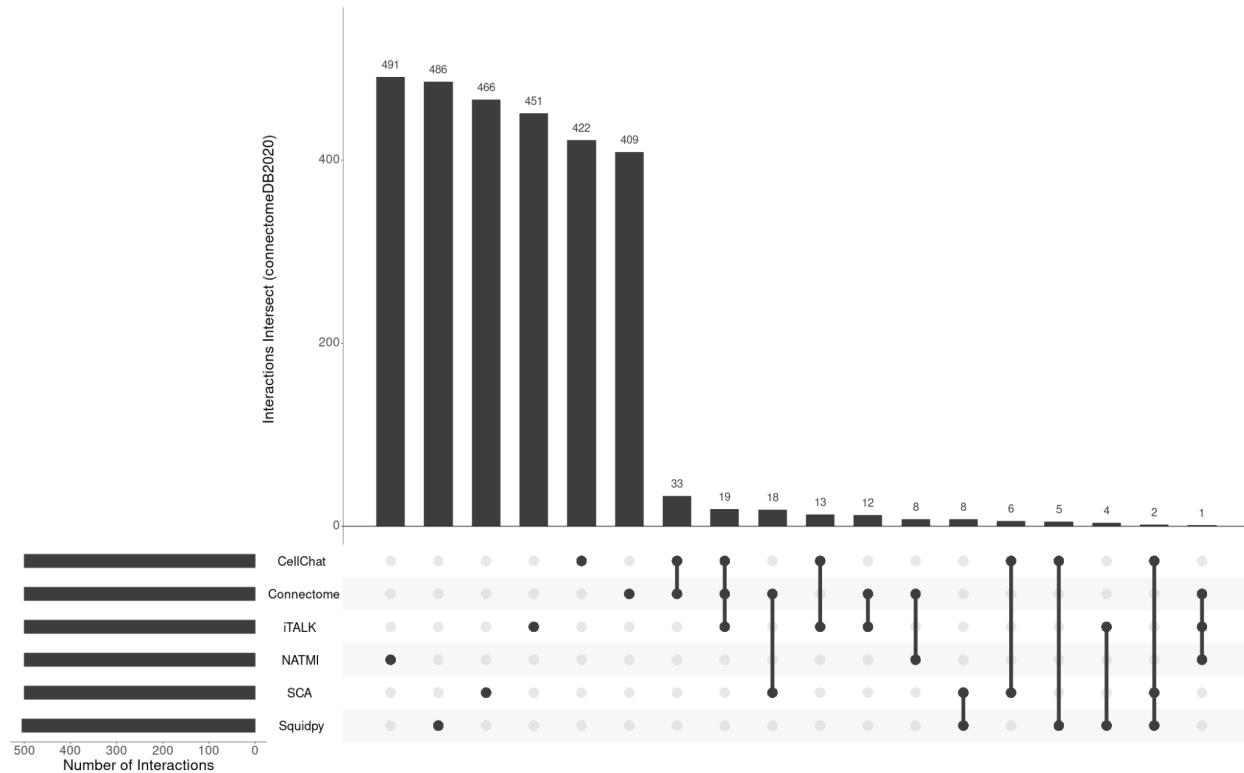
I) *Each method with the EMBRACE resource*
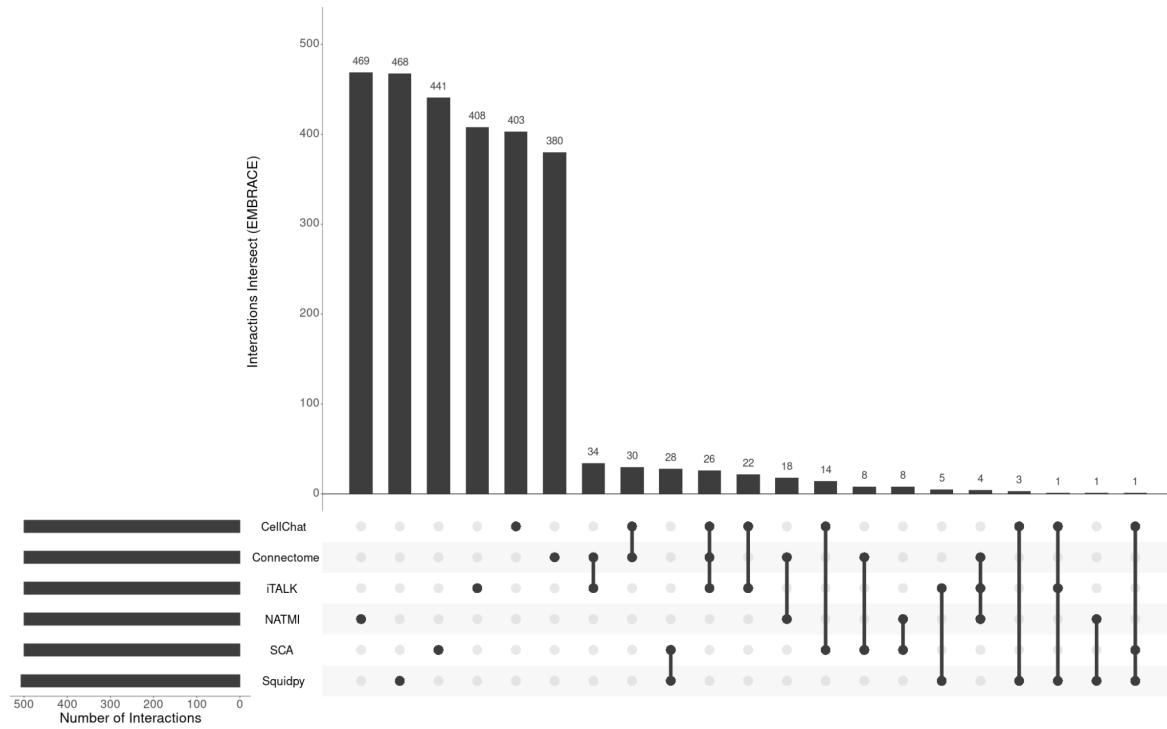
**J)** *Each method with the Guide to Pharmacology resource*

61
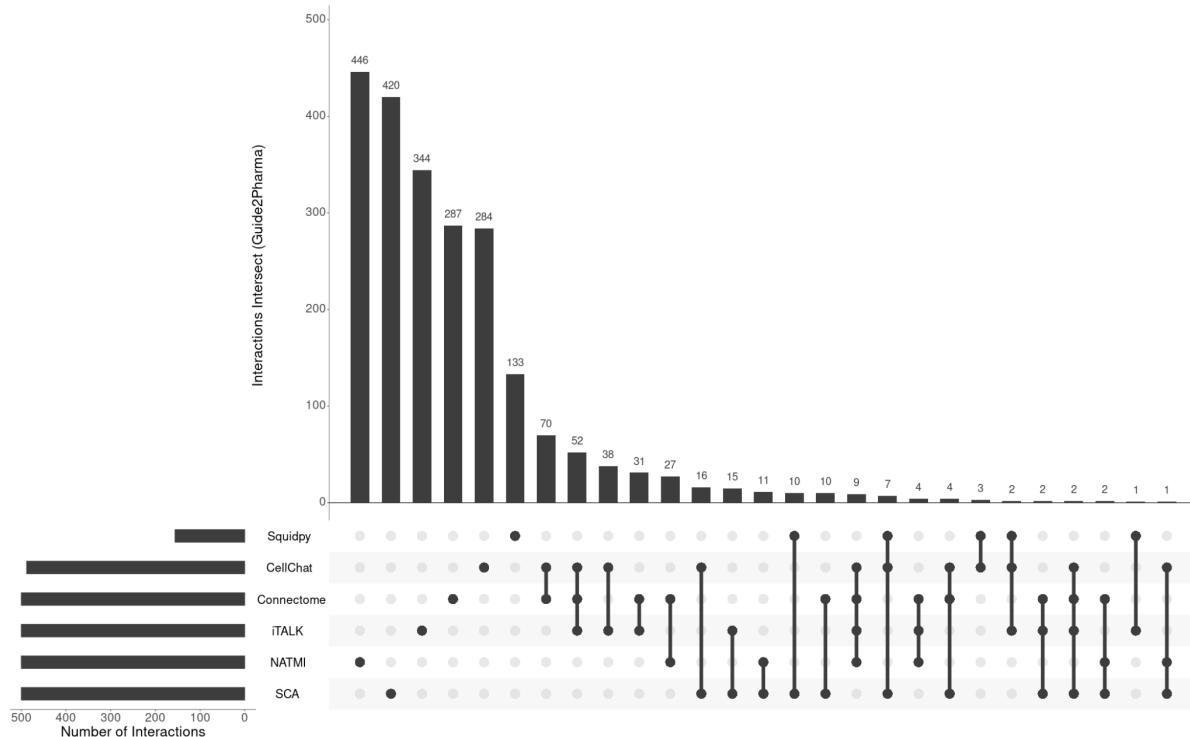
K) *Each method with the HPMR resource*
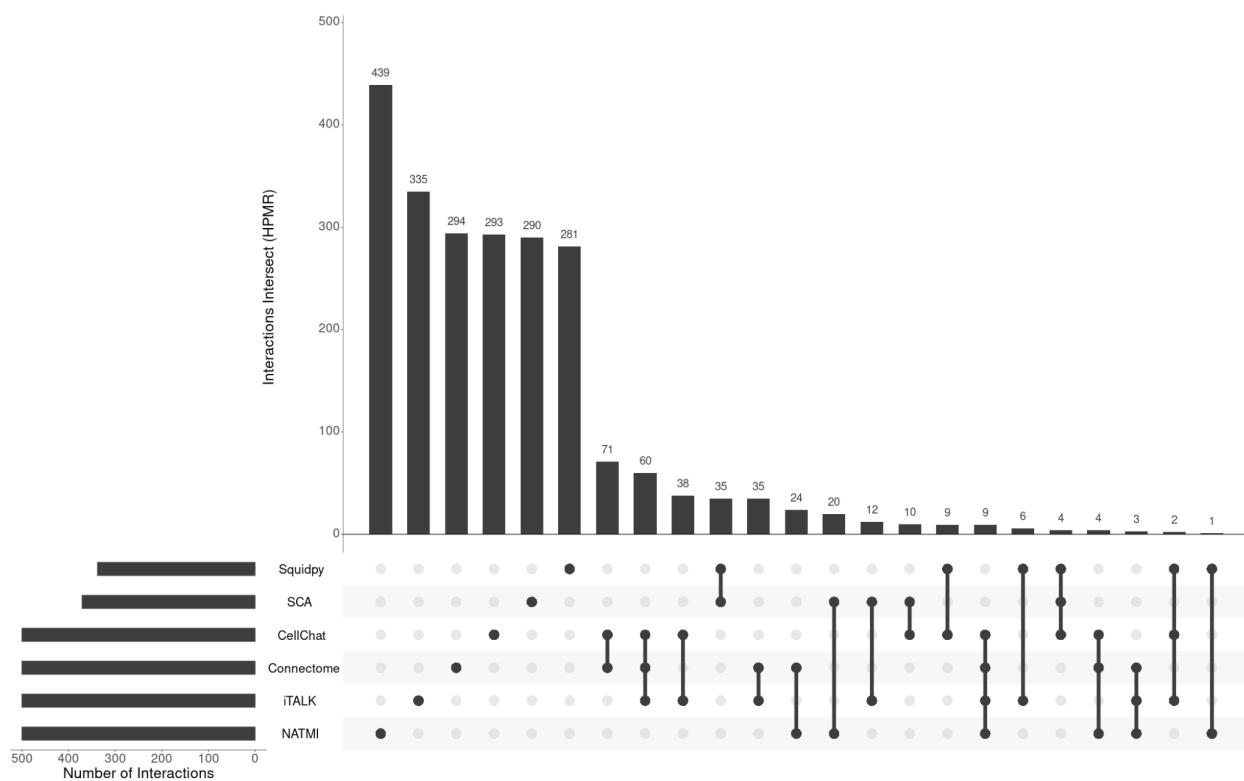

L) *Each method with the ICELLNET resource*
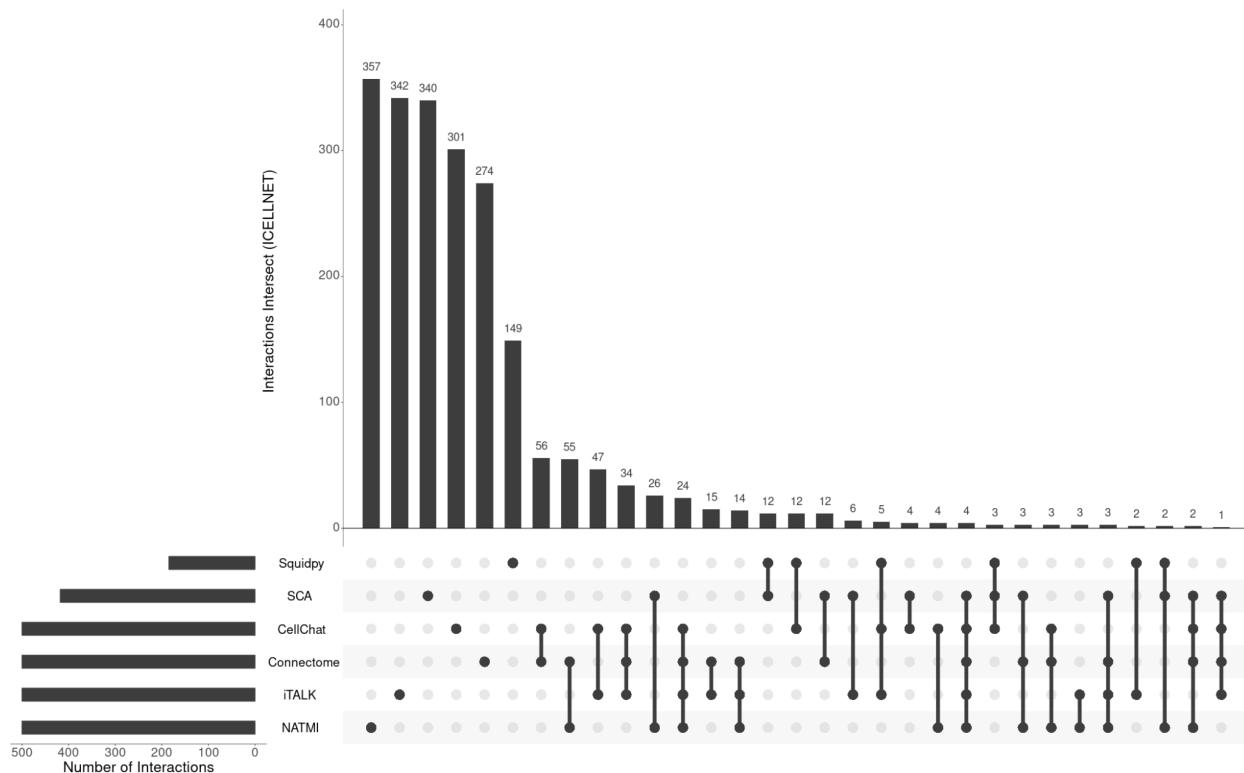
M) *Each method with the iTALK resource*
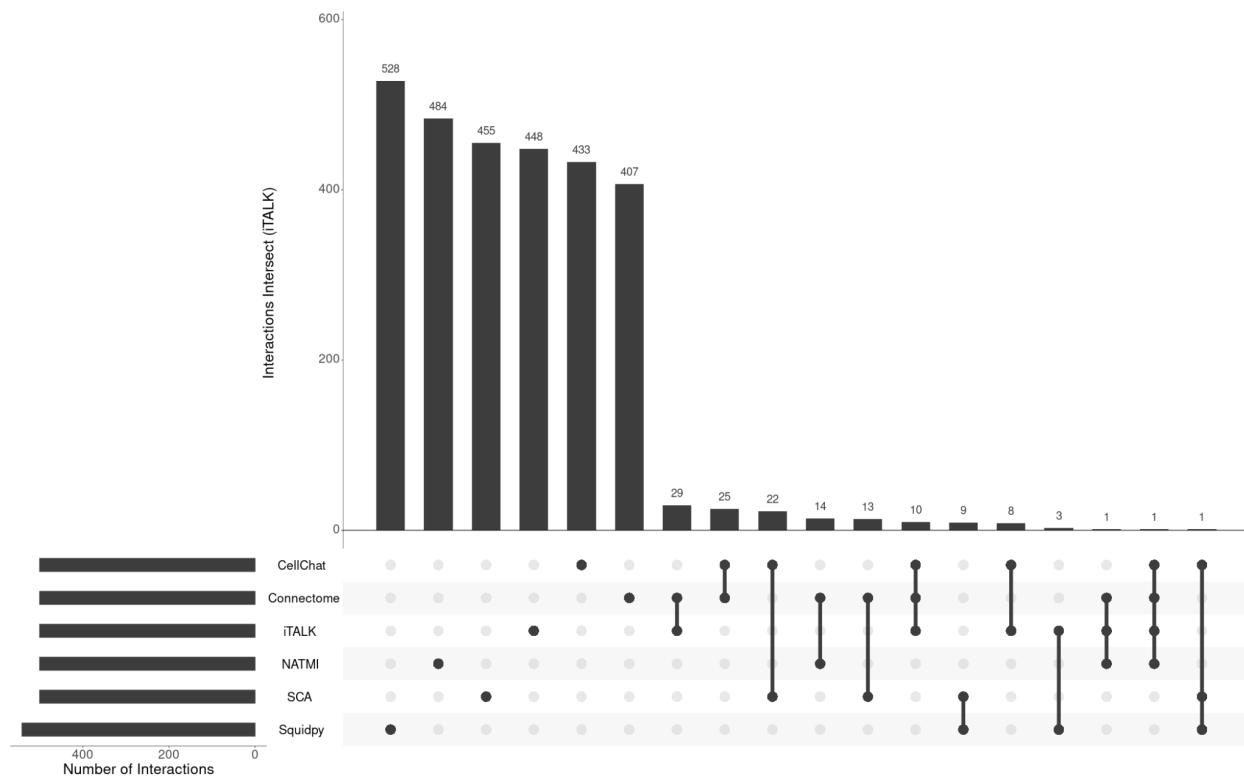
N) *Each method with the Kirouac2010 resource*


63
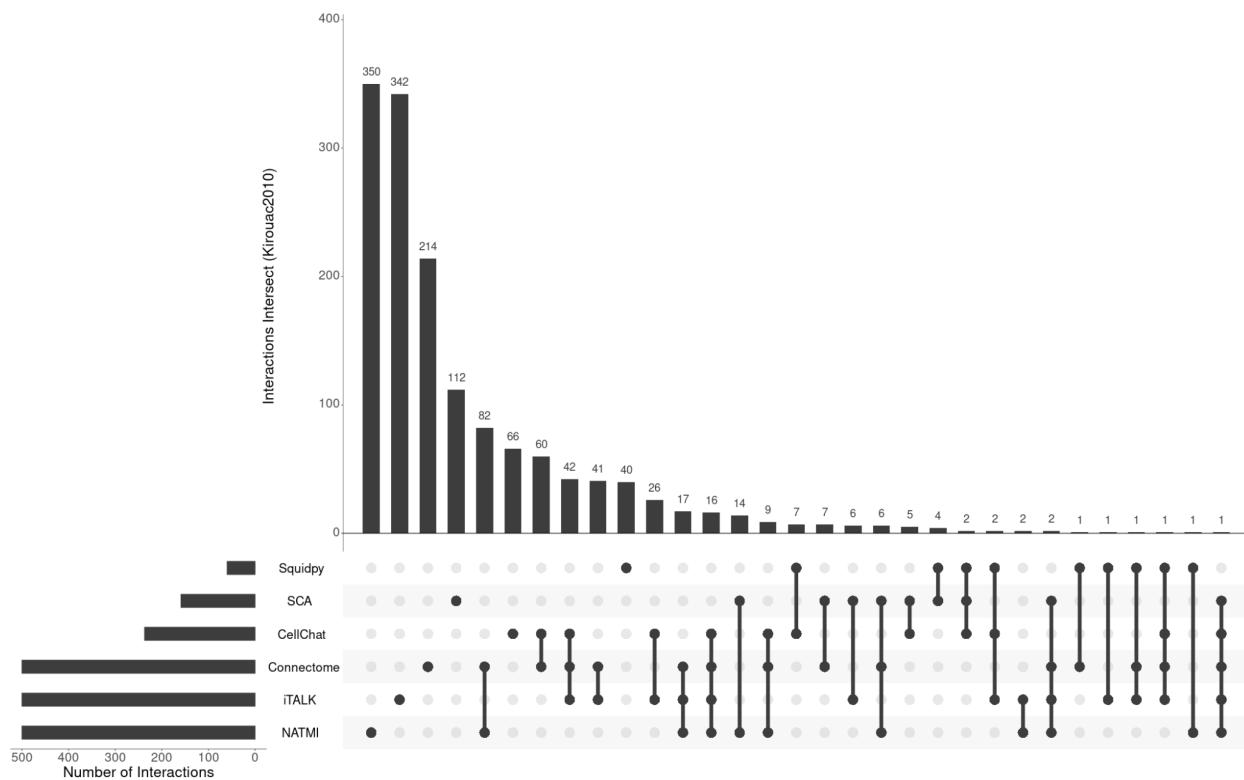
**O)** *Each method with the LRdb resource*
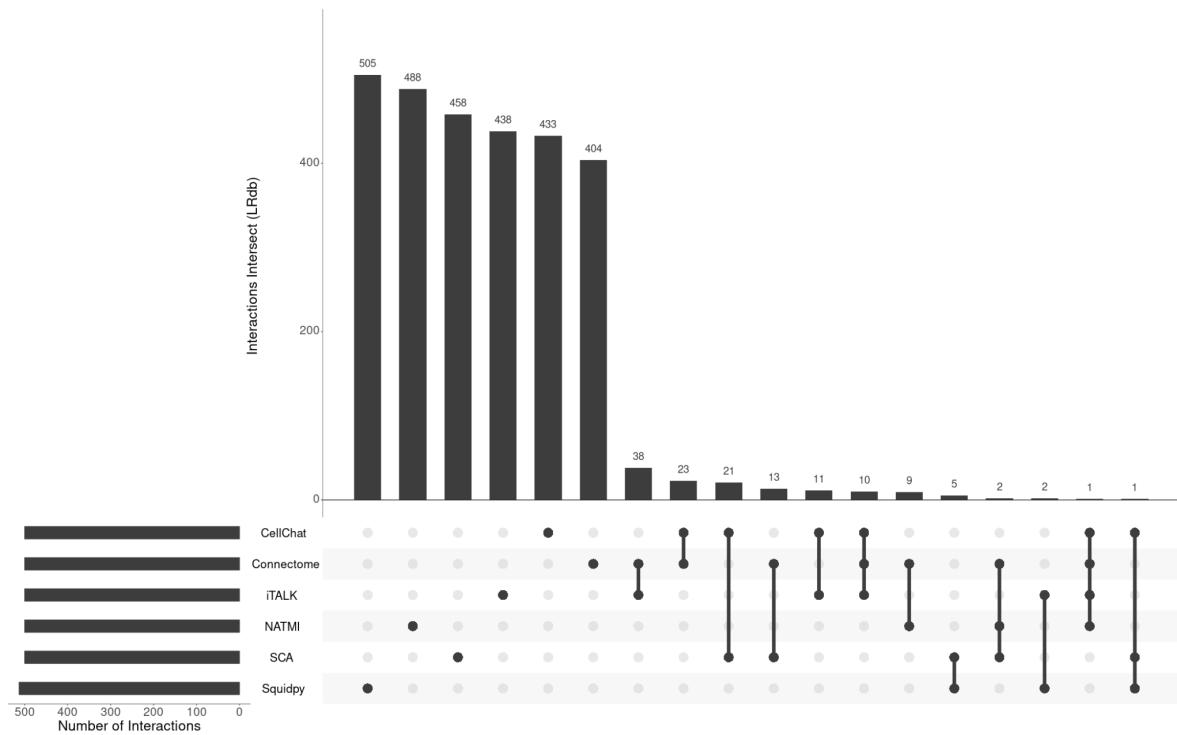
P) *Each method with the Ramilowski resource*

64
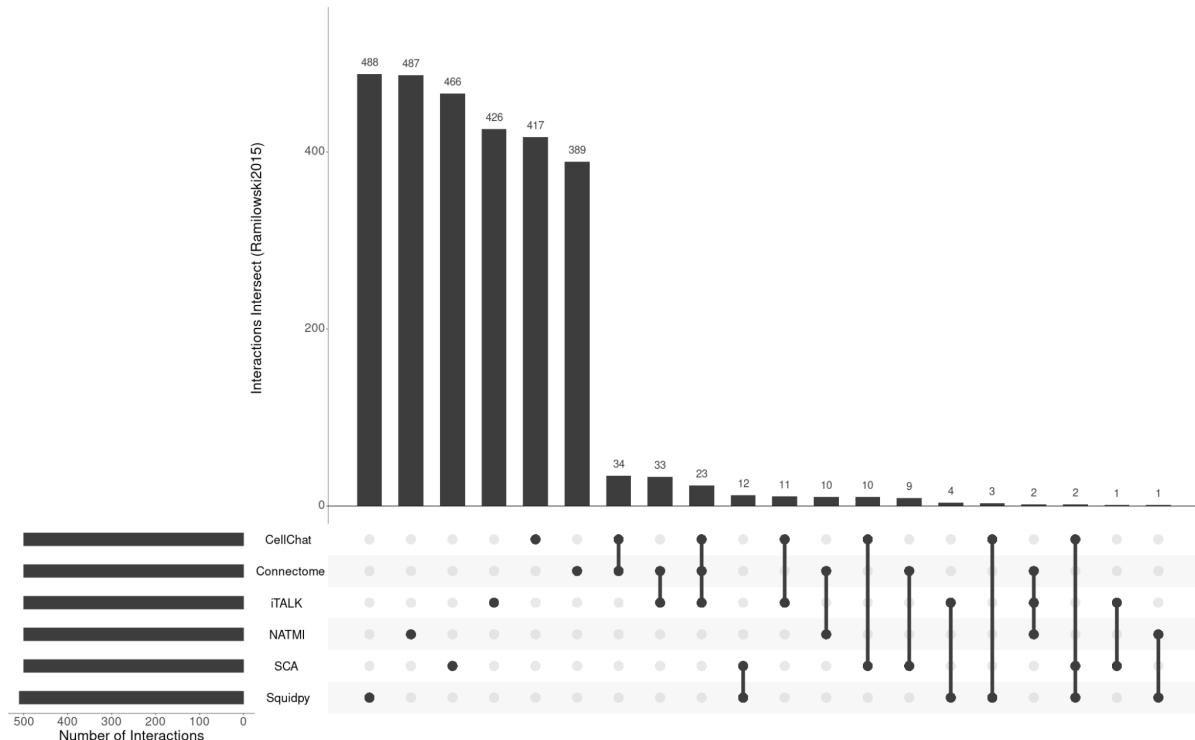
**Q)** *Each method with the talkr resource*
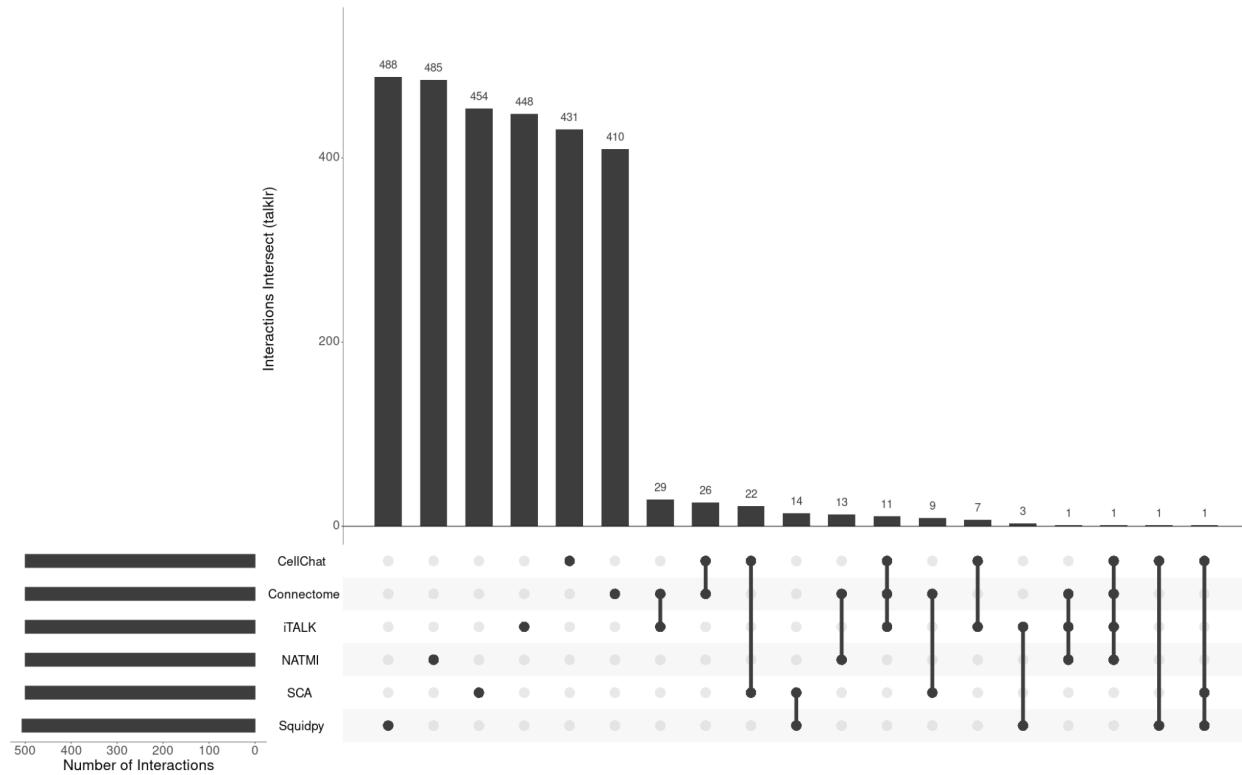
Supplementary Figure S15. *PCA of normalized average interaction rank frequencies per cell pair*
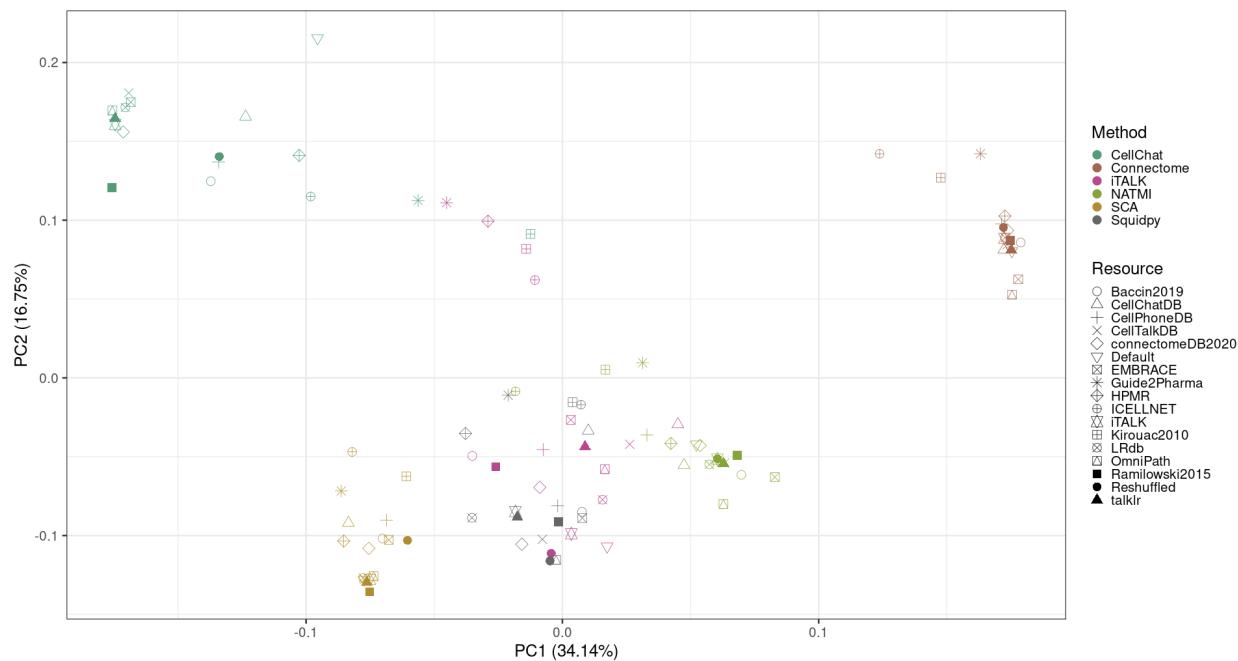
Supplementary Figure S16. *Cluster-unspecific communication agreement.* **A)** *Overlap in 500 highest ranked interactions between different combinations of methods and resources.* **B)** *Similarity among the highest ranked interactions for each method-resource combination, as measured by Jaccard index.* **C)** *Activity per Cell type, inferred as the proportion of interaction edges that stem from Source Cell types or lead to Target Cell types in the highest ranked interactions.*

A)

66
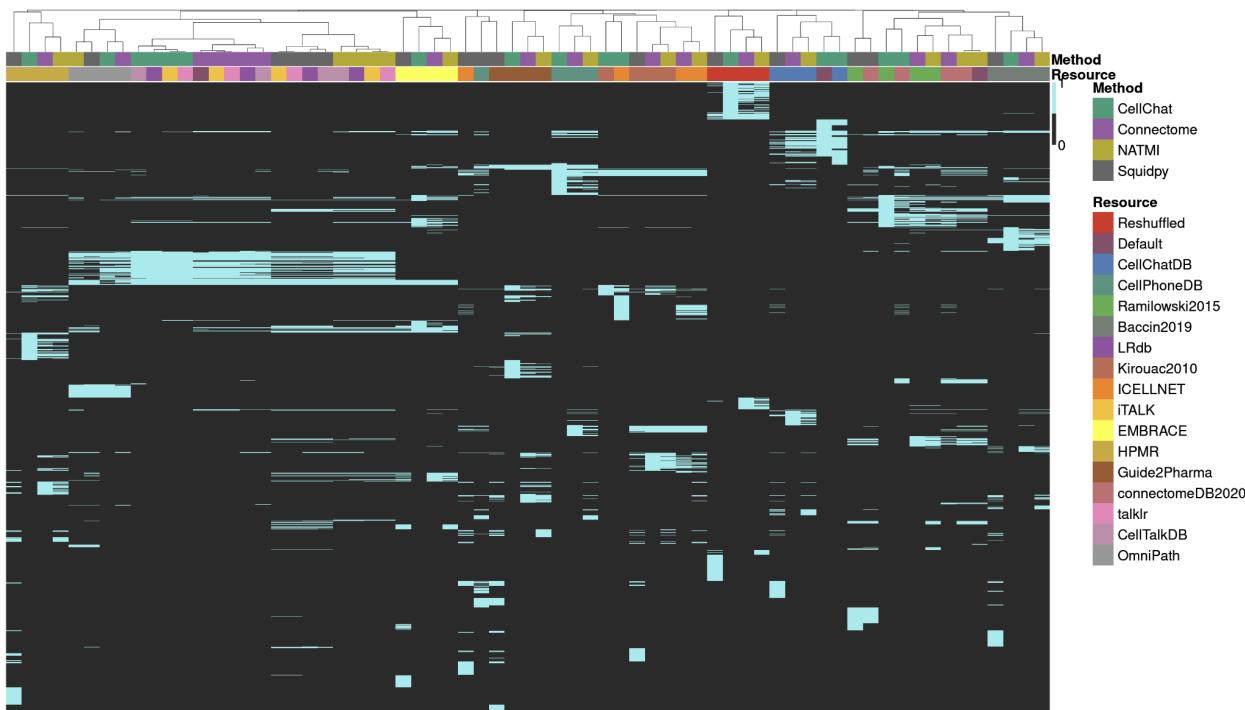
B)
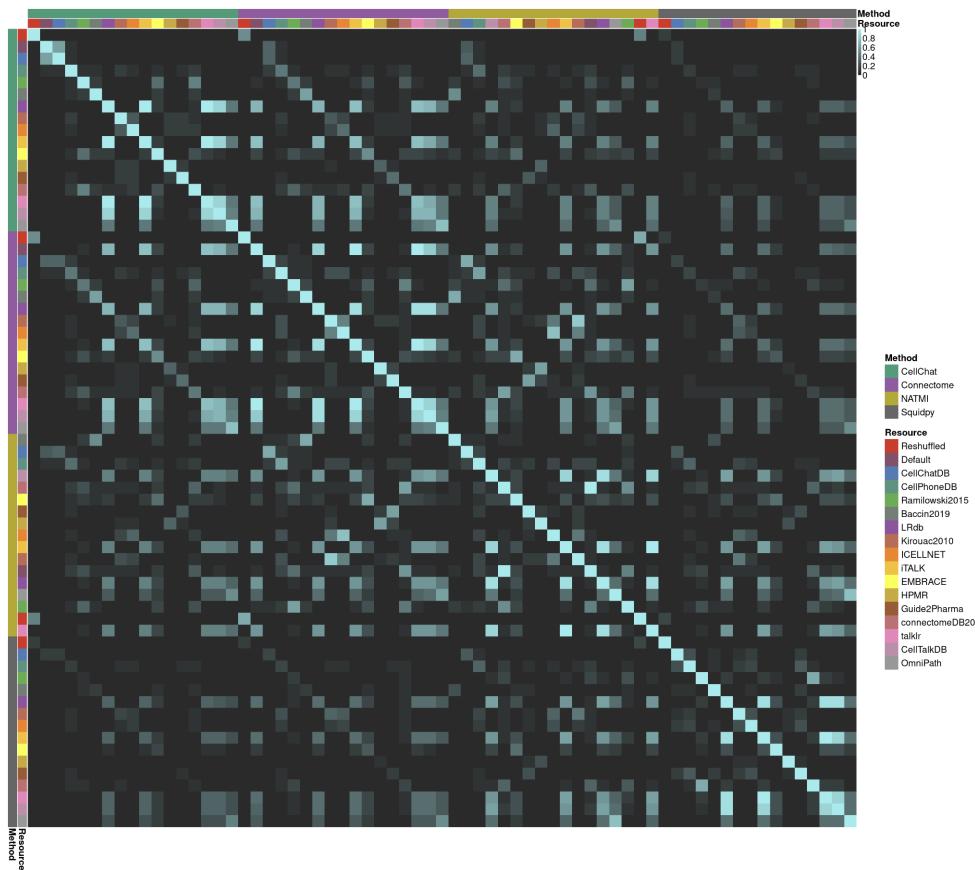
C)
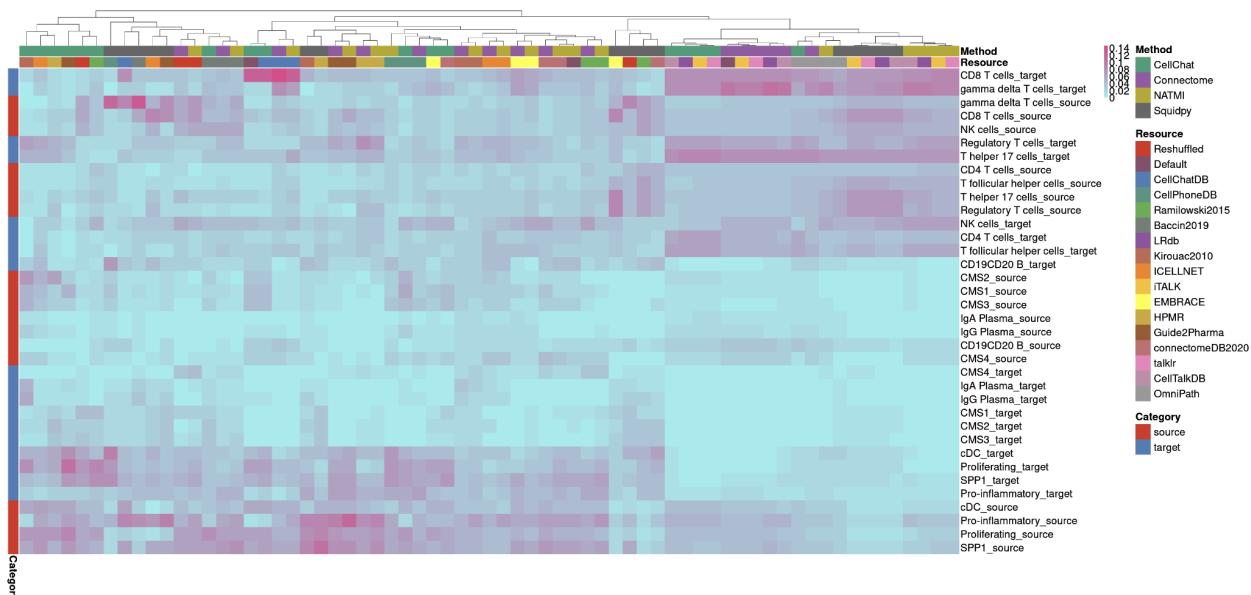
Supplementary table 1. Unique and shared Transmitters, Receivers, and interactions in each resource. We defined unique and shared interactions, receivers and transmitters between the CCC resources if they could be found in only one or at least two of the resources, respectively.

Resource	Transmitters	Receivers	Interactions
Baccin2019	10.29%	7.98%	10.52%
CellChatDB	11.73%	17.17%	50.29%
CellPhoneDB	5.55%	14.31%	15.27%
CellTalkDB	3.70%	6.93%	6.28%
ConnDB2020	7.41%	6.04%	9.14%
EMBRACE	1.16%	0.00%	3.42%
GuidePharm	0.69%	0.41%	5.89%
HPMR	10.56%	4.67%	15.84%
ICELLNET	1.38%	2.53%	3.64%
iTALK	0.00%	0.00%	0.08%
Kirouac2010	2.84%	0.00%	9.33%
LRdb	0.88%	0.00%	1.61%
OmniPath	16.28%	4.29%	45.70%
Ramilowski	0.00%	0.00%	0.00%
talklr	0.00%	0.00%	0.00%
Total	5.30%	4.57%	16.77%

Supplementary table 2. *Description of existing resources for measuring cell−cell communication.*

Resource	Further curation	Sources	Interactions*
Baccin2019(a)[22]	Murine identifiers (only), Multimeric complexes	Ramilowski2015, KEGG reactome database, Literature	1978 (1418)
CellChatDB[13]	Multimeric complexes, 229 signaling pathway families, agonists and antagonists, co-receptors, localisations, Murine identifiers	KEGG, Literature	2551 (2551)
CellPhoneDB[7,25,26]	Multimeric complexes, intercellular communication roles	Guide2Pharma, I2D, IMEx, InnateDB, IntAct, MatrixDB, MINT, UniProt, Literature	1397 (1312)
CellTalkDB[20]	Murine identifiers	STRING, Literature	3398 (3390)
ConnectomeDB2020[10]	-	Ramilowski2015, CellphoneDB, Baccin2019, LRdb, ICELLNET, Literature	2293 (2264)
EMBRACE(a)[61]	Murine identifiers	Ramilowski2015	1710 (1489)
Guide2Pharma(b)[31]	-	Literature	740 (662)
HPMR[37]	-	PubMed, GenBank	527 (461)
ICELLNET[21]	Multimeric complexes, Signalling families, Cytokine-focus	STRING, Ingenuity, BioGRID, Reactome, CellPhoneDB	380 (371)
iTALK (c) [5]	Ligand categories	Ramilowski2015, HPMR, IUPHAR-DB, Graeber2001, Griffith2014, Cameron2003, Zhou2017, Auslander2018	2648 (2565)
Kirouac2010[38]	-	Literature, COPE	270 (150)
LRdb[11]	-	cellsignal.com, Ramilowski2015, Guide2Pharma, HPMR, HPRD, Reactome, UniProt, Literature	3251 (3226)
Ramilowski2015[32]	-	DLRP, HPMR, IUPHAR, HPRD, STRING, Literature	1894 (1888)
talklr[36]	-	-	2422 (2411)
OmniPath#[2]	Combines data from more than 100 resources and contains protein-protein and gene regulatory interactions, enzyme-PTM relationships, multimeric complexes, protein	Composite resource combining all of the CCC dedicated resources listed here, along with some additional interactions.	6103

	annotations, and a CCC-dedicated composite DB		

All the resources above were retrieved from the OmniPath database (https://omnipathdb.org/). We also refer to the composite CCC resource presented here as OmniPath. The OmniPath presented in the analyses we filtered according to: **i)** we only retained interactions with literature references, **ii)** we kept interactions only where the receiver protein was plasma membrane transmembrane or peripheral according to at least 30% of the localisation annotations, and **iii)** we only considered interactions between single proteins

* The number of original interactions for each CCC dedicated resource covered in OmniPath is shown in brackets.

(a) Translated from murine identifiers to human, which accounts for the lower number of obtained interactions.

(b) Kept only the unique human-annotated interactions between transmitter and receiver proteins.

(c) Duplicates present in the original resource were excluded when imported via OmniPath

70

Supplementary table 3. *Formatted Korean CRC data set cell type counts and full names.*

Cell type	Cell subtype	Complete Name	Cell Count
B cells	CD19CD20 B	B cells_CD19+CD20+	2,049
T cells	CD4 T cells	CD4+ T cells	3,980
T cells	CD8 T cells	CD8+ T cells	4,647
Myeloids	cDC	Conventional Dendritic Cells	353
Epithelial cells	CMS1	Consensus Molecular Subtype 1	1,201
Epithelial cells	CMS2	Consensus Molecular Subtype 2	10,771
Epithelial cells	CMS3	Consensus Molecular Subtype 3	5,486
Epithelial cells	CMS4	Consensus Molecular Subtype 4	11
T cells	gamma delta T cells	$\gamma\delta$ T cells	219
B cells	IgA Plasma	IgA+ Plasma Cells	180
B cells	IgG Plasma	IgG+ Plasma	1,661
T cells	NK cells	Natural Killer Cells	948
Myeloids	Pro-inflammatory	Pro-inflammatory Macrophages	2,325
Myeloids	Proliferating	Proliferating Macrophages	165
T cells	Regulatory T cells	Regulatory T cells	2,943
Myeloids	SPP1	SPP1+ Macrophages	3,096
T cells	T follicular helper cells	T follicular helper cells	548
T cells	T helper 17 cells	T helper 17 cells	1,961

Supplementary Note 1. *Protein localisation to categorize CCC*

To estimate the localisation distributions of transmitters and receivers as well as to categorize CCC interactions according to signalling categories we obtained protein subcellular localisation annotations via OmniPath[2]. These annotations were gathered from sources such as UniProt, the Cell Surface Protein Atlas[62], and Membranome[63]. The localisation annotations were then filtered according to a consensus threshold (4.1. Descriptive analysis of resources). We then used the localisations of transmitters and receivers to approximate the categories of interactions. The largest part of interactions were those between secreted proteins targeting transmembrane proteins (S -> T), which we referred to as the secreted signalling category in Figure 3. Further, we attributed interactions between and within the transmembrane and peripheral plasma membrane proteins (T -> T, P -> T, T -> P, P -> P) to intercellular signalling events that require physical contact between cells.

As a consequence of the protein localisation annotation process, some annotations were expected to be unsuitable in the context of CCC signalling. For example, a ligand can be annotated as both secreted and membrane-bound, depending on the context of the observation. This was the case for EFNA1, a membrane-bound ligand which binds to the EPH receptors, also observed to be released as a soluble protein in breast adenocarcinoma cells[64]. Moreover, splicing variants of the same protein can have different subcellular localisations, with one variant being membrane-bound and the other secreted. For instance, FGF17 binds to membrane-bound FGFR2, but since FGFR2 also has secreted isoforms (UniProtKB - P21802), this interaction can be mislabeled as an interaction between two secreted proteins. These misannotations in the context of CCC interactions made up only a small proportion of all annotations and were grouped into the "Other" category (Figure 3), which represented interactions which did not fit as secreted or direct-contact signalling (T->S; S->S; P->S).

Supplementary Note 2. *Protein Complexes*

We further assessed the predicted proportions of interactions containing complexes from CellChat and Squidpy using Baccin, CellChatDB, CellPhoneDB, and ICELLNET resources. This analysis showed that the proportion of complexes among the highest ranked 500 hits for CellChat ranged from 1.8% (with ICELLNET) to 23.0% (with the original CellChatDB) and that of Squidpy ranged from 9.7% with ICELLNET) to 38.1% (with CellChatDB).

73

Supplementary Note 3. *Cluster Specificity and Method Dissimilarity*

As a consequence of the disagreement between methods in regards to the most actively communicating cell types, we reasoned that a possible cause was the different approaches used to assign cell cluster specificity to the interactions. To this end, we conducted the same analyses presented in the main text, but instead using the measures from each method which do not explicitly reflect the cell-type specific communication (i.e. Squidpy means; unfiltered CellChat probabilities; Connectome.weight.norm; NATMI.edge.avg.expr) (Table 1). Since SingleCellSignalR and iTALK provide a single scoring system, they were excluded from this analysis. We observed an increase in the agreement between methods (Supp. Figure 16A), as the mean Jaccard index when using the same resource with different methods ranged from 0.277 to 0.618 (mean = 0.404) (Supp. Figure 16B). The overlap between these methods when using the same resource was hence considerably higher than that observed when using cluster-specific measures (mean = 0.0247). On the contrary, the mean Jaccard index per method when using the same resource remained relatively unchanged when compared to the scoring systems that reflect cell cluster specific communication (0.118 for non-specific measures versus 0.167 for specific). Moreover, analogously to the agreement analysis, we used the cluster-unspecific measures to estimate the active cell types. As a result, methods were observed to largely agree in terms of the most active cell types (Supp. Figure 16C). Thus, this analysis suggests that the distinct approaches used to assign cell cluster specificity to the interactions explain some of the disagreement between methods for our dataset. Furthermore, when using the cluster-unspecific measures, the differences in resources were the main source of dissimilarity between the results.

Supplementary Note 4. *Single-cell CCC Benchmark Directions*

As a consequence of the observed disagreement in the results obtained when using different methods and resources, we argue that an appropriate benchmark is paramount for the future development of the CCC inference field. Some effort has already been directed to the assessment of different methods and specific directions were already proposed[19]. To this end, we also share our current benchmark ideas.

I.   Associations between CCC activity and Spatial-Adjacency

Assumptions: Cell clusters that are spatially adjacent should be communicating more actively than those that are spatially distant; Confining CCC inference to spatial adjacency should reduce false positives.

Limitations: Difficult to distinguish cell-cell communication and cellular program coregulation.

Examples: This approach was already used as a way to validate some methods[13,18], while other methods explicitly take spatial information into account for CCC inference[19]. Another example is confining CCC inference to cells that are expected to be in close contact, e.g. according to co-localizing cells in visium spots to reduce false positive interactions[25]. In a similar way, 10x Visium data can be used to identify cell types that are known to be co-located in visium spots, and are hence in close contact.

Metrics:

A benchmark focused on the relationship between Cell Pair Activity[*1] and Cell Distance[*2] composed by three main steps:

(1) Cell Pair Activity reported by different methods (***1**);

(2) Cell-cell Spatial Distance or Colocalization (***2**);

(3) Relationship between CCC method output and distance (***3**).

*1. Number of Inferred interactions between Cell Clusters; Average Cell Inference Ranks per Communicating Cell Types

*2. Physical distance, measured by Euclidean distance between the closest cell types, was already reported to be an appropriate proxy of cell pair communication activity[65]. Other measures can be the neighbourhood enrichment or spatial co-occurrence of cells[24,66]. An alternative approach would be to discretise distance according to e.g. spatially-adjacent and spatially-distant cell types[13,18].

*3. Correlation or Regression Coefficients, or any other measure used as a proxy of the relationship between the two variables.

II. **Data-driven Inference of Spatial Covariance to explain Transmitter-Receiver interactions**

Assumptions: Receiver and transmitter gene expression covariance with spatial distance is a proxy of CCC events.

Limitations: Difficult to distinguish cell-cell communication and cellular program coregulation; Possibly biased towards CCC events in which the transmitter and receiver regulate each other's expression.

Approach:
1) The expression of a receiver is spatially explainable by the expression of a transmitter and vice versa. Thus, a threshold signifying conserved spatial gene regulation between transmitters and receivers (with e.g. mistyR[67]) can be used to define putative true positive interactions.

2) Downstream signalling models to explain transmitter and receiver activity.

Some tools already utilize downstream signalling as an attempt to better model CCC interactions[15,16,18]. In a similar way prior knowledge of downstream transmitter and receiver activity models[15], can be used to spatially explain the activity of transmitter and receiver. Alternatively, one can build naïve protein-protein interaction models from existing databases[2].

Metrics:

Methods' (and Resources') coverage of the spatially explainable CCC events. In other words, we expect a method to assign preferentially high ranks to spatially explainable transmitter-receiver interactions.

AUROC can be calculated according to different thresholds of spatial covariance for transmitter/receiver genes involved in CCC interactions. Thus, a reliable tool and resource should be able to pick up spatial covariance better than a randomized resource, and better than a resource composed of genes that are not explainable by space (e.g. housekeeping genes).