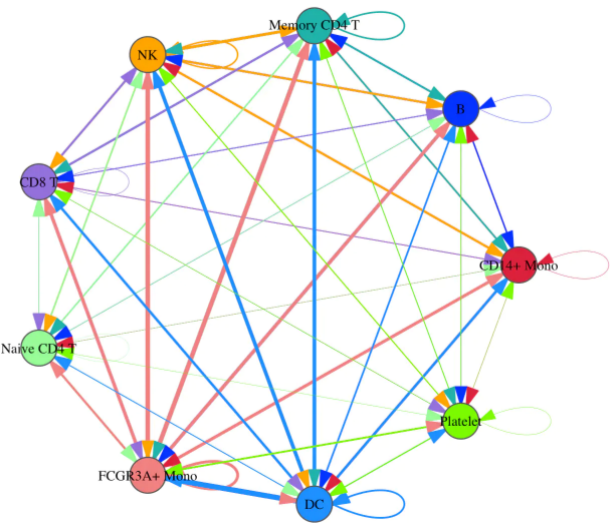


细胞通讯结果文件说明

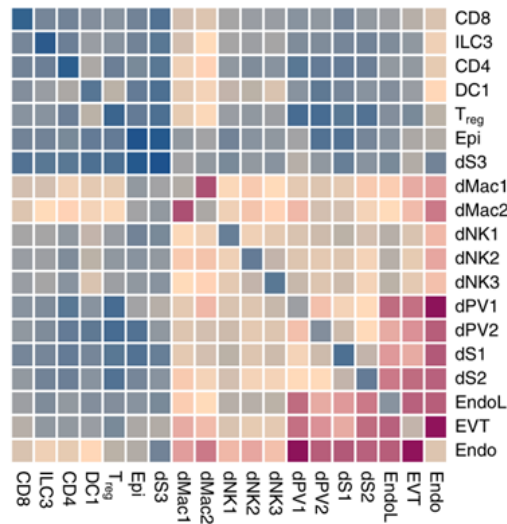
- 细胞通讯结果文件说明
- 结果目录结构
- 背景介绍
- 意义
- 分析原理

结果目录结构

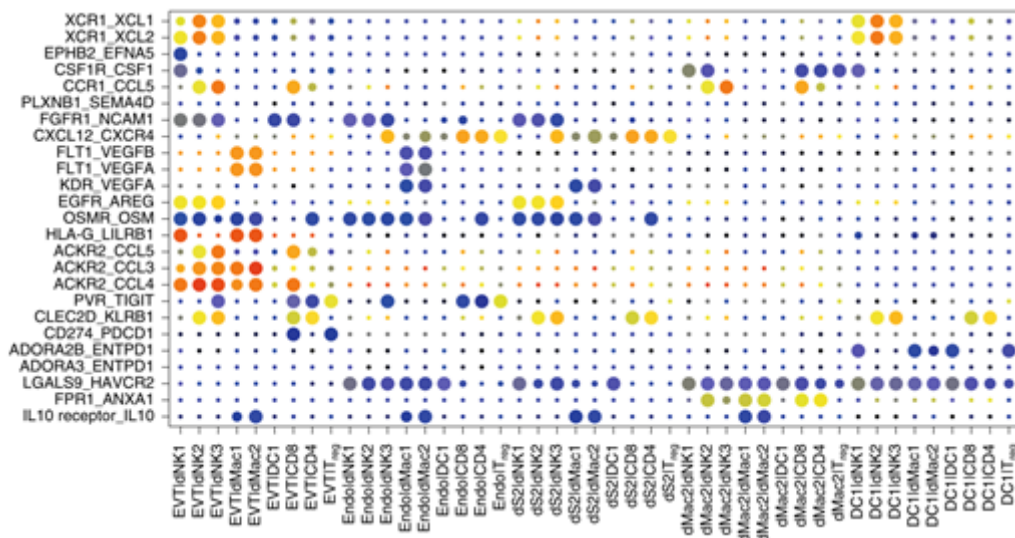
```
|--gene_count 单细胞表达矩阵输入文件每行为genesymbol号，代表基因名称；每列为barcode序列，代表不同细胞。  
|--meta_data barcode的注释信息表  
|--out 输出结果文件  
|----cell_talk.pdf 细胞间受体-配体互作网络图PDF格式  
|----cell_talk.png 细胞间受体-配体互作网络图PNG格式  
|----heatmap_log_count.pdf 细胞间受体-配体互作热图PDF格式  
|----heatmap_log_count.png 细胞间受体-配体互作热图PNG格式  
|----plot.PDF 细胞间受体-配体互作气泡图 PDF格式  
|----count_network.xls: 网络文件，source和target代表点分别代表网络的起点和终点，count代表细胞间互作的次数  
|----heatmap_count.PDF: 细胞间受体-配体互作热图 PDF格式
```



图：细胞间受体-配体互作网络图，显示不同的细胞类型之间检测到配体-受体相互作用的数量。图中分别标记了正向（从信号细胞到目标细胞）和反向信号，同时测量了每个细胞类型内的自分泌信号。网络图的节点按照细胞类型进行颜色编号，边缘按信号和目标细胞之间的相互作用的数量进行标记和缩放。



图：细胞间受体-配体交互作用热图，展示了不同细胞类型中相关作用的数量。行和列都表示为细胞类型，每格表示细胞间的相关作用的数量，颜色越红表示细胞间的相互作用数目越多，即细胞间通讯关系越强。



图：细胞间受体-配体交互作用气泡图，展示了配体-受体对在不同细胞类型中的表达强度和特异性。行表示具有细胞间通讯关系的配体-受体对，列表示发生细胞通讯的细胞类型，圆圈大小表示显著性水平，圆圈越大表示越显著，圆圈颜色越红表示细胞间通讯关系越强。

- P-value (pvalues.txt), Mean (means.txt), Significant mean (significant_means.txt)文件列名：
 - id_cp_interaction: 存储在数据库中的每个互作用的唯一 CellPhoneDB 标识符。
 - interacting_pair: 互作用对的名称，以“|”分隔。
 - partner A or B: 第一个互作用成员 (A) 或第二个 (B) 的标识符。它可能是：UniProt (前缀 simple:) 或复合体 (前缀 complex:)
 - gene A or B: 第一个相互作用伙伴 (A) 或第二个 (B) 的基因标识符。
 - secreted: 如果其中一个互作用成员是分泌的，则为真。
 - Receptor A or B: 如果第一个互作用成员 (A) 或互作用成员 (B) 在cellphonedb数据库中被注释为受体，则为真。
 - annotation_strategy: CellPhoneDB 互作用数据库来源
 - is_integrin: 如果其中一个互作用成员是整合素，则为真。
 - rank: 每个互作用的显著 p 值总数除以细胞类型-细胞类型比较的数量。(仅在 significant_means.txt 中)
 - means: 所有互作用的平均值：平均值是指相应相互作用的细胞类型对中个体成员平均表达值的总平均值。如果其中一个平均值为 0，则总平均值设置为 0。(仅在 mean.txt 中)
 - p.values: 所有相互作用成员的 p 值：p.value 是指在每个相互作用的细胞类型对中相互作用的配体-受体对的富集。(仅在 pvalues.txt 中)

- significant_mean: 所有互作成员的显著平均值计算。如果 $p.value < 0.05$, 则该值将是平均值。或者, 该值设置为 0。(仅在 Significant_means.txt 中)
- deconvoluted.txt文件列名:
 - gene_name: “means.csv”文件中定义的参与互作的亚基之一基因标识符。
 - uniprot: “means.csv”文件中定义的 UniProt 标识符。
 - is_complex: 是否为复合体的一部分
 - protein_name: “means.csv”文件中定义的亚基的蛋白质名称。
 - complex_name: 是否为复合体, 如果是为复合体名称, 否为空
 - id_cp_interaction: 存储在数据库中的每个互作的唯一 CellPhoneDB 标识符。
 - mean: 每个簇中相应基因的平均表达量。

背景介绍

意义

多细胞生物是由很多不同类型细胞组成的开放而复杂体系, 配体受体复合物介导的细胞间通讯对协调发育、分化和炎症等多种生物学过程至关重要。细胞通讯分析, 又称细胞受体-配体互作分析, 是以细胞亚群的基因表达量数据为研究对象, 通过获得细胞中配体及受体基因的表达信息, 比较细胞类型之间的配体与受体基因表达差异, 分析得到细胞亚群间的信号通讯关系, 在阐明生物学过程中细胞间通讯的复杂性、多样性和动态性方面有重要意义。

分析原理

利用CellPhoneDB软件[1]进行细胞通讯分析, CellPhoneDB开发的细胞通讯配体-受体数据库, 不仅包含公共资源注释的受体和配体, 还有人工挑选的参与细胞通讯的特定蛋白质家族。与其他软件不同的是, CellPhoneDB考虑了配体和受体的亚基结构, 可以准确揭示异聚体复合物。目前CellPhoneDB共储存了978种蛋白质, 其中501种为分泌蛋白, 466种为异聚体; 收录的存在相互作用的分泌蛋白有474种, 膜蛋白490种, 整合素250种; 参考的数据源主要来自UniProt、Ensembl、PDB、IUPHAR数据库。

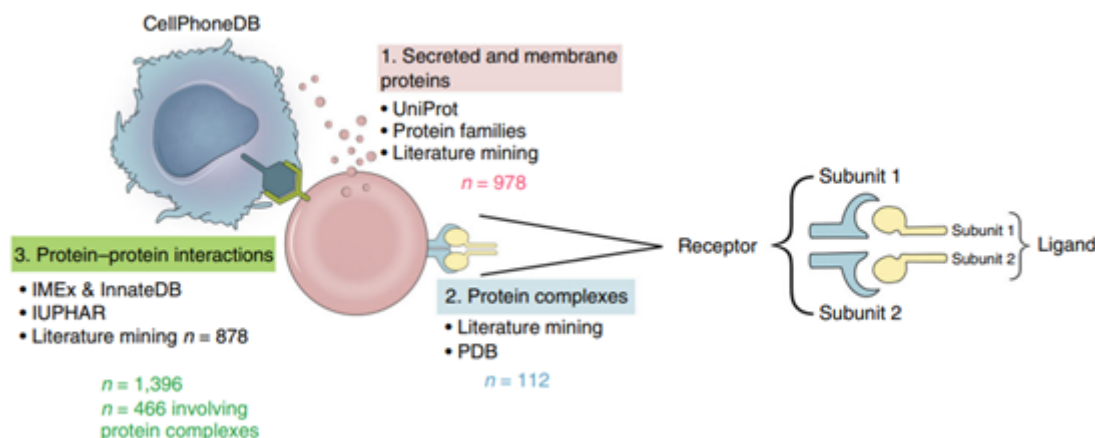
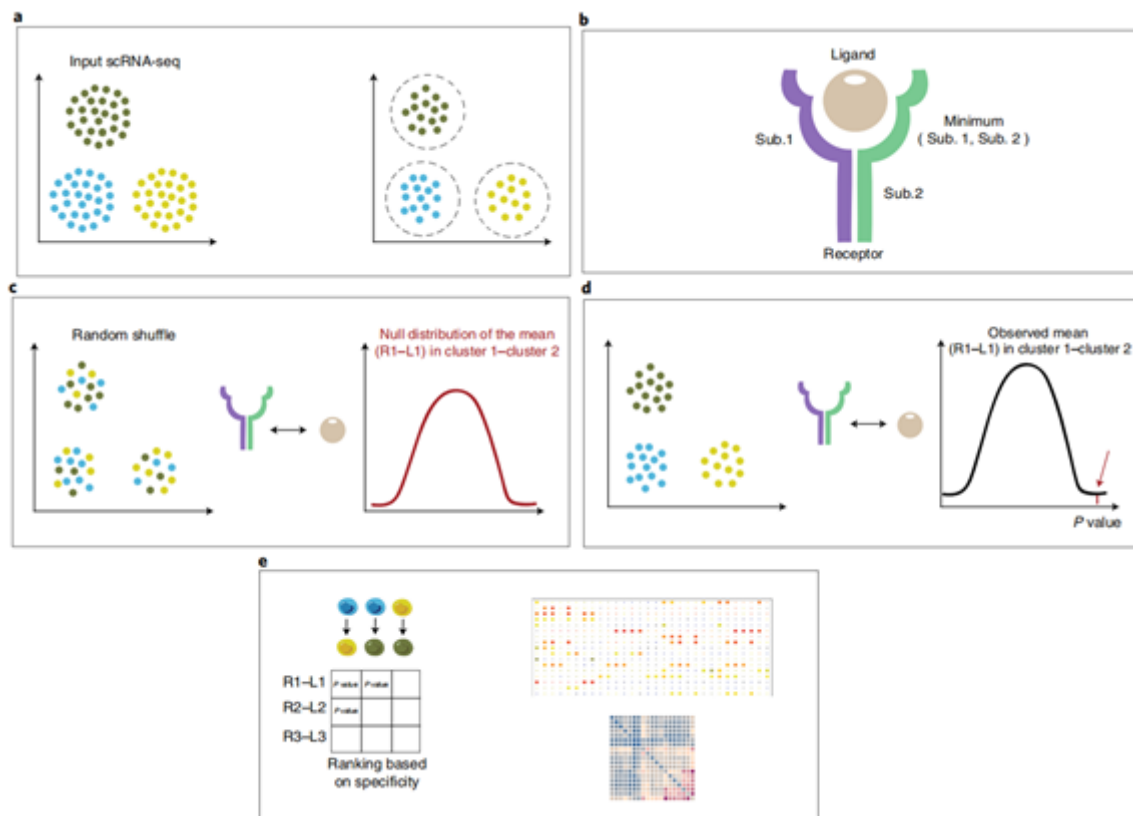


图: CellPhoneDB数据库简介。

CellPhoneDB读取包含细胞注释和基因表达水平信息的scRNA-seq数据, 具有相同细胞类型注释的细胞聚合成一种细胞状态; 根据一种细胞类型的受体表达和另一种细胞类型的配体表达, 分析两种细胞类型的受体配体相互作用。



图：a. 对于每个受体/配体基因，分别计算表达该基因的细胞百分比和基因表达平均数，细胞百分比 $\leq 10\%$ 的受体/配体基因会被过滤；对于异聚体复合物，选择亚基中较小的表达平均值作为该受体的表达值。b. 随机排列所有细胞（默认随机配列1000次），对所有细胞类型进行两两成对比较，计算随机排列后一种细胞类型中配体的平均表达水平和另一种的细胞类型中受体的平均表达水平的平均值，两种细胞类型间每个配体-受体对生成一个零分布。c. 计算两种细胞类型间配体-受体对的实际平均值，根据计算平均值等于或高于实际平均值的比例，推测该受体-配体对在这两种细胞类型中可能的显著性P值。d. 根据两种细胞类型中富集到的显著的配体-受体对的数量，对细胞类型之间高度特异的相互作用进行排序，筛选细胞类型间可能的分子相互作用，构建细胞亚群间潜在的通讯网络。