



整合受体调控基因表达信息构建细胞通信网络*

郭书旗¹⁾ 张绍武^{1)**} 李 岩¹⁾ 张世华^{2)**}⁽¹⁾ 西北工业大学自动化学院, 信息融合教育部重点实验室, 西安 710072; ⁽²⁾ 中国科学院数学与系统科学研究院, 北京 100190

摘要 **目的** 构建细胞通信网络有助于揭示细胞间协同工作机制、生物学过程和疾病发病机理。目前基于配体-受体相互作用构建细胞通信网络的方法大多只考虑配体和受体的表达信息, 忽略了受体对其调控基因的信号传递影响, 导致构建的细胞通信网络可靠性较低。鉴于此, 本文提出 IRRG 算法, 旨在构建更为准确的细胞通信网络, 并挖掘具有生物学意义的细胞通信模式。**方法** 本文提出了一种整合受体调控基因表达信息构建细胞通信网络的方法 (命名为 IRRG)。该方法通过随机游走方式计算受体对下游基因的影响得分, 进而与配体-受体共表达量结合构建细胞通信网络。**结果** 使用 IRRG 构建了小鼠滤泡间表皮 (IFE) 细胞通信网络并分析了配体-受体对的生物学意义, 验证了 IRRG 计算受体影响得分的稳定性和细胞通信网络构建的可靠性。此外, 使用 IRRG 构建了透明细胞肾细胞癌 (ccRCC) 的细胞通信网络, 挖掘并分析其肿瘤微环境细胞通信模式。**结论** IRRG 可以构建富有生物学意义并且可靠的细胞通信网络, 帮助人们从细胞通信的角度更深入地了解多种生物过程。IRRG 算法代码可从 GitHub 获取: <https://github.com/NWPU-903PR/IRRG>。

关键词 细胞通信, 配体-受体, 基因共表达, 基因调控网络, 随机游走**中图分类号** R318.04**DOI:** 10.16476/j.pibb.2022.0099

对于多细胞生物来说, 细胞通信 (也被称为细胞间相互作用) 是必不可少的。细胞通过交流和合作形成动态的通信网络, 在许多生物过程中发挥着至关重要的作用。例如, 驱动细胞分化和器官发育、维持组织器官内稳态、疾病中的免疫相互作用等^[1-3]。细胞通信所涉及的信号传递, 通常受蛋白质相互作用影响。由于 RNA 测序 (RNA-seq) 数据易获取和分析, 且基因表达水平是蛋白质丰度的主要决定因素^[4], 目前细胞通信的研究主要是以转录组学为基础。

随着单细胞 RNA 测序技术 (scRNA-seq) 技术的发展, 海量单细胞测序数据正在迅速积累。这些数据有助于揭示生物组织中潜在的细胞异质性, 并为细胞通信研究提供了重要的数据来源。利用单细胞测序数据, 从配体-受体 (L-R) 相互作用角度出发能够有效构建和分析细胞通信网络。目前, 已有一些基于配体-受体互作用构建细胞通信网络的研究。Kumar 等^[5] 通过识别和比较不同肿瘤之间的配体-受体互作用差异, 研究了 T 细胞亚群之间的细胞通信, 及与免疫过程关系; Xiong 等^[6] 根据肝细胞中高表达的配体-受体对, 构建了肝细胞配体-受体通信网络, 进而研究了患非酒精性脂肪肝

炎的患者肝脏微环境中的细胞间串扰。基于配体-受体互作用研究细胞通信的一项主要工作, 就是推断配体-受体的互作用强度。目前已经提出了一些策略用于推断配体-受体的互作用强度, 例如: 使用阈值判断配体和受体是否有效表达^[7]; 识别具有差异表达的配体和受体, 找到“活跃”的配体-受体对^[8]; 通过配体和受体各自表达量的乘积, 计算一个连续值推断配体-受体互作用强度^[9]。这些方法较为简单易行, 但存在如下潜在局限: 基因表达阈值不好选择, 配体或受体的表达量过高会导致乘积结果不准确。重要的是, 这些方法仅使用了配体和受体的表达量, 事实上, 参与细胞间通信的信号事件不仅包括细胞间的信号, 还包括细胞内的信号通路和转录调控^[10]。受体在接收到配体信号后, 会对受体下游的相关基因产生影响, 进而改变基因调控网络和转录因子活性。

鉴于此, 本文从细胞通信的生物现象角度出发

* 国家自然科学基金 (62173271, 61873202, 12126605) 资助项目。

** 通讯联系人。

张绍武 Tel: 029-88431308, E-mail: zhangsw@nwpu.edu.cn

张世华 Tel: 010-82541360, E-mail: zsh@amss.ac.cn

收稿日期: 2022-03-17, 接受日期: 2022-05-23

发,提出一种整合受体调控基因表达信息的细胞通信网络构建方法 IRRG。该方法利用 KEGG^[11] 和 Reactome^[12] 数据库建立不同细胞类型中每个受体对应的基因调控网络(即受体基因调控网络),然后采用随机游走算法计算细胞通信中受体对下游基因的影响程度得分,并将此影响得分与配体-受体共表达量结合,推断配体-受体相互作用强度。用于有效挖掘细胞间的通信现象,发现与生物过程密切相关的配体-受体对和关键通信模式。

1 数据集与方法

1.1 数据集

本文所用数据主要包括配体-受体对数据、信号通路数据、单细胞 RNA 测序数据。从 LRdb 数据库^[13] 选取配体-受体对,构建配体-受体对数据集(包含 3 251 组配体-受体对)。LRdb 数据库中的配体-受体对来自于 Reactome^[12]、HPMR^[14]、HPRD^[15]、Uniport^[16]、Fantom5^[17] 等多个数据库,这些配体-受体对可靠性较高、注释详细^[18]。为了有效整合受体调控的基因表达信息,通过 Pathway Commons^[19] 下载 KEGG 和 Reactome 信号通路,构建信号通路数据集,该数据集包含 34 万组基因相互作用的路径。此外,为了探究多个信号通路数据库对 IRRG 构建细胞通信网络结果的影响(附件文档 S1 第 1 节),本文在整合 KEGG 和 Reactome 两个通路数据库的基础上,又整合了 PID 和 Panther 两个信号通路数据库,其基因互作信息的数量达到 38 万多条。

单细胞 RNA 测序数据主要包括小鼠滤泡间表皮组织(GSE67602)^[20] 和透明细胞肾细胞癌组织(GSE159115)^[21] 单细胞测序数据,及各自的默认细胞类型标签。小鼠滤泡间表皮组织细胞数据集包含 720 个细胞,它们分为 IFE-B、IFE-DI、IFE-DII、IFE-KI、IFE-KII 5 种细胞类型^[20],构成小鼠表皮独特的分层结构,其中 IFE-B 细胞位于表皮最底部,可以增殖分化形成最外层的 IFE-K(由 IFE-KI 和 IFE-KII 两个亚型细胞组成)角化层细胞。本文使用同源基因转化将小鼠滤泡间表皮细胞的基因与配体-受体数据库以及信号通路数据库中的基因对应。肾透明细胞癌组织样本(SI_22368)和正常组织样本(SI_22369)分别包含 2 779、1 495 个细胞,肾透明细胞癌组织细胞分为肿瘤细胞(tumor)、平滑肌细胞(vascular smooth muscle cell, vSMC)、巨噬细胞(macrophage, Macro)、

T 细胞(T cell)、B 细胞(B cell)、内皮细胞(endothelial, Endo) 6 种细胞类型,正常组织细胞分为上皮细胞(epithelial)、vSMC、Macro、T 细胞、B 细胞、Endo 6 种细胞类型。

1.2 受体基因调控网络构建

构建受体在各个细胞类型中的基因调控网络是整合受体调控基因表达信息的前提。对于特定细胞类型中的受体,考虑到单细胞数据的稀疏性,本文选择该细胞类型中部分基因作为候选调控基因,用于构建受体基因调控网络。鉴于滤除特定细胞类型大部分细胞中均不表达的基因可以降低算法运行时间,且筛选候选调控基因的比例并非越大越好(大的筛选比例会使有效信息数量减少)。因此,本文按照 20% 的比例值筛选候选调控基因。然后,利用通路数据集中的基因关系,并根据表 S1 中的基因互作用简化规则,确定受体和其候选调控基因及候选调控基因间的连接。如果基因间的作用类型属于控制(Control),则基因间通过单向边连接;如果相互作用类型属于反应(Reaction)和复合物(Complex),则基因间通过双向边连接。经过上述步骤,可能会得到几个不连通的子网络,仅保留受体基因所在的基因调控子网络,进而构建出受体基因调控网络。

1.3 IRRG 算法

IRRG 算法主要由如下 3 部分组成: a. 受体影响得分计算; b. 特异性配体-受体对筛选; c. 细胞通信网络构建(图 1)。

a. 受体影响得分计算

使用一种改进的随机游走 DwanRank 算法^[22] 计算受体影响分(R_{score})以计算指定细胞类型中某一个受体 R 的受体影响得分为例,其迭代公式如下:

$$R_m^{t+1} = (1 - d_m) f_m + d_m \sum_{n=1}^N \frac{adj_{mn} r_n^t}{deg_n}, 1 \leq m, n \leq N \quad (1)$$

其中, N 表示受体 R 对应的受体基因调控网络中的基因节点数目; r_m^{t+1} 表示第 $t+1$ 次迭代后基因 m 的影响得分; adj_{mn} 表示基因 m 和基因 n 的邻接关系,如果基因 m 指向基因 n ,则 $adj_{mn}=1$; deg_n 表示基因 n 的入度; $d_m = deg_m / (deg_m + 3)$ 为一动态阻尼系数,基因节点的入度越大,其阻尼系数亦越大,进而网络中更多的连接信息被纳入计算; r_n^0 表示某类型细胞分为两组后基因 n 的差异表达值; f_m 表示将某类型细胞分为两组后基因 m 差异表达的归一化值。

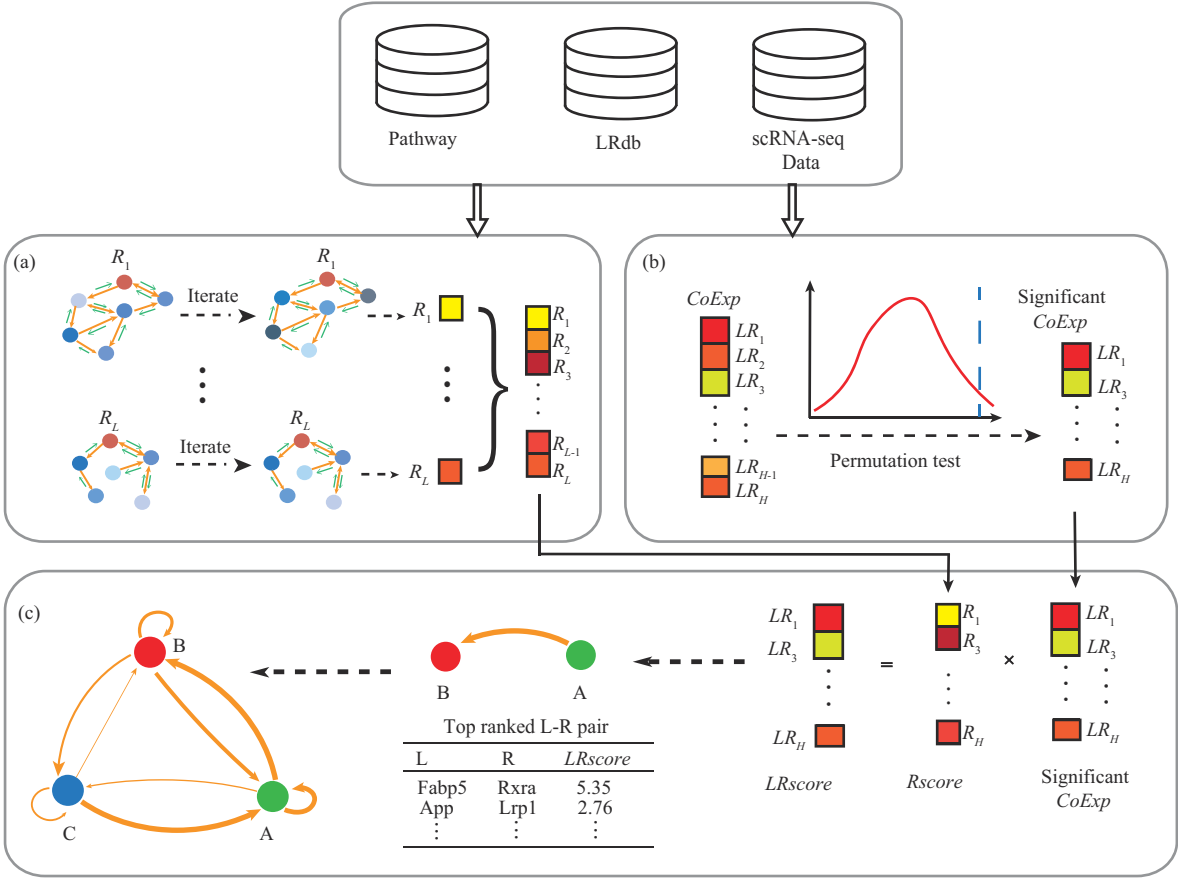


Fig. 1 Schematic diagram of IRRG

(a) Calculating the impact scores of receptors with the improved restart random walker algorithm. The green arrows represent the integration of gene expression information. (b) Screening the specific L-R pairs with Permutation test. (c) Building the cellular communication network.

r_n^0 和 f_m 的计算过程如下: 首先根据受体 R 在各个细胞中是否有效表达(本文设置基因有效表达值为0), 将某类型细胞分为受体有效表达细胞和未表达细胞两组, 然后分别计算两组细胞中受体基因调控网络的基因平均表达量 E_I 、 E_{II} , 求取二者的绝对差值计算 $r^0=[r_1^0, \dots, r_m^0, \dots, r_N^0]$, 即 $r^0=|E_I-E_{II}|$ 。根据 r^0 计算 f_m , 即 $f_m=r_m^0/\sum_{k=1}^N r_k^0$ 。

迭代更新 r^{t+1} , 直至当 $\|r^{t+1}-r^t\|<0.001$ 时, 则认为该受体基因调控网络游走已达到稳定状态, 并从此基因调控网络中获取受体 R 的稳定值。重复上述过程, 可以计算出指定细胞类型中所有受体的稳定值, 即向量 $R=[R_1, \dots, R_L]$, L 为该细胞类型中有效的受体个数。然后对向量 R 中的值进行归一化, 得到特定细胞类型中 L 个受体的影响得分。

b. 特异性配体-受体对筛选

鉴于部分配体-受体对在不同的细胞类型之间

普遍表达、不能提供关于特定细胞状态间的交流信息, IRRG采用置换检验筛选特异配体-受体对, 用于构建特异性细胞通信网络。考虑到配体和受体的基因表达是细胞的通信基础, 将配体-受体对的共表达量作为置换检验依据, 进而筛选特异性配体-受体对。下面以A、B、C 3种细胞类型为例说明如何筛选特异性配体-受体对, 首先随机置换所有细胞类型的标签, 此时, 3种细胞类型会得到新的基因表达矩阵; 然后重新计算任意二类细胞(包含同类细胞)之间配体-受体的共表达量, 重复1 000次随机置换得到零分布, 如果某配体-受体对的 $P\leq 0.05$, 则认为该配体-受体对具有特异性(图1b)。筛选出来的特异性配体-受体对用于构建细胞通信网络。

c. 细胞通信网络构建

考虑到配体-受体的共表达是细胞通信的基础, 而受体影响得分代表受体在细胞通信中对下游基因

的影响程度, 综合配体-受体共表达量 ($CoExp$) 和受体影响得分 ($Rscore$), 将二者的乘积用于度量配体-受体的互作用强度 ($LRscore$)。

$$LRscore_i = CoExp_i \times Rscore_i \quad (2)$$

其中, $LRscore_i$ 表示指细胞类型 A 和 B 之间第 i 对配体-受体的细胞通信强度; $CoExp_i$ 为第 i 对配体-受体中的配体在细胞类型 A 中的平均表达量与受体在细胞类型 B 中的平均表达量之积的算术平方根; $Rscore_i$ 表示第 i 对配体-受体中受体在细胞类型 B 中的影响得分。

将细胞类型 A 和 B 之间所有特异配体-受体对的互作用强度相加, 其加和值可作为细胞通信网络的边权重 (即通信强度), 即

$$Edge_{A \rightarrow B} = \sum_{i=1}^I LRscore_i \quad (3)$$

基于此, 可以计算出所有细胞类型间的通信强度, 进而构建出细胞通信网络。

2 结果与讨论

鉴于搭建受体基因调控网络时滤除特定细胞类型中大部分细胞中均不表达的基因, 可以降低算法运行时间, 但过大的筛选比例可能会使有效信息数量减少, 本文首先在小鼠滤泡间表皮组织数据集上通过分别设置 10%、20% 的基因筛选比例阈值构建受体基因调控网络。然后采用随机游走算法计算受体得分, 发现 10% 基因筛选比例阈值下的受体得分计算时间为 56 min 左右, 20% 基因筛选比例阈值下的受体得分计算时间为 20 min 左右。此外, 两种基因筛选比例下的各细胞类型受体得分差异较小 (图 S1)。因此, 本文按照 20% 的比例值筛选候选调控基因构建受体基因调控网络。

为验证 IRRG 算法构建细胞通信网络性能, 我们首先在小鼠滤泡间表皮细胞 (mouse interfollicular epidermal cells, IFE) 和人类肾透明细胞癌 (clear cell renal cell carcinoma, ccRCC) 的单细胞测序数据上应用本文的 IRRG 算法, 分别构建小鼠滤泡间表皮细胞通信网络和肾透明细胞癌细胞通信网络, 分析细胞通信模式。然后通过下采样策略, 评估 IRRG 算法计算的受体影响得分, 能否有效反映受体对调控基因的影响。最后, 根据细胞通信网络互作用强度大小, 统计排名靠前的配体-受体对所拥有的支持文献数量, 对比分析了 IRRG 与现有其他细胞网络构建方法的性能。

2.1 小鼠滤泡间表皮细胞和肾透明细胞癌细胞通信网络构建结果

2.1.1 小鼠滤泡间表皮细胞 (IFE) 通信网络构建结果

IFE 在保护皮肤中发挥关键作用, 鉴于其特殊的分层结构及单细胞测序数据可以公开获取, 采用 IRRG 算法构建了 5 种 IFE 细胞类型间的通信网络 (图 2a), 并通过细胞类型间通信强度和特异性配体-受体对数量 (图 2b、c), 分析 5 种细胞类型之间的细胞通信模式。从图 2a 可以看到, IFE-KII 参与的细胞通信强度较弱, IFE-B 细胞所涉及的细胞通信能力较强, 此通信模式可能与这两类细胞的功能有关。IFE-B 细胞位于表皮的基底层, 是一类增殖细胞, 它们可以增殖分化形成上基底层细胞, 而 IFE-KII 细胞是一类最终分化的角质层细胞, 它们是防止感染和受伤的重要屏障, 该类细胞逐渐被去掉细胞核并从皮肤表面脱落^[20, 23-24], 结合 IFE-KII 释放信号强度远大于接受信号强度的现象, 意味着其可能在保护小鼠皮肤中倾向担当“哨兵”角色; IFE-DII 具有出色的信号释放和接收能力, 暗示其可能担任“桥梁”角色、维持多层细胞间通信信号的正常传递; IFE-KI 细胞与其分层结构相对较远的细胞 (如 IFE-B) 有较高的通信强度, 说明小鼠滤泡间表皮不同层细胞间存在远距离细胞通信, 这些细胞通信现象有助于全面了解小鼠表皮不同层细胞间的协同工作机制。

为了帮助发现 IFE-B 细胞分化形成完整皮肤组织的过程和机理, 及 IFE-KII 细胞防止皮肤组织感染和受伤的屏障功能, 进一步分析了表皮基底层细胞 IFE-B 与其上层细胞 (即 IFE-DI、IFE-DII、IFE-KI 细胞)、角质层细胞 IFE-KII 细胞与其下层细胞 (即 IFE-KI、IFE-DII、IFE-DI、IFE-B 细胞) 间互作用强度较高的配体-受体对 (图 3)。由于 IFE-B 与 IFE-KII 之间的通信强度较低 (13.58), 图 3 中没有给出 IFE-B 与 IFE-KII 间互作用强度较高的配体-受体对统计结果。从图 3 可以看出, IFE-B 与其上层细胞的通信过程中, 普遍存在与受体 *Erbb2* 相关的配体-受体对, 这可能与受体 *Erbb2* 参与表皮组织的表皮生长因子刺激反应、上皮细胞增殖、细胞生长正向调节、伤口愈合等许多重要生物过程^[25-27] 有关, 此结果说明小鼠滤泡间表皮细胞通信中受体 *Erbb2* 发挥其重要的生物学功能。IFE-B 与 IFE-DII 的通信中, 与表皮生长因子受体 *Egfr* 相关的配体-受体对较多, 这可能与受体 *Egfr* 参与表皮发育、细

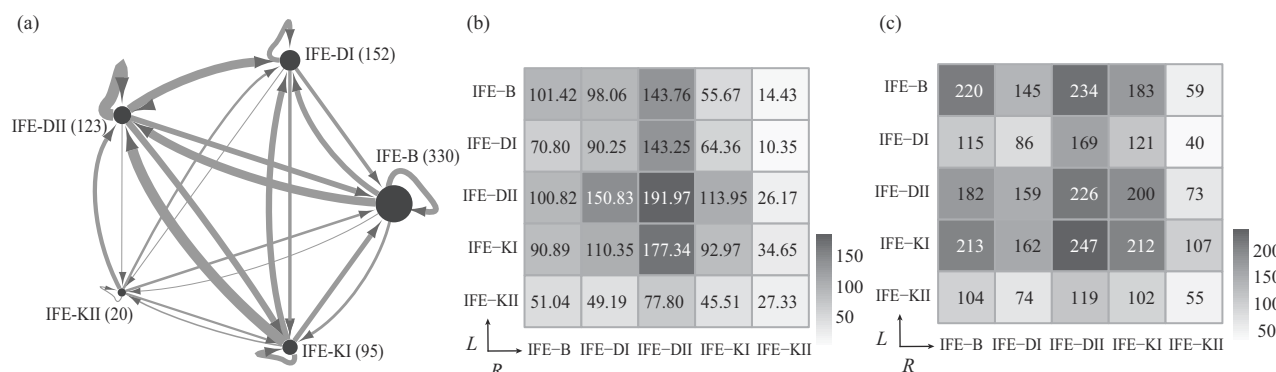


Fig. 2 Cell communication results of mouse IFE cells with IRRG algorithm

(a) Mouse cell communication network built with IRRG from the scRNA-seq gene expression profiles of 5 kinds of mouse IFE cells. The node size of every cell type depends on the number of cells it contains, and the thickness of each edge represents the communication strength between two cell types. The number in brackets after each cell type represents the number of cells contained in such cell type. (b) The strength values of cell communication among 5 kinds of mouse IFE cells. (c) Number of specific ligand-receptor pairs among 5 kinds of mouse IFE cells.

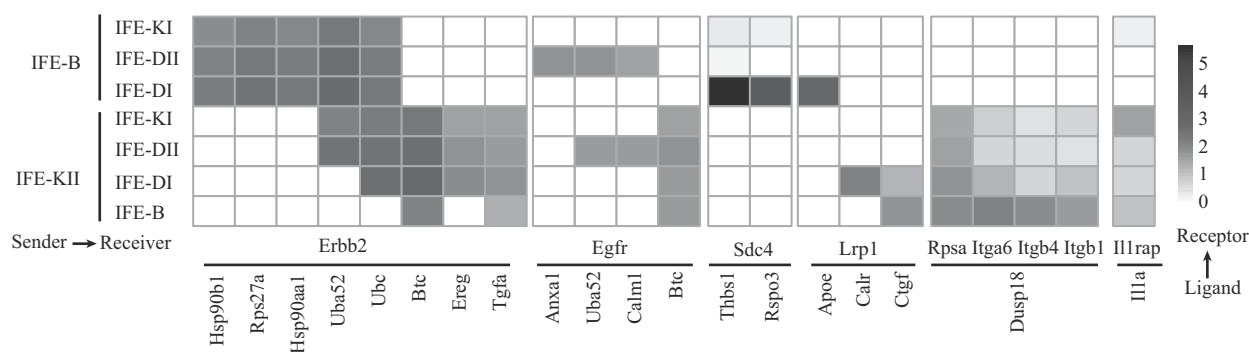


Fig. 3 Heatmap of some key ligand-receptor pairs in the communication process among five kinds of mouse IFE cells

胞迁移正向调节和细胞分化等过程^[28-29]有关。IFE-B与IFE-DI通信中,与受体Sdc4相关的配体-受体对拥有高互作用强度,这可能与受体Sdc4主要参与细胞迁移和伤口愈合等过程^[30-31]有关。

IFE-KII与其下层细胞的通信过程中,受体ErbB2和配体Btc、Ereg、Tgfa参与了细胞通信,它们均与小鼠表皮组织的生物过程相关^[32],说明IFE-KII细胞可能在促进下层细胞形成完整表皮的过程中发挥重要功能。IFE-KII与IFE-B通信过程中,受体Rpsa主要参与上皮细胞分化,受体Itga6、Itgb4在细胞迁移和皮肤形态发生中具有积极作用^[33-34]。IFE-KII和IFE-KI通信过程中,受体Il1rap可以产生白介素-1受体辅助蛋白,促进各种白介素的产生,用于激活免疫细胞或影响炎症反应^[35-36],这为IFE-K细胞作为防止入侵的屏障、发挥皮肤抗炎功能提供了有力证据。以上这些结果表明,本文的IRRG算法可以有效计算细胞通信强

度,并能够构建出具有生物学意义的细胞通信网络。另外,符合表皮生理过程不同细胞类型间的关键配体-受体对发现,有助于深入研究小鼠滤泡间表皮组织中特定层次细胞所扮演的角色,及不同类型细胞协同工作所发挥的生物学功能。

2.1.2 肾透明细胞癌细胞通信网络构建结果

肿瘤微环境中的细胞通信研究有助于发现肿瘤的生长、转移等机制。为验证IRRG算法能否有效挖掘癌细胞的通信模式,从肾透明细胞癌的单细胞RNA测序数据出发,采用IRRG算法分别构建肾透明细胞癌组织和正常组织的细胞通信网络(图S2),并分别计算肾透明细胞癌组织中,各个细胞类型间通信强度与正常组织中对应该细胞类型间通信强度的FoldChange值(图S3)及差异值(图S4),分析肾透明细胞癌细胞的通信模式。由于肾透明细胞癌组织中肿瘤细胞主要由上皮细胞癌变形成,将Tumor与Epithelial对应。从图S3, S4可以看出,

肾透明细胞癌中的肿瘤细胞 (tumor cell) 与 B 细胞 (B cell)、内皮细胞 (endothelial cell, Endo)、巨噬细胞 (macrophage, Macro)、T 细胞 (T cell)、血管平滑肌细胞 (vascular smooth muscle cells, vSMCs) 间的通信强度明显高于相应其正常组织中的上皮细胞 (epithelial) 与 B cell、endothelial、Macro、T cell、vSMCs 间的通信强度, 表明肿瘤细胞在肿瘤微环境中可通过增加与其他细胞的通信强度, 加速癌症的发展, 这些现象在许多研究中已被证明^[37-39]。作为免疫细胞的 T 细胞与 B 细胞, 在肾癌组织中的通信强度明显低于其正常组织中的通信强度, 说明 T 细胞与 B 细胞间的通信强弱与肾癌的发生发展密切相关。另外, 肾癌组织中的其他 5 种细胞 (B cell、endothelial、Macro、T cell、tumor) 与 vSMCs 间的通信强度明显高于其正常组织中其他 5 种细胞 (B cell、endothelial、Macro、T cell、epithelial) 与 vSMCs 间的通信强度, 暗示高强度的 vSMCs 通信在肾癌发展过程中扮演重要角色, 其原因可能在于 vSMC 细胞是构成血管壁组织结构及维持血管张力的主要细胞成分, 它们能够为肿瘤输送氧气和营养、转移癌细胞^[40-41]。

为了进一步研究肾癌组织中特有的细胞通信模式, 以 tumor 和 epithelial 细胞为信号释放源, 将肾癌组织和正常组织中的其他 5 种细胞 (B cell、endothelial、Macro、T cell、vSMCs) 作为受体细胞类型, 分别挑选出各个受体细胞类型中互作用强度大小排名前五的受体, 然后对这些受体在基因网络中所调控的一阶、二阶基因进行 KEGG 通路富集分析 (图 S5)。从图 S5 可以看出, 肾透明细胞癌组织中的通路富集结果在基因比例 (GeneRatio) 和对数 P 值 (LogPvalue) 上大于正常组织的通路富集结果, 且肾透明细胞癌组织中 T 细胞 (T cell) 与内皮细胞 (Endo) 的通路富集结果与其正常组织中的通路富集结果也明显不同。例如, 肾透明细胞癌组织中的 T 细胞富集信号通路 “Proteoglycans in cancer”, 主要负责在细胞膜和周围的外基质之间提供接触联系, 在调节癌细胞黏附和迁移方面起着核心作用^[42]。富集信号通路 “ECM-receptor interaction” 参与肿瘤的脱落、黏附、运动和增殖^[43]。另外, 肾透明细胞癌组织 T 细胞中缺少部分免疫相关的富集信号通路, 而 T 细胞的特异性免疫, 在对抗肿瘤发生发展过程中发挥其重要功能作用^[44], 此结果说明肿瘤细胞可能通过细胞通信抑制 T 细胞的免疫。表 S2 给出了肾透明细胞癌组织

和正常组织分别特有的富集信号通路, 例如肾癌组织中的 vSMCs 所富集的 “EGFR tyrosine kinase inhibitor resistance” 通路, 会干扰表皮生长因子抑制剂的治疗效果, 通过促进血管的形成间接帮助癌症生长。前人研究也已证实^[45-46], 透明细胞肾细胞癌作为高度血管化的肿瘤, vSMCs 被认为是肿瘤病理性过度血管生成的关键, 存在潜在的新治疗靶点。综合上述实验结果表明, 通过 IRRG 算法分别构建肾透明细胞癌组织和正常组织的细胞通信网络, 有助于揭示肿瘤细胞如何通过调控周围正常细胞完成自身生长或迁移的肿瘤发育过程, 发现透明细胞肾细胞癌中的特异富集信号通路, 深入了解肿瘤微环境的细胞通信模式。

2.2 受体影响得分计算的稳定性分析

为验证 IRRG 算法计算的 R_{score} 能否有效表示细胞通信中受体对调控基因的影响, 采用下采样策略评估受体影响得分计算的稳定性。鉴于 IFE-KI 和 IFE-KII 类细胞拥有相同的角化功能, 而 IFE-DI 和 IFE-DII 类细胞拥有相同的分化功能, 为保证下采样后每类细胞拥有一定的细胞数量, 首先将小鼠滤泡间表皮中的 IFE-KI 和 IFE-KII 细胞合并为 IFE-K 角化层细胞 (interfollicular epidermis keratinized layer cells)、IFE-DI 和 IFE-DII 细胞合并为 IFE-D 分化细胞 (interfollicular epidermis differentiated cells), 于是小鼠滤泡间表皮细胞分为 IFE-B、IFE-K 和 IFE-D 3 类, 然后按照 70%、50%、30%、10% 的采样比例分别从这 3 类细胞中随机抽样, 采用 IRRG 算法计算 3 类细胞中的受体影响得分, 并计算采样前/后共有的受体影响得分的皮尔逊相关系数。此过程重复 30 次, 图 4 为其下采样前/后受体影响得分的相关性箱状图。从图 4 可以看出, 对于 IFE-B、IFE-K 和 IFE-D 3 类细胞, 采样 70% 后计算的受体影响得分与采样前计算的受体影响得分的相关系数均在 0.9 以上。当采样率为 50% 时, 只有 IFE-D 细胞中的受体影响得分与其采样前受体影响得分的相关系数小于 0.9。当采样率为 30% 时, 3 种细胞的受体影响分数与它们采样前受体影响得分的相关系数在 0.8~0.9 之间。当采样率为 10% 时, 3 种细胞的受体影响分数与它们采样前受体影响得分的相关系数明显下降, 例如, IFE-B 细胞采样前/后的受体影响得分相关系数降低到 0.5~0.6, IFE-D 细胞降低到 0.4~0.5, IFE-K 细胞降低到 0.5~0.7。

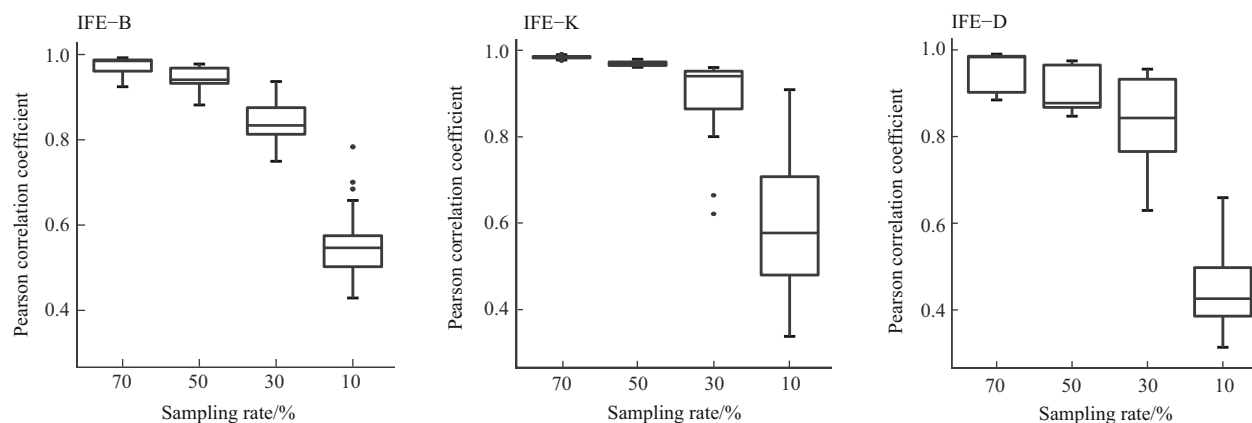


Fig. 4 Pearson correlation coefficient of receptor influence scores before and after down-sampling

综上结果分析可知, 采样率大于30%的条件下, IRRG算法计算的受体影响得分具有良好的稳定性。也就是说, 在细胞数保持一定量的条件下, IRRG计算出的受体影响得分, 仍能够有效表示细胞通信过程中受体对调控基因的影响。这说明IRRG可以合理地整合受体调控基因表达信息, 有效计算出受体对下游调控基因的影响程度得分。

2.3 IRRG算法性能分析

为了验证IRRG算法构建细胞通信网络的性能, 从小鼠滤泡间表皮细胞的单细胞RNA测序数据出发, 分别采用IRRG算法和NATMI^[47]、SCR^[13]两种方法构建小鼠滤泡间表皮5种细胞(IFE-B、IFE-DI、IFE-DII、IFE-KI、IFE-KII)间的细胞通信网络。NATMI^[38]仅使用配体-受体的共表达量计算配体-受体相互作用强度; SCR^[13]通过结合配体-受体的共表达量和归一化单细胞测序基因表达矩阵, 计算配体-受体相互作用强度, 并认为有效配体-受体对的相互作用强度大于0.5。另外, 为了验证本文的受体影响得分计算策略及特异性配体-受体对筛选策略的有效性, 分别删除IRRG算法中受体影响得分计算及特异性配体-受体对筛选单元, 设计IRRG_{RS}和IRRG_{SLR}二种消融方法, 其中IRRG_{RS}表示在IRRG算法中删除受体影响得分计算单元, IRRG_{SLR}表示在IRRG算法中删除特异性配体-受体对筛选单元。

根据LRdb数据库中配体-受体的文献注释, 在上述5种方法所构建的细胞通信网络中, 统计各细胞类型间(共25组)相互作用强度大小排名前20、30、40的配体-受体对所拥有的支持文献数量, 其统计结果如图S6所示, 每种方法在所有细胞类型

间相互作用强度排名靠前的配体-受体对所拥有的支持文献总数如图5所示。从图5可以看出, IRRG算法所构建的细胞通信网络中, 配体-受体对所拥有的支持文献总数高于NATMI和SCR两种方法。以IFE-KII细胞自通信的前20对配体-受体为例(图S6a), IRRG推断出的20对配体-受体中, 12对配体-受体共有15篇支持文献: Btc-ErbB2^[48]、Arf1-Pld2^[49-50]、Btc-Egfr^[51-52]、Scgblal-Lmbr1l^[53]、Tgfa-ErbB2^[54]、Il1a-Il1r2^[55]、Il1a-Il1r2^[56]、Btc-ErbB3^[57]、Tgfa-Egfr^[58-59]、Il1rn-Il1r2^[60]、Lamb1-Itgb1^[61]、Il1a-Il1r1^[62]; NATMI推断出的20对配体-受体中, 仅7对配体-受体共有8篇支撑文献: Il1a-Il1r2^[56]、Cdh1-Ptprf^[63]、Scgblal-Lmbr1l^[53]、Apoe-Ldlr^[64-65]、Btc-ErbB2^[48]、Hmgb1-Sdc1^[66]、Il1rn-Il1r2^[60]; SCR推断出的20对配体-受体中, 仅9对配体-受体共有10篇支撑文献: Il1a-Il1r2^[56]、Cdh1-Ptprf^[63]、Scgblal-Lmbr1l^[53]、Apoe-Ldlr^[64-65]、Btc-ErbB2^[48]、Il1rn-Il1r2^[60]、Efnb2-Rhbd12^[67]、App-Ncstn^[68]、Apoe-Lrp1^[69]。IRRG推断出的其他细胞类型间的前20对配体-受体, 所拥有的支持文献数量信息如表S3所示。为了进一步验证IRRG算法相较于NATMI和SCR算法优势的显著性, 利用Wilcoxon符号秩检验, 对3种算法在25组细胞类型间预测排名靠前配体-受体对支撑文献数量的差异性, 进行了统计显著性检验(图S7)。从图S7可以看出, IRRG的表现显著优于NATMI、SCR两种常用细胞通信网络构建方法, 意味着IRRG构建的细胞通信网络的可靠性和可解释性相对较高。另外, IRRG算法所构建的细胞通信网络中, 排名靠前的配体-受体对所拥有的支持

文献总数大于 IRRG_{RS} 和 IRRG_{SLR} 两种消融方法, 说明受体影响得分计算策略和特异性配体-受体对筛选策略能够有效提高细胞通信网络构建的可靠性。此外, 考虑到细胞生物功能的相似性, 也分析

了 IRRG 构建 IFE 3 种细胞 (IFE-K、IFE-D、IFE-B) 间细胞通信网络的性能 (附件文档 S1 第 2 节, 表 S4), 结果显示 IRRG 仍拥有最好的表现。

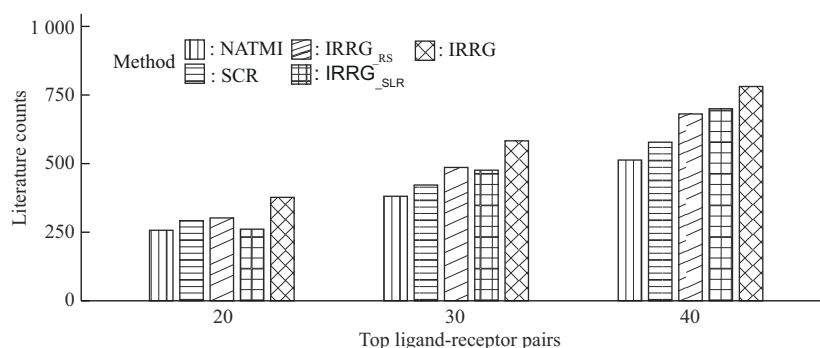


Fig. 5 Total number of the supported literatures of top 20, 30, and 40 ligand-receptor pairs among 5 cell types of IFE (*i.e.*, IFE-KII, IFE-KI, IFE-DII, IFE-DI, IFE-B)

上述结果分析表明, IRRG 算法能够发现细胞通信过程中重要的配体-受体对, 这有助于研究细胞通信模式, 从细胞通信层面研究生物过程。另外, 消融实验结果充分说明了细胞通信网络构建中, 考虑受体对下游调控基因的影响及筛选特异性配体-受体对的有效性。

3 结 论

单细胞 RNA 测序技术的快速发展为基于配体-受体的细胞通信研究提供了丰富的数据。虽然目前存在一些方法从单细胞 RNA 测序数据出发, 通过计算配体-受体互作强度构建细胞通信网络, 但這些方法忽略了细胞通信中受体会将接受到的信号传递给其他基因, 仅依据配体和受体的共表达量, 计算配体-受体的互作用强度构建细胞通信网络, 不能有效发现或挖掘细胞通信模式。鉴于此, 本文提出一个构建细胞通信网络的计算工具 IRRG。具体而言, IRRG 工具根据信号通路中的基因关系构建受体基因调控网络, 采用改进的随机游走算法计算受体对调控基因的影响得分, 并将此受体影响得分与其配体-受体的共表达量整合, 计算细胞间的配体-受体互作用强度。

小鼠滤泡间表皮细胞和肾癌组织细胞上的实验结果表明, IRRG 工具能够构建出具有一定生物学解释性的细胞通信网络, 可以发现表皮生理过程细胞间通信的关键配体-受体对及肾细胞癌的特异富

集信号通路。与现有其他细胞通信网络构建方法相比, IRRG 工具得到的细胞间配体-受体对所拥有的支撑文献数量相对较多, 说明 IRRG 工具构建的细胞通信网络比较可靠。随着空间转录组技术的发展, 如何结合单细胞空间位置信息精确构建细胞通信网络、发现新的细胞间通信模式, 未来仍然需要进一步深入研究。

附件 见本文网络版 (<http://www.pibb.ac.cn> 或 <http://www.cnki.net>)。

PIBB_20220099_Doc S1.pdf

PIBB_20220099_Tab S1.pdf

PIBB_20220099_Tab S2.pdf

PIBB_20220099_Tab S3.xlsx

PIBB_20220099_Tab S4.xlsx

PIBB_20220099_Fig S1.pdf

PIBB_20220099_Fig S2.pdf

PIBB_20220099_Fig S3.pdf

PIBB_20220099_Fig S4.pdf

PIBB_20220099_Fig S5.pdf

PIBB_20220099_Fig S6.pdf

PIBB_20220099_Fig S7.pdf

参 考 文 献

- [1] Kirouac D C, Madlambayan G J, Yu M, *et al.* Cell-cell interaction networks regulate blood stem and progenitor cell fate. *Mol Syst Biol*, 2009, 5(1): 293

- [2] Bayik D, Lathia J D. Cancer stem cell-immune cell crosstalk in tumour progression. *Nat Rev Cancer*, 2021, **21**(8): 526-536
- [3] Qi F, Qian S, Zhang S, *et al.* Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. *Biochem Biophys Res Commun*, 2020, **526**(1): 135-140
- [4] Vogel C, Marcotte E M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*, 2012, **13**(4): 227-232
- [5] Kumar M P, Du J, Lagoudas G, *et al.* Analysis of single-cell RNA-seq identifies cell-cell communication associated with tumor characteristics. *Cell Rep*, 2018, **25**(6): 1458-1468.e1454
- [6] Xiong X, Kuang H, Ansari S, *et al.* Landscape of intercellular crosstalk in healthy and NASH liver revealed by single-cell secretome gene analysis. *Mol Cell*, 2019, **75**(3): 644-660.e645
- [7] Choi H, Sheng J, Gao D, *et al.* Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep*, 2015, **10**(7): 1187-1201
- [8] Wang Y, Wang R, Zhang S, *et al.* iTALK: an R package to characterize and illustrate intercellular communication. *BioRxiv*, 2019: 507871. <http://doi.org/10.1101/507871>
- [9] Cain M P, Hernandez B J, Chen J. Quantitative single-cell interactomes in normal and virus-infected mouse lungs. *Dis Model Mech*, 2020, **13**(6): dmm044404
- [10] Sun X, Su J, Bao J, *et al.* Cytokine combination therapy prediction for bone remodeling in tissue engineering based on the intracellular signaling pathway. *Biomaterials*, 2012, **33**(33): 8265-8276
- [11] Kanehisa M, Araki M, Goto S, *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res*, 2008, **36**(Database issue): D480-484
- [12] Fabregat A, Jupe S, Matthews L, *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res*, 2018, **46**(D1): D649-D655
- [13] Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, *et al.* SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res*, 2020, **48**(10): e55
- [14] Ben-Shlomo I, Hsu S Y, Rauch R, *et al.* Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. *Sci STKE*, 2003, **2003**(187): RE9
- [15] Prasad T, Kandasamy K, Pandey A. Human protein reference database and human proteinpedia as discovery tools for systems biology. *Methods Mol Biol*, 2009, **577**: 67-79
- [16] Consortium U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*, 2019, **47**(D1): D506-D515
- [17] Ramilowski J A, Goldberg T, Harshbarger J, *et al.* A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat Commun*, 2015, **6**(1): 7866
- [18] Dimitrov D, Türei D, Gamido-Rodriguez M, *et al.* Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat Commun*, 2022, **13**: 3224
- [19] Cerami E G, Gross B E, Demir E, *et al.* Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res*, 2011, **39**(Database issue): D685-D690
- [20] Joost S, Zeisel A, Jacob T, *et al.* Single-cell transcriptomics reveals that differentiation and spatial signatures shape epidermal and hair follicle heterogeneity. *Cell Syst*, 2016, **3**(3): 221-237.e9
- [21] Zhang Y, Narayanan S P, Mannan R, *et al.* Single-cell analyses of renal cell cancers reveal insights into tumor microenvironment, cell of origin, and therapy response. *Proc Natl Acad Sci USA*, 2021, **118**(24): e2103240118
- [22] Hou J P, Ma J. DawnRank: discovering personalized driver genes in cancer. *Genome Med*, 2014, **6**(7): 56
- [23] Haensel D, Jin S Q, Sun P, *et al.* Defining epidermal basal cell states during skin homeostasis and wound healing using single-cell transcriptomics. *Cell Rep*, 2020, **30**(11): 3932-3947.e6
- [24] Rognoni E, Watt F M. Skin cell heterogeneity in development, wound healing, and cancer. *Trends Cell Biol*, 2018, **28**(9): 709-722
- [25] Li Y, Macdonald-Obermann J, Westfall C, *et al.* Quantitation of the effect of ErbB2 on epidermal growth factor receptor binding and dimerization. *J Biol Chem*, 2012, **287**(37): 31116-31125
- [26] Zhang Y, Dube P E, Washington M K, *et al.* ErbB2 and ErbB3 regulate recovery from dextran sulfate sodium-induced colitis by promoting mouse colon epithelial cell survival. *Lab Invest*, 2012, **92**(3): 437-450
- [27] Xiang B, Chatti K, Qiu H Q, *et al.* Brk is coamplified with ErbB2 to promote proliferation in breast cancer. *Proc Natl Acad Sci USA*, 2008, **105**(34): 12463-12468
- [28] Vassar R, Fuchs E. Transgenic mice provide new insights into the role of Tgf-Alpha during epidermal development and differentiation. *Genes Dev*, 1991, **5**(5): 714-727
- [29] Andl C D, Mizushima T, Nakagawa H, *et al.* Epidermal growth factor receptor mediates increased cell proliferation, migration, and aggregation in esophageal keratinocytes *in vitro* and *in vivo*. *J Biol Chem*, 2003, **278**(3): 1824-1830
- [30] Lambert J, Makin K, Akbareian S, *et al.* ADAMTS-1 and syndecan-4 intersect in the regulation of cell migration and angiogenesis. *J Cell Sci*, 2020, **133**(7): jcs235762
- [31] Bass M D, Williamson R C, Nunan R D, *et al.* A Syndecan-4 hair trigger initiates wound healing through Caveolin- and RhoG-regulated integrin endocytosis. *Dev Cell*, 2011, **21**(4): 681-693
- [32] Schneider M R, Wolf E. The epidermal growth factor receptor ligands at a glance. *J Cell Physiol*, 2009, **218**(3): 460-466
- [33] Yap L, Tay H G, Nguyen M T X, *et al.* Laminins in cellular differentiation. *Trends Cell Biol*, 2019, **29**(12): 987-1000
- [34] Feng C, Jin X, Han Y, *et al.* Expression and prognostic analyses of ITGA3, ITGA5, and ITGA6 in head and neck squamous cell carcinoma. *Med Sci Monit*, 2020, **26**: e926800
- [35] Mantovani A, Dinarello C A, Molgora M, *et al.* Interleukin-1 and related cytokines in the regulation of inflammation and immunity. *Immunity*, 2019, **50**(4): 778-795
- [36] Martin P, Goldstein J D, Mermoud L, *et al.* IL-1 family antagonists in mouse and human skin inflammation. *Front Immunol*, 2021, **12**: 652846
- [37] Choi H, Moon A. Crosstalk between cancer cells and endothelial cells: implications for tumor progression and intervention. *Arch Pharm Res*, 2018, **41**(7): 711-724
- [38] Ungefroren H, Sebens S, Seidl D, *et al.* Interaction of tumor cells with the microenvironment. *Cell Commun Signal*, 2011, **9**(1): 18
- [39] Calvo F, Sahai E. Cell communication networks in cancer

- invasion. *Curr Opin Cell Biol*, 2011, **23**(5): 621-629
- [40] Storz P. Reactive oxygen species in tumor progression. *Front Biosci*, 2005, **10**(1-3): 1881-1896
- [41] Taeger J, Moser C, Hellerbrand C, *et al.* Targeting FGFR/PDGFR/VEGFR impairs tumor growth, angiogenesis, and metastasis by effects on tumor cells, endothelial cells, and pericytes in pancreatic cancer. *Mol Cancer Ther*, 2011, **10**(11): 2157-2167
- [42] Ahrens T D, Bang-Christensen S R, Jorgensen A M, *et al.* The role of proteoglycans in cancer metastasis and circulating tumor cell analysis. *Front Cell Dev Biol*, 2020, **8**(749): 749
- [43] Bao Y L, Wang L, Shi L, *et al.* Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. *Cell Mol Biol Lett*, 2019, **24**(1): 1-20
- [44] Palmer S, Albergante L, Blackburn C C, *et al.* Thymic involution and rising disease incidence with age. *Proc Natl Acad Sci USA*, 2018, **115**(8): 1883-1888
- [45] Huang L, Fu L. Mechanisms of resistance to EGFR tyrosine kinase inhibitors. *Acta Pharm Sin B*, 2015, **5**(5): 390-401
- [46] F Bielecka Z, M Czarnecka A, Solarek W, *et al.* Mechanisms of acquired resistance to tyrosine kinase inhibitors in clear-cell renal cell carcinoma (ccRCC). *Curr Signal Transduct Ther*, 2013, **8**(3): 219-228
- [47] Hou R, Denisenko E, Ong H T, *et al.* Predicting cell-to-cell communication networks using NATMI. *Nat Commun*, 2020, **11**(1): 5011
- [48] Ishiyama N, Kanzaki M, Seno M, *et al.* Studies on the betacellulin receptor in pancreatic AR42J cells. *Diabetologia*, 1998, **41**(6): 623-628
- [49] Koch T, Brandenburg L O, Schulz S, *et al.* ADP-ribosylation factor-dependent phospholipase D2 activation is required for agonist-induced mu-opioid receptor endocytosis. *J Biol Chem*, 2003, **278**(11): 9979-9985
- [50] Lefkowitz R J. Identification of adenylate cyclase-coupled beta-adrenergic receptors with radiolabeled beta-adrenergic antagonists. *Biochem Pharmacol*, 1975, **24**(18): 1651-1658
- [51] Watanabe T, Shintani A, Nakata M, *et al.* Recombinant human betacellulin. Molecular structure, biological activities, and receptor interaction. *J Biol Chem*, 1994, **269**(13): 9966-9973
- [52] Simonson L G, Liberta A E, Richardson A. Characterization of an extracellular dextranase from *Fusarium moniliforme*. *Appl Microbiol*, 1975, **30**(5): 855-861
- [53] Zhang Z, Kim S J, Chowdhury B, *et al.* Interaction of uteroglobin with lipocalin-1 receptor suppresses cancer cell motility and invasion. *Gene*, 2006, **369**: 66-71
- [54] Xu X, Kelleher K F, Liao J, *et al.* Unique carboxyl-terminal sequences of wild type and alternatively spliced variant forms of transforming growth factor- α precursors mediate specific interactions with ErbB4 and ErbB2. *Oncogene*, 2000, **19**(28): 3172-3181
- [55] Cullinan E B, Kwee L, Nunes P, *et al.* IL-1 receptor accessory protein is an essential component of the IL-1 receptor. *J Immunol*, 1998, **161**(10): 5614-5620
- [56] Symons J A, Young P R, Duff G W. Soluble type II interleukin 1 (IL-1) receptor binds and blocks processing of IL-1 beta precursor and loses affinity for IL-1 receptor antagonist. *Proc Natl Acad Sci USA*, 1995, **92**(5): 1714-1718
- [57] Alimandi M, Wang L M, Bottaro D, *et al.* Epidermal growth factor and betacellulin mediate signal transduction through co-expressed ErbB2 and ErbB3 receptors. *EMBO J*, 1997, **16**(18): 5608-5617
- [58] Ignatz R A, Kelly B, Davis R J, *et al.* Biologically active precursor for transforming growth factor type alpha, released by retrovirally transformed cells. *Proc Natl Acad Sci USA*, 1986, **83**(17): 6307-6311
- [59] Keighley M R, Asquith P, Edwards J A, *et al.* The importance of an innervated and intact antrum and pylorus in preventing postoperative duodenogastric reflux and gastritis. *Br J Surg*, 1975, **62**(10): 845-849
- [60] Dripps D J, Verderber E, Ng R K, *et al.* Interleukin-1 receptor antagonist binds to the type II interleukin-1 receptor on B cells and neutrophils. *J Biol Chem*, 1991, **266**(30): 20311-20315
- [61] Colognato H, Yurchenco P D. Form and function: the laminin family of heterotrimers. *Dev Dyn*, 2000, **218**(2): 213-234
- [62] Dower S K, Kronheim S R, Hopp T P, *et al.* The cell surface receptors for interleukin-1 alpha and interleukin-1 beta are identical. *Nature*, 1986, **324**(6094): 266-268
- [63] Symons J R, Levea C M, Mooney R A. Expression of the leucocyte common antigen-related (LAR) tyrosine phosphatase is regulated by cell density through functional E-cadherin complexes. *Biochem J*, 2002, **365**(Pt 2): 513-519
- [64] Malby S, Pickering R, Saha S, *et al.* The first epidermal growth factor-like domain of the low-density lipoprotein receptor contains a noncanonical calcium binding site. *Biochemistry*, 2001, **40**(8): 2555-2563
- [65] Stein J M. The effect of adrenaline and of alpha- and beta-adrenergic blocking agents on ATP concentration and on incorporation of ^{32}P into ATP in rat fat cells. *Biochem Pharmacol*, 1975, **24**(18): 1659-1662
- [66] Salmivirta M, Rauvala H, Elenius K, *et al.* Neurite growth-promoting protein (amphoterin, p30) binds syndecan. *Exp Cell Res*, 1992, **200**(2): 444-451
- [67] Pascall J C, Brown K D. Intramembrane cleavage of ephrinB3 by the human rhomboid family protease, RHBDL2. *Biochem Biophys Res Commun*, 2004, **317**(1): 244-252
- [68] Yu G, Nishimura M, Arawaka S, *et al.* Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and betaAPP processing. *Nature*, 2000, **407**(6800): 48-54
- [69] Veinbergs I, Van Uden E, Mallory M, *et al.* Role of apolipoprotein E receptors in regulating the differential *in vivo* neurotrophic effects of apolipoprotein E. *Exp Neurol*, 2001, **170**(1): 15-26

Inference of Cell-cell Communication Networks by Integrating Receptor-regulated Gene Expression Information*

GUO Shu-Qi¹⁾, ZHANG Shao-Wu^{1)**}, LI Yan¹⁾, ZHANG Shi-Hua^{2)**}

¹⁾MOE Key Laboratory of Information Fusion Technology, School of Automation, Northwestern Polytechnical University, Xi'an 710072, China;

²⁾Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China)

Abstract Objective In multicellular organisms, cell communication allows multiple cells to coordinate with each other and involved in various important biological processes, and the abnormal cell communication pattern have also been demonstrated to be associated with many diseases. In the past decades, the rapid development of single-cell RNA sequencing technology makes it possible to study cell communication pattern through constructing cell communication network based on inferring ligand-receptor interactions, and many methods have been proposed. However, most of them only consider the expression of ligands and receptors to quantify the cell-cell interaction strength, but ignore the effect of specific receptor to its downstream gene regulatory network which reflects whether that receptor participates in particular biological process or not. To fill this gap, we propose a novel method (named IRRG) to construct cell communication networks at cell type level. **Methods** Our IRRG algorithm consists of the following three main steps. (1) IRRG uses the signaling pathway database to construct the gene regulatory network for each receptor in each cell type, and then calculates receptor impact score (RIS) for each receptor based on a random walk algorithm. RIS represents the degree of receptor impacts on downstream genes during cell communication. (2) IRRG uses permutation test to identify ligand-receptor pairs with significant specificity across cell types in order to discover more biologically significant cellular communication phenomena. (3) IRRG combined ligand-receptor co-expression with RIS to calculate the strength of all ligand-receptor interactions between cell types, and then constructed cellular communication networks. **Results** To validate the effectiveness of our IRRG, we construct cell communication networks of mouse epidermal tissue and human kidney cancer tissue separately. For mouse interfollicular epidermis (IFE) dataset, IRRG constructs biologically meaningful cellular communication networks, reasonably infers the different roles played by cells at different levels of epidermal tissue in cooperation, and discovers ligand-receptor pairs that are closely related to epidermal physiological processes. We also verified the robustness of the IRRG calculated receptor impact score, indicating that the IRRG calculated receptor impact score can reasonably reflect the signaling of cellular communication processes. By counting the supporting literature possessed of the top ranked ligand-receptor pairs, we also indirectly verified that the cell communication network constructed by IRRG is more reliable. In addition, we identify the cellular communication patterns of tumor tissues in clear cell renal cells (ccRCC) using IRRG and analyze how tumor cells influence normal cells to complete their own growth and development or migration process through cellular communication, further understanding the tumor microenvironment. **Conclusion** In this paper, we propose a novel method for constructing cellular communication networks named IRRG which integrating receptor-regulated gene expression information through random walk. Case studies on two datasets show that IRRG can construct biologically significant cellular communication networks, which can help us understand the mechanism of biological processes from the perspective of cellular communication. The source code and associated datasets used in this work are publicly available at <https://github.com/NWPU-903PR/IRRG>.

Key words cell-cell communication, ligand-receptor pair, gene co-expression, gene regulatory network, restart random walker

DOI: 10.16476/j.pibb.2022.0099

* This work was supported by grants from The National Natural Science Foundation of China (62173271, 61873202, 12126605).

** Corresponding author.

ZHANG Shao-Wu. Tel: 86-29-88431308, E-mail: zhangsw@nwpu.edu.cn

ZHANG Shi-Hua. Tel: 86-10-82541360, E-mail: zsh@amss.ac.cn

Received: March 17, 2022 Accepted: May 23, 2022