

基于单细胞RNA测序数据对胰腺癌肿瘤微环境中缺氧肿瘤亚群的探究

张璐, 谭建军*

北京工业大学化学与生命科学学院, 北京

收稿日期: 2024年4月23日; 录用日期: 2024年7月15日; 发布日期: 2024年7月25日

摘要

胰腺癌作为恶性程度极高的肿瘤之一, 多数患者确诊时已处于不可切除或转移性阶段。随着单细胞RNA测序(single-cell RNA-sequencing, scRNA-seq)技术的发展, 我们能够以更高分辨率深入探索肿瘤微环境(tumor microenvironment, TME)的内部异质性, 从而揭示胰腺癌进展中预后不良的关键机制。先前研究指出, 缺氧是实体瘤TME的固有特性, 能激活血管生成与转移相关的信号通路, 但缺氧TME的异质性仍需进一步阐释。通过基因集富集分析、拟时序分析和细胞间通讯分析等手段, 对胰腺癌scRNA-seq数据进行分析, 识别出具有不同生物学功能的肿瘤亚群, 特别是与缺氧密切相关的亚群。该缺氧肿瘤亚群与不良预后紧密相关, 据此构建的风险评分模型可有效预测胰腺癌患者总生存期。本研究加深了对胰腺癌TME的了解, 为胰腺癌预后的预测提供了一定参考。

关键词

胰腺癌, 单细胞RNA测序, 肿瘤微环境, 缺氧, 预后

Exploration of Hypoxia-Related Tumor Subpopulations in the Tumor Microenvironment of Pancreatic Cancer Based on Single-Cell RNA-Sequencing Data

Lu Zhang, Jianjun Tan*

College of Chemistry and Life Science, Beijing University of Technology, Beijing

Received: Apr. 23rd, 2024; accepted: Jul. 15th, 2024; published: Jul. 25th, 2024

*通讯作者。

文章引用: 张璐, 谭建军. 基于单细胞 RNA 测序数据对胰腺癌肿瘤微环境中缺氧肿瘤亚群的探究[J]. 生物医学, 2024, 14(3): 388-399. DOI: 10.12677/hjbm.2024.143043

Abstract

Pancreatic cancer is one of the most malignant tumors and most patients have unrespectable or metastatic disease at the time of diagnosis. With the development of single-cell RNA-sequencing (scRNA-seq) technology, we are able to explore the internal heterogeneity of the tumor microenvironment (TME) at a higher resolution, thereby revealing the key mechanisms underlying the poor prognosis of pancreatic cancer progression. Previous studies have shown that hypoxia is an intrinsic property of the TME in solid tumors and activates signaling pathways involved in angiogenesis and metastasis, but the heterogeneity of the hypoxic TME remains to be further elucidated. Pancreatic cancer scRNA-seq data were analyzed by gene set enrichment analysis, mimetic temporal sequencing analysis and intercellular communication analysis to identify a subpopulation of tumors with distinct biological functions, in particular one closely related to hypoxia. This hypoxic tumor subgroup was closely associated with poor prognosis, and the risk score model constructed accordingly could effectively predict the overall survival of pancreatic cancer patients. This study deepens the understanding of TME in pancreatic cancer and provides some guidance for predicting the prognosis of pancreatic cancer.

Keywords

Pancreatic Cancer, Single-Cell RNA-Sequencing, Tumor Microenvironment, Hypoxia, Prognosis

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

胰腺癌是全球癌症死亡的主要原因,发病率和死亡率迅速上升,从1990年到2017年,全球发病病例数增加了一倍多[1]。同时,胰腺癌也是一种高度致命的疾病,5年生存率仅为9% [2]。肿瘤微环境(Tumor microenvironment, TME)是一个复杂的生态系统,囊括了脉管系统、细胞外基质、细胞因子、生长因子和许多不同的细胞群,与胰腺癌肿瘤的生长和转移有着密切的关联[3]。单细胞测序技术(Single-cell RNA-sequencing, scRNA-seq)的飞速发展,为肿瘤微环境的研究提供了全新的视角。相较于传统的批量测序技术,scRNA-seq能够实现单个细胞转录组的测序,从而以单细胞分辨率揭示肿瘤内部的遗传异质性特征,有助于改善患者的预后[4] [5]。

由于肿瘤细胞的快速增殖和血管组织的供氧不足之间的不平衡,缺氧性 TME 是实体瘤的固有特征 [6]。在缺氧状态下,肿瘤细胞能够分泌一系列血管生长因子,进而促进异常血管生成,旨在增加氧气供应以维持肿瘤的生长与进展[7]。同时,这种缺氧环境还可能显著增强肿瘤细胞的侵袭性和转移性,以寻求更合适的生长环境[8]。近期研究表明,缺氧性 TME 中的缺氧相关基因(hypoxia-related genes, HRGs)是导致癌症患者预后不良的重要因素[9] [10]。然而,这些研究往往忽视了缺氧性 TME 的异质性,而这种异质性在肿瘤进展过程中可能具有至关重要的作用[11]。因此,本研究利用 scRNA-seq 数据及其相应的分析方法,探索了胰腺癌的发生发展机制以及缺氧 TME 的异质性。此外,确定了一个缺氧相关亚群,对其涉及的信号通路及其对患者预后的影响进行了研究,并基于缺氧亚群相关基因构建了风险评分模型,可以对患者的预后进行有效预测。

2. 材料和方法

2.1. 数据采集与预处理

本研究使用胰腺癌原发肿瘤的 scRNA-seq 数据对胰腺癌的肿瘤微环境进行研究, 数据收集于 Gene Expression Omnibus (GEO)数据库(<https://www.ncbi.nlm.nih.gov/geo/>), 入藏号为 GSM6603325。还收集了 Cancer Genome Atlas (TCGA)的 PAAD 的 RNA-seq 表达谱数据和生存信息, 数据通过 GDC [12] (<https://portal.gdc.cancer.gov>)进行下载。此外, 收集了 GEO 数据库中 urothelial cancer (UC)患者的数据进行验证, 入藏号为 GSE32894。

使用 R 包 Seurat [13] (version 4.1.0)对 scRNA-seq 数据的原始计数矩阵进行预处理。首先, 根据以下标准进行质量控制, 过滤低质量细胞和基因: 1) 过滤掉在至少 95% 的细胞中都没有表达的基因以及表达的基因数少于 5%的细胞; 2) 保留线粒体基因表达小于 15%的细胞; 3) 保留核糖体基因表达大于 3%的细胞; 4) 保留红细胞基因表达小于 0.1%且 $\log_{10}\text{GenesPerUMI} > 0.8$ 的细胞; 5) 使用 DoubletFinder 包对双细胞进行检测和过滤。随后, 使用 NormalizeData 函数进行标准化处理并利用 FindVariableFeatures 函数筛选出前 2000 个高度可变基因并采用 ScaleData 函数进行归一化处理。最终, 利用 PCA (Principal Component Analysis)进行降维并使用 FindNeighbors 函数和 t-SNE (t-Distributed Stochastic Neighbor Embedding) 分别对细胞进行聚类 and 可视化, 将不同类型的细胞分开并将相似的细胞聚在一起。此外, 细胞群的标志基因是利用 FindMarkers 函数鉴定的, 鉴定标准是基因表达量的倍数变化(Fold change, FC)的对数 $\log_{2}\text{FC} > 0.25$ 且 $P < 0.05$ 。根据每个亚群的标记基因的 $\log_{2}\text{FC}$ 值进行降序排列, 选取每个上皮细胞亚群 $\log_{2}\text{FC}$ 值前 10 的标志基因绘制表达量热图。

2.2. 细胞类型注释

使用 SingleR [14]软件包对细胞类型进行自动标注, SingleR 可以基于给定已知标签的样本作为参考数据集, 并根据与参考数据集的相似性来注释未知细胞。随后, 从 CellMarker 网站 (<http://bio-bigdata.hrbmu.edu.cn/CellMarker/index.jsp>)获取某些特定细胞类型的标志基因, 细胞及其标志基因分别为: T 细胞(*CD3D*, *CD3E*, *CD8A*, *CD4*); B 细胞(*CD79A*, *MS4A1*); 巨噬细胞(*CD68*, *CD163*, *CD14*); 单核细胞(*S100A9*, *S100A8*); 成纤维细胞(*FGF7*, *MME*, *DCN*); 肌成纤维细胞(*ACTA2*); 内皮细胞(*PECAM1*, *VWF*); 上皮细胞(*EPCAM*, *KRT19*, *PROM1*); 肥大细胞(*MS4A2*); 神经细胞(*MAP2*)。使用 SingleR 软件包进行注释后, 基于上述基因在不同细胞群中的表达量进行手动注释, 验证注释结果, 确保细胞注释的准确性。

2.3. 细胞良恶性鉴定

CopyKAT [15]利用细胞的基因表达数据来推断基因组拷贝数和亚克隆结构。该方法通过综合应用贝叶斯方法, 鉴定出单细胞中的全基因组非整倍性, 从而区分二倍体(Diploid)和非整倍体(Aneuploidy)细胞, 并将后者视为肿瘤细胞。将上皮细胞提取出来进行重新聚类后, 使用 CopyKAT (version 1.1.0)区分上皮细胞中的肿瘤细胞和非恶性细胞。

2.4. 单样本基因集富集分析

基因集富集分析[16] (Gene set variation analysis, GSVA)是在非参数无监督的条件下, 对基因集数据进行富集分析。GSVA 分析通常利用 MSigDB [17] (Molecular Signatures Database)数据库 (<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>)提供的基因集进行, 同时也支持用户自定义基因集。MSigDB 数据库目前包含 H 和 C1-C8 共九个预定义基因集类别, 我们选用 H 基因集进行分析, 该类别含

有 50 个基因集, 涵盖了 DNA 修复、细胞凋亡和缺氧等多个生物学过程。使用 GSVA (version 1.42.0) 软件包进行富集分析, 以 H 基因集和上皮细胞的表达矩阵作为数据输入, 计算每个上皮细胞在 50 个基因集中的富集分数, 并进行相应的可视化。

2.5. 拟时序分析

拟时序分析是一种通过单细胞转录组数据, 模拟和推断细胞在时间上的发展过程的方法, 可以揭示细胞在发育、分化或其他生物过程中的动态变化, 推断发育过程细胞的分化轨迹或细胞亚型的演化过程, 在生物学研究中具有重要意义。首先利用 Monocle3 [18] (version 1.3.1) 内置的 UMAP 聚类方法对上皮细胞进行聚类, UMAP 是一种非线性降维和数据可视化技术, 在高维数据的可视化和聚类分析中得到了广泛应用。它通过保持数据点之间的局部拓扑结构, 将高维数据映射到低维空间, 实现数据的降维和可视化。随后, 对上皮细胞的肿瘤细胞簇和非恶性细胞簇进行拟时序分析。在该过程中, 以非恶性细胞簇为起始点, 推断出上皮细胞的发育轨迹, 并利用莫兰指数(一种用于衡量基因表达变化显著性的指标)筛选出在发育轨迹中变化最显著的基因。

2.6. 细胞间通信分析

基于 R 包 CellChat [19] (version 1.6.1) 进行细胞间通信分析。利用 scRNA-seq 数据, 使用 CellChat 定量推断和分析细胞间通信网络, 通过计算每个信号通路相关的所有相互作用的通信概率, 确定主要调节不同信号通路的输出通信模式, 并根据每个相互作用对所对应的信号通路, 推断每个细胞亚群可能参与的信号通路以及相应的生物学功能。最后, 在信号通路水平上鉴定所有细胞亚群的通讯模式, 以探究不同的细胞亚群如何协同参与众多的信号通路。

2.7. 生存分析

使用 survival 软件包对 TCGA 数据库中获得的 172 个 BLCA 患者样本进行生存分析。首先, 使用单变量 Cox 比例风险回归(Univariate Cox proportional hazards regression)分析, 筛选出与患者总生存期显著相关的预后基因。随后, 使用多变量 Cox 比例风险回归(Multivariate Cox proportional hazards regression)分析构建风险评分(Risk score)模型, 如公式(1)所示:

$$\text{Risk score} = \sum_i^n \beta_i \times \exp_i \quad (1)$$

其中, β_i 和 \exp_i 分别代表基因 i 的多变量 Cox 回归系数(Multivariate Cox regression coefficient)和表达水平, n 代表预后基因的数量。接着, 利用中位数将患者分为高危组(High-risk)和 low 危组(Low-risk), 使用 survival ROC 软件包绘制受试者工作特征(Receiver operating characteristic, ROC)曲线, 并计算 ROC 曲线下面积(Area under ROC curve, AUC)。此外, 还绘制了 Kaplan-Meier (K-M)生存曲线, 并进行对数秩检验(Log-rank test), 以评估两组患者在总生存期上的差异。

3. 结果与讨论

3.1. 胰腺癌的单细胞景观

原始 scRNA-seq 数据包括 11949 个细胞, 质量控制后得到 9013 个细胞。通过对 scRNA-seq 数据进行聚类和细胞注释, 最终得到 7 种细胞类型, 如图 1(a)所示, 包括 661 个内皮细胞、940 个成纤维细胞、6498 个上皮细胞、270 个巨噬细胞、547 个 T 细胞、40 个神经细胞和 57 个肥大细胞。

由于尿路上皮膀胱癌中癌细胞主要来源于上皮细胞, 进一步提取了上皮细胞进行深入分析。上皮细

胞进行降维和聚类后, 被分为 8 个细胞亚群, 如图 1(b)所示。随后, 将上皮细胞进一步分类为肿瘤细胞和正常上皮细胞, 结果如图 1(c)所示: 在 6498 个上皮细胞中, 包含 2816 个肿瘤细胞、2717 个正常细胞和 965 个不能确定良恶性的细胞, 大部分上皮细胞呈现恶性状态。根据细胞良恶性鉴定结果, 8 个上皮细胞簇中包括 3 个正常细胞簇、4 个恶性细胞簇和 1 个不能确定良恶性的细胞簇, 如图 1(d)所示: 正常细胞簇包括正常亚群 1 (normal subpopulation 1, nor1)、正常亚群 2 (normal subpopulation 2, nor2)和正常亚群 3 (normal subpopulation 3, nor3); 肿瘤细胞簇包括肿瘤亚群 1 (tumor subpopulation 1, tum1)、肿瘤亚群 2 (tumor subpopulation 2, tum2)、肿瘤亚群 3 (tumor subpopulation 3, tum3)和肿瘤亚群 4 (tumor subpopulation 4, tum4); 不能确定良恶性的细胞簇(not.defined)。

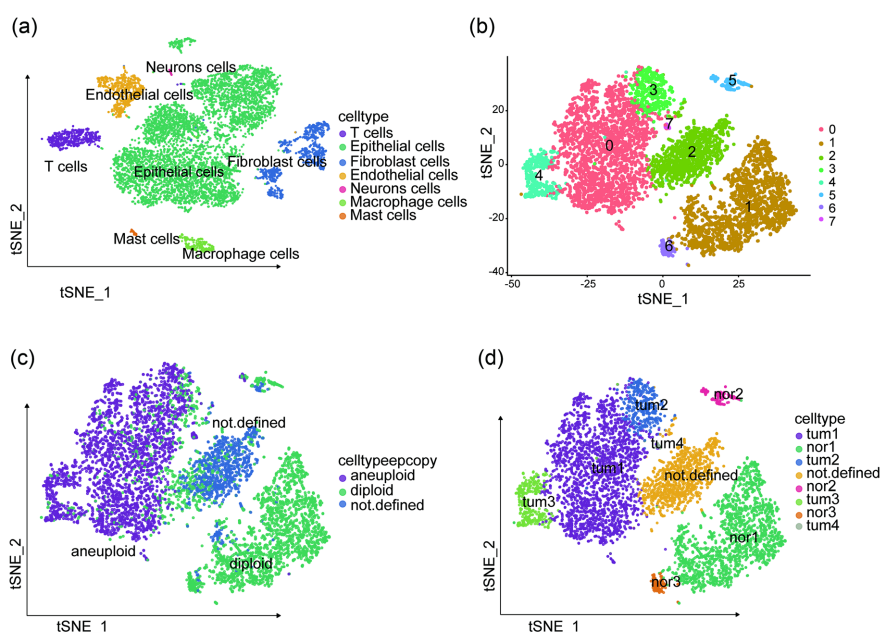


Figure 1. The single-cell landscape of pancreatic cancer, including. (a) The clustering results of all cells after annotation. (b) The clustering results of epithelial cells. (c) The identification of benign and malignant epithelial cells. (d) The clustering results of annotated epithelial cell subtypes

图 1. 胰腺癌的单细胞景观。(a) 所有细胞注释后的聚类结果; (b) 上皮细胞的聚类结果; (c) 良恶性上皮细胞的识别; (d) 上皮细胞亚群注释后的聚类结果

3.2. 上皮细胞肿瘤簇的标志基因及富集分析

为了研究上皮细胞肿瘤内的异质性, 对 4 个上皮细胞恶性细胞簇进行差异分析, 鉴定每个恶性上皮细胞亚群的标志基因, 结果如图 2(a)所示。在 tum1 中, *OLFM4* 基因和 *CRISP3* 基因的表达水平较高: *OLFM4* 基因可以促进胰腺癌细胞的增殖[20]; *CRISP3* 基因与细胞的生长和迁移有关[21]。在 tum2 中, *ERO1A* 基因、*FTH1* 基因和 *TIMP2* 基因的表达水平较高: *ERO1A* 基因具有糖酵解依赖性生长促进作用, 并可以由缺氧诱导[22]; *FTH1* 基因与癌症免疫有关[23]; *TIMP2* 基因与炎症有关[24]。在 tum3 中, *STMN1* 基因、*H2AFZ* 基因和 *TIMP2* 基因的表达水平较高: *STMN1* 基因与胰腺癌的增殖和迁移有关[25]; *H2AFZ* 基因可以促进肿瘤细胞的增殖和生长, 进而推动肿瘤的转移和上皮间质转化(Epithelial-mesenchymal transition, EMT) [26]。在 tum4 中, *CCL4* 基因和 *IL7R* 基因的表达水平较高: *CCL4* 基因与细胞免疫有关, 可以促进肿瘤的发展[27]; *IL7R* 基因具有很强的免疫调节作用, 可直接或间接作用于肿瘤细胞[28]。

通过对上皮细胞簇进行富集分析, 探究了上皮细胞亚群的特征及生物学功能, 结果如图 2(b)所示。

与促进癌细胞增殖有关的基因集(HEDGEHOG_SIGNALING, BILE_ACID_METABOLISM)在 tum1 亚群中富集, 胆汁酸和胆盐可以通过一系列信号转导途径介导炎症, 促进肿瘤增殖, 抑制肿瘤细胞凋亡[29][30]。tum2 亚群富集参与血管形成、血红素代谢和红细胞分化功能的基因集(Angiogenesis), 肿瘤细胞的生长和组织浸润需要新的血管持续形成, 有助于肿瘤获取氧气和营养, 在肿瘤的进展、侵袭和转移中起着重要作用[31]。此外, 缺氧(Hypoxia)相关基因集在 tum2 中也呈上调状态, 缺氧可以促进肿瘤表现出更具特征性的恶性行为, 可以通过诱导缺氧诱导因子 1- α (HIF-1 α)等分子促进新血管生成, 以及通过多种途径影响免疫系统[32]。在 tum3 亚群中, 与细胞增殖相关的基因集(MYC_TARGETS_V2, MYC_TARGETS_V1)呈现富集状态, MYC 是一种转录因子, 能够促进细胞的生长增殖和迁移, 还能影响细胞的凋亡过程[33][34]。在 tum4 亚群中, 炎症和免疫相关的基因集(TNFA_SIGNALING_VIA_NFK, TGF_BETA_SIGNALING)呈上调状态。核因子 κ B (Nuclear factor kappa B, NF κ B)是协调炎症反应以及细胞分化、增殖和存活的重要调节因子, 与恶性肿瘤的风险增加以及大多数类型癌症的恶性进展有关。

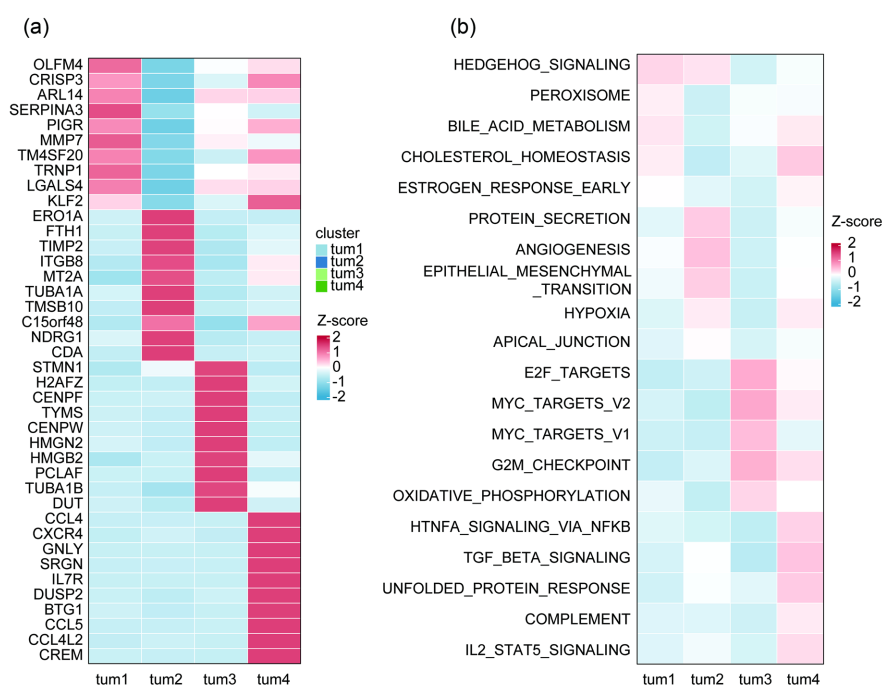


Figure 2. Differential genes and enriched pathways for epithelial cell subpopulations. (a) Top ten differential genes for each subpopulation. (b) Top five enriched pathways for each subpopulation

图 2. 上皮细胞亚群的差异基因及富集通路。(a) 每个亚群的前十个差异基因; (b) 每个亚群富集的前五个通路

3.3. 上皮细胞肿瘤簇的拟时序分析

为了深入研究肿瘤亚群的演变过程, 我们对上皮细胞进行了拟时序分析, 结果如图 3 所示。首先进行了聚类, 如图 3(a)所示。随后揭示了从非恶性细胞到恶性细胞的逐步演变过程, 如图 3(b)所示, 依照 nor3, nor1, nor2, not.defined, tum2, tum1, tum4, tum3 次序演变。

将拟时序分析结果与 GSVA 分析结果进行整合分析。发现肿瘤细胞在演变过程中呈现出从缺氧状态(tum2)转为癌细胞增殖状态(tum1), 进而转为炎症反应(tum4), 并最终向细胞迁移状态(tum3)演变。这一发现揭示了肿瘤细胞在演变过程中功能的变化, 包括从缺氧状态到形成血管以输送氧气和营养物质, 继

而发展为进行细胞增殖及细胞迁移相关的肿瘤亚群。同时, 还探究了肿瘤细胞在演化过程中变化最为显著的前十个基因, 如图 3(c)所示, 这些基因包括: *CAMK2N1*, *CFTR*, *FXYD3*, *S100A6*, *SFRP5*, *CEACAM6*, *FXYD2*, *IFI27*, *S100P* 和 *SLC4A4*。这些基因的变化可能反映了肿瘤细胞在演化过程中的关键生物学事件和调控机制, 可能是研究肿瘤发生发展的重要线索。

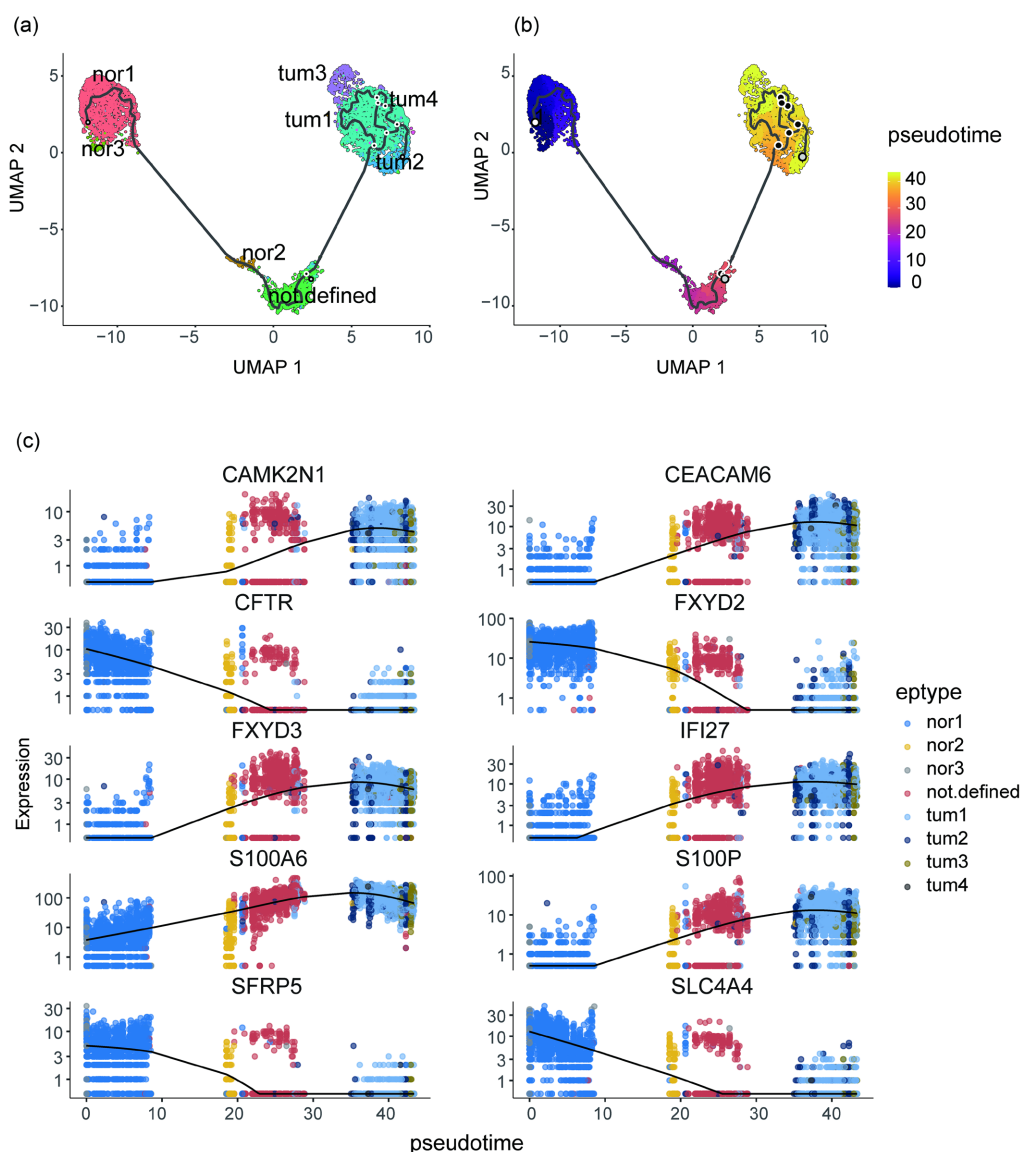


Figure 3. Results of the cell trajectory analysis of epithelial cells. (a) Distribution of epithelial cell subpopulations. (b) Evolutionary trajectory of epithelial cells. (c) Expression of the top 10 genes with the greatest change in expression during epithelial cell evolution

图 3. 上皮细胞拟时序分析结果。(a) 上皮细胞亚群的分布情况; (b) 上皮细胞的细胞轨迹分析结果; (c) 上皮细胞演化过程中表达量变化最大的前 10 个基因的表达量

3.4. 细胞间通讯情况

缺氧对于肿瘤的发生发展非常重要, 为了研究缺氧肿瘤亚群(tum2)对其他细胞亚群的潜在生物学作用和影响, 基于不同细胞亚群中配体和受体表达情况进行细胞间通讯分析, 以阐明缺氧肿瘤亚群与其他

亚群之间的细胞间相互作用情况。在所有细胞内进行细胞通讯分析, 结果如图 4(a)所示。图中每种颜色的圆代表特定的细胞亚群, 圆的大小表示相应细胞亚群中细胞数量的多少; 箭头的颜色与配体细胞颜色一致, 方向指向受体细胞, 箭头的宽度代表所对应的配体与受体细胞之间的通讯强度。从图中可以观察到, 肿瘤细胞、成纤维细胞和内皮细胞作为配体细胞时与其它细胞亚群的通讯强度较大, 而肥大细胞和 T 细胞作为配体细胞时与其它细胞亚群间的通讯强度相对较弱。

通过计算每个细胞亚群在充当配体细胞和受体细胞时的输出通讯强度(Outputting communication strength)和输入通讯强度(Inputting communication strength), 更全面地了解每个细胞亚群在通讯网络中的角色, 结果如图 4(b)所示。成纤维细胞和肿瘤细胞在通讯网络中扮演了重要的角色, 它们不仅是主要的信号发送者, 也是主要的信号接收者, 具有最高的输出通讯强度; 肿瘤亚群的通讯强度较为相近, 其中 tum4 亚群的通讯强度最高, 可能与其细胞迁移状态有关。

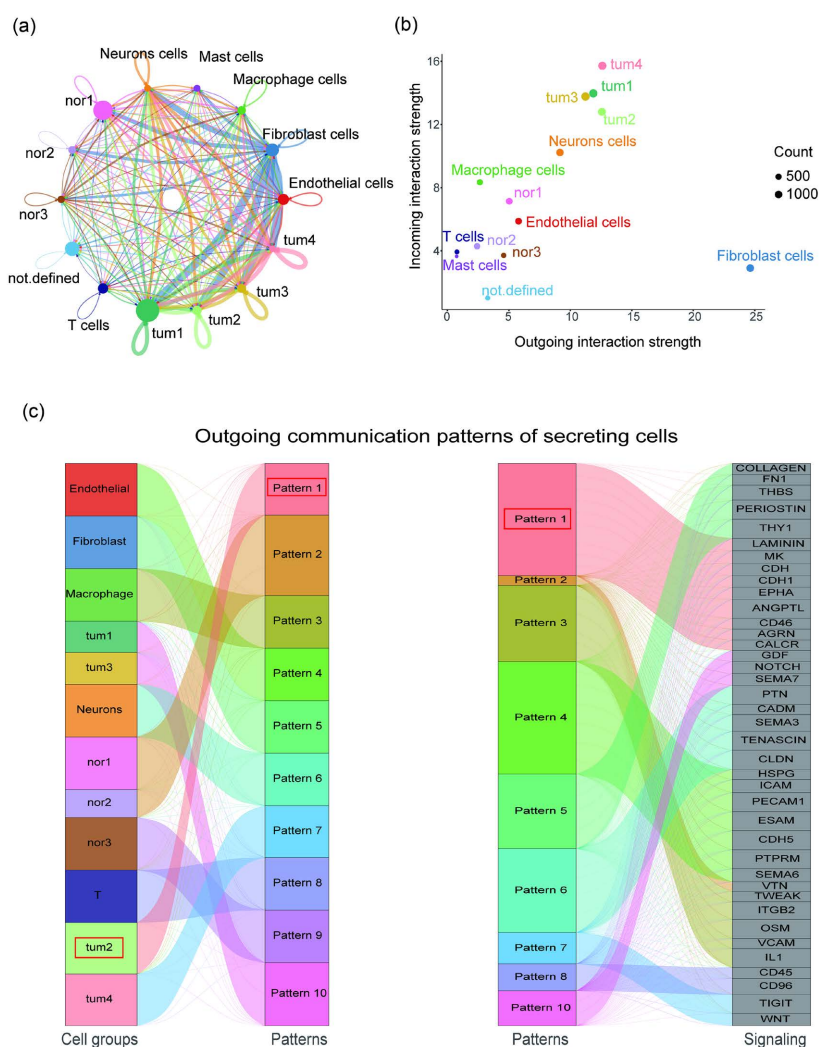


Figure 4. Results of cell-communication analysis. (a) Communication network diagram between cellular subpopulations. (b) Output/input communication strengths of all cellular subpopulations. (c) Cell-cell communication correspondence, with ten output communication modes synergistically involved in 39 signalling pathways

图 4. 细胞通讯分析结果。(a) 细胞亚群间的通讯网络图; (b) 所有细胞亚群的输出/输入通讯强度; (c) 细胞-细胞通讯对应关系, 十种输出通讯模式协同参与 39 条信号通路

细胞间通讯对应关系如图 4(c)所示。配体与受体间的交互作用与其涉及的信号通路有关, 可以推测各个细胞亚群作为配体细胞时可能介入的信号传递过程, 明确每种通讯模式中起主导作用的信号通路, 揭示细胞间通讯的复杂性与多样性。共存在八种输出通信模式并共同参与 39 条信号通路。缺氧肿瘤亚群共参与九种信号通路, 如表 1 所示。

Table 1. Nine signaling pathways primarily participated by the hypoxic tumor subpopulation

表 1. 缺氧肿瘤亚群主要参与的九个信号通路

信号通路	配体	相互作用方式
LAMININ	<i>LAMA3</i>	ECM 受体
LAMININ	<i>LAMA4</i>	ECM 受体
LAMININ	<i>LAMA5</i>	ECM 受体
LAMININ	<i>LAMC2</i>	ECM 受体
LAMININ	<i>LAMB3</i>	ECM 受体
MK	<i>MDK</i>	ECM 受体
CDH	<i>CDH1</i>	细胞间接触
CDH1	<i>CDH1</i>	细胞间接触
EPHA	<i>EFNA1</i>	细胞间接触
ANGPTL	<i>ANGPTL4</i>	分泌信号
CD46	<i>CD46</i>	细胞间接触
AGRN	<i>AGRN</i>	细胞间接触
CALCR	<i>ADM</i>	分泌信号

3.5. 预后相关基因筛选及风险评分模型的构建

通过细胞通讯分析鉴定出了缺氧肿瘤亚群主要参与的信号通路以及分泌的配体, 得到了 13 个信号通路相关的配体, 并以其作为候选基因, 利用单变量 Cox 比例风险回归分析筛选与患者的总生存期相关的预后基因, 结果如图 5(a)所示。筛选其中 $p < 0.05$ 且风险比(Hazard Ratio, HR)大于 1 的基因, 共得到 7 个基因(LAMC2, LAMB3, LAMA3, AGRN, ANGPTL4, ADM, CDH1), 这些基因与患者的不良预后相关。

利用多变量 Cox 比例风险回归分析, 通过计算出 7 个预后基因的多变量 Cox 回归系数后得到风险评分模型, 如公式(2)所示:

$$\text{Risk score} = 0.4952 \times \exp_{LAMA3} \quad (2)$$

根据患者 *LAMA3* 基因表达的中位数进行分组, 将患者划分为高风险组和低风险组, 并绘制了这两组患者的 K-M 生存曲线, 采用对数秩检验进行统计分析。结果如图 5(b)所示, 高风险组患者的生存率随时间推移而降低的速度显著快于低风险组, 这反映出高风险组患者的预后情况更差, 且两组之间的差异具有统计学显著性。此外, 我们还利用所建立的模型预测了患者在 1 年、3 年和 5 年时的生存状态, 并绘制了 ROC 曲线以评估模型的预测性能。如图 5(c)所示, 在 1 年、3 年和 5 年的时间点上, 模型的 AUC 值分别达到了 0.763、0.729 和 0.750, 这一结果表明该模型在预测患者生存状态方面具备一定的可靠性和有效性。

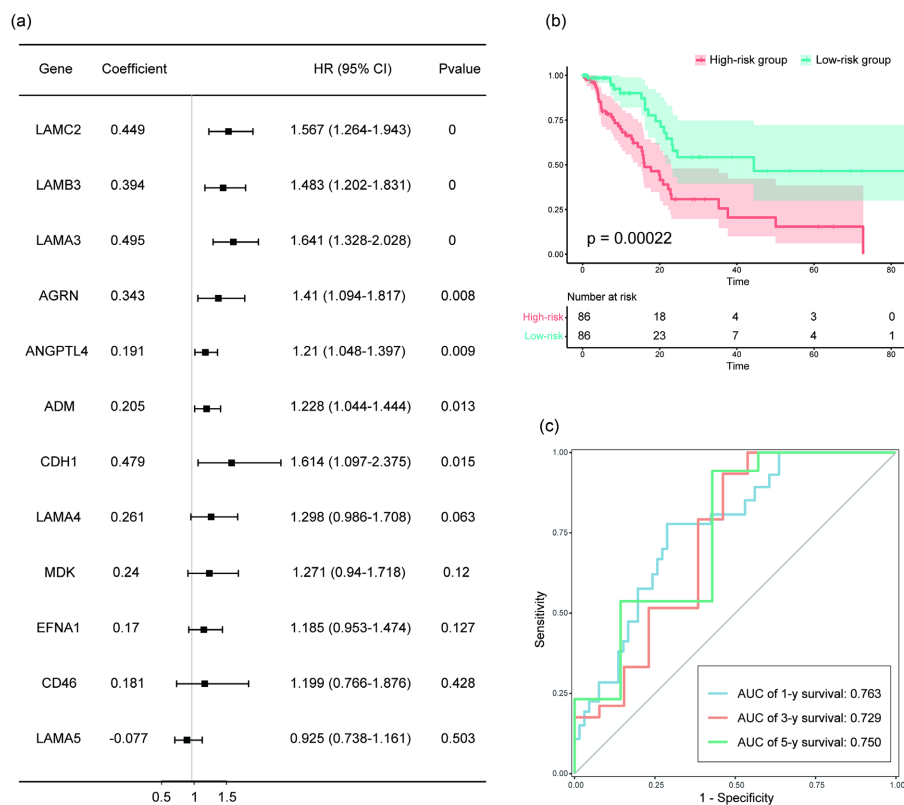


Figure 5. (a) Results of univariate Cox proportional hazards regression analysis. (b) K-M survival curves of patients in the high- and low-risk groups. (c) Time-dependent ROC curves to predict survival status of patients

图 5. (a) 单变量 Cox 比例风险回归分析结果; (b) 高/低风险组患者的 K-M 生存曲线; (c) 预测患者生存状态的时间依赖 ROC 曲线

4. 结论

本工作通过深入分析胰腺癌患者的 scRNA-seq 数据, 探讨了胰腺癌肿瘤组织的肿瘤微环境, 观察到其内部存在显著的异质性。在细胞群的初步聚类中, 鉴定出了内皮细胞、成纤维细胞、上皮细胞、巨噬细胞、T 细胞、神经细胞和肥大细胞。提取了上皮细胞, 并将其分为 3 个正常细胞簇、4 个肿瘤细胞簇和 1 个不能确定良恶性的细胞簇。通过对肿瘤亚群的标志基因以及 GSVA 结果的分析, 发现肿瘤亚群分别与肿瘤细胞的增殖、免疫和迁移等过程密切相关, 其中缺氧亚群在缺氧基因集上发生了显著富集。拟时序分析结果显示, 肿瘤细胞由缺氧肿瘤亚群演化为进行细胞增殖及迁移的肿瘤亚群。基于细胞通讯分析, 发现肿瘤亚群的细胞通讯强度很高, 主要参与九种信号通路, 在肿瘤的发展过程中具有重要作用。此外, 通过生存分析初步鉴定出了对于患者预后具有显著影响的九个基因, 构建了风险评估模型, 该模型可以可靠预测患者的总生存期, 为胰腺癌患者的预后评估和治疗策略的制定提供了重要的参考依据。尽管本文基于肿瘤缺氧亚群在胰腺癌的演进机制及预后预测取得了一定进展, 但仍有许多问题值得进一步探索: 1) 研究主要依赖于计算机模拟, 所得结果如有实验验证则更好; 2) 只关注了胰腺癌中的上皮细胞, 对于其他类型细胞对癌症发生发展的影响尚未进行深入探讨, 未来可以进一步研究肿瘤微环境中不同类型细胞对肿瘤演进和患者预后的影响; 3) 可以利用多组学数据整合分析的方法, 综合基因组、转录组和蛋白质组等数据, 更全面地揭示胰腺癌的分子特征和演进规律; 4) 还可进一步研究在其他类型癌症中的作用, 通过大规模的生物信息学分析和实验验证, 探究肿瘤缺氧亚群在不同癌症类型中的作用。

参考文献

- [1] Rahib, L., Smith, B.D., Aizenberg, R., Rosenzweig, A.B., Fleshman, J.M. and Matrisian, L.M. (2014) Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States. *Cancer Research*, **74**, 2913-2921. <https://doi.org/10.1158/0008-5472.can-14-0155>
- [2] Chen, W., Zheng, R., Baade, P.D., Zhang, S., Zeng, H., Bray, F., et al. (2016) Cancer Statistics in China, 2015. *CA: A Cancer Journal for Clinicians*, **66**, 115-132. <https://doi.org/10.3322/caac.21338>
- [3] Sherman, M.H. and Beatty, G.L. (2023) Tumor Microenvironment in Pancreatic Cancer Pathogenesis and Therapeutic Resistance. *Annual Review of Pathology: Mechanisms of Disease*, **18**, 123-148. <https://doi.org/10.1146/annurev-pathmechdis-031621-024600>
- [4] Hwang, B., Lee, J.H. and Bang, D. (2018) Single-Cell RNA Sequencing Technologies and Bioinformatics Pipelines. *Experimental & Molecular Medicine*, **50**, 1-14. <https://doi.org/10.1038/s12276-018-0071-8>
- [5] Sachs, N. and Clevers, H. (2014) Organoid Cultures for the Analysis of Cancer Phenotypes. *Current Opinion in Genetics & Development*, **24**, 68-73. <https://doi.org/10.1016/j.gde.2013.11.012>
- [6] Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of Cancer: The Next Generation. *Cell*, **144**, 646-674. <https://doi.org/10.1016/j.cell.2011.02.013>
- [7] Muz, B., de la Puente, P., Azab, F. and Azab, A.K. (2015) The Role of Hypoxia in Cancer Progression, Angiogenesis, Metastasis, and Resistance to Therapy. *Hypoxia*, **3**, 83-92. <https://doi.org/10.2147/hp.s93413>
- [8] Jing, X., Yang, F., Shao, C., Wei, K., Xie, M., Shen, H., et al. (2019) Role of Hypoxia in Cancer Therapy by Regulating the Tumor Microenvironment. *Molecular Cancer*, **18**, Article No. 157. <https://doi.org/10.1186/s12943-019-1089-9>
- [9] Sun, X., Luo, H., Han, C., Zhang, Y. and Yan, C. (2021) Identification of a Hypoxia-Related Molecular Classification and Hypoxic Tumor Microenvironment Signature for Predicting the Prognosis of Patients with Triple-Negative Breast Cancer. *Frontiers in Oncology*, **11**, Article 700062. <https://doi.org/10.3389/fonc.2021.700062>
- [10] Shi, Y., Huang, X., Du, Z. and Tan, J. (2022) Analysis of Single-Cell RNA-Sequencing Data Identifies a Hypoxic Tumor Subpopulation Associated with Poor Prognosis in Triple-Negative Breast Cancer. *Mathematical Biosciences and Engineering*, **19**, 5793-5812. <https://doi.org/10.3934/mbe.2022271>
- [11] Yang, X., Weng, X., Yang, Y., Zhang, M., Xiu, Y., Peng, W., et al. (2021) A Combined Hypoxia and Immune Gene Signature for Predicting Survival and Risk Stratification in Triple-Negative Breast Cancer. *Aging*, **13**, 19486-19509. <https://doi.org/10.18632/aging.203360>
- [12] Grossman, R.L., Heath, A.P., Ferretti, V., Varmus, H.E., Lowy, D.R., Kibbe, W.A., et al. (2016) Toward a Shared Vision for Cancer Genomic Data. *New England Journal of Medicine*, **375**, 1109-1112. <https://doi.org/10.1056/nejmp1607591>
- [13] Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., et al. (2021) Integrated Analysis of Multimodal Single-Cell Data. *Cell*, **184**, 3573-3587.E29. <https://doi.org/10.1016/j.cell.2021.04.048>
- [14] Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., et al. (2019) Reference-based Analysis of Lung Single-Cell Sequencing Reveals a Transitional Profibrotic Macrophage. *Nature Immunology*, **20**, 163-172. <https://doi.org/10.1038/s41590-018-0276-y>
- [15] Gao, R., Bai, S., Henderson, Y.C., Lin, Y., Schalck, A., Yan, Y., et al. (2021) Delineating Copy Number and Clonal Substructure in Human Tumors from Single-Cell Transcriptomes. *Nature Biotechnology*, **39**, 599-608. <https://doi.org/10.1038/s41587-020-00795-2>
- [16] Hänzelmann, S., Castelo, R. and Guinney, J. (2013) GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics*, **14**, Article No. 7. <https://doi.org/10.1186/1471-2105-14-7>
- [17] Wang, M., Chen, X., Fang, Y., Zheng, X., Huang, T., Nie, Y., et al. (2024) The Trade-Off between Individual Metabolic Specialization and Versatility Determines the Metabolic Efficiency of Microbial Communities. *Cell Systems*, **15**, 63-74.E5. <https://doi.org/10.1016/j.cels.2023.12.004>
- [18] Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., et al. (2019) The Single-Cell Transcriptional Landscape of Mammalian Organogenesis. *Nature*, **566**, 496-502. <https://doi.org/10.1038/s41586-019-0969-x>
- [19] Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C., et al. (2021) Inference and Analysis of Cell-Cell Communication Using Cellchat. *Nature Communications*, **12**, Article No. 1088. <https://doi.org/10.1038/s41467-021-21246-9>
- [20] Liu, W. and Rodgers, G.P. (2016) Olfactomedin 4 Expression and Functions in Innate Immunity, Inflammation, and Cancer. *Cancer and Metastasis Reviews*, **35**, 201-212. <https://doi.org/10.1007/s10555-016-9624-2>
- [21] Wang, L., Fu, D., Weng, S., Xu, H., Liu, L., Guo, C., et al. (2023) Genome-Scale CRISPR-Cas9 Screening Stratifies Pancreatic Cancer with Distinct Outcomes and Immunotherapeutic Efficacy. *Cellular Signalling*, **110**, Article ID:

110811. <https://doi.org/10.1016/j.cellsig.2023.110811>
- [22] Zhang, J., Yang, J., Lin, C., Liu, W., Huo, Y., Yang, M., *et al.* (2020) Endoplasmic Reticulum Stress-Dependent Expression of ERO1L Promotes Aerobic Glycolysis in Pancreatic Cancer. *Theranostics*, **10**, 8400-8414. <https://doi.org/10.7150/thno.45124>
- [23] Luo, Y., Liu, C., Yao, Y., Tang, X., Yin, E., Lu, Z., *et al.* (2024) A Comprehensive Pan-Cancer Analysis of Prognostic Value and Potential Clinical Implications of FTH1 in Cancer Immunotherapy. *Cancer Immunology, Immunotherapy*, **73**, Article No. 37. <https://doi.org/10.1007/s00262-023-03625-x>
- [24] Schoeps, B., Eckfeld, C., Prokopchuk, O., Böttcher, J., Häußler, D., Steiger, K., *et al.* (2021) TIMP1 Triggers Neutrophil Extracellular Trap Formation in Pancreatic Cancer. *Cancer Research*, **81**, 3568-3579. <https://doi.org/10.1158/0008-5472.can-20-4125>
- [25] Suzuki, K., Watanabe, A., Araki, K., *et al.* (2018) High STMN1 Expression Is Associated with Tumor Differentiation and Metastasis in Clinical Patients with Pancreatic Cancer. *Anticancer Research*, **38**, 939-944. <https://doi.org/10.21873/anticancer.12307>
- [26] Ávila-López, P.A., Guerrero, G., Nuñez-Martínez, H.N., Peralta-Alvarez, C.A., Hernández-Montes, G., Álvarez-Hilario, L.G., *et al.* (2021) H2A.Z Overexpression Suppresses Senescence and Chemosensitivity in Pancreatic Ductal Adenocarcinoma. *Oncogene*, **40**, 2065-2080. <https://doi.org/10.1038/s41388-021-01664-1>
- [27] Korbecki, J., Grochans, S., Gutowska, I., Barczak, K. and Baranowska-Bosiacka, I. (2020) CC Chemokines in a Tumor: A Review of Pro-Cancer and Anti-Cancer Properties of Receptors CCR5, CCR6, CCR8, CCR9, and CCR10 Ligands. *International Journal of Molecular Sciences*, **21**, Article 7619. <https://doi.org/10.3390/ijms21207619>
- [28] Wang, C., Kong, L., Kim, S., Lee, S., Oh, S., Jo, S., *et al.* (2022) The Role of IL-7 and IL-7R in Cancer Pathophysiology and Immunotherapy. *International Journal of Molecular Sciences*, **23**, Article 10412. <https://doi.org/10.3390/ijms231810412>
- [29] Cui, H., Lian, J., Xu, B., Yu, Z., Xiang, H., Shi, J., *et al.* (2023) Identification of a Bile Acid and Bile Salt Metabolism-Related lncRNA Signature for Predicting Prognosis and Treatment Response in Hepatocellular Carcinoma. *Scientific Reports*, **13**, Article No. 19512. <https://doi.org/10.1038/s41598-023-46805-6>
- [30] Yang, Y., Yang, C., Yang, Q., Lu, S., Liu, B., Li, D., *et al.* (2024) Elucidating Hedgehog Pathway's Role in HNSCC Progression: Insights from a 6-Gene Signature. *Scientific Reports*, **14**, Article No. 4686. <https://doi.org/10.1038/s41598-024-54937-6>
- [31] Liu, Z., Chen, H., Zheng, L., Sun, L. and Shi, L. (2023) Angiogenic Signaling Pathways and Anti-Angiogenic Therapy for Cancer. *Signal Transduction and Targeted Therapy*, **8**, Article No. 198. <https://doi.org/10.1038/s41392-023-01460-1>
- [32] Tao, G., Jiao, C., Wang, Y. and Zhou, Q. (2022) Comprehensive Analysis of Hypoxia-Related Genes for Prognosis, Immune Features, and Drugs Treatment Strategy in Gastric Cancer Using Bulk and Single-Cell RNA-Sequencing. *Scientific Reports*, **12**, Article No. 21739. <https://doi.org/10.1038/s41598-022-26395-5>
- [33] Dang, C.V., O'Donnell, K.A., Zeller, K.I., Nguyen, T., Osthus, R.C. and Li, F. (2006) The c-Myc Target Gene Network. *Seminars in Cancer Biology*, **16**, 253-264. <https://doi.org/10.1016/j.semcancer.2006.07.014>
- [34] Xu, J., Chen, Y. and Olopade, O.I. (2010) MYC and Breast Cancer. *Genes & Cancer*, **1**, 629-640. <https://doi.org/10.1177/1947601910378691>