

• 研究报告 •

王洋 哈尔滨工程大学博士，湖北大学生命科学学院副教授，研究方向为生物医学大数据组学分析，酶活性位点识别及结构分析。主持省自然科学基金 1 项，省级人才项目 1 项，近 3 年以第一作者或通讯作者（含并列）在 *Frontiers in Immunology*、*Cellular & Molecular Immunology*、*Journal of Immunology Research* 等杂志发表 SCI 论文 6 篇。



基于单细胞转录组的视网膜母细胞瘤进展特征分析

许凯龙^{1#}，聂巍巍^{2#}，童倩雯¹，马立新¹，刘洁¹，王洋¹

1 湖北大学 生命科学学院 省部共建生物催化与酶工程国家重点实验室，湖北 武汉 430062

2 哈尔滨第四医院 眼科，黑龙江 哈尔滨 150001

许凯龙，聂巍巍，童倩雯，马立新，刘洁，王洋. 基于单细胞转录组的视网膜母细胞瘤进展特征分析. 生物工程学报, 2022, 38(10): 3809-3824.

Xu KL, Nie WW, Tong QW, Ma LX, Liu J, Wang Y. Analysis of progress characteristics of retinoblastoma based on single cell transcriptome sequencing. Chin J Biotech, 2022, 38(10): 3809-3824.

摘要：视网膜母细胞瘤 (retinoblastoma, RB) 是婴幼儿最常见的眼内恶性肿瘤。在 RB 进展过程中的关键致病因素目前尚不十分清楚。因此，识别与 RB 进展密切相关的基因能为病情诊断及基因治疗提供重要信息。然而，肿瘤组织具有很强的细胞异质性，不同病理状态下的细胞，其功能及基因表达都可能呈现显著的差异。本研究从公共基因表达数据库 (gene expression omnibus, GEO) 下载了 1 例 4 个月肿瘤患者和 1 例 2 年患者的肿瘤及癌旁组织的单细胞转录组测序数据，从单细胞水平解析不同患病时长的 RB 肿瘤转录图谱，鉴定与 RB 进展有潜在关联的细胞亚群及基因集。结果显示，肿瘤组织与癌旁组织在单细胞转录图谱上具有整体的一致性，但视锥前体 G1 期细胞群、G2 期细胞群以及小胶质细胞群在肿瘤与癌旁组织中的分布比例存在明显差异。进一步分析了这 3 种细胞群在 RB 肿瘤进展过程中的作用。研究发现，在 RB 肿瘤的早期阶段，视锥前体细胞在 G1 期异常增殖，随着 RB 肿瘤的进展，视锥前体 G2 期细胞比例显著增加。同时，RB 进展过程的小胶质细胞群差异分析结果显示，主要参与免疫应答的关键基因包括 *RPL23*、*B2M*、*HLA*

Received: June 23, 2022; Accepted: September 26, 2022

Supported by: Knowledge Innovation Program of Wuhan-Shuguang Project (202210701251325); Special Foundation for Young Scientists of Hubei Educational Committee (202110701301003)

Corresponding author: WANG Yang. E-mail: yangwang@hubu.edu.cn

[#]These authors contributed equally to this study

基金项目：知识创新专项——曙光计划项目 (202210701251325); 湖北省教育厅青年人才项目 (202110701301003)

家族基因。本研究可为 RB 发病机制及进展研究提供更多新视角和数据资源。

关键词：视网膜母细胞瘤；单细胞转录组测序；肿瘤异质性；肿瘤；视锥细胞

Analysis of progress characteristics of retinoblastoma based on single cell transcriptome sequencing

XU Kailong^{1#}, NIE Weiwei^{2#}, TONG Qianwen¹, MA Lixin¹, LIU Jie¹, WANG Yang¹

1 State Key Laboratory of Biocatalysis and Enzyme Engineering, School of Life Sciences, Hubei University, Wuhan 430062, Hubei, China

2 Department of Ophthalmology, Harbin Fourth Hospital, Harbin 150001, Heilongjiang, China

Abstract: Retinoblastoma (RB) is the most common intraocular malignant tumor in infants and young children. The key causative factors in the progression of RB remain unclear. Therefore, identifying genes closely associated with RB progression may provide important clues for disease diagnosis and gene therapy. However, tumor tissues have strong cellular heterogeneity. There may be significant differences in cell function and gene expression among cells in different pathological states. In this study, we downloaded single-cell transcriptome sequencing data of RB tumors and adjacent tissues from the GEO public database. Subsequently, we analyzed RB tumor transcriptional profiles with different disease duration at the single-cell level and identified cell groups and gene sets potentially associated with RB progression. The results showed that the tumor tissue and the adjacent tissues had overall consistency in the single-cell transcriptional map, but there were obvious differences in the distribution proportions of G1 phase cells, G2 phase cells, and microglia cells of cone precursors in RB tumor and the adjacent tissues. Furthermore, the role of three cell populations in the progression of RB tumors was emphatically analyzed. We found that in the early stage of RB tumors, cone precursor cells proliferated abnormally in G1 phase. With the progression of RB tumors, the proportion of cone precursor cells in G2 phase increased significantly. Meanwhile, the results of differential analysis of microglial populations during RB progression showed that the key genes mainly involved in immune response include *RPL23*, *B2M*, and HLA superfamily genes. This study provides new perspectives and data resources for the research of RB pathogenesis and progress.

Keywords: retinoblastoma; scRNA-seq; tumor heterogeneity; tumor; cones

视网膜母细胞瘤 (retinoblastoma, RB) 是一种源于光感受器前体细胞的恶性肿瘤，多见于 3 岁以下儿童，是婴幼儿最常见的原发性眼内恶性肿瘤，在世界范围内发病率约为 1 : 2 0 000^[1]。RB 早期无临床症状，只能通过眼底检查发现较小的肿瘤。其临床症状复杂多样，多表现为结膜充血水肿、虹膜新生血管、玻璃体混浊、眼压升高及斜视等，且该病具有颅内转移

和远处转移的特点，会威胁患儿生命^[2]，治疗 RB 需按个体化治疗原则并结合 RB 病理分期。在肿瘤学上，肿瘤分期对治疗方法的选择及疗效评价极其重要^[3]。因此，探究不同进展时期 RB 的特点、识别与其进展密切相关基因，对 RB 进展研究及早期诊断具有一定的参考意义。

Cobrinik 及其同事们在早期的研究中指出，RB 肿瘤起源于视锥细胞前体，RB 肿瘤是

响应 *RBI* 单基因突变而形成^[4]。而 *RBI* 基因编码了一种肿瘤抑制蛋白 Rb, Rb 可以通过在细胞做好分裂准备之前抑制细胞周期进程来阻止过度的细胞生长^[5], 防止细胞恶性增殖进而转化为癌细胞^[6]。然而, 肿瘤的发生是一个复杂的、多因素过程, 同时, 在肿瘤整个增殖、扩散、转移的过程当中, 必然还存在其他关键因素, 进而引起组织在细胞水平发生变化, 最终发生病变。因此, 挖掘 RB 肿瘤进展过程中视锥细胞前体显著表达的一些促进增殖及抑制凋亡的关键癌相关蛋白就变得十分重要。

近年来, 单细胞转录组测序 (single-cell RNA sequencing, scRNA-seq) 技术的出现极大地推动了肿瘤发病机制研究及异质性^[7]的解析。RB 中的肿瘤细胞具有高度异质性, 从单细胞层面解析肿瘤内异质性将更好地揭示肿瘤细胞的来源, 识别 RB 进展相关的细胞组分及关键基因^[8]。因此, 本研究基于公共基因表达数据库 (gene expression omnibus, GEO) 资源, 下载 RB 肿瘤单细胞转录组测序数据, 对不同患病时长的 RB 组织及癌旁组织的转录特征进行解析, 详细阐述 RB 在不同病理时间、不同组织部位的细胞组成、比例差异, 并探索其在 RB 肿瘤进展过程中的关键作用。研究结果显示, 视锥前体 G1 期细胞群、G2 期细胞群以及小胶质细胞群在 RB 肿瘤与癌旁组织中的分布比例存在明显差异。同时, 结果显示, 随着肿瘤发病时间的增加, 小胶质细胞群显著地表达一些促进细胞增殖、转移及免疫应答相关的蛋白, 如 *B2M*、*RPL23* 及 HLA 家族基因。本研究结果可为 RB 肿瘤发病机制及进展研究提供新的数据资源。

1 材料与方法

1.1 数据的下载与整理

本课题的数据下载于 GEO 数据库, 数据编号 GSE166173^[9]。数据包括 1 例患病 4 个月

的 RB 肿瘤癌旁组织 (GSM5065164)、1 例患病 4 个月的 RB 肿瘤实质部分 (GSM5065165)、1 例患病 2 年的 RB 肿瘤癌旁组织 (GSM5065166)、1 例患病 2 年的 RB 肿瘤实质部分 (GSM5065167), 共 4 个样本。

1.2 数据分析

1.2.1 数据预处理及质控

采用 SEURAT^[10]R 包进行数据处理及相关分析, 分析软件及程序版本见表 1。设置质控参数筛选高质量细胞用于后续分析, 其中基因数目设定阈值为 500–4 000、线粒体基因<15%、核糖体基因<20%。质控结果已提交国家微生物科学数据中心 (编号: NMDCX0000144)。

1.2.2 降维聚类及细胞类型鉴定

采用 FINDVARIABLEFEATURES 函数寻找前 2 000 个高变基因, 可视化结果见附图 1C (详见《生物工程学报》网络版附件)。采用 RUNPCA 进行线性降维 ($N=20$), 聚类过程中 RESOLUTION 设定为 0.5, 对数据集进行 SCT 标准化^[11], 使用 RUNHARMONY 函数^[12]消除批次效应。使用 UMAP 非线性降维可视化聚类结果。采用 FindAllMarkers 函数识别每个细胞亚群的 Marker 基因, 应用 SingleR^[13]、Cellmarker 数据库^[14], 并结合参考文献进行人工校正, 对细胞亚群进行鉴定。

表 1 分析软件

Table 1 Analysis software

Software/package	Version
R	4.1.2
Seurat	4.0.5
ggplot2	3.3.5
dplyr	1.0.7
tidyverse	1.3.1
GSEABase	1.56.0
patchwork	1.1.1
clusterProfiler	4.2.0
KEGG.db	1.0
harmony	0.1.0
SingleR	1.8.0

1.2.3 差异表达基因筛选及富集分析

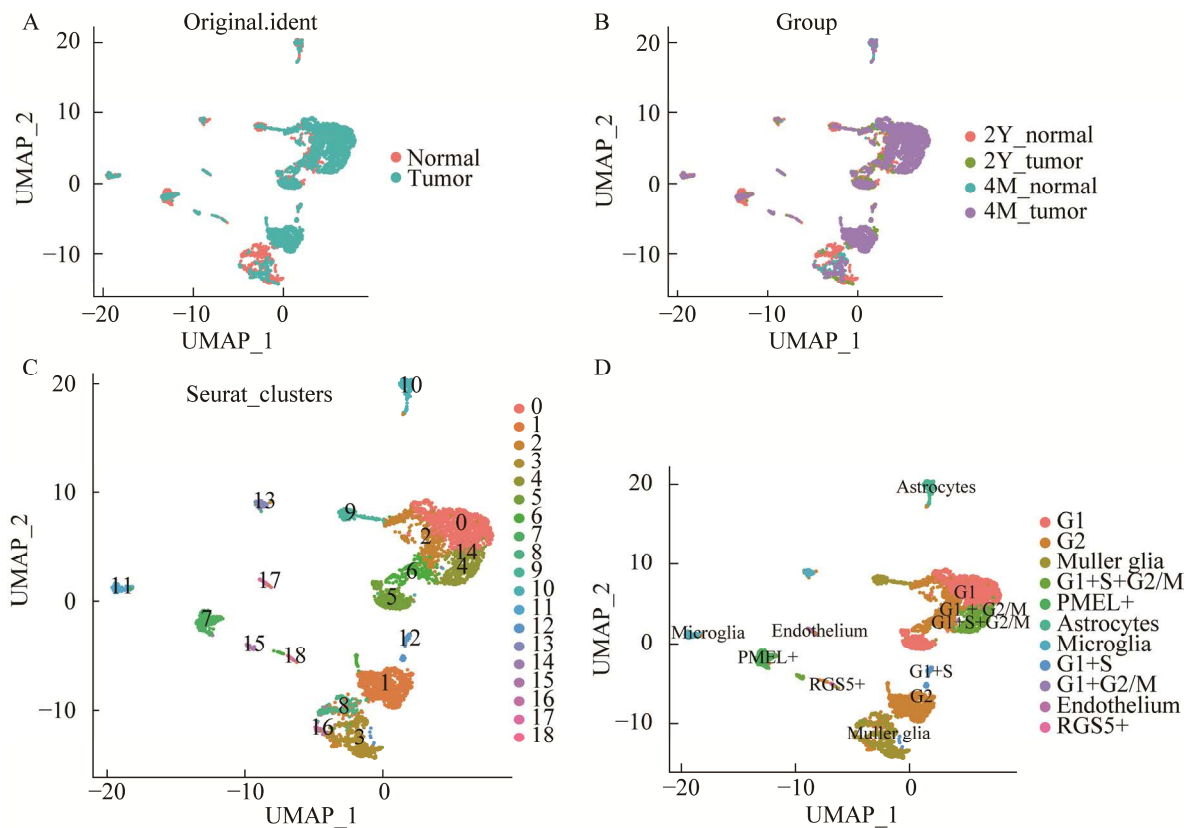
本课题最终选取了 3 组对象进行差异基因的识别及富集分析: (1) 针对患病 4 个月肿瘤实质和患病 2 年肿瘤实质部分的视锥细胞 G1 期细胞簇 (cones: G1 phase); (2) 针对患病 4 个月肿瘤实质和患病 2 年肿瘤实质部分的视锥细胞 G2 期细胞簇 (cones: G2 phase); (3) 针对患病 4 个月肿瘤实质和患病 2 年肿瘤实质部分的小胶质细胞。针对 3 组数据中两两对象之间取 marker 基因得到差异基因集, 显著性阈值设置为 $p_val_adj < 0.05$, $|avg_log_2FC| > 0.5$ 。针对 3 组数据中识别到的差异基因进行基因本体论 (gene ontology, GO)^[15]、京都基因和基因组百科全书 (Kyoto encyclopedia of genes and genomes, KEGG)^[16]和基因集富集分析 (gene set enrichment analysis, GSEA)^[17], 其中 GSEA 分析中的基因集来自 MSigDB^[18]数据库中癌症特征基因集数据 (h: hallmark gene sets, [http://www.](http://www.gsea-msigdb.org/gsea/msigdb/download_file.jsp?filePath=/msigdb/release/7.5.1/msigdb.v7.5.1.symbols.gmt)

[gsea-msigdb.org/gsea/msigdb/download_file.jsp?filePath=/msigdb/release/7.5.1/msigdb.v7.5.1.symbols.gmt](http://www.gsea-msigdb.org/gsea/msigdb/download_file.jsp?filePath=/msigdb/release/7.5.1/msigdb.v7.5.1.symbols.gmt))。

2 结果与分析

2.1 RB 肿瘤微环境单细胞图谱的刻画

通过对数据进行低质量细胞过滤及相关质控, 共保留 5 260 个细胞用于后续相关分析。其中, 3 626 个细胞来自 RB 肿瘤组织 (其中来自 4 个月 RB 肿瘤组织的共 3 044 个细胞, 来自 2 年 RB 肿瘤组织的共 581 个细胞); 1 635 个来自 RB 癌旁组织 (来自 4 个月 RB 癌旁组织的 335 个细胞, 来自 2 年 RB 癌旁组织的 1 300 个细胞), 采用统一流形逼近与投影 (uniform manifold approximation and projection, UMAP) 对细胞群体分布进行投影可视化, 结果显示整合后的细胞群不存在批次效应, 整体分布相一致 (图 1A-B)。无监督聚类分析将 5 260 个细胞划分为 19 个细胞亚群 (图 1C)。



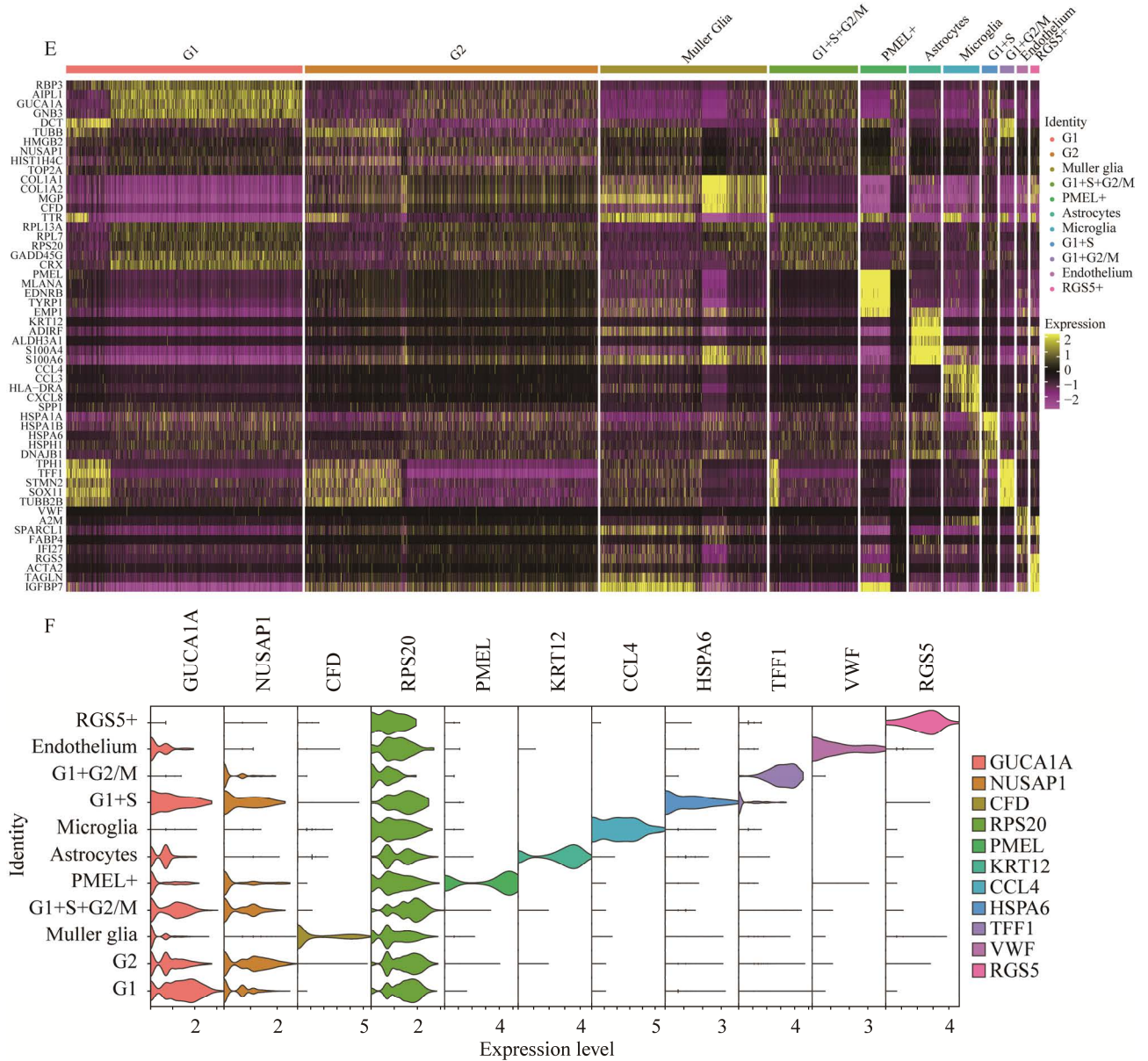


图 1 细胞分群及注释

Figure 1 Cell clustering and annotation. (A) The results after batch effect corrected (by orig.ident). (B) The results after batch effect corrected (by group). (C) Dimensional reduction and clustering on the data after QC. (D) Add annotation for clustering results. (E) The heatmap of top5 marker gene in each cell type. (F) The violinplot of top1 marker gene in each cell types.

进一步，对每个细胞亚群进行标记基因的识别，结合 SINGELR、CELLMARKER 数据库，以及相关文献中细胞的标志基因进行人工校正，最终将 19 个无监督聚类的细胞亚群注释为 11 类细胞群 (图 1D)。包括 G1、G2、

G1+S+G2/M、胶质细胞团 (MULLER GLIA)、星形胶质细胞团 (ASTROCYTES)、PMEL+、G1+S、小胶质细胞团 (MICROGLIA)、G1+G2/M、内皮细胞 (ENDOTHELIUM)、RGS5+。其中，G1、S、G2/M 分别指视锥细胞前体处于 G1

期、S 期、G2/M 期的细胞团；PMEL+ 与 RGS5+ 分别代表 *PMEL* 和 *RGS5* 基因上调表达的细胞团。结果显示，RB 肿瘤不同组织部位的细胞分布整体相似，但个别细胞群的丰度会因肿瘤发生阶段不同而存在差异，这些细胞群主要是一些细胞周期相关的细胞亚群，如 G1 亚群、G2 亚群、G1+G2/M 亚群等。各细胞亚群的前 5 个标志基因热图可视化结果及代表基因的小提琴图分布见图 1E-F。

2.2 RB 肿瘤异质性分析及特征细胞亚群的识别

本研究统计了 RB 肿瘤不同发病时间、不同组织采样部位中各类型细胞的含量，用于比较各组别间细胞组成及比例的差异，识别差异显著的细胞类型 (表 2)。

结果显示，随着 RB 肿瘤的进展，肿瘤组织及癌旁组织中各类细胞亚群比例发生了改变。其中，4 个月的癌旁组织中含量最多的是星形胶质细胞 (astrocytes)，占比 38.5%；2 年

的肿瘤组织中含最多的是穆勒胶质细胞 (Muller Glia)，占比 39.5%；4 个月的肿瘤组织中含最多的是 G1 和 G2 亚群 (占比分别为 34.9% 和 34.8%)，2 年的肿瘤组织中含最高的细胞类型是 G2 细胞亚群 (占比 49.7%)。

与此同时，本研究又根据所处细胞周期阶段对细胞状态进行了统计。结果表明，超过 50% 的视锥前体细胞处于 G2/M 期或 S 期，与 Collin 等^[9]结论一致，揭示 RB 肿瘤的发生引起了不受抑制的细胞分裂增殖，最终发生癌变。

进一步，本研究按照肿瘤组织、癌旁组织以及肿瘤的持续时间进行分组，统计不同组别的细胞所处周期状态 (表 3)。结果表明，癌旁组织中 60%–90% 的细胞处于 G1 期，而在肿瘤组织中超过 60% 的细胞处于 G2/M 期或 S 期，且这种细胞比例在不同时期会有明显差异，揭示 RB 肿瘤组织中大量细胞在异常增殖。

表 2 各组分中各细胞类型含量统计

Table 2 Statistics of cell type from different component

	4M_normal	4M_tumor	2Y_normal	2Y_tumor
G1	0 (0)	1 065 (0.349)	126 (0.096 9)	120 (0.206 5)
G2	29 (0.080)	1 060 (0.348)	247 (0.190 0)	289 (0.497 4)
Muller Glia	138 (0.410)	222 (0.072 9)	514 (0.395 4)	50 (0.086 1)
G1+S+G2/M	5 (0.010)	430 (0.141)	27 (0.020 8)	29 (0.049 9)
PMEL+	0 (0)	90 (0.029 5)	163 (0.125 4)	2 (0.003 4)
Astrocytes	129 (0.385)	35 (0.011 5)	10 (0.007 7)	3 (0.005 2)
Microglia	10 (0.030)	61 (0.020)	99 (0.076 2)	30 (0.051 6)
G1+S	3 (0.009)	51 (0.017)	17 (0.013 1)	15 (0.025 8)
G1+G2/M	0 (0)	1 (0.000 3)	46 (0.035 4)	32 (0.051 6)
Endothelium	16 (0.040)	16 (0.005 3)	22 (0.073 3)	8 (0.013 8)
RGS5+	5 (0.010)	13 (0.004 3)	29 (0.022 3)	3 (0.005 2)
Total	335	3 044	1 300	581

表 3 组间各组织源所处 G1、G2/M 和 S 期的统计分析

Table 3 Statistics analysis of cell cycle of each group between two tissue

Original.ident	Group	G1 phase	G2M phase	S phase
Normal (1 635)	4M_normal	322 (96.1%)	11 (3.3%)	2 (0.6%)
	2Y_normal	828 (63.7%)	287 (22.1%)	185 (14.2%)
Tumor (3 625)	4M_tumor	1 114 (36.6%)	742 (24.4%)	1 188 (39.0%)
	2Y_tumor	183 (31.5%)	217 (37.3%)	181 (31.2%)

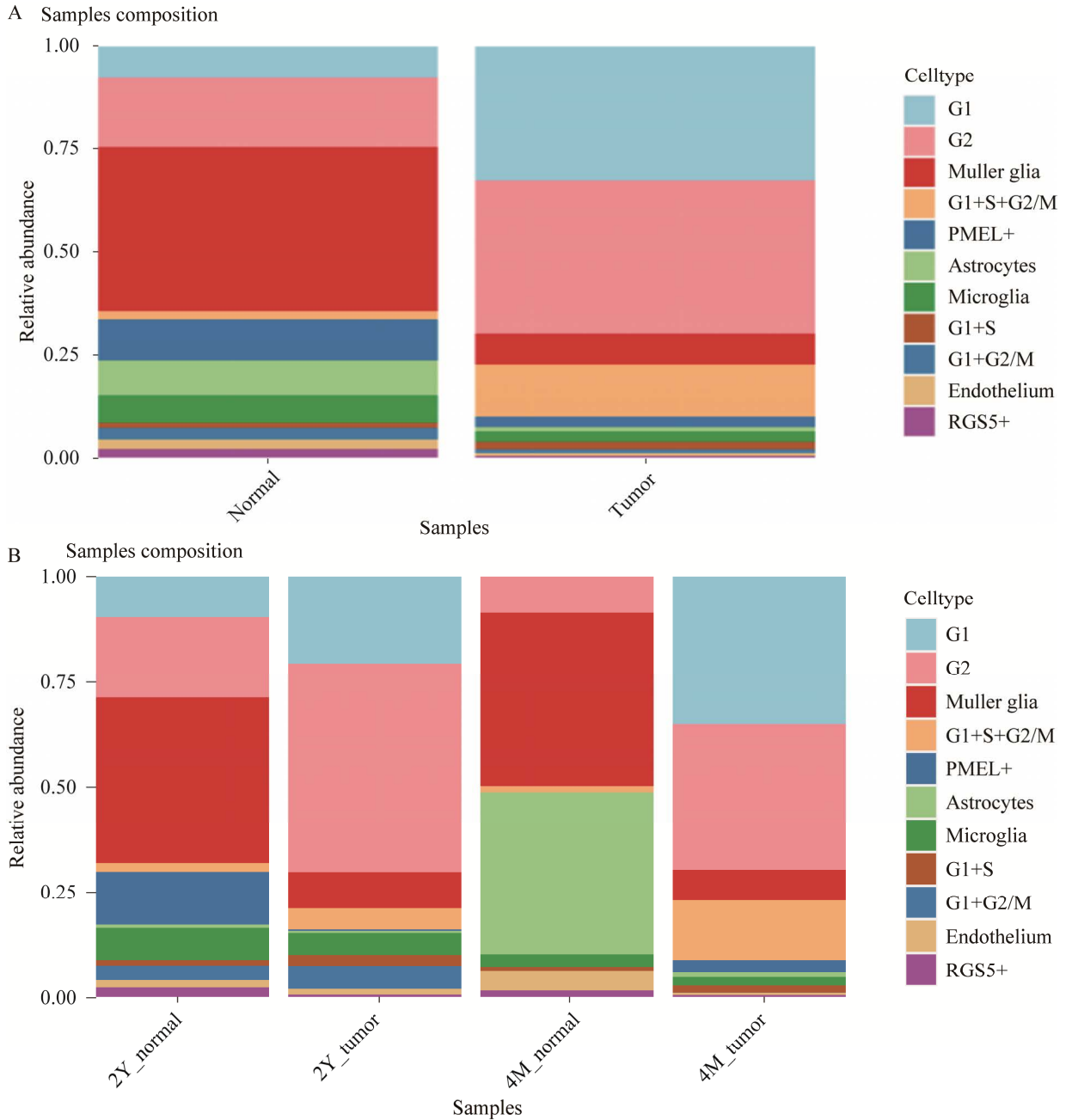


图 2 RB 细胞丰度柱状图

Figure 2 Cell abundance stackbarplot. Cell type composition and content between tumor and normal tissue (A) and among four small samples (B).

最后,本研究展示了不同组织、不同进展时间下各细胞类型的分布情况及含量差异(图 2A-B)。结果显示,相较于癌旁组织,肿瘤组织中与细胞周期相关细胞亚群(G1、G2、G1+S+G2/M 细胞亚群)大幅增多,而穆勒胶质细胞、小胶质细胞大幅减少。

对比 4 个月肿瘤与 2 年肿瘤可发现, G1 细胞亚群含量减少而 G2 细胞亚群含量增多。有研究指出^[9], RB 的发生极有可能是某些基因失活导致视锥前体细胞绕过了 G1/S 阻滞而进入 G2 乃至分裂期,引起恶性增殖,进而形成肿瘤。而这两类细胞的比例在肿瘤发生的前后分别呈现降低与升高的趋势,揭示大量处于 G1 期的视锥细胞无视抑癌基因阻滞而向 G2 期转化增殖。

2.3 G1 期细胞亚群中 RB 肿瘤进展相关特征基因的识别及筛选

提取 G1 期视锥前体细胞亚群,采用 FindMarkers 函数对 4 个月肿瘤及 2 年肿瘤进行差异分析,共识别到差异基因 1 587 个,其中 129 个基因显著上调,其中 297 个基因显著下调 ($|\text{avg_log}_2\text{FC}|>1$, $p_val_adj<0.05$)。

GO 功能分析结果显示,差异基因主要富集的生物学过程为 ATP 代谢过程、细胞质内的翻译等;主要富集的细胞组分为线粒体内膜、核糖体等;主要富集的分子功能为核糖体的结构成分等(图 3A)。这些都是与细胞增殖密切相关,进一步揭示了肿瘤进展过程中细胞功能上的差异主要体现在细胞增殖方面。

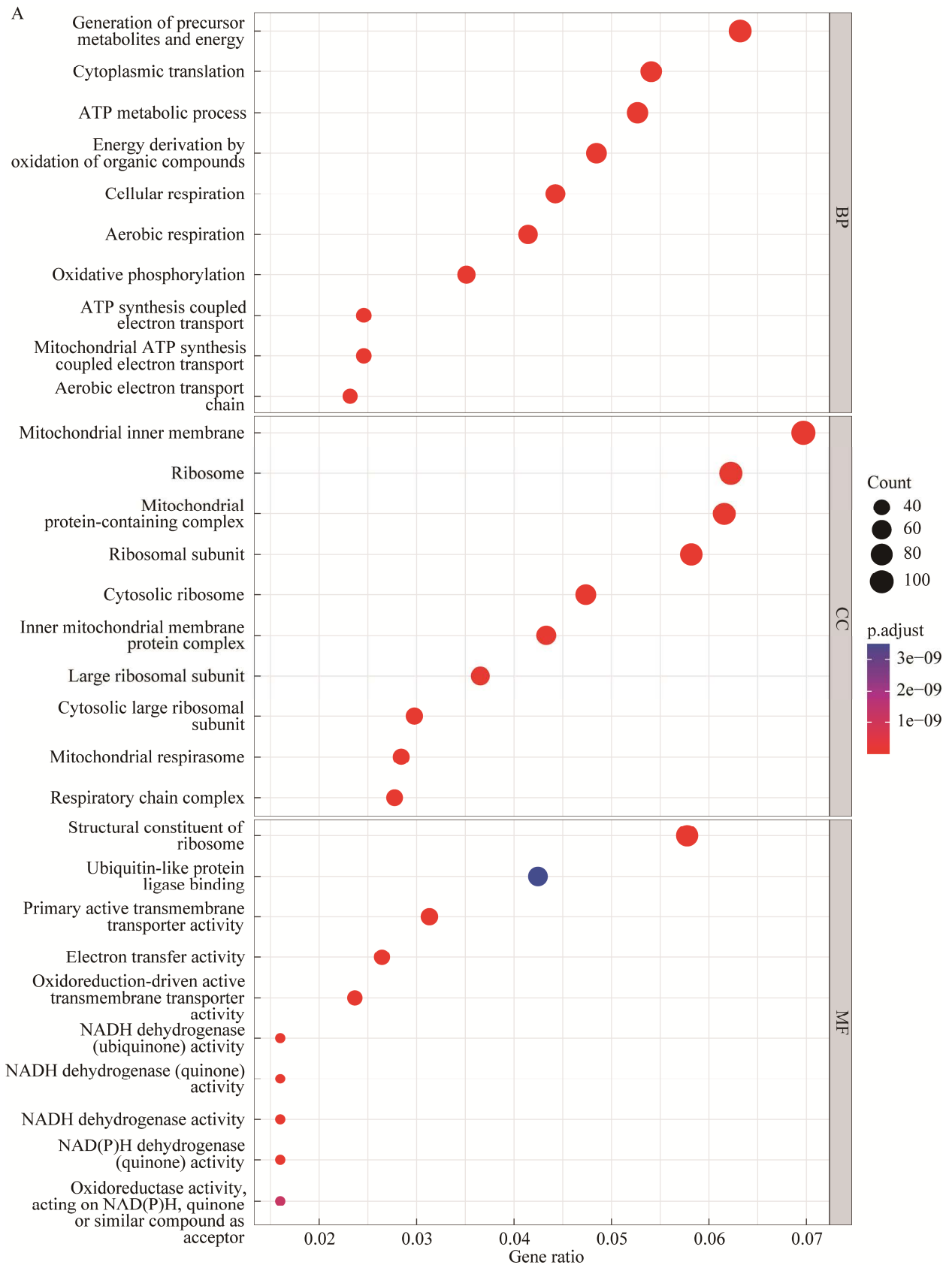
KEGG 富集分析结果显示,上调基因主要参与的通路包括抗原处理和呈递、吞噬体、甲状腺激素合成、内质网中的蛋白质加工等通路(图 3B),说明这些通路是响应 G1 期视锥前体在肿瘤发展中的关键通路;下调基因主要富集到核糖体相关通路、氧化磷酸化等生物学过

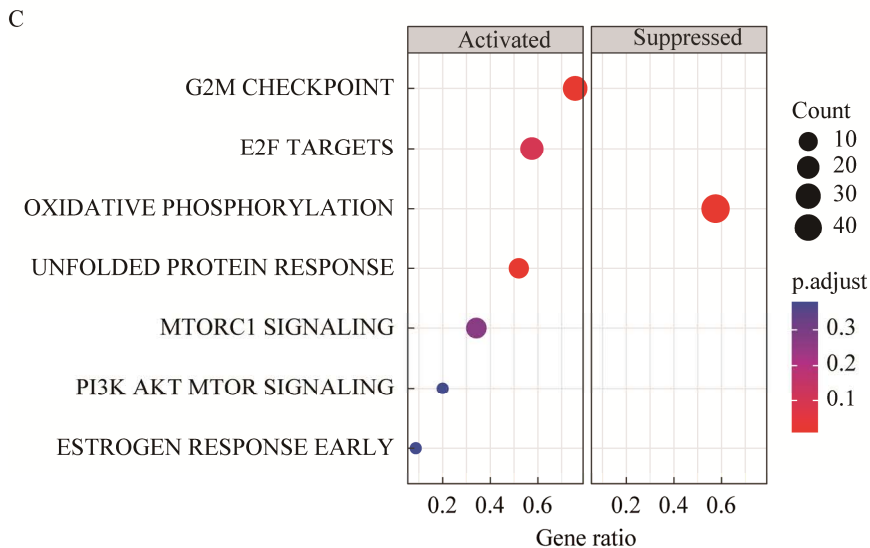
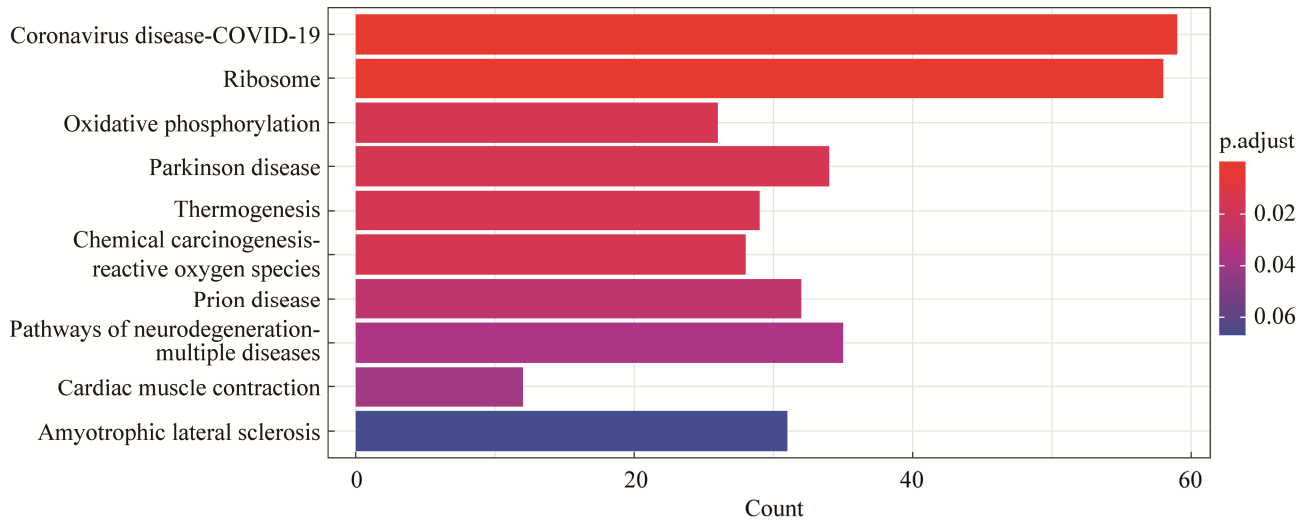
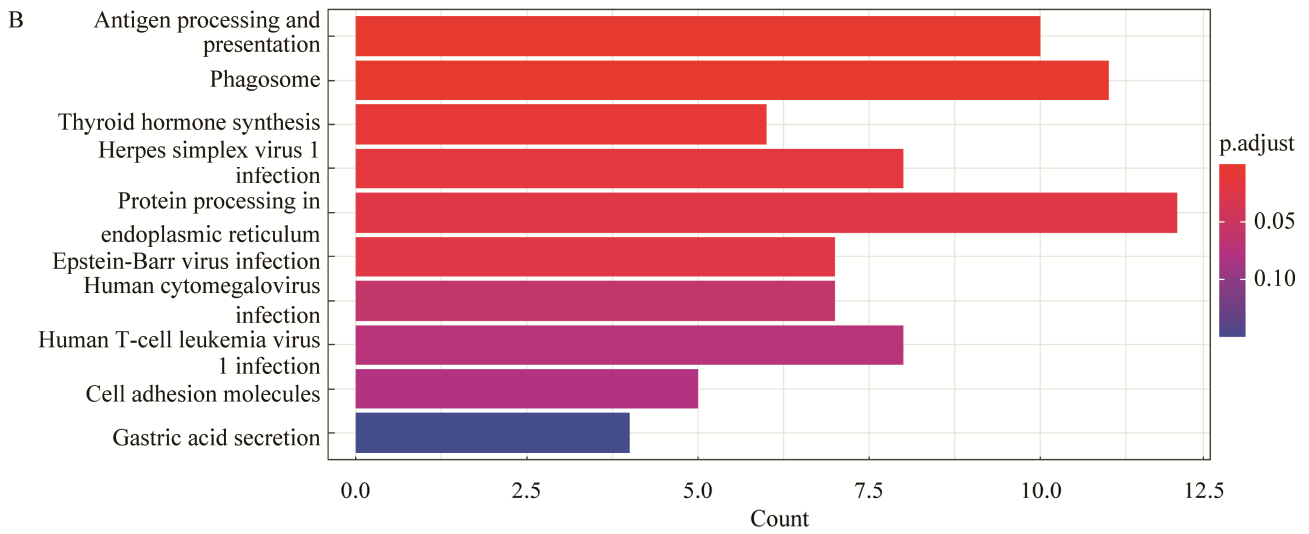
程。其中,氧化磷酸化过程能够以 ATP 形式产生 95%以上的细胞能量,如果该通路被抑制,则会造成 ATP 合成受阻,进而在一定程度上破坏能量的平衡,最终导致了肿瘤的发生和进展。

进一步, GSEA 富集基因结果发现, G1 期视锥前体细胞在肿瘤进展过程中,多数富集的通路表现为激活作用。关键被激活的通路包括 E2F 靶标、G2/M 检查点、未折叠蛋白反应等;关键被抑制的通路为氧化磷酸化通路等(图 3C),其中氧化磷酸化通路在 KEGG 富集分析的结果中亦有体现。值得注意的是, E2F 转录因子家族是参与细胞周期及细胞凋亡重要的调节因子^[19],它会以特异方式与 *RB1* 基因结合,调节 G1 到 S 期细胞周期进程,在 S 期细胞中表达最高,在 G2/M 期表达较低^[9]。本研究结果发现, E2F 靶标、G2/M 检查点在肿瘤进展过程受到激活(图 3D-E),导致患者肿瘤细胞在 G2/M 期失调,未能成功阻止细胞进入 M 期,无法使细胞的 DNA 得到修复,癌症患者的病情就得不到控制。

2.4 G2 期细胞亚群中 RB 肿瘤进展相关特征基因的识别及筛选

针对 G2 期细胞亚群进行 GSEA 富集分析,结果发现未折叠蛋白质反应、MYC TARGETS v1、G2/M_CHECKPOINT 等通路在 2 年的 RB 肿瘤组织中的是被激活的(图 4A)。而有研究指出, *MYC* 基因在 70%的肿瘤中存在过表达或突变,是最常见的高度扩增的癌基因之一^[20],揭示 G2 期的 RB 肿瘤细胞处于异常不受控制的增殖状态。KEGG 富集分析结果表明,上调基因主要参与的通路包括吞噬体、抗原处理和呈递等通路(图 4B),说明此处存在较为活跃的免疫反应,在肿瘤发展过程中其肿瘤微环境发生了变化;下调基因主要参与的通路包括松弛素信号通路、核糖体和 PI3K-Akt 信号通路等。





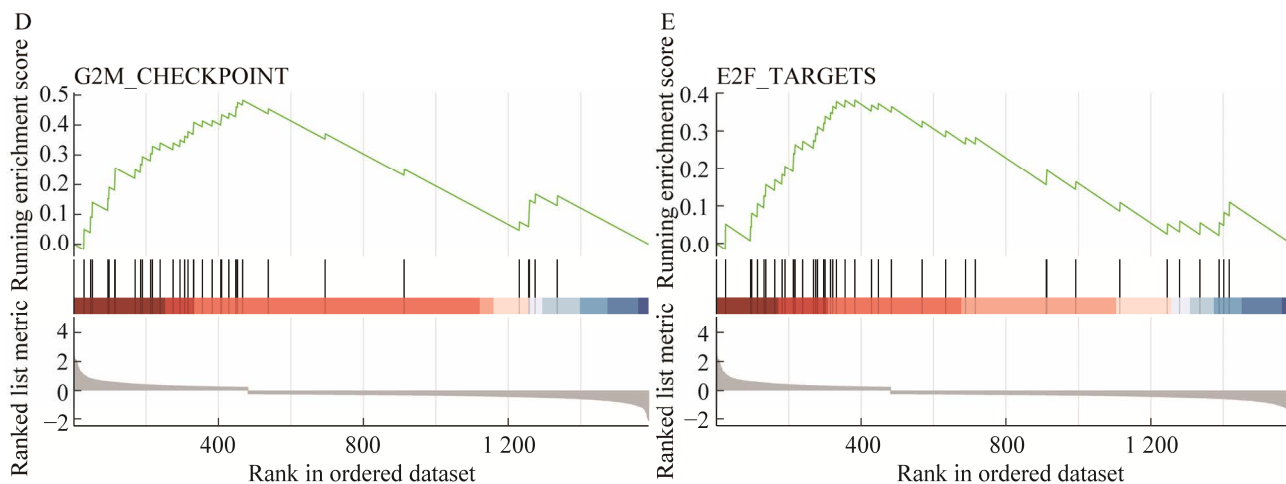


图3 视锥前体 G1 期细胞亚群在不同发育年龄的 RB 肿瘤样本中富集分析

Figure 3 Enrichment analysis for cones cell in G1 phase of different RB developmental stage. (A) Visualization for GO analysis of differential expressed genes. (B) Visualization for KEGG enrichment analysis of up-regulated and down regulated. (C) Visualization for GSEA analysis of differential exgenes. (D) Gseaplot for G2/M_CHECKPOINT. (E) Gseaplot for E2F_TARGETS.

2.5 小胶质细胞中 RB 肿瘤进展相关特征基因的识别及筛选

最后,本研究重点分析了 4 个月 RB 肿瘤与两年 RB 肿瘤组织中 microglia 细胞亚群异质性,共提取差异基因 1 410 个,其中显著上调基因 16 个,显著下调基因 17 个($|\text{avg_log}_2\text{FC}|>1$, $p_{\text{val_adj}}<0.05$)。显著上调和下调的 Top6 基因集及相关信息见表 4。

通过上调基因/下调基因的基因-通路网络图(图 5A-B)可知,上调基因富集到的通路关联最多的基因是 HLA 家族基因(如 *HLA-A* 和 *HLA-B*) 和细胞增殖、转移相关基因——*B2M* 基因,关联的通路为 MHC I 类抗原处理和肽抗原的呈递,抗原处理和内源性抗原的呈递以及肽抗原的抗原处理和呈递等(图 5A)。其中,HLA 家族基因是控制细胞间相互识别、调节免疫应答的一组紧密连锁基因群,其随着 RB 肿瘤的进展而上调表达,揭示了 HLA 家族基因在 RB 肿瘤进展过程中呈现了强烈的免疫应答过

程。同时,有研究指出,*B2M* 基因在 T 细胞杀伤肿瘤过程中发挥着关键作用,并且与肿瘤的增殖和转移密切相关^[21]。然而目前基因 *B2M* 在视网膜母细胞瘤中的相关研究非常少,并无 *B2M* 抑制肿瘤细胞迁移、侵袭的报道,因此本研究推测这可能是 RB 肿瘤发展过程中的一个重要标志物。

下调基因富集到的通路关联最多的基因是 *RPS20* 和 *RPL23*。其中 *RPL23* 编码的核糖体蛋白 L23 是已报道过的肿瘤基因治疗的靶点,是肿瘤患者潜在的临床预后因子^[22]。值得注意的是,有研究指出,*RPL23* 能够间接影响 c-MYC 的致癌活性,在正常细胞中,c-MYC 以非激活的形式存在,当 c-MYC 异常失控时,会诱导包括 *RPL23* 在内的大量核糖体蛋白转录,通过加快蛋白质合成速度以满足肿瘤细胞所需要的高增殖活性^[23]。与此同时,从 Vo 等^[24]在小脑髓母细胞瘤的研究以及 Wanzel 等^[25]的工作中可以发现,*RPL23* 通过核磷蛋白介导作用直接抑

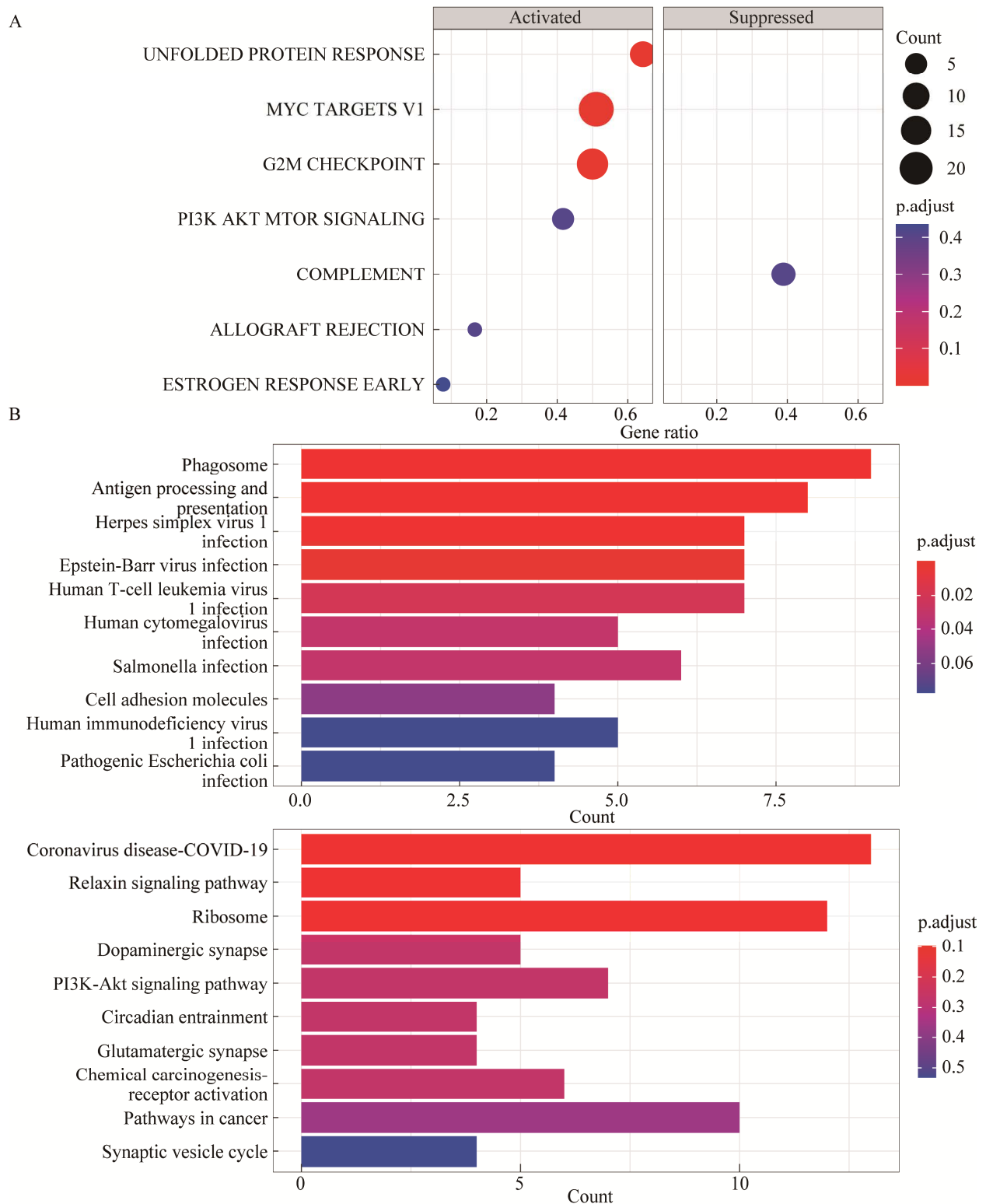


图4 视锥前体 G2 期细胞亚群在不同发育年龄的 RB 肿瘤样本中富集分析

Figure 4 Enrichment analysis for cones cell in G2 phase of different RB developmental stage. (A) Visualization for GSEA analysis of differential genes. (B) Visualization for KEGG enrichment analysis of up-regulated and down regulated genes.

制 Myc 相关锌脂蛋白活性,间接影响了 c-MYC 的表达。而 *RPL23* 是否具有反向调控的能力也存在多方争论,其临床数据并不充分,Zhou 等^[26]发现核糖体蛋白 S14 能负反馈调控 c-Myc,与此同时 *RPL23* 也拥有这样的活性,然而 Liao 等^[27]使用免疫沉淀却未能观察到 *RPL23* 能够结合 Myc-TRBP2 区域从而进一步诱导 c-Myc 凋亡的现象,认为 c-Myc 确实诱导了 *RPL23* 等核糖体蛋白的高表达,但仅仅只有核糖体蛋白 L5 和 L11 才具有特定的负反馈活性。在未来,探索 *RPL23* 对 RB 肿瘤防治也是一个值得深入的研究方向。

综合考虑分别选出 5 个最具代表意义的上调基因和下调基因做气泡图(图 5C)。已知基因 *IFI6* 和 *TGFBI*^[28]都是已报道过的与肿瘤进展或是靶向治疗有关的重要标志物,但它在不同发育年龄的 RB 肿瘤中有着不同的表达情况,因此这些基因在 RB 肿瘤进展过程中同样发挥着不可忽视的作用。

3 讨论

本研究发现,随着 RB 肿瘤的进展,3 种细胞亚群在 RB 肿瘤和癌旁组织中的比例发生了显著改变,包括视锥细胞 G1 期细胞簇、视锥细胞 G2 期细胞簇以及小胶质细胞簇。其中小胶质细胞中一些关键基因(如 *RPL23*、HLA 家族基因、*IFI6* 和 *TGFBI* 等)的异常表达,与肿瘤细胞的增殖、能量代谢以及免疫应答密切相关,可能是预测 RB 进展的重要标志基因。

为了证实我们的研究结果,我们下载另一套 RB 肿瘤单细胞转录组测序数据(GSE68434)进行了相关验证^[29]。该数据集源于 7 例不同年龄的 RB 肿瘤患者,该数据集多数细胞团有表达视锥前体 marker 基因的特征(图 6A)。提取小

胶质细胞亚群,按照患者年龄进行分组。其中,3Y+microglia 表示患者年龄在 3 年或 3 年以上,定为高年龄组;而 3Y-microglia 代表患者年龄低于 3 年,定为低年龄组。通过对本研究中表 4 的基因进行可视化的验证,发现本课题样本里在两年肿瘤患者小胶质细胞团中上调表达的基因,同样也在参考集的高年龄组中明显表达(图 6B),而下调基因在参考数据中也处于较低的平均表达量。这进一步证实我们研究的准确性,说明 HLA 家族基因、*B2M* 基因的表达的确与 RB 肿瘤进展密切相关。

表 4 小胶质细胞中 RB 肿瘤进展相关 TOP6 上调和下调基因

Table 4 TOP6 up-regulated and down regulated gene in microglia related to the progress of RB tumor

TOP6 up-regulated and down regulated gene	Avg_log ₂ FC
Up-regulated gene	
<i>HLA-C</i>	2.450 1
<i>IFI6</i>	2.430 7
<i>HLA-A</i>	2.330 2
<i>B2M</i>	2.265 1
<i>HLA-B</i>	2.144 3
<i>ANXA1</i>	1.924 3
Down-regulated gene	
<i>RPL13A</i>	-1.422 4
<i>RPS20</i>	-1.634 9
<i>HSP90AA1</i>	-1.748 3
<i>TGFBI</i>	-1.788 3
<i>ARL4C</i>	-1.921 0
<i>HSPA1A</i>	-2.285 1

FC refers to fold change, log₂ FC can represent the relative change trend of the difference gene, negative numbers represent downward regulation, positive numbers represent upward regulation.

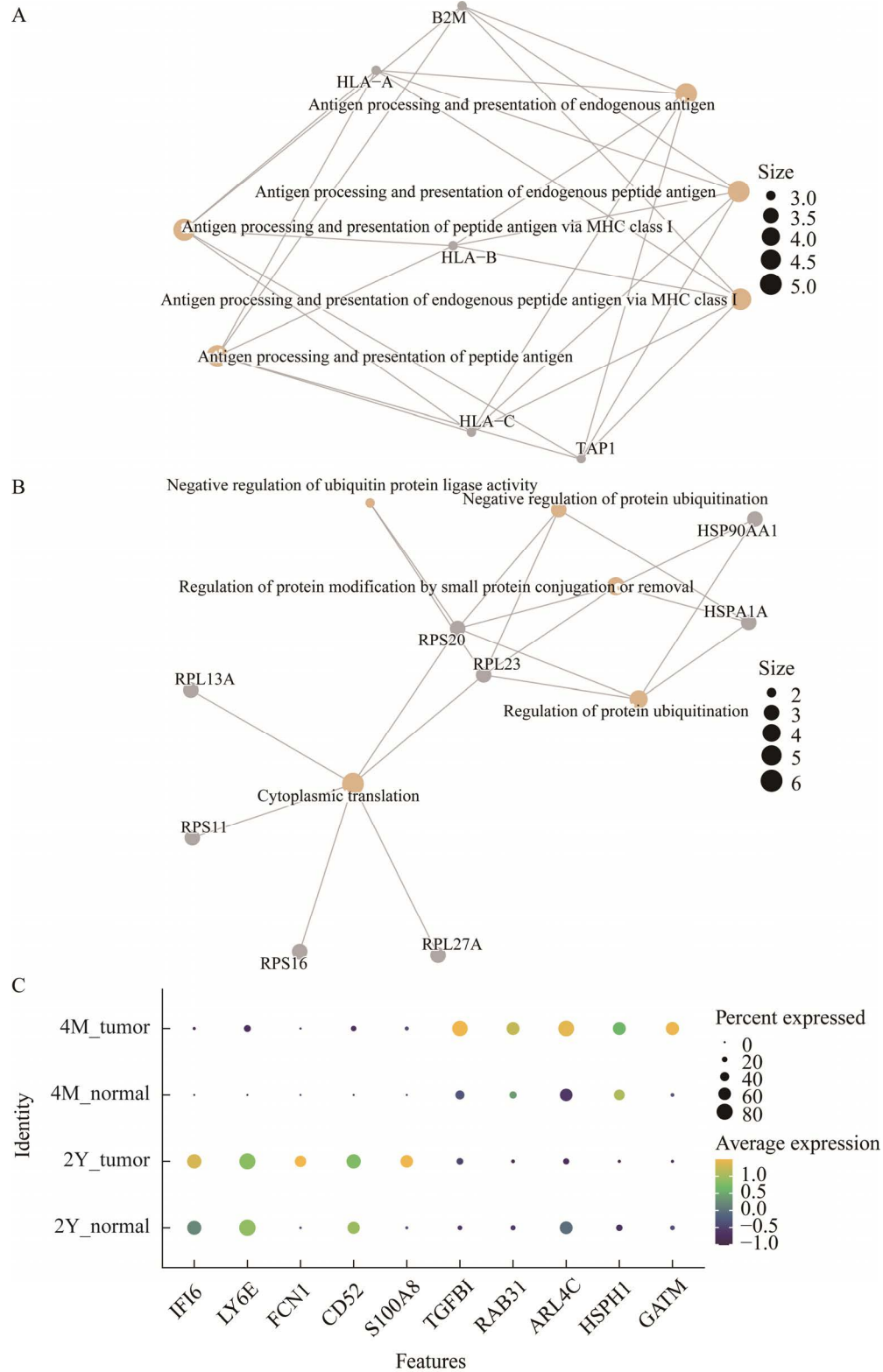


图 5 小胶质细胞亚群在不同发育年龄的 RB 肿瘤样本中富集分析

Figure 5 Enrichment analysis for microglia of different RB developmental stage. (A) Centplot for up-regulated gene. (B) Centplot for down-regulated gene. (C) Dotplot for the top 5 differential genes.

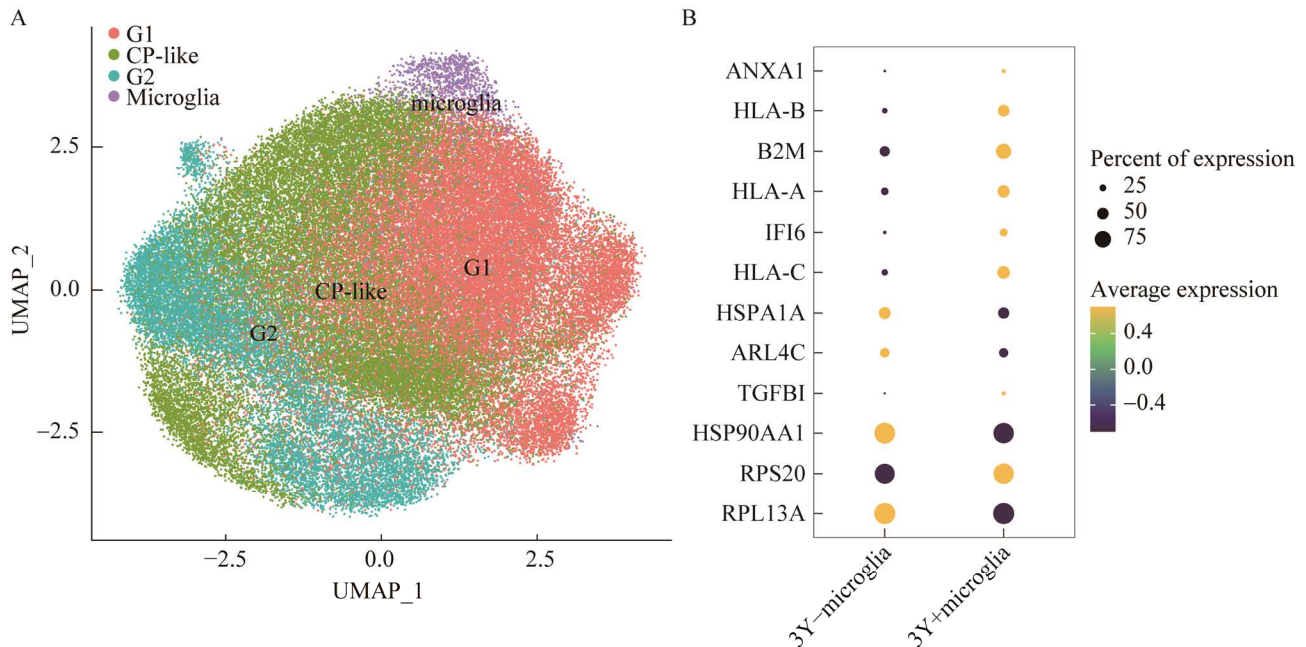


图 6 在验证数据集中对特定基因进行可视化

Figure 6 Visualization of specific genes in reference datasets. (A) ADD CELL ANNOTATION FOR reference datasets. (B) DOTPLOT FOR specific genes of microglia from different age.

本研究的创新点在于，基于原发 RB 肿瘤个体样本，选择来自 4 个不同阶段、不同部位的样本数据，解析 RB 肿瘤在同时段的不同部位、或同部位的不同时段上细胞组成及含量上的差异，从而更清晰直观地说明肿瘤异质性现象及肿瘤动态发展的特征。其次，通过锁定 3 组 RB 肿瘤进展密切相关的细胞组分做深入性的差异分析，识别与之相关的差异基因及相关生物学过程，验证了 *B2M* 及 *HLA* 家族基因 (*HLA-A*、*HLA-B* 和 *HLA-C* 等) 与 RB 肿瘤进展密切相关，可以为后续研究 RB 的关键基因和关键通路提供新的方向，也为将来 RB 肿瘤的治疗以及发病进展机制的理论研究提供了一些参考思路。

REFERENCES

- [1] Fabian ID, Onadim Z, Karaa E, et al. The management of retinoblastoma *Oncogene*, 2018, 37(12): 1551-1560.
- [2] Ortiz MV, Dunkel IJ. Retinoblastoma. *J Child Neurol*, 2016, 31(2): 227-236.
- [3] Linn MA. Intraocular retinoblastoma: the case for a new group classification. *Ophthalmol Clin North Am*, 2005, 18(1): 41-53, viii.
- [4] Xu XL, Singh HP, Wang L, et al. Rb suppresses human cone-precursor-derived retinoblastoma tumours. *Nature*, 2014, 514(7522): 385-388.
- [5] McEvoy JD, Dyer MA. Genetic and epigenetic discoveries in human retinoblastoma. *Crit Rev Oncog*, 2015, 20(3/4): 217-225.
- [6] Giacinti C, Giordano A. RB and cell cycle progression. *Oncogene*, 2006, 25(38): 5220-5227.
- [7] Prasetyanti PR, Medema JP. Intra-tumor heterogeneity from a cancer stem cell perspective. *Mol Cancer*, 2017, 16(1): 41.
- [8] Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol*, 2018, 18(1): 35-45.
- [9] Collin J, Queen R, Zerti D, et al. Dissecting the transcriptional and chromatin accessibility heterogeneity of proliferating cone precursors in human retinoblastoma tumors by single cell sequencing-opening pathways to new therapeutic

- strategies? *Invest Ophthalmol Vis Sci*, 2021, 62(6): 18.
- [10] Hao YH, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*, 2021, 184(13): 3573-3587.e29.
- [11] Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*, 2019, 20(1): 296-310.
- [12] Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods*, 2019, 16(12): 1289-1296.
- [13] Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*, 2019, 20(2): 163-172.
- [14] Zhang XX, Lan YJ, Xu JY, et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res*, 2018, 47(D1): D721-D728.
- [15] Consortium GO. The gene ontology resource: enriching a gold mine. *Nucleic Acids Res*, 2021, 49(D1): D325-D334.
- [16] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 2000, 28(1): 27-30.
- [17] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*, 2005, 102(43): 15545-15550.
- [18] Liberzon A, Subramanian A, Pinchback R, et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, 2011, 27(12): 1739-1740.
- [19] Dimova DK, Dyson NJ. The E2F transcriptional network: old acquaintances with new faces. *Oncogene*, 2005, 24(17): 2810-2826.
- [20] Schulze A, Oshi M, Endo I, et al. MYC targets scores are associated with cancer aggressiveness and poor survival in ER-positive primary and metastatic breast cancer. *Int J Mol Sci*, 2020, 21(21): 8127-8139.
- [21] Jongvilaikasem S, Sampao S, Kanjanapradit K, et al. Serum β -2 microglobulin levels are associated with distant metastasis in patients with breast cancer. *Mol Clin Oncol*, 2021, 14(6): 118-125.
- [22] Wu LY, Li X, Xu F, et al. Over-expression of RPL23 in myelodysplastic syndromes is associated with apoptosis resistance of CD34⁺ cells and predicts poor prognosis and distinct response to CHG chemotherapy or decitabine. *Ann Hematol*, 2012, 91(10): 1547-1554.
- [23] Van Riggelen J, Yetil A, Felsher DW. MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat Rev Cancer*, 2010, 10(4): 301-309.
- [24] Vo BT, Wolf E, Kawauchi D, et al. The interaction of myc with Miz1 defines medulloblastoma subgroup identity. *Cancer Cell*, 2016, 29(1): 5-16.
- [25] Wanzel M, Russ AC, Kleine-Kohlbrecher D, et al. A ribosomal protein L23-nucleophosmin circuit coordinates Miz1 function with cell growth. *Nat Cell Biol*, 2008, 10(9): 1051-1061.
- [26] Zhou X, Hao Q, Liao JM, et al. Ribosomal protein S14 negatively regulates c-Myc activity. *J Biol Chem*, 2013, 288(30): 21793-21801.
- [27] Liao JM, Zhou X, Gatignol A, et al. Ribosomal proteins L5 and L11 co-operatively inactivate c-Myc via RNA-induced silencing complex. *Oncogene*, 2014, 33(41): 4916-4923.
- [28] Zhao HY, Li ZF, Gao Y, et al. Single-cell RNA-sequencing portraying functional diversity and clinical implications of IFI₆ in ovarian cancer. *Front Cell Dev Biol*, 2021, 9: 677697.
- [29] Wu C, Yang JQ, Xiao W, et al. Single-cell characterization of malignant phenotypes and microenvironment alteration in retinoblastoma. *Cell Death Dis*, 2022, 13(5): 438-449.

(本文责编 陈宏宇)